DISCONNECTED YOUTH: PAST, PRESENT & FUTURE

SIGNE HALD ANDERSEN
DISCONNECTED YOUTH: PAST, PRESENT & FUTURE

Signe Hald Andersen
Disconnected youth: Past, present & future

Study Paper No. 116

Published by:
© The ROCKWOOL Foundation Research Unit

Address:
The ROCKWOOL Foundation Research Unit
Soelvgade 10, 2.tv.
DK-1307 Copenhagen K

Telephone +45 33 34 48 00
E-mail forskningsenheden@rff.dk
web site: www.en.rff.dk

February 2017

Abstract
Despite extensive knowledge on the background characteristics, problem load and future prospects of disconnected youth – those not in education or work between age 18 and 29 – we have little evidence on how such traits cluster in subgroups. This is problematic given the huge efforts made and extensive interventions designed to help this group (re-)connect to the labor market and pursue an education, as it may result in one-size-fits-all solutions that potentially over-treat some subgroups of disconnected youth, but under-treat others. To help fill this gap, this paper uses the rich Danish administrative registers to map the group of disconnected youth with regards to concurrent problems, their background characteristics and likelihood of (re-)connection. As expected, I find evidence of subgroups that differ widely with respect to their problem load, their background characteristics and their future prospects. Some subgroups experience only slightly elevated problem loads compared to average youth, with regards to criminal behavior, drug use, mental health problems and early pregnancy. In contrast, other subgroups seem to experience all problems at the same time and at highly elevated levels. And subgroups that have the highest problem load are also more likely to have grown up in disadvantaged environments with regard to maternal mental health, paternal crime, parents’ marital discord and low socioeconomic resources. Similarly, the most disadvantaged groups are also more likely to experience recurrent and persistent disconnection. Given this knowledge, we should think more of how to design interventions that are tailored to fit the specific needs and problems of each subgroup.
Abstract

Despite extensive knowledge on the background characteristics, problem load and future prospects of disconnected youth – those not in education or work between age 18 and 29 – we have little evidence on how such traits cluster in subgroups. This is problematic given the huge efforts made and extensive interventions designed to help this group (re-)connect to the labor market and pursue an education, as it may result in one-size-fits-all solutions that potentially over-treat some subgroups of disconnected youth, but under-treat others. To help fill this gap, this paper uses the rich Danish administrative registers to map the group of disconnected youth with regards to concurrent problems, their background characteristics and likelihood of (re-)connection. As expected, I find evidence of subgroups that differ widely with respect to their problem load, their background characteristics and their future prospects. Some subgroups experience only slightly elevated problem loads compared to average youth, with regards to criminal behavior, drug use, mental health problems and early pregnancy. In contrast, other subgroups seem to experience all problems at the same time and at highly elevated levels. And subgroups that have the highest problem load are also more likely to have grown up in disadvantaged environments with regard to maternal mental health, paternal crime, parents’ marital discord and low socioeconomic resources. Similarly, the most disadvantaged groups are also more likely to experience recurrent and persistent disconnection. Given this knowledge, we should think more of how to design interventions that are tailored to fit the specific needs and problems of each subgroup.
Introduction

One of the big challenges in most western countries today is the significant proportion of young people not in education, employment or training. These are the NEETs, the disconnected youth, or – put more positively – the opportunity youth. This group represents both an unrealized potential as well as a costly short and long term social challenge, to the extent that they fail to ever become integrated at the labor market.

Figures published in 2016 report that 14 percent of 16-24 year olds in the US belong to this category (Treskon, 2016), and similar proportions are found in the EU region (figures from 2012, Eurofound, 2012). And while shares vary across states in the US and countries in the EU, even low-frequency areas such as Denmark, have enough young people in the category that it poses a significant and costly problem (with 6.3 percent of 15-24 year old Danes respectively 10.5 percent of 25-29 year old Danes being in the category, according to the EU figures).

Due to the extent of the problem, and given the knowledge that youth unemployment does not just solve itself, but does in fact increase the probability of future unemployment and other unfavorable labor market outcomes in later life (Baker & Elias, 1991), significant efforts have been made to come up with programs, interventions and initiatives to help the group and promote their transition into either education or work.

Programs such as New Deal For Young People (UK) and Job Corps (US), show positive effects (at least in the short run), and generally, programs targeting this group seems more successful than programs targeting older groups of unemployed.

However, from e.g. life course studies in criminology (Laub & Sampson, 2003; Moffitt, 1993) and more general studies on vulnerable youth (Werner, 1993; 1995), we know that while most youth land on their feet, despite a difficult transition from adolescence to adulthood, a smaller share never makes it. This suggests that we cannot think of the disconnected youth as one homogeneous group – or on any young person as either connected or disconnected. Rather, we should think of disconnectedness as a continuum, where each young person can be more or less connected, or as an umbrella for various forms of more or less severe disconnectedness or different types of disconnectedness with different implications, all showing up in the statistics as young people not in education, employment or training. Hereby we may expect heterogeneity within the group of disconnected youth, both with regards to their problem load, their background characteristics and how they fare later in life.

While we have some indications of an overrepresentation of other types of social and individual problems in the group of disconnected youth (Meadows, 2001; Social Exclusion Unit, 2001; Social Exclusion Task Force, 2008; Coles et al., 2002; Fergusson et al, 2001; Mossakowski, 2008), of heterogeneity in background characteristics (e.g. Eurofound, 2012; Treskon, 2016) and in how they fare in later life (Caspi et al., 1998), no systematic attempts have been made to establish an empirically based typology, that clearly links the
Introduction

One of the big challenges in most western countries today is the significant proportion of young people not in education, employment or training. These are the NEETs, the disconnected youth, or – put more positively – the opportunity youth. This group represents both an unrealized potential as well as a costly short and long term social challenge, to the extent that they fail to ever become integrated at the labor market.

Figures published in 2016 report that 14 percent of 16-24 year olds in the US belong to this category (Treskon, 2016), and similar proportions are found in the EU region (figures from 2012, Eurofound, 2012). And while shares vary across states in the US and countries in the EU, even low-frequency areas such as Denmark, have enough young people in the category that it poses a significant and costly problem (with 6.3 percent of 15-24 year old Danes respectively 10.5 percent of 25-29 year old Danes being in the category, according to the EU figures).

Due to the extent of the problem, and given the knowledge that youth unemployment does not just solve itself, but does in fact increase the probability of future unemployment and other unfavorable labor market outcomes in later life (Baker & Elias, 1991), significant efforts have been made to come up with programs, interventions and initiatives to help the group and promote their transition into either education or work. Programs such as New Deal For Young People (UK) and Job Corps (US), show positive effects (at least in the short run), and generally, programs targeting this group seems more successful than programs targeting older groups of unemployed.

However, from e.g. life course studies in criminology (Laub & Sampson, 2003; Moffitt, 1993) and more general studies on vulnerable youth (Werner, 1993; 1995), we know that while most youth land on their feet, despite a difficult transition from adolescence to adulthood, a smaller share never makes it. This suggests that we cannot think of the disconnected youth as one homogeneous group – or on any young person as either connected or disconnected. Rather, we should think of disconnectedness as a continuum, where each young person can be more or less connected, or as an umbrella for various forms of more or less severe disconnectedness or different types of disconnectedness with different implications, all showing up in the statistics as young people not in education, employment or training. Hereby we may expect heterogeneity within the group of disconnected youth, both with regards to their problem load, their background characteristics and how they fare later in life.

While we have some indications of an overrepresentation of other types of social and individual problems in the group of disconnected youth (Meadows, 2001; Social Exclusion Unit, 2001; Social Exclusion Task Force, 2008; Coles et al., 2002; Fergusson et al, 2001; Mossakowski, 2008), of heterogeneity in background characteristics (e.g. Eurofound, 2012; Treskon, 2016) and in how they fare in later life (Caspi et al., 1998), no systematic attempts have been made to establish an empirically based typology, that clearly links the
young person’s degree or type of disconnectedness to his or her background characteristics and likelihood of reconnection. This knowledge could help us get a much clearer picture of the group, and significantly improve our possibilities of helping them fulfill their potential. We see a tradition of conducting such cluster analyses within other studies on adolescence – in particular those focusing on risk behavior (Mun et al., 2008; Sullivan, Childs & O’Connell, 2010; Brennan & Breitenbach, 2009; Vaughn et al., 2007) – however this tradition has not yet spread to studies on disconnected youth, just as these existing studies on risk behavior rarely include measures related to the young person’s educational aspirations or labor market status (exceptions are McGee & Newcomb, 1992; Ary et al., 1999).

My study fills this gap by linking high quality Danish registers on unemployment, educational achievements, criminal behavior, mental health problems, drug abuse, etc., and by linking children and parents. With this linked data, I map the group of disconnected youth in Denmark from 2000 to 2013, not only with regards to their representation in each cohort, but also with regard to what additional problems they experience – and how such problems cluster in subgroups - and with regards to their background characteristics and their likelihood of reconnection. Due to the nature of the data – particularly given the fact that they do not suffer from any kind of attrition – these analyses have the potential to provide a more precise picture of the clustering of problems or problem profiles than what is possible with other data, and survey data in particular (for a description of potential problems with survey data on this group, see Levitas et al., 2007).

**Background**

To motivate the search for heterogeneity and subgroups within the larger group of disconnected youth, my study departs from the vast literature on risk behavior in adolescence. This perspective offers a framework for understanding the specific social, individual and even biological factors that influence unemployment and lack of educational activities in *adolescence* and *young adulthood* as opposed to unemployment in later age. Also, there is a long standing debate among scholars studying risk behavior in adolescence, on whether risk seeking adolescents represent a uniform group, or consists in a number of subgroups with very different problem profiles, background characteristics and prospects. Given that this is exactly the debate I aim to raise in the literature on disconnected youth, the risk behavior literature is useful in my study.

According to this literature, adolescent specific risk behavior arises in modern societies where the young individual is biologically ready to be an adult several years before he or she is socially acknowledged as such. This is the maturity gap, which implies that the individual feels capable of making adult decisions but has restricted access to resources that allow him or her to carry out these decisions – for instance, there is restricted access to adult consumption goods such as cars (including drivers license), alcohol and tobacco, and independent living requires resources beyond what most individuals in this age group can afford without help from parents or other adults. This means that the immediate independence that body and mind craves, is
limited by social perceptions of what is allowed at a certain age. The resulting conflict inspires the adolescent to seek alternative routes to satisfy the need for independence, including routes suggested by antisocial influences, which may in many instances be the most efficient way to quickly reach a desired goal. Through this search, criminal behavior becomes a useful strategy for acquiring goods, such as cars, alcohol and tobacco, that are otherwise out of reach (Moffitt, 1993). In addition, behaviors which others consider unacceptable, such as e.g. using drugs or practicing unsafe sex, becomes desirable because they signal maturity and autonomy among other adolescents, possible because they are indications that the youth rejects the oppressing and restricting social rules. Importantly, the adolescent does not engage in risky behavior for the explicit purpose of taking a risk, rather this type of behavior is a natural and straightforward route given his or her current aspirations and ambitions, and the individual rarely acknowledges the degree to which this behavior jeopardizes future possibilities of finalizing normal developmental tasks, developing important skills or taking on expected social roles (Jessor, 1991).

In lieu of this framework, the labor market and educational strategies of the disconnected youth may then reflect their age-given rejection of conventional strategies for pursuing the desired independence; having a job or pursuing an education may in the long run provide the goals that the adolescent/young adult desires, however other activities (e.g. such as drug use) are likely to be more immediately rewarding in terms of coping with the maturity gap, or may simply reduce his or her resources for maintaining a job or going to school. The status as disconnected youth is then the (unanticipated/unplanned) result of adolescent specific risk behavior, and reflects “the general problem behavior syndrome”, described by Jessor and Jessor (1977). The presence of multiple problems among disconnected youth is evident from a number of empirical studies (Treskon, 2016; Micklewright, 2002, Baron, 2008; Britt, 1997; Buonanno & Montolio, 2008; Elliott & Ellingworth, 1998; Kapuscinski et al., 1998; Lin, 2007; Reilly & Witt, 1992; Witt et al., 1998).

Parts of the literature on risk behavior in adolescence end here, and settle with the claim that adolescence engaging in one type of risky behavior are highly likely to also engage in other types of risky behavior (e.g. LeBlanc & Bouthilier, 2003; Jessor, 1991). Others, such as e.g. Moffitt (1993), Farrell et al., (2000), MacDonald & Marsh (2002) and Sullivan et al. (2010), find clear evidence of heterogeneity in risk behavior among adolescents, and in particular they show that for some individuals, this type of behavior is not limited to adolescence, but starts in childhood and continues throughout life. For this group, risk behavior is strongly linked to inherent personality traits – including mental health issues such as elevated levels of aggression, schizophrenia, depression and anxiety disorders – to growing up in disadvantaged environments, and to the interaction between those two elements (Moffitt, 1993, 2005). These types of disadvantages are not age specific – they do not just show up in adolescence - rather they have also confounded childhood and will continue to affect the person way into adulthood. Multiple empirical studies find distinct subgroups within
the group of youth engaged in risk behavior, in support of this perspective (as reviewed in Sullivan et al., 2010).

Suggesting that also the group of disconnected youth consists of subgroups is not that far fetched: As I have already established, being unemployed and not pursuing an education in adolescence and early adulthood is risky behavior or the unexpected consequences of risk behavior. However, it also makes sense to imagine a subgroup among the disconnected youth, which is so troubled by inherent personality traits, and childhood disadvantages, that risk behavior is the only route they know to take and where unemployment and lack of educational achievement is just one of a substantial proportion of problems that will haunt the individual throughout life (as demonstrated in e.g. Goldman-Mellor et al., 2015; Caspi et al., 1998). This group of disconnected youth is likely to be distinctly different from disconnected youth that have the ability to conform to social rules when the haze of adolescence lifts.

Inspired by this debate, my paper tests the hypothesis that the group of disconnected youth consists of two or more subgroups with very different problem profiles, background characteristics and prospects. It does so first by identifying subgroups in the group of disconnected youth as defined by their concurrent problem load, second, by testing for differences in background characteristics between these subgroups, and third, by showing whether long term outcomes differ between these subgroups.

Data

All residents in Denmark have a unique personal identification number (similar to the social security number in the US) that is used to identify the residents in all major transactions with public authorities as well as in transactions with a range of private institutions as e.g. banks or private contractors of health services. Statistics Denmark collects an extensive part of the information registered by this unique personal identification number, and makes these data available for statistical and research purposes. Most these registers are available as a yearly panel dating back from 1980 and they contain information on dealings with the welfare state, educational level, income, criminal convictions, drug treatment etc. Importantly, the register on mental health treatment, that I will be using in the analyses only starts in 1995. The data is constructed such that it allows for linking of parents and children.

Samples of disconnected youth

From this data I identify samples for each of the years 2000 to 2013 (subsequently referred to as sample years), containing all individuals whom we may consider disconnected youth according to definitions used in other studies, both with regards to their (lack of) labor market activities, the duration of their current state and their age: As already discussed, the very acronym NEETs defines the troubled youth as someone who is not in employment, nor education or training, and disconnected youth are defined as young persons aged 16
to 24 who are not connected to work or school (Treskon, 2016). Following along these lines, I restrict my sample to include only individuals who are not employed or not in education. However, rather than focusing solely on the 16-24 year olds, as is customary in the literature on disconnected youth, I select individuals from a wider age range – those that are 18-29 in each sample year. This other age delimitation is useful in a Danish setting where eligibility for and level of welfare benefits as well as requirements of the unemployment system (e.g. with regards to participation in active labor market programs), depend on whether the claimant is above or below 30 years of age, and where citizens below 18 years of age are not eligible for such benefits. The system then treats all under 30’s, who are unemployed and who are at the same time not pursuing an education, as youth, with entitlements – and, in particular, limited entitlements - which is likely to reinforce the maturity gap and lead to the type of risk behavior predicted by the theories on adolescent risk behavior. Also, even in other studies on disconnected youth, the age group of the young people in focus varies, and does not always follow the somewhat strict 16-24 delimitation.

With regards to the duration of the state, definitions found in existing literature range from focusing on the young person’s status at a given point in time (e.g. when a cross sectional survey was collected, e.g. Eurofond, 2012) to more extensive and long-term exposure (as in Brown & Emig (1999), who defines long-term disconnection as more than 26 weeks of unemployment or educational inactivity in each of three consecutive years and not having a spouse who is employed or in education). Often the chosen definition seems to primarily reflect the availability of individual level data, i.e. whether data provide detailed information on exposure or not. The registers do not have that type of restrictions. We may therefore apply the specification that we think is most relevant, aiming to have a specification that is sufficiently narrow to not include youth that are part of the group of disconnected youth in a short, transitory phase, but broad enough to demonstrate the range of the group. For this reason, I define disconnected youth as all individuals who have been unemployed or not in education for two consecutive years, where I rely on labor market indicators measured ultimo November, each year. With this specification we exclude by-passers and cover all those with more long-term problems. ¹

In sum, each of my 14 samples then includes all individuals between age 18 and 29 in the sample year, who have not been employed or in education in the sample year and the year before.

In acknowledgement of the already very heterogeneous nature of this group with regards to age and gender, I run all analyses on four subgroups, a) men, age 18-24, b) men, age 25-29, c) women, age 18-24 and d)

¹ Importantly, using this November-measurement does create a risk that I group youth that are permanently disconnected with youth that cycle in and out of disconnectedness – and who are by chance, disconnected in two consecutive Novembers, but not between. It is, however, the assumption, that also youth in the latter group are sufficiently troubled to be relevant in this study.
women, age 25-29. Table 1 shows the number of observations in each group, by sample year. From this we learn, that women are overrepresented in the group of disconnected youth (as is also found in existing studies, see ref.), especially due to the large group of women in the older age category. Notice how numbers vary between years, and in particular, the substantial increase over the years, in the number of male disconnected youth between ages 18-24, and the general decrease in the number of all groups in the years leading up to the Great Recession (and the subsequent explosion in these numbers). This variation in the number of disconnected youth over the years does, of course, also reflect variation in the size of the included birth cohorts, which implies that the figures presented in table 1 is not a direct reflection of yearly changes in the share of disconnected youth. Still it is interesting that the groups start out being substantially different in size, but end up being fairly similar.

Table 1: Number of observations in each subgroup by sample year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Men, 18-24</th>
<th>Men, 25-29</th>
<th>Women, 18-24</th>
<th>Women, 25-29</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>6,666</td>
<td>9,699</td>
<td>10,242</td>
<td>16,039</td>
</tr>
<tr>
<td>2001</td>
<td>6,709</td>
<td>9,300</td>
<td>9,647</td>
<td>15,053</td>
</tr>
<tr>
<td>2002</td>
<td>7,270</td>
<td>9,474</td>
<td>10,015</td>
<td>14,995</td>
</tr>
<tr>
<td>2003</td>
<td>7,986</td>
<td>11,174</td>
<td>10,907</td>
<td>16,979</td>
</tr>
<tr>
<td>2004</td>
<td>7,813</td>
<td>10,691</td>
<td>10,433</td>
<td>16,284</td>
</tr>
<tr>
<td>2005</td>
<td>7,814</td>
<td>9,208</td>
<td>9,902</td>
<td>14,356</td>
</tr>
<tr>
<td>2006</td>
<td>6,816</td>
<td>7,234</td>
<td>8,569</td>
<td>11,907</td>
</tr>
<tr>
<td>2007</td>
<td>5,912</td>
<td>6,472</td>
<td>7,593</td>
<td>10,009</td>
</tr>
<tr>
<td>2008</td>
<td>5,995</td>
<td>6,320</td>
<td>7,313</td>
<td>9,204</td>
</tr>
<tr>
<td>2009</td>
<td>8,034</td>
<td>8,284</td>
<td>8,793</td>
<td>10,513</td>
</tr>
<tr>
<td>2010</td>
<td>10,541</td>
<td>10,707</td>
<td>10,383</td>
<td>11,943</td>
</tr>
<tr>
<td>2011</td>
<td>10,876</td>
<td>11,042</td>
<td>11,041</td>
<td>12,969</td>
</tr>
<tr>
<td>2012</td>
<td>11,446</td>
<td>11,708</td>
<td>11,411</td>
<td>13,665</td>
</tr>
<tr>
<td>2013</td>
<td>11,891</td>
<td>12,274</td>
<td>11,545</td>
<td>14,334</td>
</tr>
<tr>
<td>Total</td>
<td>115,769</td>
<td>133,587</td>
<td>137,794</td>
<td>188,250</td>
</tr>
</tbody>
</table>

Source: Own calculations based on data from Statistics Denmark

*Other problems*

In the analysis, I define important concurrent problems as criminal behavior, mental health problems, drug abuse, and early pregnancy, as is customary in the literature on adolescent risk behavior (Mun et al., 2008; Potter & Nenson, 2003; Pyle et al., 2016). Of these I expect especially mental health problems to be more strongly linked with persistent disconnectedness, as this is one of the elements promoted in e.g. Moffitt (1993)’s description of the life-long offenders.
Relying on the register on criminal convictions, I define criminal behavior as having spent (any) time in prison in either t or t-1 (this is anything from having spent a night in prison to getting a maximum sentence of as much as 16 years) or having gotten probation. I refrain from also including fines in my crime measure, as this type of sanction is too common to be an actual marker of crime proneness.

Relying on the register on mental health hospitalization, I define mental health problems as having been hospitalized for at least one day due to a mental health condition at any point in time, both before and after t and t-1 where we observe the individual’s status as disconnected youth (but, however only from 1995 onwards, given the yearly range of the register for mental health). This hospitalization measure is a relatively strict one, only capturing individuals with serious problems (because milder problems mainly get outpatient treatment). However, given the extent of milder problems such as ADHD, depression and stress in a general population, this is to only capture problems that are serious enough for the system to roll out the full treatment. I include hospitalization at any point in time due to an expected time lap between the onset of a mental health problem and the point of hospitalization – if I only include hospitalization in t and t-1, a lot of disconnected youth that struggle with serious mental health issues that are only discovered (or recorded) by the system later will not be included in my measure. Last, due to the large correlation between different mental health diagnoses, I do not distinguish between types of problems. However, I exclude hospitalizations for conditions related to drug abuse, to prevent an artificially large correlation between this measure and my drug abuse measure (as I specify below).

Relying on the register for treatments against drug abuse, the register on criminal charges and the register on mental health hospitalization, I define drug addiction as having received treatments against drug abuse at any point in time, having been convicted of drug related crime in t or t-1 (regardless of the sanction), and having been hospitalized for conditions related to drug abuse at any point in time. I use this composite measure due to the complex nature of drug addictions where the actual behavior is often hidden, but has a range of less disguised consequences.

Last, I define early pregnancy as becoming a parent before age 22. This might appear late by international standards, but reflects the fact that actual teenage pregnancy (defined as conceiving a child before age 19) is very rare in a Danish context (as of 2015, age of first parity in Denmark is 29.1).

---

2 Note that this imposes a data restriction that might potentially affect the cluster structure, as more individuals in the younger cohorts are then likely to appear in the mental health register prior to the year when we identify them as disconnected youth, but will also be followed for fewer years after this identification – and vice versa for the older cohorts. This means that mental health issues with early respectively late onsets will dominate younger and older cohorts to different extents.
Background characteristics

As specified, I include variables on important background characteristics in the analysis, to show (potential) differences in exposure to known risk factors between the subgroups of disconnected youth. While a range of theories, such as e.g. Bronfenbrenner’s exosystem theory (Bronfenbrenner, 1986) and the mesosystem theories described in Garmezy (1985) and Werner (1993; 1995), offer valid descriptions of where and on what levels to find risk factors, only cumulative risk theory is sufficiently specific to provide an actual guide to identify risk factors (Rutter & Rutter, 1993; Rutter, 1979; Sameroff et al., 1987; Elder, 1974). According to this theory, there are 6 (or 10 – depending on the scholar and study) risks that crucially determine how a child fares in life. These are 1) parents’ severe marital discord, 2) parents’ low socioeconomic status, 3) growing up in a large family, 4) experiencing maternal depression, 5) experiencing paternal crime and 6) having been in foster care. In addition, of the Rochester Longitudinal study further identify ethnic minority status as a risk factor. Due to data restrictions, I cannot test for the child’s foster care experience, however I focus on the remaining 6 risk factors in the empirical analyses (parents’ severe marital discord, parents’ low socioeconomic status, family size, maternal depression, paternal crime and ethnic minority status).

While the registers do not contain any information on marital discord, I know whether parents cohabit and their marital state. This allows me to construct an indicator on whether the parents live together when the child turns 18, and under the assumption that those that have split up have also experienced marital (or relationship) discord this is then a useful proxy for the presence of severe marital discord during the child’s upbringing.

Furthermore, I rely on information on parents’ educational attainment, their income and their unemployment experience to capture their low socioeconomic status. All indicators are measured the year before the child turns 18, but could also have been measured as a mean across all years in childhood or the year before the child was born. However, since many of the cohorts in the yearly samples are born before the first year in the registers (1980), this would have resulted in substantial attrition in these background characteristics. All measures are continuous, but I recode them into binary indicators, to truly capture the importance of low levels of these types of resources. First, I code an indicator which takes the value 1 if the highest educational attainment among parents is having finished compulsory school. Second, I code an indicator which takes the value 1 if both parents are in the lowest income percentile, and third, I code an indicator which takes the value 1 if both parents have been unemployed for more than 70 percent of the time within the year.

To capture large family size I construct a continuous measure for the number of children born by the biological mother at the time when the child turns 18. To capture maternal depression I construct a measure similar to the one described above for the children, which indicates whether the mother has been hospitalized due to mental health problems at any point in time. And to capture paternal crime I construct an indicator of the number of prison sentences that the fathers have received in the time period from when the child turns 9.
until he or she turns 18. It would have been more accurate to have an indicator for prison sentences throughout the childhood, however, this is not possible for all cohorts (as described above).

I define ethnic minority status as being a first or second generation immigrant, according to the registers. Obviously, third generation immigrants may also represent ethnic minorities, however I assume that this group is integrated into the Danish society to the extent where any potential problems are more likely to reflect socioeconomic resources, than ethnicity.

Observations with missing values
Despite the high quality of the Danish register data, they have important limitations. One crucial limitation which will affect my analyses is the obvious lack of information on parents of immigrants, who do not reside in Denmark (the parents that is). Since a significant share of the disconnected youth have immigrant status, and we may expect some of these to have come to Denmark without their parents, there is a substantial and distinct group of disconnected youth for whom I cannot test for risk factors. However rather than excluding this group from the analyses, I include a binary indicator taking the value 1 if information on the parents is missing. This indicator will serve the dual purpose of teasing out this problematic group and of capturing something which is likely to be an independent risk factor, namely migrating to a new country at a young age without your family to support you.

Later outcomes
The key concern with disconnected youth, and the goal of the majority of the interventions targeting this group, is that they exit the category over time and (re-)connect with “normal” society. Because of this, the later outcomes that I focus on in the analyses relate to labor market status and educational attainment. More specifically, I test whether the disconnected youth have started an education or found employment within a period of 2, 3, 4, 5, 6 and 7 year period, i.e. whether they no longer fulfill the criteria of disconnected youth that I use for my analyses. And crucially, I test whether the share who have left the category vary by disconnected youth cluster and subgroup. I test future outcomes in two ways. First, I test whether someone who is disconnected youth in year t is also disconnected youth x years later, but without conditioning the disconnected youth indicator on disconnected youth status in x-1, x-2 etc. Second, I test disconnected youth status x years later, conditioned on disconnected youth status in x-1, x-2 etc. This is to demonstrate the extent to which disconnected youth status is recurrent or persistent – and the extent to which this differs between clusters and groups.

Because I only have data on educational activities and unemployment until 2014, I cannot follow all cohorts of disconnected youth included in the analyses for all periods. This means that I have no relevant information on later outcomes for individuals who are disconnected youth in 2013. In a similar vein, only disconnected youth cohorts from the years 2000 to 2011 can be followed for 3 years and only those from cohorts 2000-
2010 can be followed for 4 years and so on. To prevent this missing information from influencing the results, I code these outcome variables as missing for individuals that I cannot observe in the relevant years, but this then also means that this part of the results presented here mainly relies on disconnected youth cohorts from the early years.

**Method**

The analysis consists of two steps. First, I identify distinct subgroups of disconnected youth through cluster analysis, and second, I rely on standard regression models to estimate the correlation between a) cluster membership and background characteristics and b) cluster membership and subsequent membership of the group of disconnected youth. Note that I run the cluster analysis for each group for each year to, on the one side, capture yearly variation in the composition of disconnected youth, and on the other, to validate the cluster structure. In contrast, I run the regression models on pooled data, to prevent drowning in results, but I control for year fixed effects.

**Cluster analysis**

Cluster analysis is a useful strategy for developing typologies and classifications and is commonly used in the literature on typologies (see Mun et al., 2008; Sullivan, Childs & O’Connell, 2010; Brennan & Breitenbach, 2009; Vaughn et al., 2007).

Cluster analysis is a data mining technique that constructs a number of groups of individuals who are similar on a range of specified parameters, and where the members of each group are more similar to each other than to the members of the other groups. It does so by calculating and grouping individuals according to one or more similarity measures – depending on the choice of cluster algorithms.

I choose the K-means algorithm, which organizes the specified variables \(x_1, x_2, \ldots, x_n\) into \(k\) numbers of clusters by minimizing the within-cluster sum of squares (WCSS), as specified in equation 1:

\[
    \arg\min_k \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2
\]

Where \(S\) is a set of clusters \((S_1, S_2, \ldots, S_k)\) and \(\mu\) is the mean of points in \(S_i\). The K-means algorithm extracts the clusters through an iterative process where it first arbitrarily groups all individuals into \(k\) number of groups – or clusters - and calculates the mean value of each relevant variable in that cluster – the resulting vector of means of the relevant variables is the cluster centroid. It then allocates each individual to the cluster with the nearest centroid, i.e. where the means of the relevant variables in the cluster are closest to the
individual’s values on the same variables. When all individuals are allocated, the algorithm recalculates the centroids, and repeats the process until no individuals change clusters (Aldenderfer & Blashfield, 1984).

I identify the optimal number of clusters using the Calinski/Harabasz pseudo-F-test, where higher test-values indicate more distinct clusters. Equation 2 shows the test, where $k$ is number of groups, $N$ is number of observations, $B$ is the between-cluster sum of squares and cross-products matrix, and $W$ is the within-cluster sum of squares and cross-products matrix:

$$[2] \quad \frac{\text{trace}(B)/(k-1)}{\text{trace}(W)/(N-k)}$$

Here, holding constant the within-clusters sum of squares, a larger value of the between-clusters sum of squares will drive up the test value meaning that larger differences between clusters suggests better fit. And vice versa, holding constant the between-clusters sum of squares, a larger value of the within-clusters sum of squares will drive down the test-value, hereby indicating that larger distances between individuals in the same cluster cause a worse model fit.

Also, given my large sample, and the option to run the cluster analysis for each sample year, my analysis has a built-in verification of the appropriate number of clusters, that allows me to not only test stability over years in number of clusters, but also in the composition of clusters with regards to the distribution of problems within and between clusters.

**Results**

Before presenting results from the more elaborated analyses described in the methods section, I will show a range of figures with raw statistics on the share of disconnected youth in Denmark in my sample years, their problem load and the extent to which these youth suffer from one or more additional problems. This is to provide an overview of the group that is easily comparable to similar statistics produced in other countries and on other data.

First, figure 1 shows the share of disconnected youth, in each of the four specified subgroups (men, age 18-24, men, age 25-29, women age 18-24 and women age 25-29) for the years 2000 to 2013. While the lines are not entirely parallel, we still see a distinct pattern across groups where the share increases from the beginning of the period until 2003/2004/2005, then drops between 2003 and 2008, only to increase again from 2008 onwards. These trends are not surprising, as they seem to largely follow the business cycle, especially with regards to the marked increase in their share following the Great Recession which set in, in Denmark in 2008.
Notice also the shares of disconnected youth in each group. In the introduction, I mentioned statistics from a Eurofound report, which suggested that 6.3 percent of all 16-24 year olds respectively 10.1 percent of all 25-29 year olds in Denmark had NEETs status, however my shares vary between 3 and 5 percent for the younger age group and 4 and 10 percent for the older age group. This difference most likely reflects the differences in the definition of NEETs/disconnected youth discussed in the data section, and the fact that the figures presented in the introduction relies on cross sectional evidence.

Figure 1: Share of disconnected youth, by gender and age. 2000-2013

![Graph showing share of disconnected youth, by gender and age. 2000-2013](image)

Source: Own calculations based on data from Statistics Denmark

Furthermore, to get an initial feel for the problem load among Danish disconnected youth, the four panels of figure 2 show the extent to which the disconnected youth engage in crime, have mental health problems, show indications that they are drug users and experience early parenthood for each year between 2000 and 2013. For comparison, figure A1 in the appendix shows similar statistics for all youth in each of the four groups. As we see, disconnected youth suffer from additional problems to a substantial degree; in the group of male disconnected youth between 10 and 15 percent have experienced incarceration or probation within a two year period – with the expected higher shares among the younger group - which is approximately four to five times more than the average young man. Between 18 and 27 percent have been hospitalized due to mental health problems – with higher shares among the older group – which is between four and seven times more than the average young man. In addition, between 20 and 27 percent have indications that they use drugs (again with higher shares in the older age group), a share that is four times higher than what is found...
among average male youth. Last, male disconnected youth are two to five times more likely than the average young man to have fathered a child before age 22.

These tendencies reappear when we focus on female disconnected youth. They are five to six times as likely to have been in prison or on probation, compared to the average young woman, they are four to five times as likely to have been hospitalized due to mental health problems, and five to six times as likely to show indications of drug use. Last, a large, though declining share of female disconnected youth have become mothers before age 22. This overrepresentation of young mothers among the disconnected youth is particularly pronounced in the early years for the younger group, where as many as a third have become mothers before age 22, reflecting an overrepresentation of a factor seven. Despite the drastic decline over the years in young mothers, this decline is paralleled by trends in young motherhood among average young women in Denmark, and in 2013 the disconnected youth are still overrepresented by a factor 5 in the group of young mothers.

All figures show yearly variation, and in particular, trends seem to be shifting in 2008, which makes sense given what we learned from figure 1 about the yearly variation in the share of disconnected youth. As intake into the group of disconnected youth vary by the business cycle, we should also expect the composition of the group to change by the business cycle: when unemployment rates are low, only the persistent and very troubled unemployed stay jobless, and when unemployment rates are high, also people with more favorable traits become unemployed. The same should be true for our disconnected youth. And in fact, the representation of problems in our four groups of disconnected youth largely follows the business cycle, with an increase in problems from the beginning of the period until 2008, and a declining problem load from 2008 onwards. However, crime among male disconnected youth follows a different pattern, which is expected given the general tendency of declining youth crime (see Andersen et al, forthcoming), and the same is true for early parenthood.

Still, when consulting figure A1 in the appendix, we see that problems also vary among the average youth, where the composition of the group is less susceptible to the business cycle, and where changes in problem load are then unlikely to reflect composition effects. This suggests that part of the variation in the problem load among the disconnected youth resorts from more general social trends in crime, mental health problems, drug use and early parenthood.
To get closer to the research question of this paper, the next figures show the share of each of the four groups of disconnected youth who suffer from one or more (figure 3a), two or more (figure 3b), three or more (figure 3c) or all (figure 3d) of the four problems in focus (crime, mental health problems, drug use and early parenthood), and how the shares change over years. Here figure 3a shows that between 32 to 50 percent of the disconnected youth have one or more problems besides from being unemployed and not in education.
And even if the share varies by year and between groups the extent of additional problems is substantial and should be a cause for concern.

Figure 3a: Share of disconnected youth suffering from one or more additional problems

Source: Own calculations based on data from Statistics Denmark

However, if we focus on youth with 2 or more problems (figure 3b) the share decreases quite dramatically, just as we see a clear gender difference; around 10 percent of male disconnected youth experience 2 or more additional problems, however among women in both age groups, the share is as low 5 percent.

Figure 3b: Share of disconnected youth suffering from two or more additional problems

Source: Own calculations based on data from Statistics Denmark
Zooming in on youth with 3 or more problems (figure 3c), as we would expect, the share decreases even further, but the gender difference persists. Around 1-1.5 percent of male disconnected youth, but less than 1 percent of female disconnected youth experience 3 or more additional problems.

Figure 3c: Share of disconnected youth suffering from three or more additional problems

Source: Own calculations based on data from Statistics Denmark

Last, we see that only very few disconnected youth in each group suffer from all additional problems – in most cases, shares are lower than 0.1 percent. Importantly, for men, the very low number of male disconnected youth who experience early parenthood drives down the share, which then also means that the gender difference disappears.
Zooming in on youth with 3 or more problems (figure 3c), as we would expect, the share decreases even further, but the gender difference persists. Around 1-1.5 percent of male disconnected youth, but less than 1 percent of female disconnected youth experience 3 or more additional problems.

Figure 3c: Share of disconnected youth suffering from three or more additional problems

Source: Own calculations based on data from Statistics Denmark

Clusters

While providing interesting information on the average problem loads among Danish disconnected youth, figures 3a-3d does not help us understand how problems cluster among individuals, and the extent to which some disconnected youth are more disadvantaged than others. As described previously, I address this issue using cluster analysis.

When I run the cluster analysis for disconnected youth in each of the years from 2000 to 2013, the Calinski/Harabasz pseudo F-test suggests that for most years, two clusters are optimal for both groups of men, and that three clusters are optimal for both groups of women. Exceptions are 2000 where only 2 clusters are optimal for women aged 18-24 and, the years 2010-2013 where only two clusters are optimal for women age 25-29. Figure 4 illustrates optimal clusters by group and years.
In addition to the validation of the cluster structure secured through the Calinski/Harabasz pseudo F-test, also the consistency over years in what is the optimal number of clusters, validates the results, and is a strong indication that relying on two-/three-cluster solutions are in fact the appropriate representation of the data.

A further validation comes from the following figures, which show the distribution of the four problems among the four subgroups, by year and cluster; here we will see a marked stability over the years, of the composition and distribution of problems within each cluster, which is again a strong sign that the two-/three-cluster solutions are valid.

**Clusters for men**

First, figure 5a shows the two clusters for the youngest men, and we see that the composition of the two clusters is quite stable over the years. Cluster 1 is the largest and comprises approximately 80 percent of the group in all years. It is characterized by young men who are engaged in crime and have indications of drug use. They have also, to some degree, experienced early parenthood, but none have been hospitalized due to mental health problems. Compared to the average 18-24 year old Danish men in the same years, disconnected youth in this first cluster are 4-5 times as likely to engage in these three types of risk behavior (see figure A1a in the appendix). Disconnected youth in the second and smaller cluster struggle with all four problems, and all have been hospitalized due to mental health problems at some point in life (equivalent to an overrepresentation of a factor 25). In addition a substantial part has indications of drug use (40 percent – an overrepresentation of a factor 7) and between a fifth and a tenth engages in crime (factor 3-6). The young men in this cluster are also overrepresented among young parents, though only with a factor 2. Given the...
massive representation of mental health problems in this second cluster, and the fact that its members experience all other problems at substantially elevated levels, this is likely to be where we find the really troubled youth.

Figure 5a Composition of clusters, men, 18-24. Share with specific problems

Cluster 1 Cluster 2

Figure 5b shows the composition of the two clusters for men aged 25-29. Interestingly, these clusters are very similar to the ones I found for the group of male disconnected youth in the younger age group. Again the first cluster consists of approximately 80 percent of the group (though with declining proportions in the years leading up to the Great Recession), and while they are engaged in both crime and drug use to a larger extent than the average Danish man in the similar age group (by a factor 4), this is probably not where we find the most troubled disconnected youth. In contrast, the male disconnected youth in cluster 2 all suffer from mental health problems, and are substantially more engaged in crime and drug use than the average 25-29 year old Danish man. As with cluster 2 among the younger male disconnected youth, it seems likely that this is where we will find the more troubled youth.
Clusters for women

Figure 5c shows the 3 clusters for female disconnected youth aged 18-24. Note that since the Calinski/Harabasz pseudo F-test suggested only 2 clusters for this group for the year 2000 (as described above), the figure combines cluster statistics for a two-cluster solution for this year, with the three-cluster solutions for all remaining years. Given how the two clusters for 2000 score on the four problems as well as their size, I place the first cluster for this year in the figure showing the composition of the first cluster for the other years, and the second cluster in the figure showing the composition of the third cluster for the other years.

Approximately 70 percent of the female disconnected youth fall in the first cluster, and as shown, the dominant problem among this cluster is mental health issues, given that between 15 and 25 percent of this group have been hospitalized for mental health problems at some point in their life. In comparison, only 4-5 percent of average 18-24 year old Danish women have had that experience. This group is also slightly overrepresented in the crime statistics, compared to other women, and for 2000 we see that a significant share has indications of drug use. The second cluster represents the smallest of the three groups. Only 6-9 percent of female disconnected youth between 18-24 years fall in this category. However, all of them have indications of drug use and more than half struggle with mental health problems (reflecting overrepresentations of factors 15 respectively 33). In addition, a significant share engages in crime and they are overrepresented among young parents (reflecting overrepresentations of factors 50 respectively 4). In the
third cluster, the overarching problem is young parenthood. The size of this group declines over the years from comprising approximately a third of female disconnected youth aged 18-24 to comprising a little less than a sixth. This probably reflects the general decline in early parenthood in Denmark in the observation period. This group does however also struggles with mental health problems to a larger degree than other youth (by a factor 2), and are more likely to engage in crime and use drugs. Of the three clusters, clearly cluster 2 is more burdened. However, clusters 1 and 3 also face problems that are difficult to overcome – having a mental health problem is not something that just goes away, and similarly, the responsibilities of parenthood will in most cases, even for this group, continue many years to come. This makes it difficult to make a clear case of what and what differences to expect with regards to background characteristics and prospects of the three groups.

Figure 5c Composition of clusters, women, 18-24. Share with specific problems

Cluster 1

Cluster 2
Figure 5d shows the three-cluster solution for female disconnected youth aged 25-29. Recall that the Calinski/Harabasz pseudo F-test suggested only 2 clusters for this group for the years 2010 to 2013, which means the three cluster solution for these years is less relevant. However, because the two-cluster solution is widely different from the three-cluster solution I refrain from combining their results in the same figures. Instead I refer to figure A2 in the appendix for the two-cluster solution. As shown, the three-cluster solution for the older female disconnected youth is very similar to what we saw for the group of younger women. The first cluster – consisting of approximately 75-80 percent of these women - primarily suffers from mental health problems, and they are 2.5 to 4 times more likely to experience such problems than the average 25-29 year old Danish woman. The second cluster (comprising of 5-7 percent of disconnected women aged 25-29) is characterized by drug use (100 percent compared to 1 percent among the comparison group), but also of massive overrepresentations of the other three problems – of factors 80 for crime, 10 for mental health problems and factor 3-4 for early parenthood. The predominant problem in the third cluster comparison of 11-20 percent of the disconnected women aged 25-29 is early parenthood, even if they also suffer from mental health issues and engage in drug use (overrepresentations of factors 2 in both cases). Similar to what we saw for the 18-25 year old female disconnected youth, cluster 1 is the largest comprising 75 to 80 percent of the group, cluster 2 is the smallest comprising 4-7 percent of the group, and the share in cluster 3 declines over the years from 20 to 10 percent. Again women in cluster 2 are the more troubled of the three groups.
Figure 5d Composition of clusters, women, 25-29. Share with specific problems

Cluster 1

Cluster 2

Cluster 3

Source: Own calculations based on data from Statistics Denmark

The two-cluster solution that I show in the appendix shows one cluster with milder problems of criminal activities and drug use (where both problems are overrepresented by a factor 4), and one with massive representations of early parenthood and mental health problems, but close to average levels of problems pertaining to crime and drug use. This last cluster seems to somewhat combine clusters 1 and 3 from the three-cluster solution.
Background characteristics

After having identified the number of subgroups in the four groups of disconnected youth, the next ambition of this paper is to test for differences in background characteristics of these groups. The purpose of such analyses is to see whether the differences in problem load between the clusters also reflect differences in background characteristics, specifically with regards to the risk factors identified in cumulative risk theory discussed earlier.

For this purpose, I run standard regression models for each group of disconnected youth as defined by age and gender (men, 18-24; men, 25-29; women, 18-24 and women, 25-29), with cluster membership as the outcome variables, and the risk factors as explanatory variables. To prevent drowning in results, I pool data for all the years (from 2000-2013), but include a year indicator to account for yearly variation. Also, since several individuals are likely to be disconnected youth in more than one year, and thus appear multiple times in the data, I cluster on person identifier in the models.

For both groups of men, the stable two-cluster solution allows me to use a standard linear probability model for these analyses. However, for women the design is potentially trickier, for two reasons. First, the preferred number of clusters changes over years, especially for the older group, and second, for most of the years, the results from the cluster analyses points to a three-cluster solution, which rules out the option of using a linear probability model for these analyses. For simplicity – and again to avoid drowning in results - I solve the first problem by using the three cluster solution for all years, but by conducting robustness checks where I exclude the “problematic” years from the main analysis, and where I conduct an analysis based on the two-cluster analysis for the same years. I solve the second problem by using a multinomial logit model. For comparability and readability of the results, I present the average marginal effects calculated from these models, rather than the raw regression coefficients.

As a benchmark for the results that I present next, table A2 in the appendix shows results of 4 linear probability models on the likelihood of being a disconnected youth, relative to being a connected youth by age and gender.

Results for men

Table 2 below shows the results for both groups of men, with the first column showing the correlation between background characteristics and cluster membership for male disconnected youth aged 18-24, and the third column showing these correlations for male disconnected youth aged 25-29. Cluster 1 takes the value 0 on the outcome variable and cluster 2 takes the value 1 for both groups. Recall that the first cluster consists of disconnected males with the fewest problems, while the second cluster are more massively affected.
Looking at the first column that holds results for the male disconnected youth aged 18-24, we see that four background characteristics have significant coefficients. First, the coefficient for mother’s mental health problems is positively correlated with membership of cluster 2, indicating that more youth in this cluster have mothers who have been hospitalized due to mental health problems at some point in their lives. Second, is the coefficient for parental cohabitation, indicating that more youth in the second cluster have broken families. Third, and possibly in contrast with the first two significant coefficients, fewer youth grew up in families where parents’ highest education attainment was compulsory school, and fourth, fewer have an immigrant background. In combination, these results suggest that disconnected youth in this group that have the most problems, did not necessarily grow up in lower resource families, however their caring environment may have been compromised by maternal mental health problems and the general distress caused by parents’ split.

When we turn to the results for the older male disconnected youth, we see a strikingly similar pattern. Youth in the second cluster are also more likely to have mothers with mental health problems, to have grown up in a broken family, and less likely to have parents with no more than compulsory school as lowest educational achievement and to have an immigrant background. But in addition, there are also fewer of the older male disconnected youth in the second cluster for whom there is no information on parents in the registers. As described earlier this is an indicator of being a first generation immigrant, and the coefficients then shows that to the extent that such individuals end up as disconnected youth, they are significantly more likely to be among the lighter, less problem-ridden group.

Table 2: Correlations of cluster membership and background characteristics, men.

<table>
<thead>
<tr>
<th></th>
<th>Men, 18-24</th>
<th></th>
<th></th>
<th>Men, 25-29</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother with mental health problems</td>
<td>0.110***</td>
<td>-0.010</td>
<td>0.124***</td>
<td>-0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother's # of children</td>
<td>-0.004</td>
<td>0.000</td>
<td>-0.003</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dad incarcerated (from child age 9-18)</td>
<td>0.000</td>
<td>-0.010</td>
<td>-0.003</td>
<td>-0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parents live together</td>
<td>0.045***</td>
<td>0.000</td>
<td>-0.039***</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parents no more than compulsory school</td>
<td>-0.030***</td>
<td>0.000</td>
<td>-0.026***</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both parents in lowest income decile</td>
<td>0.002</td>
<td>-0.010</td>
<td>0.015</td>
<td>-0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both parents unemployed for more than 70 percent of the year</td>
<td>0.003</td>
<td>-0.010</td>
<td>-0.008</td>
<td>-0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immigrant background</td>
<td>-0.015*</td>
<td>-0.010</td>
<td>-0.024***</td>
<td>-0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No parent information in registers</td>
<td>-0.019</td>
<td>-0.010</td>
<td>-0.049***</td>
<td>-0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year as disconnected youth</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.001</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.683</td>
<td>-0.890</td>
<td>1.297</td>
<td>-0.860</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: p<0.05; **: p<0.01; ***: p<0.001

Source: Own calculations based on data from Statistics Denmark
It is furthermore interesting to compare these results to the results presented in table A2 in the appendix showing the likelihood of being a disconnected youth, relative to being a connected youth. This comparison shows that in some aspects, members of cluster 2 differ significantly more from members of cluster 1 than the full group of disconnected youth differs from connected youth. For example, while male disconnected youth are 3-4 percent more likely than connected youth to have mothers with mental health problems, members of cluster 2 are 11-12 percent more likely to have this experience than members of cluster 1. And in contrast, male disconnected youth are approximately 4 percent more likely than connected youth to have parents with no more than compulsory school, however this finding only applies to members of cluster 1.

Results for women

Table 3 shows the results for women, in the form of average marginal effects from a multinomial logistical regression model. The upper panel shows results for the three clusters identified for women aged 18-24, and the lower panel shows the equivalent results for the three clusters identified for women aged 25-29. The first cluster (as defined in figures 5c and 5d) is the baseline outcome, in both cases.

Recall that the first cluster (the baseline) is dominated by female disconnected youth with mental health problems, the second by female disconnected youth with indications of drug use, but also by substantial problems pertaining to crime and mental health problem, and the third and last cluster is dominated by female disconnected youth who have become parents at an early age.

Starting with the upper panel, we see how maternal mental health problems increase the probability that a disconnected youth is in cluster 2, but reduces the probability of being in cluster 1. This risk factor does, however, not affect the probability of being in cluster 3.

Next, mother’s number of children is positively correlated with being in the third cluster, but negatively correlated with being in the second cluster and does not affect the risk of being in the first cluster. Father’s incarceration increases the likelihood of being in the second and third cluster, and decreases the likelihood of being in the first cluster, and the opposite is true for parental cohabitation – here, parental cohabitation increases the risk of being in cluster 1, but decreases the risk of being in cluster 2 and 3.

The results furthermore show that disconnected youth in cluster 1 are less likely to have parents with no schooling beyond compulsory school and who lie in the lowest income decile, but that these factors increase the probability of being in cluster 3. Parental educational attainment has no importance for membership of the second cluster, however low parental income increases the probability of such membership. And while there is no significant correlation between cluster membership and parental unemployment, having an immigrant background significantly increases the probability of being in the third cluster, and reduces the
probability of being in the first respectively second cluster. Last we see that there are more disconnected youth who lack parental information in cluster 1, but fewer in clusters 2 and 3.

An overall assessment of the observed patterns then suggests that the third cluster, which does in fact not seem to consist of the group of young women with the strongest problem load, have the most troubled background with regards to paternal crime, parents’ marital discord and their low socio-economic resources. Members of the more troubled cluster 2 have similar problems but to a smaller degree and with the crucial difference that more of their mothers experience mental health problems and fewer are immigrants.

Focusing our attention to the results for the 25-29 year old female disconnected youth, the results are again – as with men – strikingly similar, despite small differences in levels.

Table 3: Correlations of cluster membership and background characteristics, women. Average marginal effects.

<table>
<thead>
<tr>
<th>Women, 18-24</th>
<th>1. cluster</th>
<th>2. cluster</th>
<th>3. cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother with mental health problems</td>
<td>-0.032***</td>
<td>0.008</td>
<td>0.031***</td>
</tr>
<tr>
<td>Mother’s # of children</td>
<td>-0.002</td>
<td>0.002</td>
<td>-0.007***</td>
</tr>
<tr>
<td>Dad incarcerated (from child age 9-18)</td>
<td>-0.084***</td>
<td>0.009</td>
<td>0.034***</td>
</tr>
<tr>
<td>Parents live together</td>
<td>0.129***</td>
<td>0.005</td>
<td>-0.038***</td>
</tr>
<tr>
<td>Parents no more than compulsory school</td>
<td>-0.071***</td>
<td>0.005</td>
<td>-0.001</td>
</tr>
<tr>
<td>Both parents in lowest income decile</td>
<td>-0.035***</td>
<td>0.008</td>
<td>0.013**</td>
</tr>
<tr>
<td>Both parents unemployed for more than 70 percent of the year</td>
<td>-0.012</td>
<td>0.011</td>
<td>0.001</td>
</tr>
<tr>
<td>Immigrant background</td>
<td>-0.048***</td>
<td>0.008</td>
<td>-0.076***</td>
</tr>
<tr>
<td>No parent information in registers</td>
<td>0.032***</td>
<td>0.010</td>
<td>-0.011</td>
</tr>
<tr>
<td>Year as disconnected youth</td>
<td>0.009***</td>
<td>0.000</td>
<td>0.002***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Women, 25-29</th>
<th>1. cluster</th>
<th>2. cluster</th>
<th>3. cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother with mental health problems</td>
<td>-0.040***</td>
<td>0.007</td>
<td>0.029***</td>
</tr>
<tr>
<td>Mother's # of children</td>
<td>0.000</td>
<td>0.002</td>
<td>-0.006***</td>
</tr>
<tr>
<td>Dad incarcerated (from child age 9-18)</td>
<td>-0.070***</td>
<td>0.007</td>
<td>0.023***</td>
</tr>
<tr>
<td>Parents live together</td>
<td>0.114***</td>
<td>0.004</td>
<td>-0.038***</td>
</tr>
<tr>
<td>Parents no more than compulsory school</td>
<td>-0.071***</td>
<td>0.004</td>
<td>-0.000</td>
</tr>
<tr>
<td>Both parents in lowest income decile</td>
<td>-0.055***</td>
<td>0.006</td>
<td>0.016***</td>
</tr>
<tr>
<td>Both parents unemployed for more than 70 percent of the year</td>
<td>-0.009</td>
<td>0.007</td>
<td>-0.000</td>
</tr>
<tr>
<td>Immigrant background</td>
<td>0.002</td>
<td>0.007</td>
<td>-0.089***</td>
</tr>
<tr>
<td>No parent information in registers</td>
<td>0.059***</td>
<td>0.009</td>
<td>-0.022***</td>
</tr>
<tr>
<td>Year as disconnected youth</td>
<td>0.004***</td>
<td>0.000</td>
<td>0.002***</td>
</tr>
</tbody>
</table>

*: p<0.05; **: p<0.01; ***: p<0.001

Source: Own calculations based on data from Statistics Denmark
Table A1 in the appendix shows the results from the robustness check, where I rerun the multinomial logit model on a sample restricted to the years where the Calinski/Harabasz pseudo F-test suggests a three-cluster solution (2000-2009), and where I run a linear probability model for the years where the test suggests a two cluster solution. The sample restricted three-cluster analysis provides results that are very similar to what we see in table two, except that the correlation between mother’s number of children and membership of cluster 3 is no longer significant.

For the interpretation of the two-cluster solution we need to recall the cluster structure presented in figure A2. Here, the first cluster consists of disconnected youth that primarily have small indications of drug use, and the second cluster of those with multiple problems. Thus the second cluster is the most disadvantaged, something which is also, to some extent reflected in the background characteristics analysis. More female disconnected youth in cluster 2 have mothers with mental health problems and experiences of paternal incarceration, and more come from broken homes. In addition, fewer have an immigrant background and lack parental information in the registers.

Again it is interesting to compare these results to the results presented in table A2 in the appendix showing the likelihood of being a disconnected youth, relative to being a connected youth. Also this comparison shows that in some aspects, differences between the three groups of disconnected young women are larger than the differences between connected and average disconnected youth.

Prospects

As shown in previous sections, the larger group of disconnected youth consists in 2 to 3 three subgroups which are not only distinguishable with regards to the concurrent problem load, but also with regards to their background characteristics and the degree to which they have been exposed to common risk factors during their upbringing. As hypothesized, I find that some disconnected youth are significantly more troubled than others. A last, but crucial part of the empirical analysis is still missing, though; if the different groups also differ with regards to how well they manage to connect (or reconnect) to the labor market or the educational system, this is a strong signal that – as suggested in the introduction – we need to differentiate interventions targeting this group. The last part of this paper therefore tests whether the identified clusters of disconnected youth differ with respect to their probabilities of having left the disconnected youth group after 2, 3, 4, 5, 6 and 7 years, and as explained I test future disconnected youth status as a recurrent phenomenon and a persistent phenomenon.

Figure 6 shows the results related to disconnected youth status as a recurrent phenomenon, by age and gender, and, of course, cluster membership. The depressing take-home message from the figure is that for all
groups, their status as disconnected youth is largely recurrent. Even as many as a third among disconnected youth with the highest (re-)connection rate – men, aged 18-24, first cluster – experience disconnection after seven years. Among the heaviest groups – men respectively women aged 25-29, second cluster – more than 60 percent are disconnected from work and education in each of the years.

There are however, large differences between clusters, also within the subgroups as defined by age and gender, differences which follows the expectation we got from the previous analyses. For male disconnected youth in both age groups, those that belong to the second cluster are significantly more likely to also not be in education and employment after 2 to 7 years. After two years the difference between the two clusters is approximately 18 respectively 23 percent, but widens slightly over the year and ends up at 21 respectively 25 percent after 7 years.

For women, aged 18-24 we get an interesting pattern where clusters 1 and 3 have very similar reconnection rates – for both groups, approximately 60 percent are disconnected youth 2 years after and approximately 40 percent are disconnected youth 7 years after. Cluster 2 have elevated risks of experiencing recurrent disconnection in all years, and the gap doubles from 2 to 7 years (from 8 to 15 percent) after we initially observe them as disconnected youth. This is hardly surprising given the problem load of cluster 2, where all show indications of drug use, and a substantial group struggle with additional problems. In contrast, it seems more perplexing that experiencing early parenthood (cluster 3) is as or to some extent more burdensome as suffering from mental health problems (cluster 1) in a society as the Danish that provides a range of financial benefits to young (and potentially single) mothers, as well as other support related to taking care of a child. However part of the explanation could be that also a significant share of the women in cluster 3 suffers from problems related to mental health issues.

For female disconnected youth aged 25-29, the pattern is slightly different, as clusters 1 and 3 do not follow the same pattern. Here, cluster 3 is slightly more likely than cluster 1 to reappear as disconnected youth in all years (the difference is 6-7 percent), but cluster 2 is still far more likely at all points in time, to (still) be categorized as disconnected youth given their lack of educational activities and stable labor market affiliations. The difference between clusters 1 and 2 increases over the years, from 19 to 23 percent.
Figure 6: Share not employed or in education for a 2-year period, 2-7 years after initial status as disconnected youth

Source: Own calculations based on data from Statistics Denmark

Next I am interested in whether someone who is disconnected youth in both year t and year t+x has been disconnected youth for all years between t and t+x. For this purpose, figure 7 shows the extent to which disconnected youth status is such a persistent phenomenon in the clusters of the four subgroups. The figure shows dramatic drops in the share of persistently disconnected youth as we move forward in time, as only between 15 and 30 percent remain disconnected after 7 years. Differences in disconnected youth persistence between clusters follow the same pattern as we saw in figure 6, where the most troubled clusters within each subgroup were also more likely to experience recurrent disconnection. While it is reassuring that only a
minor share stick to the category, considering the evidence from figure 7 and figure 6 jointly shows that even a very large share of disconnected youth who avoid persistent disconnection, seem to be only unstably reconnected, as they cycle in and out of the category.

Figure 7: Share not employed or in education for a 2-year period, 2-7 years after initial status as disconnected youth

Source: Own calculations based on data from Statistics Denmark
Conclusion

This paper combines a range of empirical analyses to describe the extent to which problems cluster within four subgroups of disconnected youth as defined by age and gender, how such clustering reflects different background characteristics and the extent to which they are correlated with future risk of being or experiencing disconnection.

The analyses show clear and, to large extents, expected patterns. The cluster analysis suggests that male disconnected youth split into two groups – regardless of their age – where the first and largest group engages in risk behavior, such as crime, drug use and early parenthood, and at elevated levels, compared to average male youth. Also the second and smaller group engages in these activities – however at more highly elevated levels – and in addition, they all suffer from mental health problems. And while the two groups differ little with regards to what risk factors they have been exposed to during childhood, significantly more youth in the second group have mothers with mental health problems and have grown up in broken homes. Interestingly, the less disadvantaged male disconnected youth are more likely to have immigrant background, and – probably related – to have parents with compulsory school as the highest educational attainment. This then suggests that being a more disadvantaged male disconnected youth seems to be linked to compromised caring environment in childhood, while being a less disadvantaged disconnected youth is linked to fewer palpable resources. These differences reappear in the last part of the analyses where focus is on later disconnected youth status. As we may expect, the less disadvantaged disconnected youth are also more likely to (re-)connect with the labor market and the educational system later on, and less likely to experience persistent disconnection. However, the worrying finding of these analyses is that even though foreseeable shares of disconnected youth experience persistent disconnection, a substantial share – even among the less disadvantaged youth – seems to cycle in and out of the category. This means that the first indication that someone is a disconnected youth is also an indication that the person will experience labor market instability and scattered contact with the educational system in many years to come.

For women, the cluster analysis suggests three underlying clusters. The first cluster consists of female disconnected youth that primarily suffer from mental health problems, the second struggle with all problems, however with drug addictions in particular, and the dominant problem among the last cluster is early parenthood, but we also see indications of problems pertaining to mental health and drug use. Of these, the first cluster has had the most favorable upbringing, with the most resources, and the third cluster – those who have become parents early in life – has had the least favorable upbringing. However, what is special about the second, most problem ridden cluster, is the overrepresentation of maternal mental health problems – similar to what we saw for males, where also the most problem ridden cluster was more likely to have been exposed to this risk factor.
An interesting finding appears when we look at future outcomes for these groups, particularly with regards to the younger age group; while the first and the third cluster differed widely on background characteristics, their likelihood of reconnection follows almost the exact same trajectory over the years. As we may expect, the second cluster, consisting of female disconnected with the most problems, are less likely to (re-)connect; while their risk of persistent disconnection is not widely different from that of the two other clusters, they are significantly more likely to cycle in and out of the disconnected youth group. For the older group of female disconnected youth, we see larger differences between the first and the third cluster, where the first cluster is both less likely to experience recurring as well as persistent disconnection. We can only speculate what causes this difference between age groups, however a possible interpretation is that it is the more favorable background characteristics of the first cluster which kick in, in the late twenties and secures the better outcomes.

But summing up across clusters, gender and age, it seems that the risk factor most predictive of substantial concurrent problems and limited probabilities of reconnection is maternal mental health problems, over and above risk factors related to more palpable family resources such as income and education.

Limitations

As described, the analyses in this paper exploit the rich Danish registers on everything from crime and drug treatment, to education and income. And despite the great potential of the registers there are some general limitations that cannot be disregarded. Most evident, and quite importantly, in some regards, the registers merely reflect symptoms of a specific behavior, rather than evidence of that behavior; in this study, I identify drug use through drug related crime and drug treatment indicators, including hospitalization for problems pertaining to drug use. However, this is likely to be an underestimate of the actual problem given that we may expect a significant share of drug users to never show up in either systems – and given that specific types of drug users are systematically more or less likely to show up in the system. This important limitation should always be kept in mind when interpreting the results presented here.

Discussion

What then to make of these results? The question raised in the introduction is whether we should consider the group of disconnected youth as one, or as a heterogeneous group with different problems, background characteristics and prospects. The answer to the question as presented in this paper demonstrates significant heterogeneity in the group of disconnected youth on all relevant parameters, and demonstrates how a substantial share of the disconnected youth face heavy problems, and cannot be expected to just (re-)connect on their own. In particular the marked difference between the two clusters of male disconnected youth with regards to problem load and how they fare in later life is interesting; one group seems to consist in the typical risk seeking adolescent boy which is likely to age out of this behavior and for whom there is little need for
concern. However the other group seems to face substantial problems that we cannot just expect will disappear on their own. Their particular disadvantage lies in their overarching mental health problems which are likely to be life-long (to a larger or smaller extent), and that we know are linked to other types of problems that are likely to hinder reconnection, such as e.g. loneliness (see Lasgaard et al, 2010; Lasgaard et al, 2011).

When targeting disconnected youth this should be taken into consideration, such that programs are differentiated according to problem load and prospects.
However the other group seems to face substantial problems that we cannot just expect will disappear on their own. Their particular disadvantage lies in their overarching mental health problems which are likely to be life-long (to a larger or smaller extent), and that we know are linked to other types of problems that are likely to hinder reconnection, such as e.g. loneliness (see Lasgaard et al, 2010; Lasgaard et al, 2011). When targeting disconnected youth this should be taken into consideration, such that programs are differentiated according to problem load and prospects.

References

Andersen, Lars H., Anker, Anne Sofie Tegner & Andersen, Signe Hald: A Formal Decomposition of Declining Youth Crime in Denmark. Forthcoming in Demographic Research


Appendix

Figure A1: problem load among all youth, 2000-2013

A1a: Men, 18-24

A1b: Men, 25-29

A1c: Women, 18-24

A1d: Women, 25-29

Source: Own calculations based on data from Statistics Denmark
Figure A2: The two-cluster solution, women aged 25-29

![Graph showing the two-cluster solution for women aged 25-29.](image)

Source: Own calculations based on data from Statistics Denmark

Table A1: Respecification

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Two-cluster solution, years 2010-2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother with mental health problems</td>
<td>0.118***</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Mother’s # of children</td>
<td>-0.001</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Dad incarcerated (from child age 9-18)</td>
<td>0.040**</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Parents live together</td>
<td>-0.090***</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Parents no more than compulsory school</td>
<td>0.009</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Both parents in lowest income decile</td>
<td>0.020</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Both parents unemployed for more than 70 percent of the year</td>
<td>0.018</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Immigrant background</td>
<td>-0.040***</td>
<td>(0.01)</td>
</tr>
<tr>
<td>No parent information in registers</td>
<td>-0.092***</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Year as opportunity youth</td>
<td>-0.013***</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Intercept</td>
<td>25.690***</td>
<td>(3.39)</td>
</tr>
</tbody>
</table>

*: p<0.05; **: p<0.01; ***: p<0.001

Source: Own calculations based on data from Statistics Denmark
b. Three-cluster solution, years 2000-2009

<table>
<thead>
<tr>
<th>1. cluster</th>
<th>2. cluster</th>
<th>3. cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother with mental health problems</td>
<td>-0.041***</td>
<td>0.009</td>
</tr>
<tr>
<td>Mother’s # of children</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>Dad incarcerated (from child age 9-18)</td>
<td>-0.071***</td>
<td>0.008</td>
</tr>
<tr>
<td>Parents live together</td>
<td>0.117***</td>
<td>0.005</td>
</tr>
<tr>
<td>Parents no more than compulsory school</td>
<td>-0.076***</td>
<td>0.001</td>
</tr>
<tr>
<td>Both parents in lowest income decile</td>
<td>-0.065***</td>
<td>0.007</td>
</tr>
<tr>
<td>Both parents unemployed for more than 70 percent of the year</td>
<td>-0.010</td>
<td>0.007</td>
</tr>
<tr>
<td>Immigrant background</td>
<td>-0.050</td>
<td>0.009</td>
</tr>
<tr>
<td>No parent information in registers</td>
<td>0.062***</td>
<td>0.010</td>
</tr>
<tr>
<td>Year as disconnected youth</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Source: Own calculations based on data from Statistics Denmark

Table A2: Correlation between background characteristics and the risk of being a disconnected youth.

<table>
<thead>
<tr>
<th></th>
<th>Men, 18-24 years</th>
<th>Men, 25-29 years</th>
<th>Women, 18-24 years</th>
<th>Women, 25-29 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother with mental health problems</td>
<td>0.028***</td>
<td>(0.00)</td>
<td>0.039***</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Mother's # of children</td>
<td>-0.003***</td>
<td>(0.00)</td>
<td>-0.000</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Dad incarcerated (from child age 9-18)</td>
<td>0.033***</td>
<td>(0.00)</td>
<td>0.044***</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Parents live together</td>
<td>-0.023***</td>
<td>(0.00)</td>
<td>-0.038***</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Parents no more than compulsory school</td>
<td>0.037***</td>
<td>(0.00)</td>
<td>0.039***</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Both parents in lowest income decile</td>
<td>0.057***</td>
<td>(0.00)</td>
<td>0.085***</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Both parents unemployed for more than 70 percent of the year</td>
<td>0.020***</td>
<td>(0.00)</td>
<td>0.024***</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Immigrant background</td>
<td>0.002*</td>
<td>(0.00)</td>
<td>0.031***</td>
<td>(0.00)</td>
</tr>
<tr>
<td>No parent information in registers</td>
<td>-0.005***</td>
<td>(0.00)</td>
<td>-0.032***</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Year as disconnected youth</td>
<td>0.001***</td>
<td>(0.00)</td>
<td>0.001***</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.205***</td>
<td>(0.08)</td>
<td>-2.665***</td>
<td>(0.12)</td>
</tr>
</tbody>
</table>

*: p<0.05; **: p<0.01; ***: p<0.001

Source: Own calculations based on data from Statistics Denmark