ZIYAD AL BALUSHI
CANDIDATE NUMBER: 1014
SUPERVISOR: YULIIA DUDAREVA

Sibship and Beyond: Unravelling the Sibship Effect on Life's Winding Road

An Empirical Analysis Using the National Longitudinal Survey of Youth exploring the effect of a sibling's gender on educational and career choices

Master thesis, 2024

Master of Science in Business Administration

University of Stavanger Business School

Specialization: Economics

University of Stavanger

Contents

Preface

Four years ago, I was rejected by several Norwegian universities when I applied to study for a Master's degree in Psychology, a field I hadn't even pursued for my Bachelor's degree. Fast forward to today, the 10th of June 2024, and I am submitting my thesis in Economics, specifically in a topic that saw me merge psychology, behaviour, statistics, and outright pure fascination. It has been a tremendous journey of learning and a great opportunity to find discourse with like-minded individuals.

I would like to thank my advisor Yulia Dudareva who acknowledged my constant head scratching at our meetings and looks of confusion with great patience, support, and well-tailored advice. Speciale thanks go to my family back home in Oman in which our calls have kept me sane within the walls of Jernaldervien 57. Lastly, my sweet companion Maryam, you have been a true partner throughout this period, a rock upon I found periods of comfort and support when I most needed it. Thank you.

"The master-economist must possess a rare combination of gifts. He must be mathematician, historian, statesman, philosopher - in some degree. He must understand symbols and speak in words. He must contemplate the particular in terms of the general and touch abstract and concrete in the same flight of thought. He must study the present in the light of the past for the purposes of the future. No part of man's nature or his institutions must lie entirely outside his regard." – Keynes (1924)

Abstract

Understanding the links between future life outcomes and early life subtleties always presented an opportunity to discover causes and effects, one of these subtleties is family dynamics, which is a vessel upon which one's career and educational choices are moulded, and a mechanism that alters gender role attitudes. This thesis explores the effects of having a younger sister as opposed to a younger brother on the elder siblings educational and occupational choices through the National Longitudinal Surveys of Youth (NLSY79) and its subsequent Child and Young Adult datasets. The study applies linear regression models, a year fixed effects model, and an interaction model to explore such effect. The analysis reveals that having a younger sister does not significantly affect educational attainment or occupational choices. Instead, socio-economic factors such as household income and maternal education play a more pivotal role. On the other hand, elder siblings with only sisters are slightly less likely to have no degree and more likely to complete high school, both with a significance level of 10%, nonetheless, a significant gender imbalance within the groups in focus suggests a potential gender selection bias that could significantly undervalue the validity of such result. While the results for elder sibling males and females entering gender dominated jobs suggest potential gender specific effects, these results were not statistically significant. The study also explores income and job stability, finding that having a younger sister does not significantly impact earnings or joblessness. Interestingly, the effect of having a younger sister and opting into a female dominated job indicate a potential increase in compensation received when entering a female dominated job, possibly mediated through enhanced social skills, albeit these results were not significant as well.

1. Introduction

Family environment is identified as the vessel of growth for individual behaviour. An important feature of this environment is sibling composition. This thesis investigates the implications of this one factor. More specifically, the effect of the gender composition of siblings on first born individual's occupational and educational choices. The research will try to pursue a deeper understanding regarding family dynamics influencing shaped individual personality and behaviour. The exploration of career choices and their determinants presents a complex interplay of individual aspirations, societal norms, and familial influences. In general, within those influences, the role of sibling gender configuration on a career, and in the framework of opting into an artistic and creative field or occupation dominated by one or the other gender, has relatively remained unexplored within economic research. Understanding the effect of sibship gender composition can also have broader societal implications. It may influence the ways in which families are counselled and assisted, it may affect the way educational and developmental challenges are addressed in schools, and how community programs are designed to support different individuals based on their family dynamics. Additionally, it could shed some light on the role of gender dynamics beyond the family in occupational choices. For example, if it is the case that close contact with a female can alter perceptions of gender roles, then we could create tools to allow for more interactions of this type in an effort to shift gender role attitudes and impact career paths.

Since educational attainment is very closely related to career choices and since higher education usually broadens career options and life occupational preferences, choice between scientific and artistic educational fields sets the stage for later career endeavours. Thus, exploring educational attainment in the context of sibling gender composition as a prelude to career choices can provide further insights on the dynamics governing later life career outcomes.

Factors that influence career choices irrespective of gender include the income levels of the family, the job the parents have, and socio-economic factors. (Ermisch and Francesconi, 2001, Aakvik et al., 2005, Bandura et al., 2001, and Berry 1994). Gender role attitudes and peer pressure are drivers of decisions towards certain career paths. (Price, 2008, and Abe and Chikoko, 2020). Family structure, for example having a daughter, may lead to such attitudes, which in turn influences career choices. In the past, researchers have examined a number of other issues involved in educational and career pathways viewed as a function of sibling influence. Peter et al. (2018) found that besides earnings and marriage prospects, sibling gender also affects educational attainment and choices of family formation, which implies that the

gender of one's sibling plays a decisive role in life choices. Similarly, Shahbazian (2021) conducted a similar investigation regarding sibling gender composition on STEM field selection. The research reveals that sibling relationships, in particular, the gender composition of siblings, are a significant predictor for the choice of a career in a gender typical and atypical field. On the other hand, Hardie (2015) examined the predictors of young men's aspirations toward traditionally female dominated occupations, noting that roles of gender role attitudes and peer influences are significant factors. Lease (2003) explored factors influencing men to enter female dominated occupations, the study found variables such as liberal social attitudes and socioeconomic status as predictors in steering men towards female dominated jobs.

Furthermore, the influence of having a girl in the family on later life trajectories has been a subject of interest in different studies, albeit limited. For instance, it has been noticed that the presence of a sister is positively correlated with family relationship; conflict among family members is reduced, cohesion along with expressiveness has increased within a family (Cassidy et al., 2014). With such findings and the context and dynamics of family structures kept in perspective, a question arises: Would the presence of a sister within the family alter an individual's path into creative career and educational choices, and make them more open towards opting into female dominated jobs? The presence of a younger sister can influence older siblings' perceptions and attitudes towards gender roles, potentially making them more open to non-traditional career paths. Having a daughter can also alter parental attitudes, which in turn can affect the career choices of older siblings by fostering more egalitarian views on gender roles.

This thesis aims to explore the influence of familial structures on individual conduct and life choices. Namely it intends to answer the following question: **Does having a first younger sister affect the career choice of an older sibling, in particular the probability of choosing an artistic vs scientific career and the probability of choosing a female dominated job?** By integrating the concept of employment choices and educational outcomes as an indicator of artistic vs scientific pathways, this study seeks to explore an aspect of how family dynamics, in the context of sibship gender composition contributes to the formation of educational and occupational choices.

The thesis will harness the comprehensive longitudinal data provided by the National Longitudinal Survey of Youth 1979 (NLSY79) and its subsequent NLSY79 Child and Young Adult datasets which follows the biological children of the women from the original cohort (NLSY79), offering valuable intergenerational insights. By running several linear regressions

and a year fixed effect model we found that having a younger sister does not significantly impact educational attainment or career choices, with socio-economic factors like household income and maternal education being more influential. Elder siblings with only sisters are slightly less likely to have no degree and more likely to complete high school, but a significant gender imbalance suggests potential selection bias. While having a younger sister might indicate increased compensation in female dominated jobs due to enhanced social skills, these results were not statistically significant. Section 2 explains the theoretical background and the literature review behind the subject. Section 3 delves into the empirical strategy in more detail, while section 4 explains the data sources, outcome variables selection and construction. Section 5 explores the descriptive statistics of the data and the quantitative results, and section 6 discusses these findings in more detail. Section 7 includes concluding thoughts on the whole thesis.

2. Theoretical Background

As the traditional measures of human capital which include education and experience fall short of fully accounting for the discrepancy in life outcomes, economic analysts and social scientists have in recent times cast more emphasis on the influence of family dynamics on opportunities and constraints at various life stages; that is, in essence, ways by which familial structures, relationships, and parental behaviours affect or form these opportunities and constraints at different stages of life. Building on this perspective, this thesis ventures into a less explored territory by investigating the influence of a sibling's gender on labour market outcomes. It is an attempt to discover how labour market outcomes vary by the gender of a sibling and seek to know what degree gender composition in a family unit would affect a person's educational attainment, career prospects and subsequent wages, thereby trying to find an assessment of these differentials in the formation of labour market gender gap. The relationship between familial structures and individual life choices has long been a subject of scholarly interest, particularly in the context of career development.

This section of the thesis presents a literature review that would help in the exploration of the various factors that govern the career choice with regard to the role of family structure and sibling gender on the propensity for some educational outcomes and the subsequent selection of artistic and creative career paths. It also serves to bring out occupational choice determinants beyond family dynamics and labour market gender gaps as a prelude to the potential implications of the results within the study. We as well delve into the potential theoretical mechanisms underpinning occupational choices in context of sibling gender composition. By examining existing research within this domain, this review aims to understand the theoretical underpinnings that might explain later life career and educational outcomes.

2.1 Family Dynamics on Life Outcomes

The influence of familial dynamics on individual life is a broad topic that cuts across a number of dimensions covering the impact of birth order, sibling gender, household income, and general family structure on choices or decisions affecting individual career paths. A study by Peter et al. (2018) suggests that the presence and gender of siblings can significantly influence an individual's economic outcomes and life decisions. Within the study it is found that sibling gender may have some impacts on career choices by affecting the family environment and the

allocation of resources, which may consequently affect the likelihood of pursuing certain career paths, including artistic and creative careers.

Furthermore, the study by Cassidy et al. (2014) explores the relationship between family structure and psychological health in young adults. Their findings suggests that sibling relationships and parental relationships quality can have major impacts on the psychological well-being of a young adult, as such this might impact some career choices or desires. Specifically, the psychological environment and support provided by one's family can either spur or discourage pursuit of a certain career field, and educational path as personal well-being and self-expression are assumed to be leading roles.

On the other hand, research shows that parental educational attainment significantly affects the educational achievements of their children, where the former is a strong predictor of the latter (Ermisch and Francesconi, 2001, and Aakvik et al., 2005). Moreover, family structure, such as single-father and stepparent families, are more likely to be associated with lower attainments of the child's education after controlling for socio-economic status and other factors (Martin, 2012). An interactive relationship exists between educational attainment and career choices whereby educational attainment influenced by factors such as family support, career choice, and career anxiety give a grip of what type of career selections high school students opt for (Yilmaz and Gündüz, 2018). Family dynamics have been noted to influence adolescents' educational attainment; majorly, parenting styles and family involvement have been reported to account for the gender differences that are notable in adolescents' educational attainment (Hindin, 2005).

On the matter of birth order within the family Lehmann et al. (2016) offer an examination regarding differential effects of birth order on cognitive and non-cognitive outcomes. Later-born children score lower on cognitive tests than their older siblings from as early as age one. These differences only widen up to entry into school and thereafter remain constant. The study found that such differences are substantially explained by variations in parental behaviour including cognitive stimulation provided for the offspring. The study goes further from its initial findings to show that birth order affects some other aspects of life outcomes, namely, crime, and teenage pregnancy. It reveals that the relationship between birth order and educational attainment is significant. Those born later are less likely to finish high school and they tend to complete fewer years of education. For instance, second-borns are about 3 percentage points less likely to finish high school and they complete roughly half a year less of education than first-borns. This trend only worsens with each higher order birth; later-born siblings are 10 and

9

18 percentage points less likely to complete high school, respectively, and complete about 1 to 1.5 fewer years of education. Large family sizes also reduce the probability of a child's school enrolment by diluting parental resources. The reason is that siblings from large families tend to receive less parental attention and financial support, which undermines their educational attainment (Ngware et al., 2009). Such effects may naturally feed through into later in life decisions concerning occupational choices, thus, the present investigation. In that respect, this thesis examines the impact of the second child's gender on firstborns, additionally, we control for the number of siblings in the family.

Naturally we would assume that parent's income is a resource to be utilized for their children, therefore, affecting later life educational outcomes for the children and, in part, affecting choices of careers. Family income and parental education link to educational achievements, which may later shape the career choices one makes. This could be because children from better-off families have more resources or opportunities to obtain education and training in scientific fields that often require advanced degrees and specialized knowledge (Ermisch and Francesconi, 2001). Moreover, Whiston and Keller (2004) highlight that family structure factors such as parents' occupations and income, can influence career development across the lifespan of an individual. The study suggests that family income may play a role in shaping individuals' career constructs, potentially impacting their choices between artistic and scientific paths. Higher income provides greater latitude to the choice of a child to make a career in the arts or science depending on his interest and capacity. Therefore, for a more significant isolation of the young sister effect, we introduce household income as a control variable.

Exploring the gender composition of siblings and its effects on occupational and educational choices Anelli and Peri (2014) examine the relationship between the gender of siblings and the choice of college major. Their study reveals that within a mixed gender siblings composition individuals tend to choose college majors following a stereotypical gender specialization. Specifically, males are more are likely to pick highly male dominated majors like engineering; women are more likely to opt for highly female dominated majors like humanities. As a result, gender composition of siblings may indeed shape individuals' educational and occupational preferences. That said, the next section is dedicated to the impact of sisterhood and brotherhood.

Integrating these preliminary findings, it becomes clear that the gender of a sibling and the broader family dynamics not only influence economic and psychological outcomes but also potentially shape the career trajectories of individuals.

2.2 Sisterhood Effect

In the context of exploring how familial structures, particularly sibship, influence individual conduct and life choices, it is noteworthy that women tend to exhibit greater levels of social support, optimism, and personal growth, alongside a more collective approach to systems maintenance, compared to men, who often demonstrate a higher degree of locus of control (Cassidy et al., 2014). This distinction in psychological and behavioural traits between genders suggests that the presence of sisters within the family unit could enrich the familial environment with enhanced social support and optimism. Such an environment may foster personal growth and a collaborative approach to managing family dynamics, potentially shaping the career paths and life choices of siblings. These insights align with the question explored by indicating that the gender composition of siblings, particularly the influence of sisterhood, might play a significant role in determining later life choices, including the likelihood of pursuing artistic and creative careers, and getting into female dominated jobs. This section conducts an in-depth analysis of the intricate dynamics inherent in sibling relationships, with a special emphasis on the distinctive influence that sisters exert on individual growth and well-being. it investigates the contributions of female siblings to psychological resilience, social support, and conflict resolution within the familial context. The role of sisterhood in influencing life decisions, inclusive of career trajectories, is examined, with a focus on the creation of a supportive emotional environment and the reduction of psychological distress. This investigation is situated within the wider discourse on family influences, accentuating the specific manners in which bonds between siblings can shape and enhance life paths and decisions.

Cassidy et al. (2014) eludes on the importance of the gender of a sibling on certain individual characteristics and family dynamics, namely that within the study a sample of 708 young adults (294 males and 414 females) aged between 18 - 21 years were surveyed based on three distinct tests, The Locus of Control Behavioural Scale which measures ones sense of control in their lives, The Life Orientation Test (LOT) which measures optimism, and The Family Environment Scale which measures how family members perceive the family and how each member's behaviour affects the family unit during a time of crisis or transition. Participants with the least psychological distress were boys and girls with only sisters. It appears that the gender of the siblings, rather than the participants themselves, has a greater impact on levels of psychological distress. Female siblings have a positive effect, reducing distress, while male siblings have a negative effect, increasing distress (Cassidy et al., 2014). Within this framework we can expect later life impacts on occupational choices and endeavours.

Unlike men, women display a definite trend as far as sibling influence is concerned. Having a sister as opposed to a brother leads to reduced years of education (Butcher and Case, 1994 as cited in Peter et al., 2018). Income on the other hand also induces a negative relationship but fails to be significant in this research. Surprisingly, sibling gender does not affect family formation probability but family formation age. Women with sisters have their first child earlier. As a consequence, having a sister is associated with earlier childbearing and lower educational attainment for women (Peter et al., 2018). Having a sister also seems to moderate unemployment slightly, probably by the use of common job search networks, pointing towards a supportive role of female siblings. The theory derived from this study is that the noted differences in family formation may be a consequence or result of differential parental treatment of sons and, perhaps, competition among sisters. What all this suggests is that females might be more supportive of siblings than males, through social support networks perhaps, which may mediate outcomes that are observed.

Furthermore, the presence of sisters tends to steer females towards fields of study that are not centred around science (Tao and Cheng, 2022). Research by Oguzoglu and Ozbeklik (2016) as cited in Tao and Cheng (2022) indicates that daughters of fathers employed in STEM fields are less inclined to pursue STEM majors in college if they have one or more brothers. Similarly, analysed data from twins in North Carolina and found that twins of opposite sexes were more inclined to select majors typical for their gender compared to twins of the same sex. This suggests that sisters play a significant role in guiding their siblings' educational and career choices away from traditionally male-dominated fields (Shi, 2018 as cited in Tao and Cheng, 2022). Additionally, individuals pursuing non-science fields reported a greater influence from sisters compared to brothers (Tao and Cheng 2022), similarly and in terms of educational decisions female participants noted that their sisters were more influential than their brothers. A study by Cools and Patacchini (2017) indicated that males with sisters tend to choose more traditionally male disciplines than those with brothers. However, the study does not find any influence of sibling gender on female major choice. Shahbazian (2021) found that younger siblings are more likely to major in STEM field if their older sibling has ever attended or is currently attending a STEM program. Also, there is a clear gender difference in the choice of a STEM field among younger siblings, namely that females are more likely to choose a STEM field if they have an older sister who has attended a STEM program, compared to those that have an older brother in a similar program. The corresponding results are not found for males.

Such a strong male domination at the tertiary level of STEM fields, therefore, implies the presence of a same-sex role model for young girls in gender-atypical educational choices.

2.3 Brotherhood Effect

While the prior section investigated the effect of a sisters on one's life choices and characterises this section delves on the effect of brothers.

Existing literature also indicates that male siblings tend to increase conflict and decrease cohesion within sibling relationships (Weiss et al., 2001 as cited in Cassidy et al., 2014). Furthermore, the study by Cassidy et al. 2014 and as mentioned previously revealed that participants with brothers experienced the highest levels of psychological distress. When considering the gender of the participant, it was found that boys with brothers were the most distressed. Girls with brothers ranked second in terms of psychological distress, followed closely by both boys and girls with a mix of brothers and sisters. Given that sibling support is a crucial source of social support during family issues (Dunn, 1996 as cited in Cassidy et al., 2014), it logically follows that female siblings would provide more support than male siblings. This support is likely the mediating factor through which the gender of siblings influences health outcomes, and other individual characteristics.

Previous studies show that the gender of siblings doesn't affect men's educational attainment (Peter et al., 2018), albeit in a study by Tao and Cheng (2022) male participants indicated that their brothers had a more significant influence on their educational decisions than their sisters. However, income seems to be impacted, the findings reveal that men with brothers, compared to those with sisters, tend to have higher earnings and are more likely to form families (through marriage and having children), thus as a result have more children (Peter et al., 2018). The effects observed for women are in a similar direction but are less pronounced and not as consistently robust across measures. The study suggests that the increased earnings among men with brothers might stem from competitive dynamics leading them to select into higher-paying occupations, hinting at the influence of non-cognitive factors.

In the realm of occupational choices, the likelihood of daughters, whose fathers are engaged in STEM professions, opting for a STEM major in college diminishes when they have brothers (Oguzoglu and Ozbeklik, 2016 as cited in Tao and Cheng, 2022). Families with siblings of both genders, children are more likely to follow educational paths that conform to gender stereotypes, as opposed to families with siblings of the same gender (Anelli and Peri, 2015 as

cited in Tao and Cheng, 2022). Furthermore, among participants studying science-related subjects, brothers played a more significant role in shaping educational choices than sisters did (Tao and Cheng 2022).

2.4 Occupational Choice Determinants Byond Family Composition

Beyond the interplay of family dynamics and later life choices comes a range of self-determining factors that might influence career and educational choices. These factors might have a greater effect on one's educational and career choices away from the family dynamics narrative.

According to Bandura et al. (2001) self-efficacy is one such factor. Children's perceived self-efficacy, instead of actual academic achievement, appears to be more crucial. The findings of the study with 272 children tested a structural model for the understanding of social cognitive influences over career aspirations and trajectories. The research found that familial socioeconomic status (SES) impacts children's career paths indirectly by affecting parents' perceived efficacy and academic aspirations. These parental influences shape children's career efficacy and choices through their perceived academic, social, and self-regulatory efficacy. Perceived occupational self-efficacy turned out to be a reliable determinant of the career activities children would consider and reject, indicating that children's beliefs about their abilities, rather than their actual academic achievements, are key determinants of their occupational choices. Gender differences were noted, with perceived occupational self-efficacy predicting the traditionality of career choices. Thus, an element of subjectivity and self-selection is a factor in career choices. Subjective expectations and non-financial preferences also affect career choices, as discussed by Arcidiacono et al. (2014). Different views of individuals concerning expectations of wages paid and their interest developed towards different occupations provide a view and, in fact help explain what motivates some individuals to choose some careers. Non-monetary factors that associate with personal interests and values play a big role in guiding people towards or away from specific career paths.

Furthermore, Occupational choice models, ones proposed by Berry (1994), help connect subjective data on earnings and choice probabilities with individuals' preferences for various occupations. These models consider the trade-offs individuals make between different job characteristics and non-pecuniary aspects when selecting a career. Factors like job differentiation and individual preferences contribute to the decision-making process. Applying

this to the context of sibling dynamics, having a younger sister might predispose an individual to choose a female dominated job. The close sibling relationship could help develop a skillset suited to such roles, potentially leading to higher compensation due to a better fit with job demands and interpersonal dynamics typical in female dominated fields. This aligns with Berry's model by showing how individual preferences and skills shaped by family dynamics influence occupational choices and earnings potential.

Naturally, uncertainty and experience also play their role in influencing career choices, as explained, as explored by Connelly (1989). While selecting any career, workers need to consider the future earnings potential in every occupation. The factors such as stability of the job, growth and opportunities for advancement influence the decisions of individuals respect to career paths. Moreover, the career choices can be affected by internal and external hindrances, cognitive styles, and career development variables, as studied by Creed et al. (2003). As per various researchers, personality variables, career adaptability, and career decision making self-efficacy are some other aspects of influencing the career choices of individuals (Bi, 2023; Pang et al., 2021; Chui et al., 2020). Additionally, from the study, factors like resilience, adaptability, and decision-making skills affect how individuals explore career options, make choices, and navigate their professional journeys.

In conclusion, a combination of personal attributes, environmental factors, cognitive processes, and individual preferences shapes individuals' occupational choices. By considering these major determinants, the study notes that beyond family dynamics effects in general and sibship gender in particular there are complex other factors in play to determine one's career and educational choices.

2.5 Labor Market Gender Gaps

One cannot delve into occupational choices without understanding how gender gaps play a role in the labour market. As our study delves into the specific effect of sibling gender as opposed to the gender of the individuals, it is crucial to understand the gender dynamics behind the current market gender gaps, especially in the U.S as our thesis explores a dataset that is based in the U.S.

Research has shown that the gender division of jobs is articulated by societal expectations, thus becoming a major factor that even perpetuates unequal conditions for males and females at work. This translates even to variable pay, upward mobility on the career ladder, to overall job

outcomes, which contribute as factors to the ideology surrounding continuous pay gap between genders (Palffy et al., 2023). According to Palffy et al. (2023), young men in areas with strong traditional views on gender are more inclined to choose careers that are typically seen as 'men's work'. This pattern isn't as strong for young women, which suggests that getting men into jobs usually done by women, like those in healthcare, might do more to close the gender gap in workplaces than just encouraging women to enter STEM fields. It's also interesting to note that having a younger sister might make a boy more open to non-traditional job choices for his gender. This could help even out the number of men and women in different professions, as men who grew up with sisters may gain skills and interests that fit well with jobs mostly done by women, leading to a more balanced workforce.

The choices we make in education, shaped by what society expects of us, can lead to men and women having different experiences in the job market, as Gayle and Golan (2011) have pointed out. The level of education we reach, our chosen field of study, and the skills we pick up along the way can all steer us towards certain careers and affect how much we might earn, which in turn plays into the differences in wages between genders.

To tackle these differences in the job market, we need to look closely at and question the social norms, biases, and hurdles that shape how people decide on their careers and what chances they have in the workplace. This study is looking into a new aspect: how having brothers or sisters might change how gender gaps show up in the job market, and what that means for making things more equal. For instance, having a sister might mean a guy ends up earning more in a job field where there are mostly women or just has a better shot at getting such a job. Digging into this could give us new insights into how to achieve a more balanced representation of genders in the workplace. So, policies that encourage kids to dream big about their careers from an early age, considering their family setup, might be a smart move. This could include educational activities that build skills without gender bias, like playdates with kids of the opposite sex, and career advice that takes into account how family can shape our choices.

2.6 Theoretical Mechanisms

This thesis investigates two main themes, educational and occupational outcomes. Educational outcomes arise as a prelude to the main part of the analysis which is occupational choices. It aims to investigate the effect of having a younger first sibling sister as opposed to a brother on certain educational outcomes such as having no degree, completing high school, completing

college, completing an advanced degree, having a Bachelor of Art (BA) or Bachelor of Science (BS). With BA and BS acting as an identifier for creative or scientific educational endeavours. Within this section we aim to investigate the potential theoretical mechanisms that might explain the career choices in the context of having a first younger sister. Please see section 5.1 on the results of this exploration within the dataset.

The segment of the study concerning occupational outcomes delves into the narrative of how the presence of a younger sister may influence certain life choices. The Compensation Differentials Model, thoroughly examined by Rosen (1986), offers a theoretical framework for comprehending how wage differentials serve as compensation for various job characteristics. This model suggests that wage disparities among occupations are not merely arbitrary but serve to balance out non-monetary aspects of employment, including the negative attributes associated with specific types of work. This conceptual framework is particularly pertinent to our investigation, which explores the potential impact of sibling gender on occupational preferences.

Research suggests that individuals often exhibit a preference for jobs dominated by their own gender, which could stem from social and cultural norms that define gender roles within occupational contexts (Charles and Grusky, 2004). Thus, we can assume that getting into an opposite sex job may provide an extra amenity to compensate for entering an unfavourable work situation, that extra amenity is represented in wages. Therefore, the basic assumption suggests that an individual will be compensated higher if entering an opposite sex job. Please see figure 5 for the results exploring this assumption.

On the other hand, the Roy model (Roy, 1951) illustrates that people choose careers where their skills yield the highest returns, leading to occupational sorting. This self-selection mechanism helps explain the distribution of earnings and occupational choices, showing that differences in individual abilities and preferences significantly influence labour market outcomes. Within this framework it is reasonable to assume that for instance, those with a younger sisters might be more attuned to the challenges faced by women and thus more likely to enter female dominated professions if these are perceived to offer better work-life balance and a higher compensation relative to the skills they have developed as opposed to those who have a younger brother.

Thus, given this framework and theoretical mechanisms governing occupational choices the hypothesis behind the occupational choices is formalized as such:

- Having a first younger sister, an elder sibling is:

A. More likely to choose female dominated job as the individual is more familiarized with the potential working environment of a female dominated job.

B. Less likely to get into a STEM related job as these types of jobs are more male dominated (Pew Research Centre, 2022). Additionally, STEM related jobs selection works as a variable for a scientific career path.

C. Less likely to get fired as having a younger sister might act as a mediator towards higher empathy and less risky behaviour (See section 2.2).

D. More likely to receive higher compensation when entering a female dominated job as the individual is more familiarized with the behaviours associated with female dominated jobs, thus has the skills set that pushes the higher compensation assumption on the framework of the Roy self-selection model. This logic takes the form:

$$Into\ a\ Female\ Dominated\ Job\ with\ a\ Younger\ Brother: U_{yb} = W_{yb}$$

$$Into\ a\ Female\ Dominated\ Job\ with\ a\ Younger\ Sister: U_{ys} = W_{ys} + \sigma$$

$$Where\ \sigma > 0\ Higher\ Compensation$$

$$Indifferent\ if\ W_{ys} = W_{ys} + \sigma - \triangle\ and$$

$$\triangle = W_{ys} - (W_{yb} - \sigma)$$

Where U_{ys} is the utility received getting into a female dominated job when having a younger sister, U_{yb} is the utility received in a female dominated job while having a younger brother, W_{ys} denotes the wages of younger sister group getting into a female dominated job, W_{yb} is the wages of the female dominated job while having a younger brother, and σ denotes our measure of interest. Where we expect to see an extra compensation if its above zero. Through interacting adjusted income received for the reported job with the likelihood of getting into a female dominated job we can measure the effect of the assumed higher compensation when entering a female dominated job while having a younger sister (See Table 11 for the results of this interaction).

3. Empirical Strategy

This section outlines the empirical strategy utilized to examine the influence of having a younger sister on educational attainment and career choices. We conceptually frame our models

to compare having a first younger sister as opposed to a younger brother on the outcome variables. To capture the effects, we rely on the assumption of no self-selection which is validated in section 5.2. Having a first younger sister is our main treatment group across the models, while having a first younger brother would be the control group.

3.1 Educational Outcomes

Central to our analysis is the construction of linear regression models that isolate the effect of having younger sister as opposed to a younger brother on various educational outcomes (See section 5.3), controlling for a set of pre-treatment households, parental, and individual characteristics. The linear model takes the form:

Model 1 Main Educational Outcomes Regression Model

$$EducationalOutcome_i$$
$$= \beta 0 + \boldsymbol{\beta 1 YoungerSister_i} + \beta 2 AdjustedIncome_{10ki}$$
$$+ \beta 3 MotherEmployed_i + \beta 4 MotherOutOfLaborForce_i$$
$$+ \beta 5 LivingWithMother_i + \beta 6 FathersPresence_i$$
$$+ \beta 7 MothersEducation_i + \beta 8 Hispanic_i + \beta 9 Black_i + \beta 10 Male_i$$
$$+ \beta 11 TotalSiblings_i + \epsilon_i$$

In examining the relationship between sibling gender and educational attainment, our primary variable of interest is the presence of younger sister, $\beta 1$. This is denoted by a binary indicator that takes the value of 1 if the elder sibling has a first younger sister and 0 for a first younger brother. The model also includes the household income ($\beta 2$) around either one or two years before the birth of the younger siblings, the income is adjusted for 2020 dollars; if the input for the year before the younger sibling is missing then we opt for the year after, maintaining a two-round interview timeframe. This ensures limiting the missing variables for this variable. The employment status of the mother one or two years before the birth of the younger sibling is split into two distinct variables: one indicating whether the mother was employed ($\beta 3$), capturing economic activity, and another indicating whether the mother was out of the labour force ($\beta 4$), reflecting non-economic engagement or potential caregiving responsibilities, where is the reference group are those mothers that were unemployed around the younger sibling's birth. The familial environment is further characterized by variables indicating whether the elder sibling lived with the mother ($\beta 5$) one year before the birth of the younger sibling or two years before the birth of the younger sibling, and the presence of the father in the years surrounding

the birth of the younger sibling ($\beta6$). These variables are pivotal in capturing family stability and dynamics. Additionally, the mother's years of education ($\beta7$) one or two years before the birth of the younger sibling is incorporated to control for parental influence on educational aspirations and support. Race and ethnicity are accounted for with binary indicators for Hispanic ($\beta8$) and Black ($\beta9$) backgrounds, compared against all other racial classifications of older siblings. The variable 'Male' ($\beta10$) is included to control for gender-specific differences in educational outcomes. Lastly, the total number of siblings ($\beta11$) is integrated into the model to consider the effects of family size on resource allocation and individual attention, which may influence educational achievement.

The main linear regression model illustrated in model 1 above seeks to capture the net effect of having a younger sister on the likelihood different educational attainment outcomes while controlling for other factors as seen in table 4.

Another angle to the study is the exploration of having only sisters or having only brothers for the elder sibling in comparison to a mixed gender siblings' group. This approach sets two treatment groups which are having only brothers or having only sisters to the elder siblings as opposed to the control group of mixed gender younger siblings. This allows us to delve into the effect of having only brothers or sisters as opposed to having a mixed gender sibling composition. The methodology follows a similar rout to the main equation illustrated above, with the only difference being is changing the having a younger sister variable to the only sister's variable and adding the only brother's variable to the overall equation (See table 8 for the results) as below.

Model 2 Same Sex Composition Educational Outcomes Regression Model

$$
\begin{aligned}
EducationalOutcome_i \\
= \beta0 + \boldsymbol{\beta1 HasOnlySisters_i} + \beta2 AdjustedIncome_{10ki} \\
+ \beta3 MotherEmployed_i + \beta4 MotherOutOfLaborForce_i \\
+ \beta5 LivingWithMother_i + \beta6 FathersPresence_i \\
+ \beta7 MothersEducation_i + \beta8 Hispanic_i + \beta9 Black_i + \beta10 Male_i \\
+ \beta11 TotalSiblings_i + \boldsymbol{\beta12 HasOnlyBrothers} + \epsilon_i
\end{aligned}
$$

This estimation equation segregates the effects of having only sisters and only brother to elder sibling as opposed to a mixed sibling gender composition. However, the results are to be treated with caution as a gender selection bias between the two groups is present, as well as the mixed gender siblings control group (Refer to the Section 5.3 for the further details).

3.2 Occupational Outcomes

The second aspect of the strategy is on the occupational choices of the elder siblings, it is divided into two segments. Initially, we will discuss the general model applied for the occupational outcomes and then we will discuss the adjusted income model which would help us test the theoretical background outline in to measure the effect of having a younger sister on income when entering into a female dominated job.

3.2.1 Occupational Outcomes

A year fixed effect model is applied to measure the effect of having a first younger sister on several occupational outcomes. The model takes the form:

Model 3 Occupational Outcomes Regression Model

$$OccupationalOutcome_i$$
$$= \beta_0 + \boldsymbol{\beta_1 YoungerSister} + \beta_2 AdjustedIncome_{10ki}$$
$$+ \beta_3 MotherEmployed_i + \beta_4 MotherOutOfLaborForce_i$$
$$+ \beta_5 LivingWithMother_i + \beta_6 FathersPresence_i$$
$$+ \beta_7 MothersEducation_i + \beta_8 Hispanic_i + \beta_9 Black_i + \beta_{10} Male_i$$
$$+ \beta_{11} TotalSiblings_i + \sum_{t=2004}^{2004-2019} \beta_t yearfactor_{t,i} + \epsilon_i$$

Similar to the educational outcomes approach we control for several key variables in the same way as noted in . The main difference between the two models is the inclusion of the year factor variables (dummy variables for each survey year from 2004 to 2019) with 2004 as the base year of comparison for the occupational outcomes model. Implementing a year fixed effect model is crucial as it accounts for temporal effects that could influence occupational choices, such as economic cycles and changes in labour market conditions. These factors ensure the model adjusts for external temporal influences, providing a robust analysis of how personal and contextual factors converge to shape occupational outcomes over time. This approach allows for an exploration of how family dynamics, personal characteristics, and broader economic and social contexts interact to influence career paths, particularly in relation to gender-conforming and non-conforming occupational choices.

The occupational outcomes variables are firstly, the likelihood of getting into female dominated job, the likelihood of getting into a STEM related job, the number of times reported jobless

above the age of 30, and the income (Adjusted to 2020 dollars for each induvial) for the reported job. However, jobless count model is not a year fixed effect model. It simply counts the number of times an individual has reported jobless after the age of 30. Thus, in the estimating equation as represented above we simply remove the year factor variable. Please find more details in section 4.3 on how the outcome variables were constructed.

3.2.2 Adjusted Income Outcomes

In order to explore the effect of having a younger sister on income as explored in section 2.6 we introduced an interaction within the variables of having a younger sister and getting into female dominated jobs within the year fixed effects model, this takes the form:

Model 4 Occupational Interaction Regression Model

$$
\begin{aligned}
Adjusted\ Income_i &= \beta 0 + \boldsymbol{\beta 1 Has\ Younger\ Sister_i} + \beta 2 Adjusted Income_{10ki} \\
&+ \beta 3 Mother Employed_i + \beta 4 Mother Out Of Labor Force_i \\
&+ \beta 5 Living With Mother_i + \beta 6 Fathers Presence_i \\
&+ \beta 7 Mothers Education_i + \beta 8 Hispanic_i + \beta 9 Black_i + \beta 10 Male_i \\
&+ \beta 11 Total Siblings_i + \boldsymbol{\beta 12 Female\ Dominated_i} \\
&+ \boldsymbol{\beta 13 (Has\ Younger\ Sister_i \times Female\ Dominated_i)} \\
&+ \sum_{t=2004}^{2004-2019} \beta t year factor_{t,i} + \epsilon i
\end{aligned}
$$

The adjusted income variable is a measure of income from the occupation as reported by the individual in the interview round, we adjust the figure for inflation representing 2020 dollars. See section 5.4.3 for the results of this model.

As suggested before, under the Roy model theory (See section 2.6), one would expect that if a trait (like having a younger sister) alters an individual's preferences or behaviours in a way that impacts job choice or performance, it might lead to different compensation outcomes. In female dominated jobs, which typically offer lower compensation (Glynn and Boesch, 2022), the theory would suggest that individuals with younger sisters might either have the necessary skillset to deal with the challenges of female dominated jobs which gives an advantage that translates to higher compensation. The interaction term allows us to examine whether the combination of these two conditions affects income differently than either condition alone. It specifically tests whether the impact of having a younger sister on income is different for those in female dominated jobs compared to those in other jobs.

3.3 Limitations to the Empirical Strategy

The framework we applied for the educational outcomes model and the occupational outcomes model introduces several limitations. Firstly, missing influential variables such as peer influence, fathers' education level around the younger sibling's birth, and school quality can lead to biased estimates. Such variables and other potential factors might affect educational outcomes and as such career choices, these control variables couldn't be found in the NLSY datasets. Secondly, the assumption of linearity in regression models implies a straight-line relationship between the dependent and independent variables. In the context of our study, this assumption may not hold true, which can lead to inaccurate results. For instance, the relationship between having a younger sister and educational or occupational outcomes might be more complex than a linear model can capture. Non-linear relationships, interactions between variables, or threshold effects might better explain how sibling gender composition influences career choices and educational attainment. Ignoring these complexities can result in misestimating the true effects and overlooking important dynamics.

4. Data

4.1 Primary Source

This thesis uses the extensive data available from the National Longitudinal Surveys of Youth, combining the main NLSY and the NLSY79 Children and the Young Adult data. It focuses mainly on sibling relationships and their outcomes on various educational and occupational measures. Being longitudinal in nature, the NLSY offers rich detail by gathering comprehensive demographic, socioeconomic, job, and education histories for its respondents over the span of several decades.

A subset of the NLSY 79 children and young youth was prepared, focusing on siblings within the families surveyed This involved filtering the families with at least two children in order to study the dynamics between elder siblings and their immediate younger sibling. Special focus was attached to birth years of both elder and younger siblings in a bid to investigate the pre-treatment conditions likely to affect educational attainment and occupational choices. Various pre-birth and post-birth factors were considered, including parental employment status, household income adjusted for inflation in 2020, and the years of education attained by the

mother around the time of the younger sibling's birth which is represented as a variable that goes from 0 years to above. Variables were constructed to capture the essence of these familial and socio-economic contexts, providing a framework for analysing the potential impacts on educational and occupational trajectories while controlling for different covariates that might affect the variables under investigation.

4.2 Secondary Sources

Furthermore, the IPUMS 5-year survey from 2004 upwards to 2019 were utilized to identify the female dominated occupations in the United States. The collective data was downloaded with four main variables, the year of the survey, individuals' gender, individuals weight which indicates how many persons in the U.S. population are represented by a given person in the sample, and the occupation reported for the respective year (Census Occupation 2010 codes). We then aggregated the results for each reported occupational code by gender, calculated how many males and females opted into the occupation and weighting the gender ratios. If an occupational is reported to be above 50% dominated by females, then it will be assigned as a female dominated job. This established the necessary metric to identify if the selection of an individual in the study was into either female or male dominated occupations. We believe that using the survey years from 2004 to 2019 is a fair representation of the U.S labour market and is in tandem with the study period examined for the female dominated model (2004 to 2020).

Additionally, to construct the binary variable identifying if an occupation is STEM related or not, we have utilized the Census classification of STEM related occupations as published by the U.S Statistics Bureo in 2010 (see appendix table 14 for the complete list).

4.3 Outcome Variables and Sample Selection

There are two main datasets that are utilized for the models applied. The educational outcomes dataset and the occupational outcomes dataset. The main difference between the two, besides the educational outcomes model being a linear regression model and the occupational outcomes being a year fixed effects model, is the outcome variables.

For the educational attainment outcome variables, we constructed categories such as no degree, high school completion, college completion, advanced degree attainment, and holding a Bachelor of Arts (BA) or Bachelor of Science (BS) degree. These variables were constructed

using the highest degree earned variable from the National Longitudinal Survey of Youth (NLSY) Children and Young Adults dataset, this variable is a categorical variable where it showcases options from 0 (having no degree) to 8 (professional degree), in between is also a selection of 4 (having a BA) and 5 (having a BS) which we used for our selection into a BA or BS outcomes. For our advanced degree outcome, we used the inputs of 6 (having a Master's degree), 7 (having a PhD), and 8 (professional degree, eg, MD, LLD). As for the completing college outcomes we used the aggregate of selecting 3 (Associate's degree), 4 (having a BA), and 5 (having a BS). Where is having no degree was taken as is, the same goes for completing high school.

On the other hand, the occupational outcomes variables are, getting into a female dominated job, getting into a STEM related job, the number of times reported jobless above the age of 30, and the reported income (Adjusted to 2020 dollars for each induvial) for the reported job. To select the represented job to be identified as female dominated or STEM related in the model, we initially take in consideration all the reported jobs from age 30 and above by the elder sibling, secondly, take the most repeated consecutive job reported under the assumption that this is the job choice in which the individual has settled within their career. The NLSY 79 Children and Young Youth dataset reports five jobs per individual for a given year, this creates several reported ties in which an individual had two or more jobs represented for the model. To break this tie, we take the earliest reported job by year. To ensure robustness of the results and that taking either case won't alter the results we ran another model considering the latest reported job if a tie is represented, the results showed no noticeable difference. Reported income, is a self-reported income measure that asks the individual how much they earned in income from wages, salary, commissions, or tips from all jobs in the reference year. Given that the NLSY datasets does not connect the occupation reported each year with the income received exactly from that occupation, we had to refer to the same year when the occupation was selected for the female dominated and STEM related jobs models; adjusting it for 2020 dollars.

The different outcome variables between the two models caused discrepancies in terms of the final samples for the models. Consequently, variations in missing values among these control outcome variables and the control variables lead to differing sample sizes for the two models. This discrepancy necessitates the presentation of separate descriptive statistics tables for educational and occupational outcomes. See section 4.4 for the missing values.

Several filters were conducted to derive the final two datasets for the models (Table 3 and Table 5), the analysis focuses on families with exactly two siblings and above to delve into the

specifics of sibling gender effects on education. While initially we wanted to study exactly two sibling families to eliminate the effect of later born siblings and focus only on the effect of second-born children on first-borns, we encounter a selection bias represented by the 2-sample equality of proportions test between females and males composition in the two groups of having a younger sister and having a younger brother. Such bias might exist when limiting the study to only two sibling families as we are more likely to observe siblings of different genders due to a preference for diversity among parents. When parents have one girl and then a boy, they are more likely to stop having more children compared to parents with two girls, who might try for a third child in hopes of having a boy. This presented a selection bias in the data, as families with mixed-gender siblings may be overrepresented, skewing the analysis of sibling gender composition effects on educational and occupational outcomes.

The treatment introduced within the two models is having a first younger sister, thus, the control group would be having a first younger brother. Please note that any refence from here on to the treatment versus control groups would follow that the treatment group is having a first younger sister to the elder sibling, while the control group would be having a first younger brother to the elder sibling.

Note that similar representations with very minimal slight unnoticeable deviations between the two datasets exist on the figures representing the mothers age at first birth, the older sibling's year of birth, the younger sibling's year of birth, and the total number of siblings. Thus, these figures were shown once as they were calculated from the educational outcomes dataset.

An age difference of no more than five years was imposed between the elder and younger siblings to ensure the potential effects of the gender to be present within development years of the siblings. Furthermore, an age limit of above or equal to 30 for the individuals by 2020 was enforced on the assumption that by that age individuals would have mostly completed their educational endeavours and settled within certain occupations or industries. Another restriction was the age of the mother at first birth, this was set to be above or equal to 18 years (See figure 1 for the age distribution of mothers at first birth). This decision is informed by literature that has consistently found early motherhood to be associated with a range of adverse outcomes for both the mother and the child, including lower educational achievement and economic disadvantages (Furstenberg et al., 1987). By focusing on mothers who were at least 18 years old at the time of their first childbirth, the analysis seeks to mitigate these potential confounders and focus the effects of sibling gender composition within a more standardized family structure.

Mothers Age at First Birth
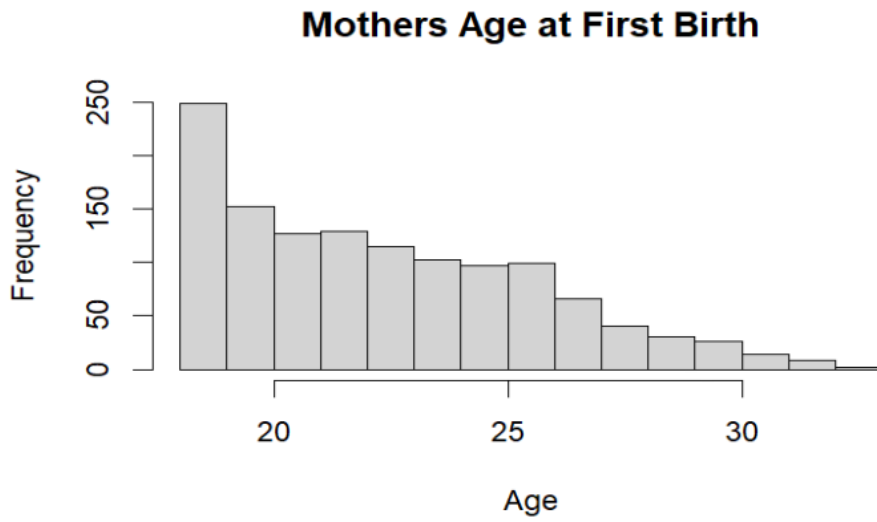
Figure 1 Age Distribution of Mothers at First Birth

The distribution of the elder sibling's years of birth is shown in figure 2, peaking around the mid-1980s, which suggests a significant number of the study's elder siblings were born in that era.

Distribution of Elders Year of Birth
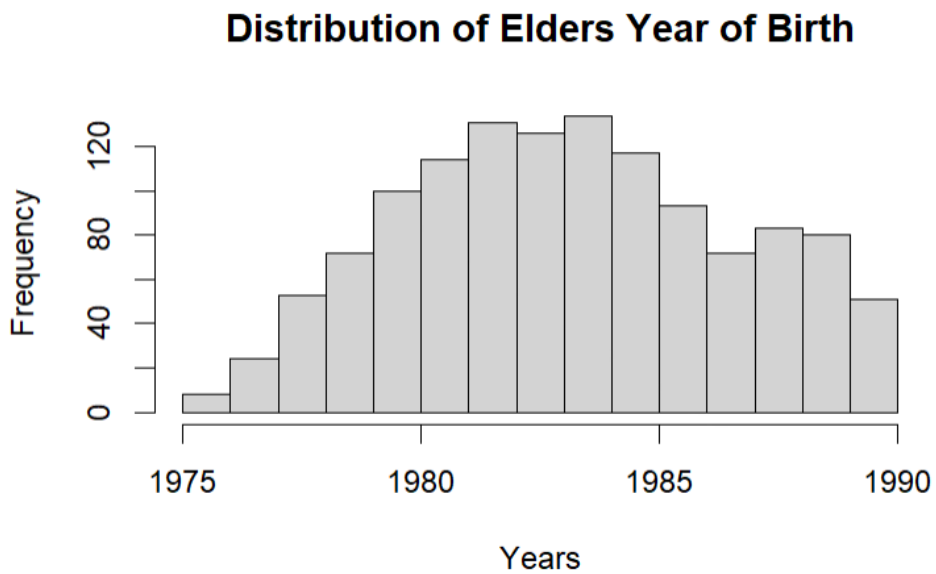
Figure 2 Histogram of Elders Year of Birth

Figuer 3 shows the distribution of younger siblings' years of birth with a similar mid-1980s peak, indicating a close age range (mean of 2.97 years) between elder and younger siblings in the dataset.

Distribution of Youngers Year of Birth
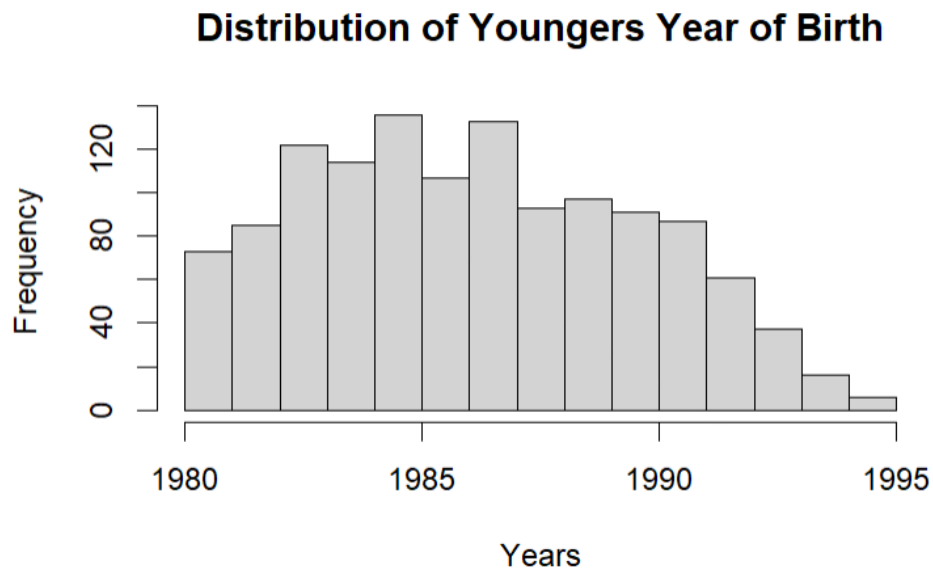
Figure 3 Histogram of Youngers Year of Birth

Figure 4 illustrates the distribution of total siblings in families, with the majority having two children, and a rapidly decreasing frequency as the number of siblings increases.

Distribution of Total Siblings
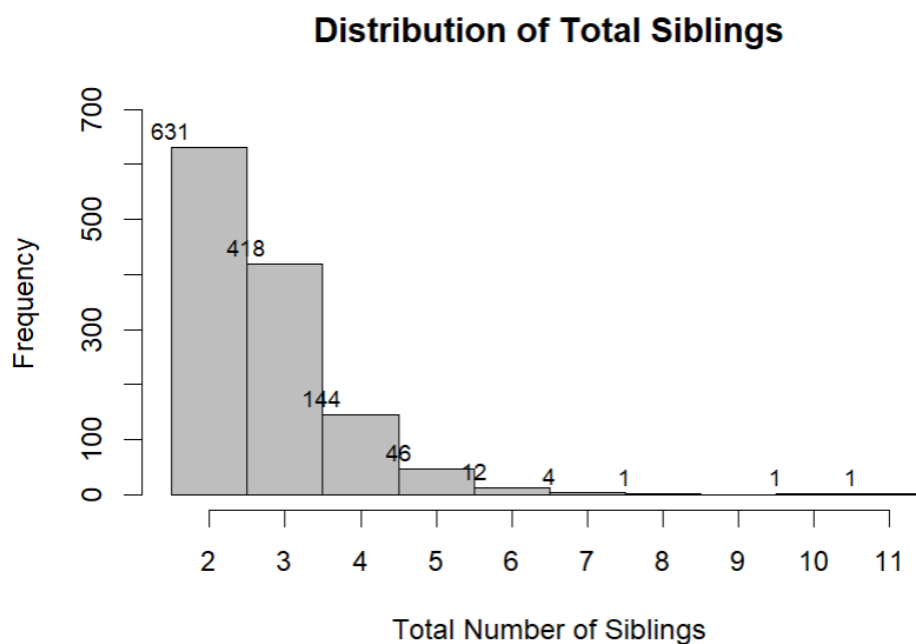
Figure 4 Total Number of Siblings Per Family

4.4 Missing values

Given the difference in outcome variables which resulted in a different share of missing variables, we ended up with two datasets to run the models on. See section 5.2 for the descriptive statistics of these datasets.

Missing share of variables as shown in table 1 were removed when conducting the analysis. While the final dataset above in table 3 shows 1258 individuals in total, it was originally 2181 individuals before accounting for the missing values below. By incorporating any of the two year-round interviews before the younger siblings' birth we tried to minimize the proportion of missing values. The number of missing values especially for our outcome variable (26.69%) represents a high percentage relative to the rests, this is beyond our control as the missing values comes from the source itself. While the reduction in the total number of participants in the study reduces the significance power of the results, we don't believe this might introduce bias, and lead to inaccurate estimates of relationships between variables as we already established that the pre-treatment covariates are equal between the control group (having a younger brother) and treatment group (having a younger sister).

Table 1 Missing Shares of Variables for the Educational Dataset

At Birth of Younger Sibling	Missing
Household Income Adjusted for 2020 Dollars	15.50%
Mothers Employed any of the Prior 2 Years	3.85%
Mothers Out of Labour Force any of the Prior 2 Years	3.85%
Living with Mother any of the Prior 2 Years	9.67%
Fathers Present in HH any of the Prior 2 Years	4.95%
Other Predictor Variables	**Missing**
Race	0.00%
Age gap	0.00%
Mother's Highest Degree	3.39%
Gender	0.00%
Mothers Age at First Birth	0.00%
Dependent Variables	**Missing**
Highest Degree Earned	26.69%

As for the occupational outcomes dataset the missing share of values is shown in table 2 below.

Table 2 Missing Values for the Occupational Outcomes Dataset

At Birth of Younger Sibling	Missing
Household Income Adjusted for 2020 Dollars	8.54%

Mothers Employed any of the Prior 2 Years	2.23%
Mothers Out of Labour Force any of the Prior 2 Years	2.23%
Living with Mother any of the Prior 2 Years	7.47%
Fathers Present in HH any of the Prior 2 Years	3.73%
Other Predictor Variables	**Missing**
Race	0.00%
Age gap	0.00%
Mother's Highest Degree	2.15%
Gender	0.00%
Mothers Age at First Birth	0.00%
Dependent Variables	**Missing**
Into Female Dominated Jobs	10.41%
Into STEM related Jobs	0.00%
Jobless Claims	0.00%
Adjusted Income (2020)	8.54%

In comparison to the educational outcomes dataset the occupational outcomes dataset shows fewer missing inputs across the board, hence the discrepancy as stated before between the number of available datapoints between the models. The variation of different missing inputs across the dependent variables in the occupational outcome's dataset creates discrepancies in the overall number of individuals available to run the models on. For example, the number of complete observations available for the female dominated model is 980, while it is 1092 for both the STEM related jobs and Jobless claims, and it is 1001 complete cases for the adjusted income model. The different number of complete observations for the models doesn't change the 2-sample equality of proportions results for all the variables between the treatment and control groups. Thus, we decided to simply report the uncomplete cases dataset which has 1393 individuals.

5. Results

Overall, we cannot reject the null hypothesis that there is no sisterhood effect on the studied outcomes, this implies that the presence of a younger sister does not significantly influence the educational and occupational choices of older siblings. This result suggests that sibling gender composition may not play a pivotal role in shaping the studied outcomes, contrary to the initial hypothesis. The following sections delves into the results of the study. Initially describing the results from the educational outcomes model and ending with the occupational outcomes model

results. We will provide a detailed interpretation of the of the results in the discussion section (See section 6). Additionally, multicollinearity was not an issue in any of the regression models used in this study.

5.1 Mechanisms Explored Quantitatively

Testing the assumptions and mechanisms we explored in section 2.6 we ran the income distribution between those individuals going into same sex occupations and those getting into opposite sex occupations with figure 5 below showcasing the results.
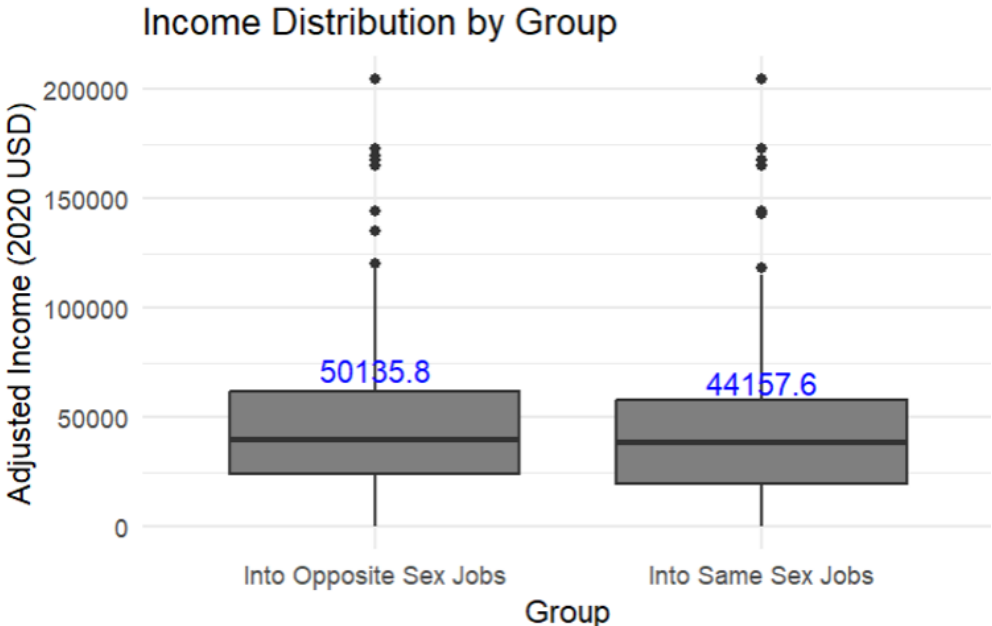
Figure 5 Wage Differences Between the Groups in Focus

There are 310 elder siblings in the opposite sex group with 153 males and 157 females, on the other hand, there are 832 older siblings in the same sex group with 393 males and 439 females. This represents a percentage split of 27% and 73% respectively between the two groups, showcasing a preference for entering a same sex job. The t-test (0.027*) shows a significant difference between the means of the two groups backing our assumption that individuals entering an opposite sex job are compensated higher than entering a same sex job. Albeit gender differences between males and females might arise given that male dominated jobs tend to pay higher (Blau and Khan 2017).

Building on the principles of the Compensation Differentials Model (Rosen 1986) and the evidence we found on differences between getting into same and opposite sex jobs, the framework behind the occupational part of the thesis seeks to investigate the effect of an

younger sister as opposed to a younger brother on income, choice of male or female dominated jobs, STEM related jobs, and the jobless count after the age of 30.

Following the same analogy as the opposite sex and same sex groups mean adjusted income comparison, having a younger sister as opposed to a younger brother could influence an individual's perception of gender roles and consequently their openness or aversion to entering occupations traditionally dominated by the opposite sex. While the Compensation Differential theory provides a framework for understanding wage differentials from a utility versus disutility standpoint, the Roy model of self-selection suggests that individuals choose jobs based on their skills and characteristics, leading to a match between worker attributes and job requirements. It describes how individuals self-select into different occupations based on their comparative advantages.

Figure 6 below shows the comparison between the adjusted income getting into a female dominated job while having a younger sister (262 elder females and 88 elder males) and getting into a female dominated job while having a brother (222 elder females with 76 elder brothers). The split between the two groups is 54% to 46% respectively. Although the compensation received in the younger sister groups is higher, there is no statistical significance between the means of the two groups.
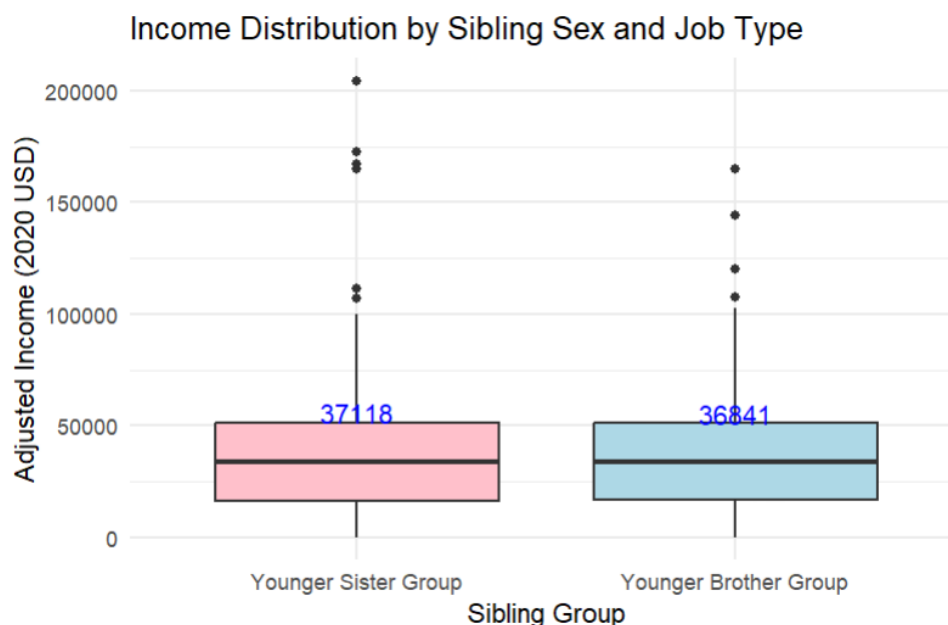
Figure 6 Income Distribution by Groups Getting into Female Dominated Jobs

5.2 Descriptive Statistics

This section provides an overview of the two datasets extracted from the Longitudinal Survey of Youth 1979 (NLSY79) and its subsequent NLSY79 Child and Young Adult datasets to run the necessary models for the analysis and to showcase that there is no random selection into the treatment as well as ensuring no difference in the characteristics of older siblings before the birth of the first younger sibling. It covers the educational and occupational datasets. As noted prior in section 4.4, the existence of two datasets for the analysis is due to the variation in missing values for the different outcome variables between educational and occupational choices.

5.2.1 Descriptive Statistics for the Educational Outcomes Dataset

The descriptive statistics table below provides a detailed comparison of elder siblings with younger brothers against those with younger sisters on various demographic and socio-economic factors. We compare older siblings with the first younger brother and older siblings with the first younger sister. To make sure that there is no selection in the gender of the first younger sibling we conduct a 2-sample equality of proportions test. Notably, the proportion of elder siblings who have younger brothers is almost equally split with those who have younger sisters, suggesting a balanced gender distribution. The 2-sample tests for equality of proportions offer statistical evidence on whether the observed differences in proportions are random, essentially measuring the probability of having a younger sister to an elder male or female, the 2-sample equality of proportions tests are statistically insignificant, with results indicating no significant gender bias in the composition of the sibling groups, thus, avoiding any potential gender selection bias that might arise into the data. Additionally, we want to make sure that there is no difference in the characteristics of older siblings before the birth of the first younger sibling. Table 1 shows that there are no differences and that the covariates prior to the treatment of the younger siblings' birth are equal between the two groups for the educational outcomes model.

Table 3 Descriptive Statistics for Educational Outcomes Dataset

Pre-Treatment Groups at Time of Birth of Younger Sibling - Complete Cases				
Characteristics of the Elder Sibling	**With Younger Brother**	**With Younger Sister**	*2-sample test for equality of proportions*	
Number of Males	334 (49.1%)	306 (52.9%)	*0.55*	*0.052*
Number of Females	346 (50.9%)	272 (47.1%)		
Total	**680 (54.1%)**	**578 (45.9%)**		
Blacks	157 (23.1%)	144 (24.9%)	*NA*	
Hispanics	158 (23.2%)	107 (18.5%)		
Other	365 (53.7%)	327 (56.6%)		
Number of Observations	**1258**		***Welch Two Sample t-test***	
Household Variables	**Mean (SD)**		*p-value*	*t-value*
Age in 2020	36.35 (3.45)	36.43 (3.42)	*0.68*	*-0.41*
Age of Elder Sibling (Age Gap)	3.01 (1.15)	2.93 (1.22)	*0.20*	*1.30*
Household Income Adjusted for 2020 Dollars	60,259 (93,382)	60,751 (97,708)	*0.93*	*-0.09*
Age of Mother at First Birth	22.78 (3.38)	22.72 (3.40)	*0.77*	*0.29*
Parents Characteristics	**Count (Percentage)**		***Welch Two Sample t-test***	
Mothers Completed High School or Above	639 (94.0%)	541 (93.6%)	*0.79*	*0.27*
Mothers Completed 4th Year College or Above	157 (23.1%)	117 (20.2%)	*0.22*	*1.22*
Fathers Present in HH any of the Prior 2 Years	463 (68.1%)	380 (65.7%)	*0.38*	*0.88*
Mothers Employed any of the Prior 2 Years	324 (47.6%	274 (47.4%)	*0.93*	*0.09*
Mothers Out of Labor Force any of the Prior 2 Years	325 (47.8%)	268 (46.4%)	*0.61*	*0.51*
Living with Mother any of the Prior 2 Years	674 (99.1%)	574 (99.3%)	*0.70*	*-0.38*

In the dataset analysed, educational attainment varies between elder siblings with younger brothers and those with younger sisters as seen in table 4 below. Advanced degrees are Masters, PhD, and professional degrees. The statistical significance of these variations is seen in the main educational outcomes model regression results within table 6.

Table 4 Educational Attainment Statistics of Educational Outcomes

Educational Attainment		
Education	**Younger Brother**	**Younger Sister**
No Degree	46 (6.8%)	37 (6.4%)
Only High School	321 (47.2%)	268 (46.4%)

Only College	234 (34.3%)	200 (34.6%)
Completed HS and Above	634 (93.2%)	541 (93.6%)
Completed College and Above	313 (46.0%)	273 (47.2%)
Has a BA	109 (16.0%)	93 (16.1%)
Has a BS	136 (20.0%)	107 (18.5%)
Advanced Degree	79 (11.6%)	73 (12.6%)

5.2.2 Descriptive Statistics for the Occupational Outcomes Dataset

Similarly, for the occupational outcomes dataset we ran the same 2-sample test for equality of proportions to test if there are any significant differences between the treatment group (having a younger sister) and the control group (having a younger brother). The results as shown below indicate no significant difference in the characteristics of the older siblings when the younger sibling was born, setting the scene for our occupational outcomes models.

Table 5 Descriptive Statistics for the Occupational Outcomes Models

Pre-Treatment Groups at Time of Birth of Younger Sibling - Uncomplete Cases				
Characteristics of the Elder Sibling	**With Younger Brother**	**With Younger Sister**	**2-sample test for equality of proportions**	
Number of Males	367 (50.8%)	325 (50.2%)	*0.570*	*0.910*
Number of Females	379 (49.2%)	322 (49.8%)		
Total	**746 (53.6%)**	**647 (46.4%)**		
Blacks	185 (24.8%)	171 (26.4%)	*NA*	
Hispanics	180 (24.1%)	129 (19.9%)		
Other	381 (51.1%)	347 (53.6%)		
Number of Observations	**1393**		**Welch Two Sample t-test**	
Household Variables	**Mean (SD)**		*p-value*	*t-value*
Age in 2020	36.82 (3.76)	36.98 (3.68)	*0.42*	*-0.80*
Age of Elder Sibling (Age Gap)	2.81 (1.23)	2.79 (1.27)	*0.77*	*0.29*
Household Income Adjusted for 2020 Dollars	58,473 (94,994)	58, 651 (100,085)	*0.97*	*-0.03*
Age of Mother at First Birth	22.47 (3.41)	22.38 (3.43)	*0.60*	*0.52*
Parents Characteristics	**Count (Percentage)**		**Welch Two Sample t-test**	
Mothers Completed High School or Above	619 (85.0%)	539 (84.9%)	*0.94*	*0.07*
Mothers Completed 4th Year College or Above	105 (14.4%)	87 (13.7%)	*0.70*	*0.38*

Fathers Present in HH any of the Prior 2 Years	423 (58.9%)	362 (58.2%)	*0.79*	*0.26*
Mothers Employed any of the Prior 2 Years	339 (47.6%)	288 (44.2%)	*0.20*	*1.28*
Mothers Out of Labor Force any of the Prior 2 Years	353 (46.9%)	308 (50.4%)	*0.20*	*-1.27*
Living with Mother any of the Prior 2 Years	680 (98.6%)	591 (98.7%)	*0.86*	*-0.17*

5.3 Educational Outcomes Results

The results section provides an analysis of the impact of various socio-economic factors, family dynamics, and sibling gender on educational attainment. Leveraging a dataset with 1,258 observations (table 3), regression models set out how these variables correlate with achieving different levels of education, ranging from no degree to advanced degrees.

The educational attainment regression model results below show 6 columns, each of the columns represents the estimates of the impact of having a younger sister on the various outcomes, for example, columns 5 and 6 present estimates of the impact of having a first younger sister on choosing a BA or BS, where there is no significant effect. Similarly, there is no significant effect of having a first younger sister on not having a degree, completing high school, completing college, and earning an advanced degree.

Variables such as the mother's employment status around the time of the younger sibling's birth and the number of total siblings lowers the chances of completing high school and ending up with no degree, indicating that these factors contribute differently depending on the education level being considered. Household income around the younger sibling's birth seem to increase the likelihood of earning a college degree and an advanced degree. While the main variable in question, having a younger sister in comparison to having a younger brother doesn't show a significance in any of the educational outcomes.

Table 1 Educational Outcomes Regression Model

Educational Outcomes Regression Models Summary

	No Degree (1)	Completed HS (2)	Completed College (3)	Advanced Degree (4)	BA (5)	BS (6)
			Dependent variables:			
Has Younger Sister	-0.004	0.004	0.02	0.01	0.002	-0.02
	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)
Adjusted Household Income 10k (2020)	-0.001	0.001	**0.003****	**0.003*****	**0.003****	**0.002***
	(0.001)	(0.001)	**(0.001)**	**(0.001)**	**(0.001)**	**(0.001)**
Mother Employed Around Younger Birth	-0.06*	0.06*	0.04	0.01	0.03	0.03
	(0.03)	(0.03)	(0.06)	(0.04)	(0.05)	(0.05)
Mother Out of Labor Force Around Younger Birth	-0.03	0.03	0.04	0.01	0.03	0.01
	(0.03)	(0.03)	(0.06)	(0.04)	(0.05)	(0.05)
Living With Mother Around Younger Birth	0.12	-0.12	-0.19	-0.13	-0.12	-0.08
	(0.08)	(0.08)	(0.15)	(0.10)	(0.12)	(0.12)
Fathers Presence Before Younger Birth	-0.02	0.02	**0.08*****	0.02	**0.04***	0.02
	(0.02)	(0.02)	**(0.03)**	(0.02)	**(0.02)**	(0.02)
Mothers Education Around Younger Birth	-0.003	0.003	**0.03*****	**0.01*****	**0.01*****	**0.02*****
	(0.002)	(0.002)	**(0.004)**	**(0.003)**	**(0.003)**	**(0.003)**
Hispanic	**0.05*****	**-0.05*****	**-0.10*****	-0.01	**-0.06****	-0.05
	(0.02)	**(0.02)**	**(0.04)**	(0.02)	**(0.03)**	(0.03)
Black	0.01	-0.01	**-0.11*****	-0.01	**-0.08*****	**-0.06****
	(0.02)	(0.02)	**(0.03)**	(0.02)	**(0.03)**	**(0.03)**
Male	**0.03****	**-0.03****	**-0.15*****	**-0.07*****	**-0.05*****	-0.02
	(0.01)	**(0.01)**	**(0.03)**	**(0.02)**	**(0.02)**	(0.02)
Total Siblings	**0.02****	**-0.02****	-0.02	-0.01	-0.01	-0.004
	(0.01)	**(0.01)**	(0.01)	(0.01)	(0.01)	(0.01)
Constant	-0.01	**1.01*****	**0.33***	0.07	0.13	0.03
	(0.09)	**(0.09)**	**(0.18)**	(0.12)	(0.13)	(0.14)
Observations	1,258	1,258	1,258	1,258	1,258	1,258
R^2	0.03	0.03	0.11	0.05	0.05	0.05

Note: *p<0.1; **p<0.05; ***p<0.01

Similarly, no significance is shown for the main dependent variable of having a sister on achieving only a high school degree or only a college degree as seen below.

Table 2 Only HS or College Model

Secondary Regression Models Summary

	Dependent variables:	
	HS Only	College Only
	(1)	(2)
Has Younger Sister	-0.01	0.004
	(0.03)	(0.03)
Adjusted Household Income 10k (2020)	-0.002	0.0003
	(0.001)	(0.001)
Mother Employed Around Younger Birth	0.02	0.03
	(0.06)	(0.06)
Mother Out of Labor Force Around Younger Birth	-0.01	0.03
	(0.06)	(0.06)
Living With Mother Around Younger Birth	0.06	-0.06
	(0.16)	(0.15)
Fathers Presence Before Younger Birth	**-0.06****	**0.06****
	(0.03)	**(0.03)**
Mothers Education Around Younger Birth	**-0.02*****	**0.01*****
	(0.004)	**(0.004)**
Hispanic	0.05	**-0.09*****
	(0.04)	**(0.03)**
Black	**0.10*****	**-0.10*****
	(0.03)	**(0.03)**
Male	**0.12*****	**-0.09*****
	(0.03)	**(0.03)**
Total Siblings	-0.0003	-0.01
	(0.01)	(0.01)
Constant	**0.69*****	0.26
	(0.18)	(0.17)
Observations	1,258	1,258
R^2	0.07	0.04
Note:	*$p<0.1$; **$p<0.05$; ***$p<0.01$	

On the other hand, table 8 below showcases the results of the effect of having only sisters (380 elder siblings) or only brothers (479 elder siblings) as opposed to having a mixed gender siblings group (399 elder siblings). Please find the descriptive statistics of the dataset used for this model in the appendix table 14. The results indicates that elder siblings with only sisters are slightly less likely to have no degree, with a significance level at 10%, while the same group appears more likely to have completed high school, again at the significance level of 10%. Similar to prior models, it suggests that household income and mothers' education level around the younger sibling's birth are significant predictors of educational attainment, particularly for advanced degrees. Ethnic background shows varying impacts, with Hispanic individuals less likely to have completed college but not significantly different in achieving an advanced degree. Elder male siblings have a consistently lower likelihood of educational achievement across most categories.

However, these results should be treated with caution due to the significant gender imbalance as indicated by the 2-sample tests for equality of proportions within the three groups in focus, only sisters, only brothers, and the mixed gender group. The proportion of males with brothers (45.3%) is significantly lower than that of females (54.7%) with brothers (p-value of 0.004***). Similarly, the proportion of males with sisters (55%) is significantly higher than that of females (45%) with sisters (p-value of 0.007***). Furthermore, the mixed gender group shows a similar result with a distribution of females 53.6% to males 46.3% with a 2-sample test score of 0.047*. The significance test score for the only brothers and only sisters groups suggests a selection bias within these treatment groups, while a similar selection bias presents its self even within the control group (mixed gender siblings). This gender difference could influence the outcomes and may suggest underlying gender dynamics within the household that affect educational attainment. It points out to a potential selection bias into the gender of the younger sibling based on the gender of the elder sibling which in turn suggests that within this sample the gender of the second born child is not random, thus, the treatment cannot be guaranteed to be random. It's important to consider this gender effect when interpreting the results of the regression, as it may confound the relationships between sibling composition and educational outcomes.

Table 3 Same Sex Sibling Composition Model Results

Only Brothers or Sisters Models Summary

	Dependent variable:					
	No Degree	Completed HS	Completed College	Advanced Degree	BA	BS
	(1)	(2)	(3)	(4)	(5)	(6)
Has Only Sisters	**-0.04***	**0.04***	0.05	0.03	0.02	-0.01
	(0.02)	**(0.02)**	(0.04)	(0.03)	(0.03)	(0.03)
Has Only Brothers	-0.03	0.03	0.04	0.02	0.02	0.01
	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Adjusted Household Income 10k (2020)	-0.001	0.001	**0.003****	**0.003*****	**0.003****	**0.002***
	(0.001)	(0.001)	**(0.001)**	**(0.001)**	**(0.001)**	**(0.001)**
Mother Employed Around Younger Birth	**-0.06***	**0.06***	0.04	0.01	0.03	0.03
	(0.03)	**(0.03)**	(0.06)	(0.04)	(0.05)	(0.05)
Mother Out of Labor Force Around Younger Birth	-0.03	0.03	0.04	0.01	0.03	0.01
	(0.03)	(0.03)	(0.06)	(0.04)	(0.05)	(0.05)
Living With Mother Around Younger Birth	0.12	-0.12	-0.18	-0.12	-0.12	-0.09
	(0.08)	(0.08)	(0.15)	(0.10)	(0.12)	(0.12)
Fathers Presence Before Younger Birth	-0.02	0.02	0.08**	0.01	0.04	0.02
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Mothers Education Around Younger Birth	-0.003	0.003	**0.03*****	**0.01*****	**0.01*****	**0.02*****
	(0.002)	(0.002)	**(0.004)**	**(0.003)**	**(0.003)**	**(0.003)**
Hispanic	**0.05*****	**-0.05*****	**-0.10*****	-0.01	**-0.06****	-0.05
	(0.02)	**(0.02)**	**(0.04)**	(0.02)	**(0.03)**	(0.03)
Black	0.01	-0.01	**-0.11*****	-0.01	**-0.08*****	**-0.06****
	(0.02)	(0.02)	**(0.03)**	(0.02)	**(0.03)**	**(0.03)**
Male	**0.03****	**-0.03****	**-0.15*****	**-0.06*****	**-0.05*****	-0.02
	(0.01)	**(0.01)**	**(0.03)**	**(0.02)**	**(0.02)**	(0.02)
Total Siblings	0.01	-0.01	-0.002	0.001	0.001	-0.004
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Constant	0.03	**0.97*****	0.27	0.04	0.10	0.03
	(0.09)	**(0.09)**	(0.18)	(0.12)	(0.14)	(0.15)
Observations	1,258	1,258	1,258	1,258	1,258	1,258

R^2	0.04	0.04	0.11	0.05	0.05	0.05

In summary the analysis in general points to no significant effect of having a younger sister as opposed to a brother on the education attainment of the elder siblings, with elder siblings with only sisters are slightly less likely to have no degree, and more likely to have completed high school, albeit these results are subject to further scrutiny due to the selection bias mentioned earlier. Further similar regression result tables on only two sibling's families and three and above sibling families where no significant effect is shown can be seen in the appendix (Table 12 and Table 13).

5.4 Occupational Outcomes Results

The second part of the analysis investigates the effect of having a younger sister on several occupational outcomes as indicated by model 3.

5.4.1 Occupational Outcomes Results

The results show insignificance, so we cannot reject null hypothesis that there is no effect of having a younger sister. The different columns in table 9 below presents an estimate of the impact of having a first younger sister on the different occupational outcomes presented in the column head.

Although there is no significant effect across the different outcome variables, the signs of the coefficients suggest that having a sister influences an older sibling's career choices in several ways. Firstly, they are more likely to enter female dominated occupations. Secondly, these occupations are less likely to be in STEM fields. Lastly, these positions tend to offer slightly lower pay compared to other fields.

Table 4 Occupational Outcomes Regression Results

Occupational Outcomes Regression Models Summary

	Dependent variables:			
	Into Female Dominated Job	Into STEM related Jobs	Jobless Count[1]	Adjusted Income
	(1)	(2)	(3)	(4)
Has Younger Sister	0.01	-0.01	-0.01	-945.60
	(0.03)	(0.02)	(0.11)	(2,325.58)
Adjusted Household Income 10k (2020)	0.002	0.001	-0.003	149.83
	(0.001)	(0.001)	(0.01)	(113.40)
Mother Employed Around Younger Birth	**0.15****	-0.01	0.15	1,043.08
	(0.07)	(0.04)	(0.25)	(5,552.48)
Mother Out of Labor Force Around Younger Birth	**0.13***	-0.05	**0.42***	295.85
	(0.07)	(0.04)	**(0.25)**	(5,497.08)
Living With Mother Around Younger Birth	0.16	**-0.17***	**0.98***	-1,707.75
	(0.14)	**(0.10)**	**(0.55)**	(13,125.16)
Fathers Presence Before Younger Birth	0.002	0.02	**-0.51*****	-72.54
	(0.04)	(0.02)	**(0.12)**	(2,875.80)
Mothers Education Around Younger Birth	-0.001	**0.01****	**-0.05*****	**1,058.16*****
	(0.005)	**(0.003)**	**(0.02)**	**(380.95)**
Male	0.03	**-0.06****	**0.50*****	**-8,017.78*****
	(0.04)	**(0.02)**	**(0.14)**	**(3,069.23)**
Hispanic	0.06	-0.03	**0.37*****	**-10,351.95*****
	(0.04)	(0.02)	**(0.13)**	**(2,952.18)**
Black	**-0.46*****	0.004	**-0.37*****	**13,799.01*****
	(0.03)	(0.02)	**(0.10)**	**(2,327.42)**
Total Siblings	0.02	0.001	0.03	176.60
	(0.02)	(0.01)	(0.06)	(1,254.89)
Constant	**0.63***	0.15	0.79	6,525.31
	(0.36)	(0.24)	(0.65)	(29,764.29)
Observations	980	1,092	1,092	1,001

---

[1] Unlike the other variables, the jobless count model is not a year fixed effects model. The variable is a continuous variable that reports the number of times an individual has reported jobless after the age of 30.

R^2	0.23	0.04	0.08	0.11

In column 1, which predicts entry into female dominated jobs, the coefficient for having a younger sister is slightly positive but statistically insignificant, suggesting that having a younger sister does not significantly influence the likelihood of an elder sibling entering female dominated professions. This outcome indicates that the mere presence of a younger sister does not alter occupational choices towards female dominated fields in a statistically meaningful way. Conversely, focusing on entry into STEM related jobs, presents a slightly negative but similarly insignificant coefficient for having a younger sister. This suggests that having a younger sister does not deter elder siblings from pursuing STEM related occupations, which are traditionally male dominated. However, the negative sign in front of the coefficient does suggest a negative relationship between the two variables. This probes further investigation into exactly how a younger sister effect elder male and female siblings does separately (See Table 10 those results).

Our findings suggest that having a first younger sister doesn't not have a significant direct impact on the occupational decisions of elder siblings in terms of entering female dominated or STEM related fields. This lack of significant influence could indicate that other factors such as socio-economic, maternal educational levels, household income, or educational backgrounds, play a more pivotal role in such decisions.

On the other hand, the mothers education around the younger siblings birth does seem to be show consistent significance across the different dependent variables, the higher the mothers level of education the more likely an elder sibling to enter into STEM related positions, this is consistent with prior literature suggesting parental influence on career choices, it also lower the chance of being jobless after the age of 30 and is correlated with higher income for the individual.

5.4.2 Results by Sex of the Elder Sibling

To dive deeper into the results presented in table 9 we ran a similar model but focusing specifically the potential gender differences between the siblings, this helps us uncover potential heterogenous effects as to focus on specific gender differences. The model aims to see if there are any different effects of having a younger sister to an elder male or female sibling.

Table 5 Focused look into the Males vs Females Occupational Outcomes

Males and Females Occupational Regression Models Summary

	Dependent variables:			
	Elder Males into Female Dominated Jobs	Elder Males Into STEM related Jobs	Elder Females into Female Dominated Jobs	Elder Females Into STEM related Jobs
	(1)	(2)	(3)	(4)
Has Younger Sister	0.02	-0.03	-0.002	0.01
	(0.04)	(0.03)	(0.04)	(0.03)
Adjusted Household Income 10k (2020)	**0.01****	0.003	0.001	0.001
	(0.01)	(0.004)	(0.001)	(0.001)
Mother Employed Around Younger Birth	0.15	0.05	0.12	-0.07
	(0.11)	(0.07)	(0.09)	(0.06)
Mother Out of Labor Force Around Younger Birth	0.16	0.01	0.09	-0.09
	(0.11)	(0.07)	(0.09)	(0.06)
Living With Mother Around Younger Birth	0.32	-0.25	0.06	-0.13
	(0.23)	(0.16)	(0.19)	(0.12)
Fathers Presence Before Younger Birth	0.01	0.04	0.01	0.01
	(0.05)	(0.03)	(0.05)	(0.03)
Mothers Education Around Younger Birth	0.01	**0.01***	**-0.01***	0.005
	(0.01)	**(0.005)**	**(0.01)**	(0.004)
Hispanic	0.08	-0.03	0.02	**-0.08****
	(0.06)	(0.03)	(0.05)	**(0.03)**
Black	**0.15*****	-0.03	0.003	-0.03
	(0.06)	(0.03)	(0.05)	(0.03)
Total Siblings	0.02	0.01	0.02	-0.01
	(0.02)	(0.01)	(0.02)	(0.01)
Constant	-0.16	0.13	**0.80*****	0.27
	(0.42)	(0.28)	**(0.26)**	(0.17)
Observations	464	539	516	553
R^2	0.05	0.07	0.03	0.04

Our main variable of interest does not show a statistically significant impact on occupational choices either for males or females across all models. This suggests that the presence of a younger sister does not significantly influence whether elder siblings enter female dominated or STEM related jobs. However, it is crucial to note the difference in terms of the signs in front of the coefficients and the magnitude thereof of them. The positive sign for elder males who have a younger sister in comparison to the negative one for elder females entering a female dominated job suggests that maybe elder males are more effected by the presence of younger sister than the elder females. The magnitude difference of 10 times (0.02 compared to -0.002) between the two groups does point in that direction as well. Similarly, an elder male getting into a STEM related job (Mainly male dominated) shows a negative sign when having a younger sister while it is positive for elder females who have a younger sister.

Interestingly, mother's education has a positive significant impact on males entering STEM related jobs, indicating that higher maternal education levels might encourage or enable sons to pursue careers in these technically complex fields. This aligns with literature suggesting that parental education, particularly that of mothers, plays a crucial role in shaping children's educational aspirations and achievements (Kean 2005).

5.4.3 Adjusted Income Interaction

Model 4 results are illustrated below. The coefficient for having a younger sister is positive but not statistically significant, suggesting having a younger sister does not, in isolation, affect income. Furthermore, the coefficient for female dominated jobs indicates that such jobs generally pay less, aligning with existing literature that notes wage gaps in gender-typed occupations. As for our main interaction variables, the effect is not strong enough to be statistically significant in this model. This outcome may imply that any familial gender composition effects are overshadowed by the broader economic structures that dictate wage levels in female dominated fields.

Table 6 Occupational Interaction Regression Results

Has a Younger Sister and Into Female Dominated Job Interaction Regression Models Summary

	Dependent variable:
	Adjusted Income
Has a Younger Sister	2,320.97
	(3,297.72)
Female Dominated Job	**-16,967.09*****
	(3,094.49)
Adjusted Household Income 10k (2020)	6,398.21
	(5,362.09)
Mother Employed Around Younger Birth	3,835.50
	(5,299.32)
Mother Out of Labor Force Around Younger Birth	-12,273.75
	(9,569.38)
Living With Mother Around Younger Birth	-364.94
	(2,832.28)
Fathers Presence Before Younger Birth	826.82**
	(364.90)
Mothers Education Around Younger Birth	**-10,086.54*****
	(3,003.76)
Male	**-9,246.06*****
	(2,843.37)
Hispanic	468.26
	(1,223.80)
Black	-11,025.00
	(16,036.53)
Total Siblings	-17,652.34
	(15,391.36)
Has a Younger Sister*Female Dominated Job	-3,682.47
	(4,582.51)
Constant	64,249.56***
	(18,900.89)
Observations	1,018
R^2	0.14
Note:	*p<0.1; **p<0.05; ***p<0.01

Nevertheless, the less negative value of the interaction term relative to the coefficient for the female dominated job alone makes for an interesting perspective to interpret, even though it is not statistically significant. This may imply that the interaction effects of having a younger sister and holding a female dominated job would attenuate the negative wage effect associated with female dominated jobs alone. This perhaps might imply that in the presence of a younger sister somehow negates the typical wage penalty found within female dominated professions, by taking advantage of interpersonal skill sets or personality traits better valued within female dominated jobs, like empathic or communication and caregiving skill sets.

While it might be interesting to see how this smaller negative coefficient might be interpreted, it should not be forgotten that, in any case, the interaction itself is not significant. Thus, interpretations would run even more speculative than those above and would require further evidence to support such interpretations as valid. Further research with an even larger sample size or different granularity of data might be necessary to further pursue and establish if any such effects, real and reliable, are present.

6. Discussion

Our findings suggest no significant effect of having a first younger sister on an elder sibling's educational and occupational outcomes. Aligning partially with some existing studies while differing from others. This outcome invites a deeper exploration into the potential mechanisms and broader implications of familial dynamics on individual life choices. Here, we discuss these results in the context of existing literature and introduce new ideas and concepts that may provide further insights.

Studies have shown that parental education and income of the household have been identified as strong predictors of children's education outcomes (Haveman and Wolfe, 1995). This thesis's findings reinforce this notion, whereby maternal education and income of the household at the time of birth of the younger sibling emerged as strong predictors of educational attainment. Our analysis reveals the intricate ways in which family dynamics affect education, although the genders of siblings can have an effect, socioeconomic status and parental education are far more influential.

Furthermore, the findings in our thesis suggests that sibling gender composition does not significantly affect the likelihood of entering female dominated or STEM related jobs, other

factors, probably personal interests, expectations of the society and economic incentives, are of much greater importance. Notably, the mother's education level around the younger sibling's birth is significant int terms of getting into a STEM related jobs, less joblessness count, and a higher adjusted income. According to the theory of gender socialization (Eccles, 2011), gender disparity in occupations related to math and physical science, females are less likely than males to enter these fields. This is partly because women have lower confidence in their abilities in these areas and place less subjective value on these fields compared to other potential careers. Additionally, gendered socialization practices at home, in schools, and among peers significantly shape these differences in self-perceptions and perceived task values.

Going back to prior findings in terms of the sisterhood effect, Butcher and Case (1994) reported that having a sister reduces years of education, which contrasts with our result showing elder siblings with only sisters are more likely to complete high school. Tao and Cheng (2022) showcased that the presence of sisters tends to steer females towards fields of study that are not centred around science, while we didn't see any significance in terms of a similar result, the negative coefficient on the effect of having a first younger sister on opting for a Bachelor of Science seems to suggest a similar finding.

Conversely, Cools and Patacchini (2017) indicated males with sisters choose more traditionally male disciplines, while our findings suggest no significant impact of sibling gender on male educational choices. Furthermore, Peter et al. (2018) found little bearing of sibling gender on educational attainment, resonating with our findings. However, Tao and Cheng (2022) observed that brothers influence male educational decisions more significantly, and men with brothers have higher earnings, which was not significantly supported by our data, albeit a positive coefficient is present for the adjusted income outcome when having a first younger brother.

Our findings suggest that both the presence of a younger sister and the gendered nature of work environments are relevant but perhaps not as significant as one might infer from the Roy self-selection theory. This may be due to other factors that determine pay scales in female-dominated jobs. Future studies can be done on more intricate family relationships or even broader aspects of society that can provide a clearer explanation of the association. Interestingly enough, our findings show a less negative impact on earnings due to having a younger sister when joining an occupation mostly composed of females, though not significant, this might indicate how family relationships, more specifically sisterhood, can affect earning in traditionally gender-defined occupations in very nuanced ways.

Interestingly, the non-significant but differently signed coefficients for elder males and females in terms of the entering into female dominated jobs and STEM related jobs respectively suggest that sibling gender composition might influence males and females differently, albeit not strongly enough to be statistically significant in our study. This could be further investigated through the lens of social role theory, which argues that gender differences in behaviour and career choices arise from societal expectations and the roles individuals are socialized into from a young age (Eagly and Wood, 2012), as well as potentially through the lens of gender socialization theory proposed by Eccles (2011) and as discussed above. Future research could examine how these social roles interact with sibling dynamics to influence career choices, potentially using qualitative methods to capture more nuanced family interactions.

The study's outcomes, particularly on the wage penalty minimization in female-dominated occupations due to having a younger sister, can be explained and understood through social capital theories. Social capital is what people gain from the relationship they have established or the ties that they have invested, in this case, with their family relationships (Coleman, 1988). Older siblings who have younger sisters may become more social, emphatic, and have better developed communication skills, which is highly appreciated in female dominated professions. This is in line with findings that suggest social competencies are becoming more crucial in the job market and can have a notable impact on income (Deming, 2017). Nonetheless, the results are indicated by their lack of statistical significance.

6.1 Limitations to the Method

A primary limitation to the study is the potential mixed gender composition of the siblings. The thesis measures the effect of having a first younger sister as opposed to a brother but does not account for the gender composition of the other younger siblings and their birth order. It solely focuses on the elder sibling and the immediate first younger sibling within a five-year time frame. If we were to limit the study on only two sibling families, a selection bias among the gender composition of the siblings emerges as we remove three and above sibling families. Considering the families that have exactly two children only, the probability of having at least one sibling of each sex is higher than if no such consideration were made. The assumption must be that parents want their children to be both sexes. Consequently, parents with one daughter and one son are more likely to stop having more children than parents with two daughters, who are more likely to try again for a third child in the hopes of having a son. Hence, this restriction

does not enable us to study the segregated impact of one sibling of a particular gender upon the elder siblings.

The NLSY 79 Children and Young Youth dataset currently conducts an interview every two years, when prior to the 1990's the interviews were every two years. This limits the granularity of the data post 1990 where most of our control variables are reported. To limit the number of missing values we ensured to cover at least two years of prior available datapoints upon the birth of the younger sibling. If one interview year was missing, we went to the prior interview round to retrieve the data and ensure the control variables has less missing datapoints. As an example, some individuals we had a datapoint for the highest grade achieved by the mother one interview round prior to the younger sibling's birth and for other it was two rounds prior.

The NLSY 79 cohort started with a specific sample of individuals born between 1957 and 1964, which may not be fully representative of later generations. The children and young adults of these respondents, whom our study focuses on, inherit this lack of representativeness, potentially limiting the generalizability of the findings to all current young adult populations. The data set is representative of the experience of people at one point in history. While it reflects the historical social circumstances surrounding the time when the data were collected, it may not be representative of change in social norms, economic circumstance and labour market relevant to the research about current or future times. Therein lies a potential limit to the validity of application of findings to the current, or future condition. Finally, findings from this data set will not transfer to other groups not within the sample, for example to people from a different culture, or to other socioeconomic statuses than those represented primarily in the NLSY79.

Another limitation to our study is that a BS, or a BA is not an exact way of measuring a choice to enter a scientific field, or an artistic field of study. The NLSY does not have a variable in the dataset regarding an exact degree, thus, a BS and BA can refer to several fields, not exclusively scientific or artistic ones. This lack of specificity may lead to misclassification and an inaccurate assessment of the influence of sibling gender composition on choosing scientific versus non-scientific careers.

Furthermore, the NLSY 79 Children and Young Youth doesn't report the start and end dates of the occupations reported. There is no link between the occupational code reported for each interview round and the employer. This creates a lack of information that might be important for our study in terms of occupational choices, namely the choice of entering a female dominated job or a STEM related job. The availability of a start to end date for each occupation

reported would've created a more accurate representation for the longest position held as a represeter for the individual in terms of the variables instead of calculating the most repeated occupational codes every two interview rounds.

As with many longitudinal studies, missing data can be a significant issue in NLSY79. Participants might not answer every question or participate in every survey wave, leading to incomplete data, this has been a recurrent issue in the thesis limiting the sampling power of the results. However, across the datasets we ensured that the missing data doesn't alter the control conditions for the treatment (having a younger sister) and control group (having a younger sister). See section 4.4 for missing data details.

6.2 Broader Implications

Our findings suggest that effects of sibling gender composition are highly contextual and should thus be viewed as a function of the broader socio-economic and societal environment. Whereas sibling gender composition itself seems to play a rather limited direct role, the significant role of parental education and household income suggests that interventions aimed at improving educational and occupational outcomes should focus on supporting families holistically.

For instance, policies that support educating the parents, particularly the mother, and economically assisting low-income families would have a very broad impact on the future educational success of children. Additionally, programs organized in schools for career counselling with regard to students should be targeted to address the wider problem of the socioeconomic context in a manner that helps smooth the social expectations and individual interest needs.

6.3 New Pathways for the Study

Non-cognitive skills such as resilience, self-control, and social skills are critical for success in education and the labour market (Heckman and Kautz, 2012). Future research could investigate how sibling gender composition influences the development of these skills, using longitudinal data to track changes over time. The NLSY 79 Children and Young Adult datasets provides several cognitive tests throughout the development phases of children. A study by Lehman et al. (2016) used the cognitive tests withing the dataset and found that later-born children tend to score lower on cognitive tests compared to their older siblings as early as age one, with this gap

widening until school entry and remaining statistically significant thereafter. A substantial portion of these birth order differences in cognitive abilities can be attributed to variations in parental behaviour, particularly the cognitive stimulation provided by mothers. These findings imply that broad shifts in parental behaviour are plausible explanations for the observed disparities in educational and labour market outcomes related to birth order. Thus, a similar pathway towards sibling gender composition rather than birth order can be further researched. Early childhood interventions aimed at enhancing parental involvement and family cohesion could be studied to see if they moderate the effects of sibling gender composition on long-term outcomes.

Furthermore, peer and school environments significantly impact educational and occupational choices (Wentzel and Caldwell, 1997). Research could examine how sibling gender composition interacts with these external influences

7. Conclusion

This thesis set out to probe into the influence of sibling gender composition on the educational and occupational outcomes of first-borns, specifically, on the effect of a first younger sister on elder siblings. The thesis utilized the comprehensive longitudinal dataset from the National Longitudinal Surveys of Youth (NLSY79) and its subsequent Child and Young Adult datasets. Our findings point out that family and socio-economic factors overshadow the direct impact of having a younger sister.

The results indicate that having a younger sister does not significantly affect the educational attainment of elder siblings or that of pursuing an artistic educational path through a Bachelor of Arts or a scientific one through a Bachelor of Science, nor through a scientific career path through a STEM related occupation. The minimal impact observed suggests that other factors, notably socio-economic background and parental education, play a more pivotal role in shaping educational outcomes which in turn points to potential similar effect towards later career choices.

In the realm of occupational outcomes, the analysis similarly found no significant influence of having a younger sister on the likelihood of entering female dominated or STEM related jobs. This finding aligns with the gender socialization theory (Eccles, 2011), which suggests that broader societal norms and expectations about gender roles are more influential in career

choices than other dynamics, namely for our study, sibling dynamics. While the coefficients for elder males and females differed in sign, suggesting potential gender-specific effects, these were not statistically significant, indicating that other socio-economic and educational factors are more decisive in occupational decisions.

The exploration of income and job stability revealed that having a younger sister does not significantly impact earnings or joblessness. However, the interaction term for female dominated jobs and having a younger sister showed a less negative coefficient, hinting that the presence of a younger sister might mitigate the typical lower compensation witnessed in female dominated job. This could be interpreted through the lens of social capital (Coleman, 1988), where enhanced interpersonal skills developed through sibling interactions might be valued in female dominated jobs. Yet, the lack of statistical significance in our findings suggests that these effects are subtle and require further investigation.

The broader implications of the study point out the need for comprehensiveness in approaches to understanding those factors that influence educational and occupational outcomes. While sibling gender composition is not associated with direct effects, the strong influence of socio-economic background and maternal education suggests the need to holistically support families. Programs related to education and career counselling consider the broader family background and socio-economic context in effectively helping individuals in their career and educational life courses.

Future research should explore the gender-specific effects of sibling composition more deeply. Additionally, examining cultural variations and the interplay of sibling dynamics with peer and school influences could provide a richer understanding of the mechanisms at play. Investigating the impact of early childhood interventions aimed at enhancing parental involvement and family cohesion could also offer valuable insights into how familial dynamics shape long-term outcomes.

In conclusion, the contribution of this thesis finds limited direct influence of sibling gender composition, however, high influence from broader socio-economic context and high influence from maternal education on one's educational and occupational later life outcomes. The findings therefore point to the complexity of family dynamics as well as the need for further research to uncover the subtle mechanisms operating in a context-dependent way to shape life courses of individuals.

8. Appendix

Table 7 Two Siblings Families Educational Outcomes Regression Results

Two Siblings Model Summary

	Dependent variable:					
	No Degree	Completed HS	Completed College	Advanced Degree	BA	BS
	(1)	(2)	(3)	(4)	(5)	(6)
Has Younger Sister	-0.01	0.01	0.03	0.01	-0.03	0.03
	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Adjusted Household Income 10k (2020)	-0.0003	0.0003	0.002	0.003^{**}	0.002	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Mother Employed Around Younger Birth	-0.09^{**}	0.09^{**}	0.10	0.02	0.003	0.07
	(0.04)	(0.04)	(0.10)	(0.07)	(0.07)	(0.08)
Mother Out Of Labor Force Around Younger Birth	-0.08^{*}	0.08^{*}	0.16^{*}	0.03	0.04	0.05
	(0.04)	(0.04)	(0.10)	(0.07)	(0.07)	(0.08)
Living With Mother Around Younger Birth	0.07	-0.07	-0.21	0.09	-0.19	-0.19
	(0.13)	(0.13)	(0.29)	(0.20)	(0.22)	(0.23)
Fathers Presence Before Younger Birth	-0.02	0.02	0.04	0.03	0.01	0.02
	(0.02)	(0.02)	(0.05)	(0.03)	(0.04)	(0.04)
Mothers Education Around Younger Birth	-0.001	0.001	0.02^{***}	0.01^{***}	0.01^{**}	0.02^{***}
	(0.002)	(0.002)	(0.01)	(0.004)	(0.004)	(0.004)
Hispanic	0.02	-0.02	-0.06	0.01	-0.08^{*}	-0.03
	(0.02)	(0.02)	(0.05)	(0.04)	(0.04)	(0.04)
Black	-0.004	0.004	-0.02	0.04	-0.10^{***}	0.03
	(0.02)	(0.02)	(0.05)	(0.03)	(0.04)	(0.04)
Male	0.03^{*}	-0.03^{*}	-0.12^{***}	-0.07^{**}	-0.04	-0.03
	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Constant	0.08	0.92^{***}	0.34	-0.19	0.27	0.09
	(0.14)	(0.14)	(0.31)	(0.21)	(0.23)	(0.25)
Observations	631	631	631	631	631	631
R^2	0.02	0.02	0.06	0.06	0.04	0.04

Table 8 Three Siblings and Above Families Educational Outcomes Regression Results

Three Siblings and Above Model Summary

	Dependent variable:					
	No Degree	Completed HS	Completed College	Advanced Degree	BA	BS
	(1)	(2)	(3)	(4)	(5)	(6)
Has Younger Sister	-0.002	0.002	-0.01	0.01	0.03	-0.06**
	(0.02)	(0.02)	(0.04)	(0.02)	(0.03)	(0.03)
Adjusted Household Income 10k (2020)	-0.01**	0.01**	0.01**	0.003	0.01***	0.01**
	(0.003)	(0.003)	(0.005)	(0.003)	(0.004)	(0.004)
Employed	-0.02	0.02	-0.02	0.002	0.04	-0.02
	(0.05)	(0.05)	(0.08)	(0.05)	(0.06)	(0.07)
Out Of Labor Force Around Younger Birth	0.01	-0.01	-0.05	-0.002	0.02	-0.03
	(0.05)	(0.05)	(0.08)	(0.05)	(0.06)	(0.06)
Living With Mother Around Younger Birth	0.16	-0.16	-0.19	-0.23*	-0.09	-0.05
	(0.10)	(0.10)	(0.17)	(0.12)	(0.13)	(0.14)
Fathers Presence Before Younger Birth	-0.01	0.01	0.10***	0.005	0.05	0.001
	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Mothers Education Around Younger Birth	-0.005	0.005	0.04***	0.01***	0.01**	0.02***
	(0.005)	(0.005)	(0.01)	(0.01)	(0.01)	(0.01)
Hispanic	0.07***	-0.07***	-0.12***	-0.03	-0.02	-0.05
	(0.03)	(0.03)	(0.05)	(0.03)	(0.04)	(0.04)
Black	0.01	-0.01	-0.17***	-0.07**	-0.02	-0.13***
	(0.03)	(0.03)	(0.05)	(0.03)	(0.04)	(0.04)
Male	0.03	-0.03	-0.18***	-0.06**	-0.06**	-0.01
	(0.02)	(0.02)	(0.04)	(0.02)	(0.03)	(0.03)
Total Siblings	0.02*	-0.02*	-0.01	-0.004	-0.01	-0.01
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Constant	-0.06	1.06***	0.20	0.20	-0.02	0.03
	(0.13)	(0.13)	(0.22)	(0.15)	(0.17)	(0.18)

Observations	627	627	627	627	627	627
R^2	0.06	0.06	0.20	0.05	0.08	0.09

Note: **p<0.1; **p<0.05; ***p<0.01*

Table 9 Same Sex Sibling Composition Dataset

Pre-Treatment Groups at Time of Birth of Younger Sibling - Complete Cases			
Characteristics of the Elder Sibling	**With Only Sisters**	**With Only Brothers**	**Mixed Gender Composition**
Number of Males	209 (55.0%)	217 (45.3%)	214 (53.6%)
Number of Females	171 (45.0%)	262 (54.7%)	185 (46.4%)
Total	**380 (30.20%)**	**479 (38.10%)**	**399 (31.72%)**
Blacks	100 (26.3%)	102 (21.3%)	99 (24.8%)
Hispanics	59 (15.5%)	100 (20.9%)	106 (26.7%)
Other	221 (58.2%)	277 (57.8%)	194 (48.6%)
Number of Observations	**1258**		
Household Variables	**Mean (SD)**		
Age in 2020	36.03 (3.45)	36.20 (3.49)	36.95 (3.29)
Age of Elder Sibling (Age Gap)	3.01 (1.22)	3.14 (1.09)	2.74 (1.22)
Household Income Adjusted for 2020 Dollars	64,215 (115,365)	62,276 (107,302)	54,783 (46,583)
Age of Mother at First Birth	23.16 (3.44)	23.03 (3.40)	22.03 (3.20)
Parents Characteristics	**Count (Percentage)**		
Mothers Completed High School or Above	333 (87.6%)	421 (87.9%)	336 (84.2%)
Mothers Completed 4th Year College or Above	58 (15.3%)	77 (16.1%)	52 (13.0%)
Fathers Present in HH any of the Prior 2 Years	271 (71.3%)	337 (70.4%)	235 (58.9%)
Mothers Employed any of the Prior 2 Years	189 (49.7%)	242 (50.5%)	167 (41.9%)
Mothers Out of Labor Force any of the Prior 2 Years	169 (44.5%)	217 (45.3%)	207 (51.9%)
Living with Mother any of the Prior 2 Years	377 (99.2%)	477 (99.5%)	394 (98.7%)

Table 10 STEM related Occupations as Reported by the U.S Census Bureau (2010)

Occupation	2010 Census	2010 SOC
Computer and information systems managers	0110	11-3021

Architectural and engineering managers	0300	11-9041
Natural sciences managers	0360	11-9121
Computer and information research scientists	1005	15-1111
Computer systems analysts	1006	15-1121
Information security analysts	1007	15-1122
Computer programmers	1010	15-1131
Software developers, applications and systems software	1020	15-113X
Web developers	1030	15-1134
Computer support specialists	1050	15-1150
Database administrators	1060	15-1141
Network and computer systems administrators	1105	15-1142
Computer network architects	1106	15-1143
Computer occupations, all other	1107	15-1199
Actuaries	1200	15-2011
Mathematicians	1210	15-2021
Operations research analysts	1220	15-2031
Statisticians	1230	15-2041
Miscellaneous mathematical science occupations	1240	15-2090
Surveyors, cartographers, and photogrammetrists	1310	17-1020
Aerospace engineers	1320	17-2011
Agricultural engineers	1330	17-2021
Biomedical engineers	1340	17-2031
Chemical engineers	1350	17-2041
Civil engineers	1360	17-2051
Computer hardware engineers	1400	17-2061
Electrical and electronics engineers	1410	17-2070
Environmental engineers	1420	17-2081
Industrial engineers, including health and safety	1430	17-2110
Marine engineers and naval architects	1440	17-2121
Materials engineers	1450	17-2131
Mechanical engineers	1460	17-2141
Mining and geological engineers, including mining safety engineers	1500	17-2151
Nuclear engineers	1510	17-2161
Petroleum engineers	1520	17-2171
Engineers, all other	1530	17-2199
Drafters	1540	17-3010
Engineering technicians, except drafters	1550	17-3020
Surveying and mapping technicians	1560	17-3031
Agricultural and food scientists	1600	19-1010
Biological scientists	1610	19-1020
Conservation scientists and foresters	1640	19-1030
Medical scientists	1650	19-1040
Life scientists, all other	1660	19-1099
Astronomers and physicists	1700	19-2010
Atmospheric and space scientists	1710	19-2021
Chemists and materials scientists	1720	19-2030
Environmental scientists and geoscientists	1740	19-2040

Physical scientists, all other	1760	19-2099
Economists	1800	19-3011
Survey researchers	1815	19-3022
Psychologists	1820	19-3030
Sociologists	1830	19-3041
Urban and regional planners	1840	19-3051
Miscellaneous social scientists and related workers	1860	19-3090
Agricultural and food science technicians	1900	19-4011
Biological technicians	1910	19-4021
Chemical technicians	1920	19-4031
Geological and petroleum technicians	1930	19-4041
Nuclear technicians	1940	19-4051
Social science research assistants	1950	19-4061
Miscellaneous life, physical, and social science technicians	1965	19-4090
Sales engineers	4930	41-9031
Medical and health services managers	0350	11-9111
Architects, except naval	1300	17-1010
Chiropractors	3000	29-1011
Dentists	3010	29-1020
Dietitians and nutritionists	3030	29-1031
Optometrists	3040	29-1041
Pharmacists	3050	29-1051
Physicians and surgeons	3060	29-1060
Physician assistants	3110	29-1071
Podiatrists	3120	29-1081
Audiologists	3140	29-1181
Occupational therapists	3150	29-1122
Physical therapists	3160	29-1123
Radiation therapists	3200	29-1124
Recreational therapists	3210	29-1125
Respiratory therapists	3220	29-1126
Speech-language pathologists	3230	29-1127
Exercise physiologists	3235	29-1128
Therapists, all other	3245	29-1129
Veterinarians	3250	29-1131
Registered nurses	3255	29-1141
Nurse anesthetists	3256	29-1151
Nurse midwives	3257	29-1161
Nurse practitioners	3258	29-1171
Health diagnosing and treating practitioners, all other	3260	29-1199
Clinical laboratory technologists and technicians	3300	29-2010
Dental hygienists	3310	29-2021
Diagnostic related technologists and technicians	3320	29-2030
Emergency medical technicians and paramedics	3400	29-2041
Health practitioner support technologists and technicians	3420	29-2050
Licensed practical and licensed vocational nurses	3500	29-2061
Medical records and health information technicians	3510	29-2071
Opticians, dispensing	3520	29-2081

Miscellaneous health technologists and technicians	3535	29-2090
Other healthcare practitioners and technical occupations	3540	29-9000

References

- Aakvik, A., Salvanes, K., and Vaage, K. (2005). Educational attainment and family background. German Economic Review, 6(3), 377-394.

- Abe, I. and Chikoko, V. (2020). Exploring the factors that influence the career decision of stem students at a university in south africa. International Journal of Stem Education, 7(1).

- Anelli, M. and Peri, G. (2014). Gender of siblings and choice of college major. Cesifo Economic Studies, 61(1), 53-71.

- Arcidiacono, P., Hotz, V., Maurel, A., and Romano, T. (2014). Recovering ex ante returns and preferences for occupations using subjective expectations data. SSRN Electronic Journal.

- Bandura, A., Barbaranelli, C., Caprara, G., and Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. Child Development, 72(1), 187-206.

- Berry, S. (1994). Estimating discrete-choice models of product differentiation. The Rand Journal of Economics, 25(2), 242.

- Bi, Y. (2023). The relationship between professional self-concept and career decision-making difficulties among postgraduate nursing students in china: the mediating role of career decision-making self-efficacy. Frontiers in Psychology, 14.

- Blau, F. D., and Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. Journal of Economic Literature, 55(3), 789-865

- Cassidy, T., Wright, E., and Noon, E. (2014). Family structure and psychological health in young adults. Psychology, 05(10), 1165-1174.

- Charles, M., and Grusky, D. B. (2004). Occupational Ghettos: The Worldwide Segregation of Women and Men. Stanford University Press.

- Chui, H., Hui, L., and Ngo, H. (2020). Linking protean career orientation with career optimism: career adaptability and career decision self-efficacy as mediators. Journal of Career Development, 49(1), 161-173.

- Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. American Journal of Sociology, 94, S95-S120.

- Connelly, R. (1989). Occupational choice under uncertainty when experience is a determinant of earnings. Econometrica, 57(5), 1215.

- Cools, A. and Patacchini, E. (2017). Sibling gender composition and women's wages. SSRN Electronic Journal.

- Creed, P., Patton, W., and Bartrum, D. (2003). Internal and external barriers, cognitive style, and the career development variables of focus and indecision. Journal of Career Development, 30(4), 277-294.

- Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. Journal of Family Psychology, 19(2), 294-304.

- Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. Quarterly Journal of Economics, 132(4), 1593-1640.

- Eagly, A. H., and Wood, W. (2012). Social Role Theory. In P. A. M. Van Lange, A. W. Kruglanski, and E. T. Higgins (Eds.), Handbook of Theories of Social Psychology (Vol. 2, pp. 458-476). SAGE Publications.

- Eccles, J. S. (2011). Gendered Educational and Occupational Choices: Applying the Eccles et al. Model of Achievement-Related Choices. International Journal of Behavioral Development, 35(3), 195-201.

- Ermisch, J. and Francesconi, M. (2001). Family matters: impacts of family background on educational attainments. Economica, 68(270), 137-156.

- Feliciano, C., and Rumbaut, R. (2005). Gendered paths: Educational and occupational expectations and outcomes among adult children of immigrants. Ethnic and Racial Studies, 28, 1087-1118.

- Furstenberg, F. F., Brooks-Gunn, J., and Morgan, S. P. (1987). Adolescent Mothers in Later Life. Cambridge University Press.

- Glynn, S. J., and Boesch, D. (2022). Connecting the Dots: "Women's Work" and the Wage Gap. U.S. Department of Labor Blog.

- Hardie, J. H. (2015). Women's Work? Predictors of Young Men's Aspirations for Entering Traditionally Female dominated Occupations. Sex Roles, 72, 349-362.

- Haveman, R., and Wolfe, B. (1995). The Determinants of Children's Attainments: A Review of Methods and Findings. Journal of Economic Literature, 33(4), 1829-1878.

- Heckman, J. J., and Kautz, T. (2012). Hard Evidence on Soft Skills. Labour Economics, 19(4), 451-464.

- Holland, J. L. (1997). Making vocational choices: A theory of vocational personalities and work environments (3rd ed.). Psychological Assessment Resources.

- Hung-Lin Tao, Hui-Pei Cheng (2022), Parental and sibling influence on study field choice: Gender-stereotypical or field preference transmission, Journal of Asian Economics.

- Lehmann, J. K., Nuevo-Chiquero, A., and Vidal-Fernández, M. (2016). The early origins of birth order differences in children's outcomes and parental behavior. Journal of Human Resources, 53(1), 123-156.

- Martin, M. (2012). Family structure and the intergenerational transmission of educational advantage. Social Science Research, 41(1), 33-47.

- Ngware, M. W., Ciera, J., Musyoka, P., and Oketch, M. (2009). Family structure and educational attainment in the slums of Nairobi. International Journal of Educational Development, 29(5), 444-453.

- Palffy, P., Lehnert, P., and Backes-Gellner, U. (2023). Social norms and gendered occupational choices of men and women: time to turn the tide?. Industrial Relations a Journal of Economy and Society, 62(4), 380-410.

- Pang, L., Xiu, W., Liu, F., Fang, T., Chen, H., and Wen, Y. (2021). The relationship between college students' resilience and career decision-making difficulties: the mediating role of career adaptability. Psychology, 12(06), 872-886.

- Peter, N., Lundborg, P., Mikkelsen, S., and Webbink, D. (2018). The effect of a sibling's gender on earnings and family formation. Labour Economics, 54, 61-78.

- Petts, R. J. (2009). Family and Religious Characteristics' Influence on Delinquency Trajectories from Adolescence to Young Adulthood. *American Sociological Review, 74*, 465-483.

- Pew Research Center. (2022). STEM Jobs See Uneven Progress in Increasing Gender, Racial and Ethnic Diversity.

- Price, S. (2008). Becoming a nurse: a meta-study of early professional socialization and career choice in nursing. Journal of Advanced Nursing, 65(1), 11-19.

- Rosen, S. (1986). The theory of equalizing differences. In O. Ashenfelter and R. Layard (Eds.), Handbook of Labor Economics, Volume 1 (pp. 641-692). Elsevier Science Publishers.

- Roy, A. D. (1951). Some thoughts on the distribution of earnings. Oxford Economic Papers, 3(2), 135-146.

- Shahbazian, R. (2021). Under the influence of our older brother and sister: The association between sibling gender configuration and STEM degrees. Social science research, 97, 102558.

- Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rogers, and Megan Schouweiler. IPUMS USA: Version 15.0 [2009 ACS 5yr, 2014 ACS 5yr, and 2019 5yr]. Minneapolis, MN: IPUMS, 2024.

- U.S. Census Bureau. (2010). STEM, STEM-related, and Non-STEM Occupation Code List 2010.

- Wentzel, K. R., and Caldwell, K. (1997). Friendships, Peer Acceptance, and Group Membership: Relations to Academic Achievement in Middle School. Child Development, 68(6), 1198-1209.

- Whiston, S. and Keller, B. (2004). The influences of the family of origin on career development. The Counseling Psychologist, 32(4), 493-568.

- Yilmaz, F. and Gündüz, H. (2018). Career indecision and career anxiety in high school students: an investigation through structural equation modelling. Eurasian Journal of Educational Research, 18(78), 1-20.