FEDERAL UNIVERSITY OF ITAJUBÁ - UNIFEI
PROGRAM OF
ELECTRICAL ENGINEERING

Estimating the Family Bias to Autism: A Bayesian Approach

Emerson Assis de Carvalho

Itajubá/MG, Brazil, Feb., 2022

FEDERAL UNIVERSITY OF ITAJUBÁ - UNIFEI
PROGRAM OF
ELECTRICAL ENGINEERING

Emerson Assis de Carvalho

Estimating the Family Bias to Autism: A Bayesian Approach

Doctoral Dissertation presented to the Program of Electrical Engineering in partial fulfillment of the requirements for the Degree of Doctor in Science in Electrical Engineering.

Field: Automation and Industrial Electrical Systems

Advisor: Prof. Dr. Guilherme Sousa Bastos

Feb., 2022
Itajubá/MG, Brazil

Emerson Assis de Carvalho

Estimating the Family Bias to Autism: A Bayesian Approach

Doctoral Dissertation presented to the Program of Electrical Engineering in partial fulfillment of the requirements for the Degree of Doctor in Science in Electrical Engineering.

Approved. Itajubá/MG, Brazil, Feb. 21, 2022:

Prof. Dr. Roberto Hirochi Herai

Prof. Dr. Lucelmo Lacerda de Brito

Prof. Dr. Ricardo Zorzetto N. Vêncio

Prof. Dr. Edmilson Marmo Moreira

Prof. Dr. Joao Paulo Reus R. Leite

Prof. Dr. Guilherme Sousa Bastos
Advisor

Itajubá/MG, Brazil
Feb., 2022

Acknowledgements

First, I would like to thank my advisor, professor Dr. Guilherme Sousa Bastos, for supporting me over these years and giving me so much freedom to explore both areas of autism and technology. I also must thank him for sharing such personal experiences, allowing me to work with such a relevant topic, and for becoming a great friend. Finally, I would like to congratulate him on being the best father that the lovely Alvina could have.

My other committee members have also been very supportive. Lucelmo Lacerda de Brito has long provided several reviews and has kept me attentive about all the details of autism. Professor Dr. Edmilson Marmo Moreira has long been an inspiration since my graduate degree. His classes have proved to be one of my best learning experiences. Special thanks for the collaboration of professors Dr. João Paulo Reus Rodrigues Leite, Dr. Ricardo Zorzetto Nicoliello Vêncio, and Dr. Roberto Hirochi Herai.

I want to thank my many colleagues at Unifei with whom I have enjoyed working over these years. These include Caio Pinheiro Santana, Gustavo Leite Lopes, José Renato Castro Milanez, Luiz Fernando Nunes, Vinícius de Almeida Paiva, and all the Automation and Information Technology Group (GATI) members.

In particular, I would like to thank my friends Fábio Júnior Alves, Igor Rodrigues Duarte, Lênio Oliveira Prado Júnior, and Ricardo Emerson Julio for providing a welcome distraction from work, for their help, friendship, and support during the development of this project. I am very grateful to all and for each one of these people.

Finally, I would like to thank my parents, Rosânia Fátima de Carvalho and Noé Lourenço de Carvalho, and my daughter Luiza Oliveira Carvalho for putting up with my absences, for listening about my work, which they often did not understand, for giving me the motivation to finish this thesis, and for all love they offer to me, especially my little girl.

"If you can dream it, you can do it. Always remember that this whole thing was started with a dream and a mouse."

(Walt Disney)

Abstract

Autism is an age- and sex-related lifelong neurodevelopmental condition characterized primarily by persistent deficits in core domains such as social communication. It is estimated that $\approx 2\%$ of children have some ASD trait. The autism etiology is mainly due to inherited genetic factors (>80%). The importance of early diagnosis and interventions motivated several studies involving groups at high risk for ASD, those with a greater predisposition to the disorder. Such studies are characterized by evaluating some characteristics of the individual itself or the family members of diagnosed individuals, mainly aiming to predict a future diagnosis or recurrence rates. One of the primary goals of Artificial Intelligence is to create artificial agents capable of intelligent behaviors, such as prediction problems. Prediction problems usually involve reasoning with uncertainty due to some information deficiency, in which the data may be imprecise or incorrect. Such solutions may seek the application of probabilistic methods to construct inference models. In this thesis, we will discuss the development of probabilistic networks capable of estimating the risk of autism among the family members given some evidence (e.g., other family members with ASD).

In particular, the main novel contributions of this thesis are as follows: the proposal of some estimates regarding parents with ASD generating children with ASD; the highlighting regarding the decrease in the ASD prevalence sex ratio among males and females when genetic factors are taken into account; the corroboration and quantification of past evidence that the clustering of ASD in families is primarily due to genetic factors; the computation of some estimates regarding the risk of ASD for parents, grandparents, and siblings; an estimate regarding the number of ASD cases in a family sufficient to attribute the ASD occurrences to the genetic inheritance; the assessment of some estimates for males and females individuals given evidence in grandparents, aunts-or-uncles, nieces-or-nephews and cousins; and the proposition of some estimates indicating risk ranges for ASD by genetic similarity.

Key-words: Autism Spectrum Disorder Prevalence. Autism Spectrum Disorder Etiology. Probabilistic Graphical Models, Bayesian Networks, Markov Models.

List of Tables

List of Figures

List of Abbreviations and Acronyms

ADI	Autism Diagnostic Interview
ADI-R	ADI–Revised
ADOS	Autism Diagnostic Observation Schedule
ADOS-G	ADOS-Generic
ADOS-T	ADOS-Toddler Module
AI	Artificial Intelligence
APA	American Psychiatric Association
API	Application Programming Interface
ARR	Absolute Recurrence Risk
ASD	Autism Spectrum Disorder
ASSQ	Autism Spectrum Screening Questionnaire
BN	Bayesian Network
CNV	Copy Number Variation
CPQ	Conditional Probability Query
CPT	Conditional Probability Table
CTM	Classical Twin Model
DAG	Directed Acyclic Graph
DAWBA	Development and Wellbeing Assessment
DBN	Dynamic Bayesian Network
DNA	Deoxyribonucleic Acid
DSM	Diagnostic and Statistical Manual of Mental Disorders

DSM-III	DSM, Third Edition
DSM-III-R	DSM-III, Revised
DSM-IV	DSM, Fourth Edition
DSM-IV-TR	DSM-IV – Text Revision
DSM-V	DSM, Fifth Edition
FM	Falconer Model
GLMM	Generalized Linear Mixed Model
HMM	Hidden Markov Model
ICD	International Statistical Classification of Diseases and Related Health Problems
ICD-8	ICD, Eight Revision
ICD-9	ICD, Ninth Revision
ICD-10	ICD, Tenth Revision
LMM	Linear Mixed Model
LTM	Liability Threshold Model
LUI	Language Use Inventory
MAP	Maximum a Posteriori
MC	Markov Chain
MM	Markov Model
MPE	Most Probable Explanation
MSEL	Mullen Scales of Early Learning
PDD	Pervasive Developmental Disorder
PDD-NOS	Pervasive Developmental Disorder Not Otherwise Specified

PRS	Polygenic Risk Score
RRR	Relative Recurrence Risk
SCQ	Social Communication Questionnaire
SEM	Structural Equation Modeling
SNP	Single Nucleotide Polymorphisms
SP	Stochastic Process
STAT	Screening Tool for Autism in Two-Year-Olds
TD	Typical Development
VABS	Vineland Adaptive Behavior Scales

Table of Contents

1 Introduction

1.1 Motivation

Composed by the Greek words “autos” (self) and “ismos” (action), the term autism was initially borrowed from the description of schizophrenia given by Eugene Bleuler in 1911, characterizing the withdrawal from reality (SZATMARI, 2000). In 1943, Leo Kanner adopted the term autism to describe children with an extreme inability to relate to others, primarily due to severe difficulties in using language to communicate (KANNER et al., 1943). In 2013, after a list of nomenclatures and classifications, the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) included autism in a section named Autism Spectrum Disorder (ASD) (APA, 2013). From then on, ASD is described as an age- and sex-related lifelong neurodevelopmental disorder characterized primarily by persistent deficits in core domains as social communication across multiple contexts, in addition to restricted and repetitive patterns of behavior, interests, or activities (WANG et al., 2018; APA, 2013; RAPIN; TUCHMAN, 2008).

The ASD symptoms are mainly manifested in the early developmental period, under three years of age, and impair the individuals’ everyday lives. ASD has a wide range of severity. It is dimensionally defined, with borders that overlap normality on the one hand and profound intellectual impairment caused by brain malfunctions on the other hand. Thus, people who meet the criteria for ASD are diagnosed as having ASD Level 1, ASD Level 2, or ASD Level 3. These three levels are based on a person’s strengths and limitations and indicate how much support ASD people need. Level 1 indicates Requiring Support; Level 2 indicates Requiring Substantial Support, and Level 3 indicates Requiring Very Substantial Support (APA, 2013).

The most recent prevalence estimates indicate that we are witnessing an increase in positive ASD diagnoses. It was estimated one case for every 150 United States children in 2000, whereas it was estimated one case for every 54 children in 2016 (MAENNER et al., 2020), and one case for every 44 children in 2018 (MAENNER et al., 2021). Estimated increases in the ASD diagnoses, from 6-7% to 15% per year, make ASD the fastest-growing developmental disability in the United States (BONIS, 2016; ÖZERK, 2016). Although, a global report highlights that the prevalence increase may be affected by changes in diagnostic concepts, service availability, and awareness about the disorder (ELSABBAGH; JOHNSON, 2010). Some statistics from other countries indicate one case for every 64 children in the United Kingdom (BARON-COHEN et al., 2009), one case for every 38 children in South Korea (KIM et al., 2011), one case for every 52 children in Sweden (XIE et al., 2019), one case for every 83 children in a multi-national population-based study[1] (HANSEN et al., 2019), and one case for every 160 global children, no matter the racial, ethnic or socioeconomic characteristics (ELSABBAGH et al., 2012). Although these estimates range widely between genders, ASD is about three to four times more common among boys than girls (BAIO et al., 2018; LOOMES; HULL; MANDY, 2017).

The economic cost estimates of ASD on individuals, their families, and society are substantial, with a high financial burden in several domains such as medical, healthcare, therapeutic, special education, and productivity loss (ROGGE; JANSSEN, 2019; KOGAN et al., 2008). The amount of money invested in supporting ASD individuals during his/her lifetime is about US$1.4 million in the United States and the United Kingdom. The costs with ASD in these countries can reach US$2.4 million and US$2.2 million if the individuals present intellectual disability. Such charges include special education services and parental productivity loss when children, plus special care, sheltered work, and individual productivity loss when adults (BUESCHER et al., 2014). Children and adolescents with ASD have medical expenses up to 6.2 times greater than those with Typical Development (TD), with general costs from 8.4 to 9.5 times greater than the average (SHIMABUKURO; GROSSE; RICE, 2008). Leigh and Du (2015) estimated an annual total cost of US$268 billion with direct medical, non-medical, and productivity loss in the United States for 2015, forecasting an expense of US$461 billion for 2025.

In addition to medical expenses, intensive behavioral interventions needed for the ASD treatment has costs from US$40 thousand to US$60 thousand per child per year (AMENDAH et al., 2011). ASD Children are more likely to have more significant health care needs and difficulties accessing health care than children with other emotional or behavioral disorders (e.g., anxiety, depression, attention-deficit, hyperactivity disorder, developmental delay, Down syndrome, intellectual disability, learning disability) and children without these conditions, even in high-income countries (KOGAN et al., 2018).

[1] California, Denmark, Finland, Israel, Sweden and Western Australia

Besides the economic costs, parents of children with ASD show higher stress levels than other parents. Parental emotional overload is one of the main difficulties suffered by families. The main stress factors for families with ASD children are diagnostic delay, trouble dealing with the diagnosis, and reduced access to health assistance and social support (GOMES et al., 2015). The daily care tasks affect all aspects of the children's life and the parent's mental health (BONIS, 2016). The stress level experienced by mothers of teenagers with ASD is comparable to that of combat soldiers or parents of children with cancer (SELTZER et al., 2010).

Early diagnosis and proper interventions are critical factors in improving autistic behaviors in children. Early treatments may result in improved communication and social interaction, stereotyped and repetitive behaviors, and fixed and restricted interests, allowing an evolution that may lead to healthy adult life, as well as significant long-term societal costs reductions (LANDA, 2018; HAZLETT et al., 2017; EMERSON et al., 2017; ZWAIGENBAUM et al., 2015). The importance of early diagnosis motivated several studies regarding groups of people at high risk for ASD, those with a greater predisposition to the disorder. Such studies are characterized by evaluating some characteristics of the individual itself that could predict a future diagnosis, the family members of diagnosed individuals, the most common are siblings, in addition to extensive studies to explain the ASD etiology.

Despite the importance of early diagnosis and interventions, there are no low-cost automated tests to identify the disorder. The ASD diagnosis is performed through clinical observation, which is challenging to accomplish in young children (Brazil's Ministry of Health, 2014). Due to ASD heterogeneity, the process involving its cycle of diagnoses and treatment is not at a very advanced stage of effectiveness. Consequently, there are limited intervention options to improve the ASD core symptoms, including mental or medical comorbidities (MASI et al., 2017).

The ASD has a multi-factorial etiology: neurobiological, genetic, and environmental (LYALL et al., 2017). In addition to other causes, several studies have related some parents' characteristics or the gestation environment with an ASD risk increase in their descendants. However, the most evident characteristic concerning the ASD risk increase is, for now, some genetic factors (GROVE et al., 2019; IAKOUCHEVA; MUOTRI; SEBAT, 2019; WANG et al., 2017; SANDIN et al., 2016; TICK et al., 2016; SANDIN et al., 2014; HALLMAYER et al., 2011). According to Almandil et al. (2019), Bai et al. (2019), Sandin et al. (2017) and Kroncke, Willard and Huckabee (2016), from 80% to 90% of the ASD cases is caused by hereditary factors, with a small environmental contribution. In early 2020, an analysis of the genetic sequencing of more than 35 thousand people with ASD, and their families, identified 102 genes as the primary genes associated with autism (SATTERSTROM et al., 2020).

All these aspects motivated several studies that involve the use of the technology applied to autism, such as to characterize autistic symptoms (PALMER; LAWSON; HOHWY, 2017), diagnostic methods (HEINSFELD et al., 2018; BHAUMIK et al., 2018; KHOSLA et al., 2018; LIAO; LU, 2018; ZHAO et al., 2018; DVORNEK; VENTOLA; DUNCAN, 2018; DEKHIL et al., 2018b; DEKHIL et al., 2018a; HAZLETT et al., 2017; EMERSON et al., 2017; DVORNEK et al., 2017; YAHATA et al., 2016), the use of robots and other Artificial Intelligence (AI) techniques applied to the therapy processes (ALVES et al., 2020), early prediction approaches (BUSSU et al., 2018), and several studies involving groups at high risk for ASD (Sections 2.4, 2.5 and 2.6).

Nowadays, science and technology are omnipresent in our everyday lives, becoming the new basis for belief and, together, bring new ways to improve the quality of life of our entire society (FEENBERG, 2006). However, most real-world events are unpredictable, demanding that intelligent applications need to handle partial observability, nondeterminism, or any eventuality (RUSSELL; NORVIG, 2020). The uncertainty arises from some information deficiency. Thus the information may be incomplete, imprecise, incorrect, or contradictory (KLIR, 2006).

AI is the study of intelligent behaviors. Its primary goal is a theory of intelligence that explains the behavior of natural intelligent entities and guides the creation of artificial agents capable of intelligent behaviors, such as prediction problems and decision support. Prediction problems usually involve reasoning with uncertainty. Reasoning under uncertainty has a long history and is a significant issue in AI. Many problems require solutions for a better decision-making process. Such solutions may seek the application of probabilistic methods to construct inference models (RUSSELL; NORVIG, 2020; GENESERETH; NILSSON, 2012).

Probabilistic methods may involve, for example, graphical probabilistic models such as Bayesian Networks (BNs) and Markov Models (MMs), which are among the best methods for reasoning about uncertainty (NEIL; FENTON; NIELSON, 2000). These probabilistic networks allow inter-causal reasoning, a vital aspect that distinguishes them from other inference systems (KJAERULFF; MADSEN, 2013). In inter-causal reasoning, taking evidence about a hypothesis decreases the belief in the competing unobserved hypotheses automatically (KJAERULFF; MADSEN, 2013), establishing a safe and complete inference mechanism (PEARL, 1988).

BNs are graphical representations of causal relationships in a particular domain (HOLMES; JAIN, 2008). A BN is a directed and acyclic graph in which each node corresponds to a random variable. Directed edges connect pairs of nodes, indicating a direct influence of one node over the other. In addition, each node has a conditional probability table that quantifies the effect of its parent nodes over it. By using inference algorithms over BNs, it is possible to estimate beliefs in the context of observed pieces of evidence (RUSSELL; NORVIG, 2020). Employing rigorous formalism and practical algorithms for probabilistic reasoning, BNs support any reasoning with causal variables, such as diagnosis, prediction, or causal explanation (RUSSELL; NORVIG, 2020; WILLIAMSON, 2002).

BNs have applications in engineering such as monitoring power generators (MORJARIA; SANTOSA, 1996), displaying time-critical information at NASA's mission control systems (HORVITZ; BARRY, 1995), the field of network tomography (CASTRO et al., 2004), and diagnosis-and-repair tools (BREESE; HECKERMAN, 1996; HORVITZ et al., 1998). BNs also have practical applications for medicine (SAHEKI, 2005), such as diagnosing (ANDERSEN et al., 1989; POURRET; NAÏM; MARCOT, 2008), pathology finder (HECKERMAN, 1990), genetic models (SILBERSTEIN et al., 2013; POURRET; NAÏM; MARCOT, 2008), and clinical support (POURRET; NAÏM; MARCOT, 2008).

Hidden Markov Models (HMMs) are a double stochastic process, with a non-visible stochastic process (not observable) that can be observed/predicted through another stochastic process that produces the sequence of observations. The hidden processes are a set of states connected by transitions with probabilities. In contrast, the observable (non-hidden) processes are outputs or observable states, each one emitted by each not observable state according to some output of a probability density function. HMMs allow the designing of systems to predict a sequence of related states through a sequence of observations (RABINER, 1989). For example, HMMs were used to propose data-driven tools to predict Power Quality disturbance based on past weather conditions (XIAO; AI, 2018).

Also used for modeling several different problems in medical researches (KROGH; MIAN; HAUSSLER, 1994; MEYER; DURBIN, 2002; TESTA et al., 2015), HMMs have been applied in several different areas of AI, such as Computer Vision (GHAHRAMANI, 2001), Robotics (BERG et al., 2018), Speech and Face Recognition (MUSTAFA; ALLEN; APPIAH, 2019; RAHUL et al., 2019), and Computational Biology (TAMPOSIS et al., 2019).

The causal nature of the ASD etiology combined with the possibility of structuring the probabilistic networks as trees (like a pedigree chart) motivates the probabilistic graphical models as a reasonable alternative to investigate the familiar bias to ASD.

1.2 Research Question

According to the assumptions/propositions presented in which: (1) there is an increase in the ASD prevalence/diagnosis nowadays; (2) early diagnosis leads to better outcomes for autism treatment and to a long term individual and societal costs reduction, which corroborates the importance of investigating groups at high risk for ASD; (3) the genetic nature of the ASD etiology with high heritability estimates, which makes it predominantly causal (from parents to children); and (4) the ability of probabilistic networks to build transparent and efficient inference models in inter-causal domains combined with the availability of statistical data in the literature, especially related to the prevalence, recurrence, and heritability of ASD; the problem addressed by this work is: *are probabilistic networks a suitable approach to model the family bias to autism*?

We intend to evaluate whether probabilistic networks are a suitable approach by:

- Simulation: in which we will validate the proposed inference model results compared to the ASD heritability estimates and the ASD recurrence rate among siblings available in the state-of-the-art literature; and

- Quality Analysis: in which we will analyze the quality of the proposed model according to the construction techniques of BNs.

Thus, we will consider probabilistic networks a suitable model if they estimate the likelihood of autism in family members with a proper efficiency concerning the known ASD prevalence and recurrence rates.

1.3 Objectives

This work's primary objective is to develop probabilistic networks capable of estimating the risk of autism among family members. Given some evidence, for example, the ASD diagnosis of one family member, these models will estimate the risk of ASD among other family members.

To achieve our primary goal, we also accomplish other steps:

1. Gather, assess, estimate, and summarize statistical information regarding the ASD prevalence;
2. Gather, assess, estimate, and summarize statistical information regarding the ASD etiology, specially heritability and recurrence rates among siblings;
3. Propose a probabilistic model to infer the general probability of parents with ASD characteristics generating ASD children;
4. Propose a causal probabilistic model to infer the probabilities of ASD in family members, given some evidence of ASD in the family genealogical tree;
5. Evaluate the proposed models with the ASD heritability and recurrence data existing in the state-of-the-art literature; and
6. Introduce some ASD probabilities estimates from predefined family compositions to alert about the likelihood of ASD in other family members, both below and above in the family tree.

1.4 Methodology

Most surveys regarding ASD etiology and recurrence rates explored first-degree relatives, especially twins, full and half-siblings. Few surveys worked with second-degree relatives, such as grandparents and cousins. Despite presenting deeply relevant results, these studies have general and static probabilities. For example, it is almost a consensus that an autistic person has inherited the condition in $\approx 80\%$ of the cases. However, to the best of our knowledge, no estimate says how likely a person is to be autistic, given that it has one, or maybe two, diagnosed maternal cousins.

We defined a Model approach for this work, a broadly used methodology to define an abstract model for a natural system. Model approaches are usually used in combination with other methods, such as experimental methodologies. Experiments based on a model are named simulations. A model checking might be necessary if there is a formal description of the model to verify the system's functionality or correctness (AMARAL et al., 2011).

As a model and experimental methodology, this work was divided into three main phases: an exploratory phase, a building phase, and an evaluation phase:

- In the exploratory phase, non-systematic bibliographic reviews will gather results from previous works that 1) studied the association of genetics and environmental factors with the ASD risk; 2) studied the ASD recurrence risk among siblings and other family members; and 3) studied the association of other mental and neurological disorders with the ASD risk (Chapter 2), we also will explore the most suitable probabilistic graphical models (Chapter 3);
- In the modeling phase, the statistics gathered in the previous phase will base both the study and building process of proper probabilistic models capable of inferring the ASD risk given some specified family compositions (Chapters 4 and 5);
- In the evaluation phase, simulations of familial structure will estimate specific probabilities that we must validate based on the empirical research outcomes gathered in the initial phase (Chapters 6 and 7).

1.5 Thesis Outline

The general outline and the dependence among the chapters of this thesis are depicted in Figure 1.

Chapter 2 brings a background on ASD, emphasizing its classification history, prevalence rates, and etiology nature. We first introduce autism and some of its main

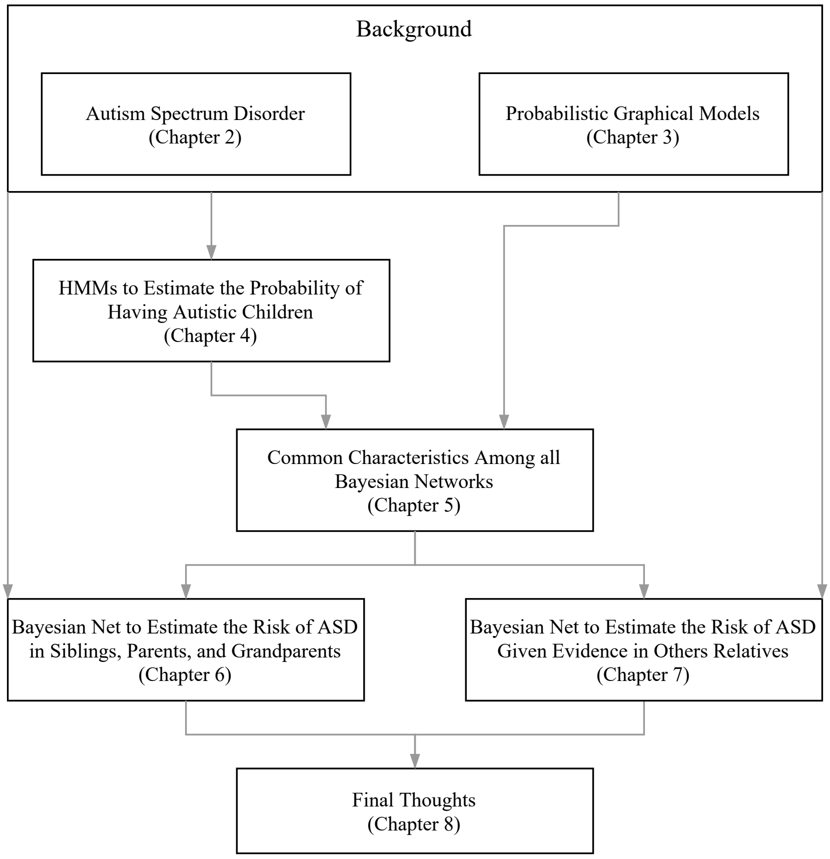

Background
Autism Spectrum Disorder
(Chapter 2)
Probabilistic Graphical Models
(Chapter 3)
HMMs to Estimate the Probability of
Having Autistic Children
(Chapter 4)
Common Characteristics Among all
Bayesian Networks
(Chapter 5)
Bayesian Net to Estimate the Risk of ASD
in Siblings, Parents, and Grandparents
(Chapter 6)
Bayesian Net to Estimate the Risk of ASD
Given Evidence in Others Relatives
(Chapter 7)
Final Thoughts
(Chapter 8)

Figure 1 – Outline and dependence among the chapters of this thesis.

classification. We then introduce the disorder prevalence data, which contains essential prior probabilities to model the inference system. We then investigate the multifactorial ASD etiology, in which genetic and environmental factors and their interactions contribute to ASD risk factors. The chapter provides essential statistical information related to the ASD etiology, such as 1) genetic factors (additive and dominant); 2) environmental factors (shared or non-shared); and 3) recurrence rates among siblings. These statistical data constitute the probabilistic basis for several conditional probabilities necessary to construct the proposed models.

Chapter 3 provides an overview of probabilistic models applied to reasoning under environments of uncertainty. It introduces some necessary concepts of probability theory and presents two of the main probabilistic graphical approaches. Section 3.4 introduces the BNs, highlighting their syntax, semantics, and methods of learning and inference. Section 3.5 introduces the MMs, highlighting the HMMs and their ability to model conditional dependencies of hidden states. Also, the chapter presents the justification to use these approaches to model causal domains. Moreover, the chapter shows related works on AI applied to medical researches.

Chapter 4 presents our proposed HMM used to estimate the likelihood of ASD parents generating ASD children. Once genetic factors have been pointed out as the primary root associated with the ASD risk, we used HMMs in conjunction with the ASD heritability and recurrence rates among siblings to develop a model capable of estimating the potential causal effect of ASD parents regarding their children. The chapter presents our assumption regarding the statistical information used, as well as the model building and validation process.

Chapter 5 introduces the design process used to model our BNs. The design process of a BN requires a well-defined problem to be solved, careful identification and selection of the relevant variables, a detailed description of independence relationships among the selected variables, and a proper elicitation of the required prior and conditional probabilities. The chapter presents and describes all aleatory variables we used, the domain of the variables, and the base set of conditional and prior probabilities.

Chapters 6 and 7 define the problems to be investigated, which are the risk of ASD in siblings, parents, grandparents, grandchildren, aunts, uncles, nieces, nephews, and cousins. Then, they present the networks structures, the results of the inferences performed, and discuss such results examining the literature.

Finally, Chapter 8 summarizes the achievements of our work, discussing its strengths, limitations, and future works.

2 Autism Spectrum Disorder

Autism and ASD are general terms for a collection of complex neurodevelopmental disorders earlier classified as distinct subtypes (e.g., Autistic Disorder, Childhood Disintegrative Disorder, Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS), and Asperger Syndrome). Autism is a complex lifelong neurodevelopmental disability currently merged into an umbrella of ASD diagnosis. Symptoms typically emerge during early childhood (between 1 and 3 years of age) and affect communication and interaction with others. Defined by a particular set of behaviors, autism is considered a spectral condition that affects individuals differently and in varying degrees (SOCIETY, 2020).

The first autism studies started at the beginning of the 20th-century (SZATMARI, 2000; KANNER et al., 1943). Since then, its classification, prevalence, recurrence rates, and etiology have undergone many changes over time, especially in the last few decades.

2.1 Autism Spectrum Disorder Classification

The category of autism diagnosis was not immediately recognized as a distinct category. The Diagnostic and Statistical Manual of Mental Disorders (DSM) of the American Psychiatric Association (APA) has already classified autism as:

- A psychiatric condition, an autistic sign in children with psychosis marked by a detachment from reality (DSM-I) (ASSOCIATION, 1952);
- Infantile schizophrenia, understood as a behavior of schizophrenia in childhood (DSM-II) (ASSOCIATION, 1968);
- A syndrome within the Global Developmental Disorders established with its separate diagnosis and described as a Pervasive Developmental Disorder (PDD), distinct from schizophrenia (DSM-III) (ASSOCIATION, 1980); and
- Invasive Developmental Disorders, characterized by a triad: impaired communication, impaired social interaction, and stereotyped and repetitive behavior, and further divided into sub-conditions such as Classic Autism, Rett Syndrome, Asperger's

Syndrome, Childhood Disintegrative Disorder, and Global Development Disorder with no other specification (DSM-IV) (ASSOCIATION, 1994).

The DSM fifth edition (DSM-V) (APA, 2013) named the disorder as Autism Spectrum Disorder. ASD was defined as a neurodevelopmental disorder categorized by the dyad:

- **Communication and social interaction**: showing deficits in socioemotional reciprocity, non-verbal communicative behaviors, and in the development, maintenance, and understanding of relationships; and
- **Repetitive and stereotyped behaviors with fixed and restricted interests**: showing deficits in motor movements, stereotyped or repetitive speech or use of objects, fixed and highly restricted interests that are abnormal in intensity or focus, strong adherence to routines, ritualized patterns of verbal or non-verbal behavior, hyper/hypo reactivity to sensory stimuli or unusual interest in the environment sensory aspects.

Alternatively, clinicians in many countries use the International Statistical Classification of Diseases and Related Health Problems (ICD), a global standard for health data, clinical documentation, and statistical aggregation. Released in the 1990s, ICD already classified autism as:

- Infantile Autism, listed under the Schizophrenia group (ICD-8) (OUSLEY; CERMAK, 2014; LEEKAM et al., 2002);
- Infantile Autism, Disintegrative Psychosis, Other, and Unspecified, listed under the Psychoses with Origin Specific to Childhood group (ICD-9) (OUSLEY; CERMAK, 2014; LEEKAM et al., 2002);
- Childhood Autism, Atypical Autism, Rett Syndrome, Other Childhood Disintegrative Disorder, Overactive Disorder with Mental Retardation and Stereotyped Movements, Asperger Syndrome, Other PDDs, and PDD Unspecified, listed under the PDD group (ICD-10) (ORGANIZATION et al., 1992);

The ICD eleventh edition (ICD-11) (ORGANIZATION et al., 2018) also named the disorder as Autism Spectrum Disorder, placing ASD inside the Neurodevelopmental Disorders group. ICD-11 characterizes ASD as persistent deficits in initiating and sustaining reciprocal social interaction and communication. In addition to a range of restricted, repetitive, and inflexible behavioral patterns, interests, or activities, usually atypical or excessive for the individuals' age and socio-cultural context. The disorder's onset occurs during the developmental period, typically in early childhood, but symptoms may not become fully manifest until later when social demands exceed limited capacities. Deficits are sufficiently severe to cause impairment in personal, family, social, educational, occupational, or other important areas of functioning and are usually a pervasive feature of the individuals functioning observable in all settings. However, they may vary according to social, educational, or another context. Individuals along the spectrum exhibit a full range of intellectual functioning and language abilities.

ICD-11 classifies the disorder as:

- ASD without disorder of intellectual development and with mild or no impairment of functional language;
- ASD without disorder of intellectual development and with impaired functional language;
- ASD with disorder of intellectual development and with mild or no impairment of functional language;
- ASD with disorder of intellectual development and with impaired functional language; and
- ASD with disorder of intellectual development and with absence of functional language.

The DSM and ICD manuals, especially the latest versions, are currently the guides used by specialized health professionals to provide ASD diagnosis.

2.2 Autism Spectrum Disorder Prevalence

Prevalence, or prevalence rate, is the proportion of individuals in a population who have a particular disease or attribute at a specified point in time or over a specified period. Prevalence differs from incidence because it includes all cases, new and preexisting, in the population at the specified time, whereas incidence is limited to new cases only (DICKER et al., 2006).

The first autism epidemiological surveys indicated a prevalence from 0.4 to 2 cases for every 1000 people (0.04%-0.2%). Recent works show a higher ASD prevalence than previously estimated, although there is no standardization of autism survey methodology. Relying on refined research methodologies, including large populations, from multiple geographical locations, stratified samples with detailed screening activities, and well-known diagnostic procedures, some recent epidemiological surveys suggest an ASD prevalence ranging from 1% to 2% in many countries and regions, although estimates over 2% also have been presented (FOMBONNE, 2018).

In 2010, it was estimated 52 million ASD cases worldwide (BAXTER et al., 2015), while in 2016, it was estimated 62.2 million ASD cases (VOS et al., 2017). Despite the heterogeneity between ASD survey methodologies, there is a shared trend toward an increasing ASD prevalence. Table 1 shows reputable ASD prevalence surveys published in the last decade. We concentrated on recent surveys (previous ten years), with large sample sizes (preferably diversified in terms of the subjects' age and geographical area), and the definition of ASD cases based on modern diagnostic tools and certified health professionals.

Most studies concentrated on children and adolescents (0-18 years). Dietz et al. (2020) simulated ASD prevalence in adults using ASD prevalence data from children and adolescents (3-17 years). Grønborg, Schendel and Parner (2013) also worked with adults, although $\approx 75\%$ of their diagnosed ASD cases were adolescents (under 18 years old). The systematic review of Baxter et al. (2015) included samples up to 27 years old, despite not finding population-representative data for adults. Indeed, the ASD diagnosis usually occurs up to 17 years of age (OFNER et al., 2018).

Most studies have had follow-up intervals ending in the previous 5-6 years, with a few works with follow-up intervals ending in the latest ten years. These recent follow-up

Table 1 – ASD prevalence.

General (M:F) %	**Sex Ratio (M:F)**	**Site(s)**	**ASD Criteria**	**Sample Size**	**Age (years)**	**Follow Up Interval**	**Reference**
2.21 (3.6:0.9)	4.5:1	US	⊕	◁	18-84	⊘	(DIETZ et al., 2020)
1.85 (3.0:0.7)	4.3:1	US∗	DSM-V	275 419⋆	8	2008-2016	(MAENNER et al., 2020)
1.74 (2.7:0.8)	3.4:1	US	•	88 530	3-17	1997-2017	(ZABLOTSKY et al., 2019)
1.20 (NA:NA)	NA	Multi⊙	DSM-IV ICD-9/10	2 551 918	4-17	1998-2015	(HANSEN et al., 2019)
1.10 (1.6:0.5)	3.2:1	Multi⋄	DSM† ICD†	2 001 631	0-16	1998-2018	(BAI et al., 2019)
1.14 (1.8:0.4)	4.3:1	Qatar	DSM-V	133 781	6-11	2015-2018	(ALSHABAN et al., 2019)
1.92 (2.7:1.1)	2.5:1	Sweden⊖	DSM-IV ICD-10	567 436	0-17	1984-2011	(XIE et al., 2019)
2.50 (3.9:1.0)	3.5:1	US	∨	43 021	3-17	1999-2016	(KOGAN et al., 2018)
1.50 (2.4:0.6)	4:1	Canada	DSM-IV/V ICD-9/10	≀	5-17	2003-2015	(OFNER et al., 2018)
2.47 (3.6:1.2)	2.9:1	US	±	30 502	3-17	1999-2016	(XU et al., 2018)
1.25 (2.0:0.5)	3.9:1	US∘	ICD-9	3 166 542	4-18	1998-2016	(PALMER et al., 2017)
0.80 (1.2:0.3)	4.2:1	Global△	DSM ICD	50 378 584▽	NA-27	1980-2009	(BAXTER et al., 2015)
1.18 (NA:NA)	NA	Denmark	ICD-8/10	1 546 667	06-30	1980-2010	(GRØNBORG; SCHENDEL; PARNER, 2013)
1.15 (1.6:0.7)	2.5:1	Sweden⊖	DSM-IV ICD-10	444 154	0-17	1990-2007	(IDRING et al., 2012)
1.20 (2.2:0.5)	4.4:1	NL×	⊗	62 505	4-16	NA	(ROELFSEMA et al., 2012)
2.64 (3.7:1.5)	2.5:1	South Korea	∓	55 266	7-12	2006-2009	(KIM et al., 2011)
1.57 (NA:NA)	NA	UK‡	ICD-10	11 700	5-9	NA	(BARON-COHEN et al., 2009)

$^{\text{(M:F)}}$Male:Female; $^{\oplus}$Reported by parents on the National Survey of Children's Health (NSCH); $^{\triangleleft}$Estimated the 2017 national and state ASD prevalence by simulation; $^{\oslash}$2016-2018 state ASD prevalence of male and female children ages 3-17 (born 1999-2015), an estimate of the state populations in 2017, and state mortality rates from 1999 to 2017; *Arizona, Arkansas, Colorado, Georgia, Maryland, Minnesota, Missouri, New Jersey, North Carolina, Tennessee, and Wisconsin; *Monitors ASD among children aged 8 years in participating communities; $^{\bullet}$National Health Interview Survey; $^{\odot}$California (US), Denmark, Finland, Israel, Sweden and Western Australia; $^{\diamond}$Denmark, Finland, Sweden, Israel, and Western Australia; †DSM-III-R/IV/IV-TR/V or ICD-8/9/10 according to the period, and in-person diagnosis by psychiatrists or pediatric neurologists with expertise in neurodevelopmental disabilities in Israel before age three; $^{\ominus}$Stockholm County; $^{\vee}$Parent reported ASD children at the NSCH who ever received an ASD diagnosis by a care provider; $^{\wr}$40% of all children and youth aged 5–17 years across Canada (based on 2011 Canadian census); $^{\pm}$Parent report of a physician diagnosis at the NSCH; $^{\circ}$A de-identified database from Aetna; $^{\triangle}$Tables S3 and S4 in the (BAXTER et al., 2015) supplementary material; $^{\triangledown}$Of ASD cases; $^{\times}$Netherlands (Eindhoven, Haarlem, and Utrecht); $^{\otimes}$Diagnoses made by a clinical professional (e.g. psychologists or psychiatrists); $^{\mp}$Autism Spectrum Screening Questionnaire (ASSQ), ADOS, ADI-R, Korean WISC-III, Leiter International Performance Scale-Revised (ASD diagnoses met DSM-IV criteria); ‡Cambridge City, East Cambridgeshire, South Cambridgeshire, and Fenland districts; $^{\text{NA}}$Could not be determined with sufficient precision.

intervals guided the adoption of modern, widely used, and well-known ASD diagnostic tools, especially DSM and ICD.

These recent researches investigated samples of multiple sizes (small to large), geographically dispersed across states, countries, and even continents. Such a set of characteristics contributes to the reliability of their results, mainly if analyzed collectively.

According to these studies' results, the overall ASD prevalence has a mean of ≈ 1.6%.

Table 2 shows some measures of the centrality of the ASD prevalence data presented in Table 1. Aiming to reduce outliers, we excluded results with overall ASD prevalence less than 1% or greater than 2%. In addition, we also excluded results that did not present prevalence by gender. The mean and the median of ASD prevalence data are close. Applying the standard deviation to the mean/median, we obtain an overall ASD prevalence from ≈ 1.1% to ≈ 1.76%, an ASD prevalence among males from ≈ 1.71% to ≈ 2.73%, an ASD prevalence among females from ≈ 0.44% to ≈ 0.86%, and a male:female sex ratio of ≈ 3-4:1.

Table 2 – Central tendencies of ASD prevalence from Table 1.

Measure	**General (%)**	**Male (%)**	**Female (%)**	**Sex Ratio (M:F)**
Mean	1.43	2.22	0.65	3.57
Standard Deviation	0.33	0.51	0.21	0.79
Median	1.25	2.20	0.60	4.00

ASD occurs globally irrespective of culture, geography, or degree of industrialization. Some prevalence data from developed countries appear to be more comprehensive and reliable than those from developing countries. In developed countries, the greater availability of screening and diagnostic services usually increases ASD diagnosis. Anyway, the global ASD prevalences are rising, even when considering data from both developed and developing countries (ONAOLAPO; ONAOLAPO, 2017; ROTHOLZ et al., 2017; JANVIER et al., 2016).

However, much remains necessary to figure out the ASD prevalence trend, especially in developing countries. While recent ASD prevalence among developed countries tends to percentage values that approach 2%, epidemiological researches in developing countries point to percentage values quite below (e.g., 0.11% in Ecuador, 0.15% in India, 0.27% in Brazil, 0.53% in Caribbean Islands, 0.68% in Uganda, 0.80% in Nigeria, and 0.87% in Mexico) (ONAOLAPO; ONAOLAPO, 2017). Even some recent epidemiological researches in developed countries point to ASD prevalences under 1% (e.g., 0.51% in Israel (1-12 years-old, 2010), 0.63% in Australia (6-12 years-old, 2005), 0.71% in Denmark (5-6 years-old, 2006), 0.60-0.80% in Norway (0-11 years-old, 2010-2011), and 0.90% in United Kingdom (5-16 years-old, 2004)) (ÖZERK, 2016). These short ASD prevalences in some countries/regions may explain the low ASD prevalences values from global estimates (BAXTER et al., 2015; ELSABBAGH et al., 2012).

Methodological differences in case definition (CHIAROTTI; VENEROSI, 2020), diagnostic criteria (KING; BEARMAN, 2009; MATTILA et al., 2007), sample size (ELSABBAGH et al., 2012), surveyed areas due to educational or health care systems (MATSON; KOZLOWSKI, 2011), the strategy for targeting risk individuals or groups (case-finding procedures) (CHIAROTTI; VENEROSI, 2020; WAZANA; BRESNAHAN; KLINE, 2007), socio-economic factors (DURKIN; WOLFE, 2020; DURKIN et al., 2017), ASD awareness (HERTZ-PICCIOTTO; DELWICHE, 2009) and even cultural influence (TAYLOR; JICK; MACLAUGHLIN, 2013), jointly, affect the ASD prevalence estimates.

The evolution of the manuals, the ASD diagnosis categories and subcategories, and the differences between multiple versions of ICD and DSM over the past decades also have had an evident impact on the ASD prevalence statistics, estimates, and rates (ÖZERK, 2016). Among others, some characteristics attributed to the rise of the ASD prevalence are the ability to diagnose, with a possible reflection of the success in identifying children who were previously not diagnosed (HANSEN; SCHENDEL; PARNER, 2015; NEVISON, 2014), and changes in awareness, earlier diagnosis, and redefinition of diagnostic criteria (ZABLOTSKY et al., 2015).

Even with ASD awareness and more well-defined epidemiological studies, ASD prevalence estimates still vary across and within geographical areas and countries, years of research, and source of data used. Such variation conducts to the large variability of prevalence estimates worldwide (CHIAROTTI; VENEROSI, 2020). For example, the ASD prevalence varies widely across geographic areas in one of the most recent surveys (MAENNER et al., 2021), from 1.65% (Missouri) to 3.89% (California), with an overall ASD prevalence of 2.27%.

Some studies aimed to estimate the ASD prevalence in Brazil (BECK, 2017; PAULA et al., 2011), despite they only evaluated a specific country area and obtained extremely low prevalence estimates even compared to the most modest estimates nowadays ($\approx 0.04\%$ and $\approx 0.3\%$ respectively). Thus, there is no reliable estimate of the population-related prevalence of ASD in Brazil or any other Latin American country (HINBEST; CHMILIAR, 2021; PAULA et al., 2011; ELSABBAGH et al., 2010). Considering the worldwide ASD prevalence estimated by the World Health Organization of 1% (ORGANIZATION, 2019) and the estimated Brazilian population of ≈ 211 million people (IBGE, 2020), we could estimate that approximately 2 million individuals in Brazil have autism.

2.3 Autism Spectrum Disorder Etiology

Etiology refers to the study and determination of the causes of the diseases. Models of etiology try to explain the processes that initiate a particular disorder. The necessary conditions for developing the diseases are known as etiological factors. However, etiological factors are only the causes that directly start the disease process, and necessarily such causes have to precede the onset of the disease in terms of time. Several different conditions (biological, immunological, environmental, etc.) may contribute to defining a particular disease etiology (GELLMAN; TURNER et al., 2013).

Many complex mental and physical disorders (e.g., autism, depression, obesity) have partially unknown etiology. The ASD etiology is multifactorial: neurobiological, genetic, and environmental. Although incompletely understood, genetic and environmental factors and their interactions contribute to ASD etiology (BÖLTE; GIRDLER; MARSCHIK, 2019; LYALL et al., 2017; BAUMAN; KEMPER, 2005). Twin and family studies have shown a predominant genetic contribution to the ASD etiology. A complex interaction between common and rare genetic variants constitutes the genetic composition of ASD, with common genetic variants accounting for almost all ASD heritability (ROSTI et al., 2014; GAUGLER et al., 2014; HALLMAYER et al., 2011; GARDENER; SPIEGELMAN; BUKA, 2011).

2.3.1 Environmental Autism Spectrum Disorder Risk Factors

Environmental factors can be due to an aggregation of factors. Climatic, nutritional, and the interaction of individuals with their environment are usually the most critical factors. Several environmental agents have been pointed as possible ASD risk factors. We can highlight advanced parental age, the significant age difference between parents, preterm birth, pre, peri, and post-natal factors, cesarian delivery, short interpregnancy interval, exposure to air pollution, and still some events during pregnancy, such as the use of valproic acid, maternal infections, and exposure to environmental toxins (HODGES;

FEALKO; SOARES, 2020; BAI et al., 2019; BÖLTE; GIRDLER; MARSCHIK, 2019; LYALL et al., 2017; SEALEY et al., 2016).

Some studies suggested that shared environmental factors have at least equal or even more significant influence than genetic factors in ASD risk. The ASD liability variation in a clinical sample pointed primarily to shared environmental factors (58%) than genetic effects (38%) (HALLMAYER et al., 2011). Frazier et al. (2014) suggest an even higher estimate of shared environmental risk factors (64-78%). However, Frazier et al. (2014) did not systematically collect probands from the general population; thus, affected people had no equal chance to be selected.

Despite those previous associations in ASD risk, shared environmental factors have accounted for a tiny percent of increases in ASD diagnosis. Maternal and paternal age contributed to ≈ 2.7% of the 143% ASD prevalence increase among 0-3 years-old children from 1994 to 2001 (QUINLAN et al., 2015). Cesarean delivery, multiple births, changes in preterm delivery, assisted reproductive technology, and small for gestational age fetuses contributed to less than 1% of the almost 60% ASD prevalence increase among eight-year-old children born in 1994 contrasted to those born in 1998 (SCHIEVE et al., 2011).

Furthermore, a contrasting study using a sample cohort of ≈ two million people revealed that the individual ASD risk increased with genetic relatedness, with no shared environment effects (SANDIN et al., 2014). A meta-analysis comprising twin studies showed that ASD heritability is due to genetic effects. Previous studies that pointed to significant shared environmental factors are probably statistical artifacts due to the assumptions regarding ASD prevalence and dizygotic concordant pairs' oversampling. Thus, shared environmental effects seem unable to explain the majority of the ASD variance (TICK et al., 2016). Similarly, a novel study pointed out that shared environmental factors are unlikely to explain the rise in ASD prevalence. Once the ASD etiology consistently reveals a more significant genetic role (TAYLOR et al., 2020).

2.3.2 Genetic Autism Spectrum Disorder Risk Factors

Human genetics is both a fundamental and applied science that studies the inherited traits, their variations, and how they are transmitted in human beings. The transmission of features and biological information from parents to offspring is known as heredity.

A key role in genetics is understanding the relative contribution of genetic and environmental factors to phenotypic variance. The phenotypic (visible characteristic or effect on health) variance of a trait due to genetic differences in a specific population at a given time is known as heritability, also known as the proportion of the phenotypic variation that is not explained by the environment or random chance. The heritability of human traits is usually estimated based on the inference of genetic factors shared among relatives (BASELMANS et al., 2020; LEWIS, 2018; BISWAS; SINGH; REDDY, 2017).

The hereditary units transmitted from parents to offspring are known as genes. Consisting of the long molecules of deoxyribonucleic acid (DNA), human genes instruct our cells how to produce specific proteins that regulate the characteristics that comprise our individuality. The DNA sequences of the human genome (the entire collection of genetic instructions of a human) are dispersed among 23 pairs of structures called chromosomes and transmit information in its sequence of building blocks (like an alphabet) (LEWIS, 2018; BISWAS; SINGH; REDDY, 2017). Several levels constitute our genetics (genome, DNA, genes, chromosomes). Figure 2 shows the structure of human genetics.

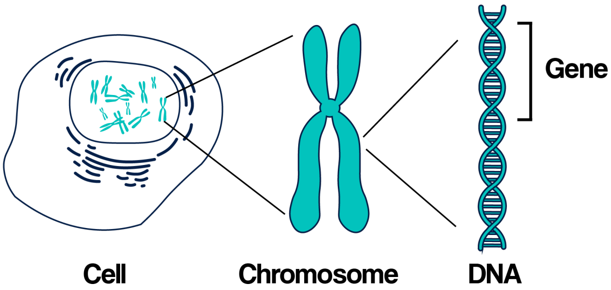


Figure 2 – Human Genetics Structure

Genetics defines a trait as single-gene (also know as Mendelian or monogenic) or polygenic. Single-gene traits are rare and caused by DNA changes in one particular gene. Polygenic traits are more common and express actions of one or more genes, usually including the contribution of environmental factors. Each human cell contains two copies of the genome. They differ in appearance and activities according to the genes they use once a cell uses only some of its genes. Environmental conditions (inside and outside the body) determine which genes a cell uses at any given time. The environment can influence single-gene and polygenic traits, which indicates that both can be multifactorial. The more factors (inherited or environmental) contribute to a disease, the more difficult it is to estimate the incidence risk (LEWIS, 2018).

Twins and family studies support the genetic contribution to ASD etiology by exposing high ASD heritability estimates and ASD recurrence rates among siblings (MAENNER et al., 2020; PALMER et al., 2017). However, research groups differ significantly on their assessments for the number of genes associated with ASD, ranging from a few to hundreds (SCHAAF et al., 2020).

The ASD heritability relies on a complex combination of genes, mutations, and chromosomal abnormalities. The ASD genetic architecture ranges from a rare single gene mutation to a polygenic risk. The main components of ASD genetic risk include: 1) *de novo* mutations, which occur spontaneously in offspring; 2) rare inherited single-gene disorders, occurred relatively recently in humans; and 3) polygenic variation, genetic changes in one or many genes widespread in humans. Many of the ASD risk genes operate as regulators of neurodevelopment or neural activity (IAKOUCHEVA; MUOTRI; SEBAT, 2019; BOURGERON, 2015).

2.3.2.1 Rare Inherited Variants and *De Novo* Mutations

Human genomics has identified a range of DNA sequence variations, including insertions and deletions of nucleotides and translocations of various chromosome segments. These mutations have been named Copy Number Variants (CNVs). CNVs are a DNA segment present at a variable copy number compared to a reference genome (ZARREI et al., 2015). CNVs are associated with certain human diseases' etiology (GIRIRAJAN; CAMPBELL; EICHLER, 2011; STANKIEWICZ; LUPSKI, 2010), although they are also present in healthy individuals (CONRAD et al., 2010). Genetic analyses like genome sequencing may expose which single-gene disease a person has, carry, or may develop. Tests with infected children and their parents can indicate if the disease cause is a mutation inherited from carrier parents or due to a dominant *de novo* mutation (LEWIS, 2018).

Part of the genetic risk for ASD consists of rare CNVs inherited from parents. Most variants occur as recurrent *de novo* mutations transmitted from a parent, with mild or no symptoms, due to variable levels of cognitive impairments. Thus, part of the ASD genetic architecture consists of rare CNVs inherited from parents who do not meet ASD diagnosis (IAKOUCHEVA; MUOTRI; SEBAT, 2019). *De novo* CNVs rates are up to ten times higher among ASD individuals (RUBEIS et al., 2014; SANDERS et al., 2011; XU et al., 2008), suggesting a substantial role of *de novo* CNVs in ASD. Advanced parental age could be associated with the increased risk of *de novo* spontaneous mutations (usually paternal) (ATSEM et al., 2016; KONG et al., 2012). Approximately 70% of *de novo* mutations originate from the father, and the rate of new mutations increases with the father's age (1-2 mutations per year of age) (MICHAELSON et al., 2012; KONG et al., 2012).

Recent studies aiming to identify ASD susceptibility of *de novo* genes have implicated 102 genes in ASD risk, with 53 genes having a greater frequency in ASD (SATTERSTROM et al., 2018; SANDERS et al., 2015). Despite some estimates that *de novo* mutations contribute to ≈ 30% of ASD cases (IOSSIFOV et al., 2014), rare variants seem to explain at most 17% of ASD heritability (GAUGLER et al., 2014), with rare *de novo* and inherited CNVs limited to 10% of children with nonsyndromic autism (TORRE-UBIETA et al., 2016).

Although *de novo* mutations are considered genetic factors, they do not contribute to the ASD heritability once they are present only in the affected descendant (excluding rare germinal mosaicisms present in parental germline and transmitted to offspring). Thus, *de novo* could be considered environmental causes of ASD acting on the DNA. From 500 to 1000 genes could account for these monogenic forms of ASD, reinforcing the high level of genetic heterogeneity (HUGUET; BOURGERON, 2016).

2.3.2.2 Common Polygenic Risk

As polygenic disorders result from the joint contribution or interaction of several independent genes and occur more frequently in humans, their genetic variance is essentially due to the additive effects of recessive alleles of different genes. Few dominant alleles can significantly affect the phenotype for some traits, but they do not contribute to heritability considerably because they are rare (LEWIS, 2018; LVOVS; FAVOROVA; FAVOROV, 2012).

A person carries nearly three million genetic variants compared with a genome of reference. Most of these variants (≈ 95%) are called common variants (FU et al., 2013). The most prevalent type of genetic variant in humans is called Single Nucleotide Polymorphisms (SNPs). Each SNP expresses a variation in a single nucleotide (a DNA building block). For example, an SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a specific DNA stretch. Common variants represent a key role in ASD susceptibility and the severity of symptoms, where numerous alleles contribute additively to the overall ASD risk (HUGUET; BOURGERON, 2016).

Polygenic Risk Score (PRS) is a genetic measure that summarizes all common SNPs' contributions to a trait (DUDBRIDGE, 2013). There are pieces of evidence for the additive contribution of rare and common genetic variants to ASD risk. Thus, common polygenic variation can also influence the diagnosis of individuals who carry a rare variant of significant effect. Compared with typically developing controls, ASD individuals with *de novo* mutations have significantly increased PRSs for ASD (WEINER et al., 2017).

All three categories of ASD risk genes are similar once a gene regulatory network broadly distributes their effects. For example, a genetic effect from a single gene mutation can influence other ASD genes' functions and spread extensively (IAKOUCHEVA; MUOTRI; SEBAT, 2019). Iakoucheva, Muotri and Sebat (2019) suggest that the genetics of ASD is typically compatible with an Omnigenic model (BOYLE; LI; PRITCHARD, 2017), in which the genetic basis of a complex trait is highly polygenic, being challenging to distinguish core genes with direct effects from several marginal genes with indirect effects.

2.3.2.3 Females Protective Effect

As with the ASD prevalence studies, increased male prevalences have been reported in other studies regarding neurodevelopmental disorders, supporting the theory of a female protective model. Such protective theory is supported mainly due to the excess of deleterious autosomal CNVs in females compared to males concerning the molecular basis of neurodevelopmental disorders. Females with ASD have more CNVs related to autism than males with ASD (LEVY et al., 2011), and autistic females are three times more likely to carry deleterious autosomal CNVs, besides having an excess of deleterious SNPs (JACQUEMONT et al., 2014).

This female protective effect raises the hypothesis that females require a more significant etiologic load to manifest the same level of impairment as males. A recent study reinforces this hypothesis that girls may need a higher familial etiologic load to manifest the ASD phenotype (ROBINSON et al., 2013). Other studies that support this hypothesis point that ASD girls have a higher proportion of affected relatives (parents and siblings) than ASD boys (WERLING; GESCHWIND, 2015; TSAI; STEWART; AUGUST, 1981).

Despite sharing autistic features, such as impairments in social and communication areas, repetitive or restricted interests and behavior, individuals with ASD are clinically vastly heterogeneous and differ in their developmental course, the pattern of symptoms, as well as in cognitive and language abilities. As advances in genome testing continue to expose the complexity of ASD etiology, the efficient translation of this genomic information needs to be incorporated into the clinical environment (HOANG; CYTRYNBAUM; SCHERER, 2018).

Given this broad clinical spectrum and the diverse and complex genetics associated with ASD, Hoang, Cytrynbaum and Scherer (2018) proposed a communication model to facilitate communication and understanding regarding the clinical and genetic heterogeneity of ASD known as the cup model. The cup model uses an analogy with cups and balls, both of different sizes, and is widely accepted to explain the complexity of ASD etiology and communicate genetic testing results comprehensively (CAMELI et al., 2021; LUCAS et al., 2021).

According to the cup model, an individual will develop ASD if the cup is filled with enough risk factors (balls) to reach a critical threshold. Risk factors can be highly penetrant (e.g., strong genetic variants) or lightly penetrant (e.g., weak genetic variants or environmental factors). Balls of different sizes represent risk factors. Bigger balls represent higher risk factors, while smaller balls represent lower risk factors.

Thus, each individual has an ASD risk cup with balls representing risk factors. Individuals without ASD may have some risk factors in their cups, but not enough to develop ASD. Conversely, individuals with ASD have enough risk factors to exceed the threshold to develop ASD.

The cup model points out that a genetic variant that contributes to the ASD diagnosis of a person can be inherited from a parent without ASD. Thus, the cup model could be used to demonstrate the difference regarding ASD genetic effects among siblings and between males and females. The ASD male:female sex ratio suggests a lower penetrance in females than males, meaning females have a higher threshold than males. The cup model illustrates this by using a larger cup for females, assuming females require more risk factors than males to reach the threshold for ASD (HOANG; CYTRYNBAUM; SCHERER, 2018).

2.4 Autism Spectrum Disorder Heritability Estimates

High levels of heritability characterize the ASD etiology, with a genetic factor estimated up to 98%, with a small environmental contribution. Although genetics is already a widely accepted risk factor for ASD, there is no consensus on the percentage of autism caused by genetic factors. Researches point to percentages ranging from 38% to 98% (BAI et al., 2019; ALMANDIL et al., 2019; SANDIN et al., 2017; TICK et al., 2016; KRONCKE; WILLARD; HUCKABEE, 2016; HALLMAYER et al., 2011; BAILEY et al., 1995; FOLSTEIN; RUTTER, 1977). In part, these discrepancies can be explained by the variation of the research methods. Thus, the ASD heritability estimates are sensitive to the research methods, once these methods require several and often untestable assumptions (SANDIN et al., 2017).

Table 3 presents reputable researches regarding ASD etiological origins. Most studies decompose the ASD liability variance into four components: (A) additive genetic effects, which means inherited additive effects from different alleles; (D) nonadditive genetic (dominance) factors, usually due to the interaction effects between alleles at the same locus; (C) shared environmental effects, which means nongenetic influences contributing to similarity within relatives; and (E) nonshared environmental effects, which make relatives dissimilar. This liability model is usually known as the ACDE model (NEALE; CARDON, 2013). Since most studies have emphasized additive genetics, total heritability correlates with the additive component, except for works by Sandin et al. (2017), Gaugler et al. (2014), and Sandin et al. (2014).

Some studies also decompose the ASD liability variance into maternal effects (M). Maternal effects indicate the effect of mothers on the environment of their offspring (i.e., noninherited genetic influences originating from mother beyond what is inherited) (NEALE; CARDON, 2013). However, such studies reveal a modest contribution, if any exists, of maternal effects to the ASD liability (BAI et al., 2019; YIP et al., 2018).

Because of a time-to-event approach concerning the ASD diagnosis, the work of Sandin et al. (2014) underestimates the sibling pairs concordant for ASD (possibly missing about half of the concordant pairs). This underestimate may have reduced their heritability estimates (SANDIN et al., 2014). Years later, Sandin et al. (2017) demonstrated their hypothesis regarding underestimated heritability by performing a reanalysis of the same

Table 3 – ASD Etiology Measures.

Heritability (%)			Environmental (%)		Sample Size (Total)	Statistical Model	Reference
Total	**Additive (A)**	**Nonadditive (D)**	**Shared (C)**	**Nonshared (E)**			
81	81-83	NE	≈ 0.3	≈ 18	(2001631) 1392096⊙ 1748450⊘	GLMM	(BAI et al., 2019)
85	73-87	NE	≈ 0.2	≈ 15	(776212) 98570⊙ 11780± 14865∓ 650997⊘	GLMM	(YIP et al., 2018)
83	79-87	≈ 10	≈ 4	≈ 17	(2049973) 37570⊕ 2642064⊙ 445531± 432281∓	LTM	(SANDIN et al., 2017)
≈ 81⋆	64-93	NE	6-35	1-3	(21-7982) ⊕	LTM	(TICK et al., 2016)
≈ 61	47-75	NE	NE	≈ 40	(75) ⊕	CTM	(DENG et al., 2015)
≈ 60	52	7	•	•	(3046) ⊕ ⊙ ± ∓ ⊘	GCTA	(GAUGLER et al., 2014)
50	33-50	≈ 16	≈ 5	≈ 48	(2049973) 37570⊕ 2642064⊙ 445531± 432281∓ 5799875⊘	LTM	(SANDIN et al., 2014)
21-35	21-35	NE	64-78	NE	(1136) ⊕	DF LTM	(FRAZIER et al., 2014)
38	14-67	NE	≈ 58	NE	(384) ⊕	CTM	(HALLMAYER et al., 2011)
≈ 80	73-87	NE	0-15	13-27	(90) ⊕	SEM	(TANIAI et al., 2008)
57	43-68	NE	NE	≈ 43	(464) 370⊕ 94⊙	SEM	(HOEKSTRA et al., 2007)

$^{\mathrm{NE}}$Not estimated; $^{\odot}$Full siblings; $^{\oslash}$Cousins; $^{\pm}$Paternal half-siblings; $^{\mp}$Maternal half-siblings; $^{\oplus}$Twins; *An approximate median/average of the additive genetic effects from the six different meta-analyses configurations; $^{\bullet}$41% for environmental (shared + nonshared); $^{\mathrm{GCTA}}$A software for Genome-wide Complex Trait Analysis based on Linear Mixed Models. (YANG et al., 2011); $^{\mathrm{DF}}$DeFries-Fulker Regression (DEFRIES; FULKER, 1985); $^{\mathrm{SEM}}$Structural Equation Modeling.

population. However, they used an alternative methodology to define sibling pairs as concordant or discordant for ASD. Their new heritability estimate increased ≈ 66% (from 50% (SANDIN et al., 2014) to 83% (SANDIN et al., 2017)).

Liability collectively defines both the genetic and environmental factors that contribute to the development of multifactorial diseases. A person will be affected by a condition when it accumulates a specific liability. The diagnosis of several human disorders results in a set of binary (e.g., affected or unaffected) or ordered (e.g., mild, moderate, or severe) values. There are four primary methods to estimating the etiology of complex and quantitative binary traits: Liability Threshold Model (LTM), Classical Twin Model (CTM), Falconer Model (FM), and (Generalized) Linear Mixed Model (GLMM) (BONNET, 2016; TENESA; HALEY, 2013).

Based on the correlation of the disease status among pairs of relatives of a specific type extracted from a random sample of the population (known as tetrachoric correlation) (PEARSON; LEE, 1900), LTMs assume that the disease's liability is (or can be transformed in) a normal distribution with a threshold above which all subjects manifest the disease and below which no individuals manifest the disease. The disease prevalence is the metric that usually defines the threshold estimates, and the model variables are all genes and environmental conditions protecting or increasing the risk of disease (NEALE, 2005). This normal distribution of the liability is supported by the complex etiology of most human diseases (VERHULST; NEALE, 2021) and by Genome-wide Association Studies (GWAS) results, suggesting that the more complex a disease, the more polygenic it is (BOYLE; LI; PRITCHARD, 2017). As much as this model description of the disease liability appears simplistic, LTMs design has proven valuable, and no empirical data have shown a reason to discard it (BASELMANS et al., 2020). Besides autism, LTMs using family members also have been used to describe the etiology of other traits as skin cancer (LINDSTRÖM et al., 2007), preeclampsia (NOH et al., 2006), and schizophrenia.

CTMs analyze the similarity among monozygotic and dizygotic twins (BOOMSMA; BUSJAHN; PELTONEN, 2002). CTMs strength comes from the similarity in genetic sharing of monozygotic twins (100% of genetic sharing) and dizygotic twins (50% of additive genetic sharing and 25% of dominant genetic sharing). Such genetic sharing allows partitioning the phenotype variance into an ACDE model, assuming that these combined sources result in the phenotypic variance. In CTMs, dominant effects tend to reduce the dizygotic correlation relative to the monozygotic correlation, while the shared environment increases the dizygotic correlation close to the monozygotic correlation. Thus, dominant and shared effects are negatively confounded, and CTM studies usually estimate shared or dominant effects (LITTLE, 2014). Indeed, this bias due to the common environment and dominant effects is often a concern in full siblings studies (TENESA; HALEY, 2013).

The GLMMs (HOPPER, 1993) are the most flexible approach to estimate etiology variance in families (TENESA; HALEY, 2013). GLMMs allow handling complex family trees of diverse size and structure, which are a limitation of the previous methods once the data are structured into defined families of the same size. Broadly used in several areas (e.g., agriculture, biology, and genetics), LMMs also have been used for heritability estimates of binary human diseases (BONNET et al., 2015), as well as to measure genotypes in GWAS with large samples using a large number of SNPs to identify genetic variants that explain phenotypic variances (BASELMANS et al., 2020; YANG et al., 2011).

LMMs also support splitting the phenotypic variance into separate components: a genetic variance, usually separated into additive, dominance, and epistatic; an environmental, traditionally divided into common, maternal influence, and the stochastic; and possibly a gene-environment interaction component. LMMs perform heritability estimates by the portion of total variation attributed to additive genetic components and the amount of total variation attributed to other variance components similarly estimated (BASELMANS et al., 2020; PAWITAN et al., 2004). Despite being computationally hard to fit, GLMMs have shown exemplary performance in experimental groups. The statistical and computing advancements, allied with the ability to exploit complex family trees, make GLMMs the preferred approach in practice (TENESA; HALEY, 2013).

Disorders with low prevalence rates (affecting one in a hundred) require huge samples to estimate heritability and recurrence rates among relatives (BASELMANS et al., 2020; HILKER et al., 2018). Small sample sizes can lead to heterogeneous heritability estimates (TENESA; HALEY, 2013). Many twin and family studies regarding ASD heritability usually run with samples of small size, which is generally recognized as a limitation (TICK et al., 2016; DENG et al., 2015; GAUGLER et al., 2014; FRAZIER et al., 2014; HALLMAYER et al., 2011; TANIAI et al., 2008, 2008; HOEKSTRA et al., 2007). Sample ascertainment bias and different measurement tools also are potential causes for the heterogeneity in ASD heritability (COLVERT et al., 2015).

However, recent studies (Table 3) that explored large and more diverse samples (twins, full- and half-siblings, and cousins) and applied more flexible and robust heritability estimation methods (GLMMs, LTMs, and SEMs), point to ASD heritability estimates ranging from $\approx 80\%$ to $\approx 85\%$ (BAI et al., 2019; YIP et al., 2018; SANDIN et al., 2017; TICK et al., 2016). It is important to note that ASD heritability may vary across populations, environments, sub-groups of people with different characteristics (e.g., age) and may change over time, even in these more elaborated studies (VISSCHER; HILL; WRAY, 2008).

2.5 Autism Spectrum Disorder Recurrence Rate Among Siblings

The risk of ASD recurrence in siblings of an ASD child is an important measure of the genetic contribution to the ASD etiology (GRØNBORG; SCHENDEL; PARNER, 2013). The ASD recurrence among relatives of affected family members is high in comparison to the overall ASD prevalence. Both the level of relatedness and the individuals' gender seem to be determinant factors to the recurrence extent among family members (HANSEN et al., 2019).

The estimated ASD recurrence rates in siblings of an ASD proband (usually named high-risk siblings) who do not manifest other diseases or syndromes range from ≈ 9% to ≈ 25%, varying according to the individual's gender (Table 4). The ASD risk in younger siblings ranges from ≈ 32% to ≈ 50% if there are two or more ASD children in the family (WOOD et al., 2015; SCHAEFER; MENDELSOHN, 2013; OZONOFF et al., 2011; SIMONOFF, 1998). This high ASD recurrence risk in affected families also reflects the heritable nature of ASD (GIRAULT et al., 2020).

Table 4 – ASD Recurrence Among Siblings.

Sample Size	ASD Criteria	ASD ARR (ASD RRR)							Reference
		Male ↓ Female	Female ↓ Female	Male ↓ Male	Female ↓ Male	Both ↓ Female	Both ↓ Male	Both ↓ Both	
20 882⊕	ICD-9	4.2	7.6	12.9	16.8	4.9	13.7	9.3	(PALMER et al., 2017)
13 997⊕	DSM-IV ICD-9/10	3.8⋆ (7.5)	5.1⋆ (10.2)	13.0⋆ (6.6)	19.3⋆ (9.8)	4.1⋆ (8.2)	17.5⋆ (8.9)	10.1● (8.4)	(HANSEN et al., 2019)
13 533⊖	DSM-III-R DSM-IV	5.1	6.7	14.1	17.0	5.3	14.5	10.1	(RISCH et al., 2014)
592⊙	DSM-IV-R	6.1	6.1	15.4	19.3	6.2	16.1	11.3	(XIE; PELTIER; GETAHUN, 2016)
319⊕	DSM-IV-R	18.5	25.0	33.8	26.5	19.6	32.4	26.6	(ZWAIGENBAUM et al., 2012)
385⊕	DSM-IV-R	-	-	-	-	12.8	30.1	23.1	(GIRAULT et al., 2020)
1 241⊕	ADOS DSM-IV	-	-	-	-	10.3	26.7	19.5	(MESSINGER et al., 2015)
664⊕	ADIR ADOS SCQ	-	-	-	-	9.1	26.2	18.7	(OZONOFF et al., 2011)
19 710⊕	DSM-III-R DSM-IV	-	-	-	-	-	-	10.1	(HOFFMANN et al., 2014)
13 164⊕	ICD-8/10	-	-	-	-	-	-	6.1	(GRØNBORG; SCHENDEL; PARNER, 2013)
1 235⊖	ADIR ADOS	-	-	-	-	-	-	14.2	(CONSTANTINO et al., 2010)
299⊖	DSM-IV ICD-10	-	-	-	-	-	-	24.7	(WOOD et al., 2015)

$^{\text{ARR}}$Absolute Recurrence Risk; $^{\text{RRR}}$Relative Recurrence Risk; $^{\ominus}$Families having one or more child with ASD; $^{\oplus}$Infants with an older sibling with ASD; $^{\odot}$Infants with ASD with at least one older sibling; $^{-}$Not available; $^{\bullet}$Estimated based on the overall ASD prevalence of Hansen et al. (2019) (1.2%); *Estimated based on the ASD prevalence by sex of Palmer et al. (2017) (male: 2.0%; female: 0.5%);

The ASD Relative Recurrence Risk (RRR) quantifies the ASD risk increase among individuals who have one or more family members with ASD compared to the overall ASD risk (prevalence) among individuals who do not have any family member with ASD (HANSEN et al., 2019). The ASD Absolute Recurrence Risk (ARR) rate quantifies the ASD probability among individuals with one or more family members with ASD (PALMER et al., 2017).

Table 4 presents reputable researches that investigated the ASD recurrence among siblings. The data are ordered by the level of information detail (sex-specific) and sample size. Those works that investigated large sample size, Palmer et al. (2017) (≈ 21 thousand siblings), Hoffmann et al. (2014) (≈ 20 thousand siblings), Hansen et al. (2019) (≈ 14 thousand siblings), Risch et al. (2014) (≈ 13.5 thousand siblings), showed similar recurrence rates estimates (overall: ≈ 9-10%; male: ≈ 14-18%; female: ≈ 4-5%). Except for the work conducted by Grønborg, Schendel and Parner (2013), which investigated a large sample size (≈ 13 thousand siblings), although exposed an overall ASD recurrence lower than works mentioned above (≈ 6%, ranging from 4.5% to 10.5% over time). The remainder of the studies explored relatively small sample sizes (from ≈ 300 to ≈ 1 300 siblings) and presented significantly high recurrence rates than those with larger sample sizes (overall: ≈ 14-27%; male: ≈ 26-32%; female: ≈ 9-20%). Except for the work conducted by Xie, Peltier and Getahun (2016), the recurrence rate tends to increase as the samples become too small (GIRAULT et al., 2020; WOOD et al., 2015; ZWAIGENBAUM et al., 2012).

2.6 The Broader Autism Spectrum Disorder Phenotype

The genetic risk to family members of ASD people extends not only to a possible ASD diagnosis but also to less or milder expressions of the social and communication impairments seen in the disorder. Such lesser expressions are usually below the threshold for an ASD clinical diagnosis (SZATMARI et al., 2000). First-degree relatives of ASD people are at increased risk for ASD-related characteristics. Such sub-clinical features, behaviors, and traits, conceptually similar to ASD core symptoms but insufficient to meet diagnostic criteria, have been referred to as the Broader Autism Phenotype (BAP) (GANGI et al., 2021). Since the BAP is strongly associated with ASD, it may be considered a marker of genes contributing to the risk of ASD (LOSH et al., 2009; DAWSON et al., 2002).

The BAP has been associated with difficulties in social relationships and poor mental health outcomes, such as language difficulties or delays, emotion recognition, social functioning deficits, less social interests, restricted or repetitive patterns of behaviors with higher rigidity and intense interests, difficulties in initiating and maintaining friendships in emerging adulthood, lower efficiency in planning, less expressiveness in nonverbal communication, attention shifting, and poorer conversational skills and verbal fluency (JAMIL; GRAGG; DEPAPE, 2017; PISULA; ZIEGART-SADOWSKA, 2015; SUCKSMITH; ROTH; HOEKSTRA, 2011).

More common among the family members of ASD individuals than in the general population (RUBENSTEIN; CHAWLA, 2018), the BAP studies also investigate the genetic mechanisms involved in ASD etiology. Most of the BAP measurement studies, although they vary, have in common a focus on sub-clinical versions of ASD symptoms (less functionally impairments) and reported several cognitive deficits in siblings of ASD children (GANGI et al., 2021). Whereas the BAP and ASD symptoms share a commonality, these symptoms' structures may differ. Therefore, instead of the severity degree, the number of confirmed symptoms may better differentiate BAP traits (RANKIN; TOMENY, 2019).

Studies suggest different developmental pathways to ASD in children with an older sibling with ASD (high-risk siblings). Assessing how ASD develops from birth is crucial to understanding the ASD developmental mechanisms and providing more precise objectives for genetic research (CHAWARSKA et al., 2014). Furthermore, as a complement of studies regarding ASD children, investigations of the BAP in children can specifically inform intermediate developmental trajectories that are often the most difficult to distinguish from typical development (KELLERMAN et al., 2019). In addition, a detailed understanding of ASD developmental pathways can help identify the need for early intervention and improve the range of available intervention options (JONES et al., 2014).

The BAP measurements are difficult due to the variety of functioning levels and countless risk factors combinations. Thus, there are no current standardized criteria for the BAP (KELLERMAN et al., 2019; PISULA; ZIEGART-SADOWSKA, 2015). The main difficulty in studies involving siblings of ASD individuals is to distinguish clearly the ASD symptoms from BAP traits. Mainly because the risk of ASD rather than the BAP characteristics is the primary concern (PISULA; ZIEGART-SADOWSKA, 2015). Usually defined using different domains, measurement tools, and report techniques, the

BAP estimates vary significantly across studies (RUBENSTEIN; CHAWLA, 2018). At three years of age, high-risk siblings present higher ASD symptomatology or lower developmental functioning levels than children without a family history of ASD, despite not receiving an ASD diagnosis (CHARMAN et al., 2017; MILLER et al., 2015; MESSINGER et al., 2013). However, atypical development in cognition, language, motor coordination, and especially social communication may emerge before three years (OZONOFF et al., 2014).

Table 5 presents reputable research on BAP in siblings of ASD individuals in the last decade. As social impairment diagnoses such as ASD tend to be more stable after 30 months of age (TURNER; STONE, 2007), our central focus was to place estimates of BAP effects in younger siblings of ASD individuals and the ASD recurrence in some cases, around three years.

Thus, we excluded studies with a wide age range because it may compromise the measure of developmental levels across multiple domains or the capability in specific functions. In addition, a wide age range makes it difficult to determine subgroups functioning levels by age, mainly due to small sample sizes, which limits statistical analysis. Studies on siblings at preschool age or older were also avoided, mainly because of the lack of longitudinal studies. Besides, subjects at such age may already overcome some of the difficulties previously identified, losing the diagnosis condition (PISULA; ZIEGART-SADOWSKA, 2015). ASD and BAP measuring diagnoses around three years of age are especially meaningful. Such research on infant siblings of ASD probands can provide valuable information on early ASD characteristics, start the investigation of BAP features, and further clarify the ASD genetic mechanisms (JONES et al., 2014).

In most works, ASD diagnosis in older siblings was confirmed via clinical best estimates, mainly using the appropriate version of one or more of the following diagnostic and measurement tools: ADI, ADI-R, ADOS, DAWBA, MSEL, SCQ, and VABS. We can split these diagnostic tools into two classes: parents' interviews and direct observations.

Interviews methods can be applied by parents, caregivers, or teachers. Such methods include Autism Diagnostic Interview (ADI) (COUTEUR et al., 1989), Autism Diagnostic Interview-Revised (ADI-R) (RUTTER et al., 2003), Social Communication Questionnaire (SCQ) (RUTTER; BAILEY; LORD, 2003), and Development and Well-Being

Table 5 – ASD and BAP Among Siblings.

Sample Size•	Tools⊕	Age⊙	ASD O(M:F) (%)	BAP⊖ O(M:F) (%)	ASD+BAP O(M:F) (%)	Reference
859	ADI-R, ADOS DSM-IV, DSM-V ICD-10, MSEL VABS	36	19	29(33:25)	48	(CHARMAN et al., 2017)
719	ADOS, MSEL	36	22(29:12)	25(28:20)	47(57:32)	(CHAWARSKA et al., 2014)
447	ADOS, MSEL	36	14	21	35	(MESSINGER et al., 2013)
294	ADOS, DSM-IV-TR MSEL	36	17(26:06)	28(31:25)	46(57:31)	(OZONOFF et al., 2014)
288	ADOS, DSM-IV MSEL, VABS	36	36	19	55	(D'ABATE et al., 2019)
204	ADOS, MSEL VABS	36	25(36:11)	15(19:09)	41(55:20)	(LANDA et al., 2012)
188	ADOS, DSM-IV LUI, MSEL	36		31		(MILLER et al., 2015)
135	ADOS, DSM-IV MSEL	36	13(24:03)	27(34:19)	40(57:22)	(SCHWICHTENBERG et al., 2010)
81	ADOS, DSM-IV MSEL	36	17(24:10)	10(12:8)	27(37:18)	(HUTMAN et al., 2012)
58	ADOS, DSM-IV MSEL, SCQ	18-36	29(42:12)	21(18:24)	50(61:36)	(CHRISTENSEN et al., 2010)
53	ADI-R, ADOS-G ICD-10, MSEL SCQ	38	32(53:19)	23(14:28)	55(67:47)	(HUDRY et al., 2014)
53	ADI, ADOS-G DSM-IV	24	23(25:17)	55(63:34)	77(88:52)	(MACARI et al., 2012)
47	ADI-R, ADOS-G ICD-10, MSEL	36	36(22:55)	26(33:15)	62(56:70)	(GLIGA et al., 2014)
45	ADOS, MSEL VABS	24-36	≈ 13	≈ 49	≈ 62	(KELLERMAN et al., 2019)
43	ADI-R, ADOS ICD-10, MSEL	12-36	28(47:15)	26(18:31)	53(65:46)	(WAN et al., 2013)
43	ADI-R, ADOS DSM-IV-TR MSEL, STAT	34	15	20	35	(YODER et al., 2009)
42	ADOS, DSM-IV	30-42	14(21:06)	21(31:09)	36(52:16)	(NICHOLS et al., 2014)
38	ADI-R, ADOS DSM-IV-TR	18-36	21(30:06)	32	53	(CORNEW et al., 2012)
35	ADI-R, ADOS-G ICD-10, MSEL	36	34	26	60	(BEDFORD et al., 2012)
24	ADOS-T, MSEL	24	29	25	54	(PAUL et al., 2011)
20	ADI-R, ADOS DSM-IV, MSEL	33	40	20	60	(DAMIANO et al., 2013)

•Infants at high risk for ASD (have at least one older sibling with an diagnosis of ASD); ⊕ASD/BAP diagnostic and measurement tools; ⊙Age of infants in months when the measurement was performed; ⊖At least one BAP trait or ASD-related behavioral characteristic; O(M:F)Overall(Male:Female).

Assessment (DAWBA) (GOODMAN et al., 2000). SCQ was designed to evaluate anyone over age four, ADI for children of at least five years, ADI-R for children of at least 18 months, and DAWBA for children from 5 to 16 years old.

The second class refers to methods where there is direct observation of the children during pre-modeling activities, specially elaborated to access domains related to

ASD. Such methods include Autism Diagnostic Observation Schedule (ADOS) (LORD et al., 1989), the Autism Diagnostic Observation Schedule Generic (ADOS-G) (LORD et al., 2000), Mullen Scales of Early Learning for the assessment of young children (MSEL) (MULLEN et al., 1995), Vineland Adaptive Behavior Scales (VABS) (SPARROW; CICCHETTI, 1989), Language Use Inventory (LUI) (O'NEILL, 2009), and the Screening Tool for Autism in two-year-olds (STAT) (STONE; OUSLEY, 2004).

Younger siblings of ASD children who do not receive an ASD diagnosis themselves present a high risk for developing BAP than siblings with no history of ASD in the family. The BAP estimates among these high-risk siblings range from 10% (HUTMAN et al., 2012) to 55% (MACARI et al., 2012) at ≈ 36 months of age. Adding the ASD estimates to the BAP estimates, the overall risk for developmental concerns in high-risk siblings ranges from ≈ 27% to ≈ 77%, an average of ≈ 50%, with the average for males nearly to 60% and the average for females nearly to 32% when excluding uncommon cases in which ASD plus BAP estimates in females surpassed the estimates in males (GLIGA et al., 2014). Added together, these ASD and BAP estimates point to a male:female sex ratio of ≈ 1.9:1, which is lower than expected if compared to the ASD prevalence sex ratio, but similar to that reported by D'Abate et al. (2019), which suggests a decrease in sex ratio as diagnostic criteria become more rigorous and detailed.

Results showed more elevated severity of ASD traits in younger siblings of ASD individuals than individuals with no family history of ASD. In addition, the ASD severity is even likely higher in multiplex ASD families (those with two or more ASD children). ASD and BAP traits such as less expressiveness in nonverbal communication, less social interest, poorer conversational skills, higher rigidity, and intense interests are more pronounced in siblings from multiplex ASD families than in siblings from simplex ASD families (those families with only one ASD child) (GERDTS et al., 2013; SCHWICHTENBERG et al., 2010).

Some studies assessed ASD symptomatology only in non-affected siblings of ASD children. Even in non-affected siblings, their results suggest a higher incidence of deficits in at least one ASD typical domain. However, it is noteworthy that these results are not entirely consistent in terms of the affected domains nor the depth of the deficits (PISULA; ZIEGART-SADOWSKA, 2015).

Other neurodevelopmental abnormalities also are more common among unaffected siblings of ASD children. For example, compared to control groups (no history of ASD in the family), unaffected siblings of ASD children are more likely to develop developmental delays, such as developmental coordination disorder, developmental speech or language disorder, attention-deficit hyperactivity disorder, anxiety disorders, unipolar depression, intellectual disability, and disruptive behavior disorders (LIN et al., 2021).

Similar BAP results for social and communication domains in parents showed that different genetic transfer mechanisms might operate in simplex ASD families compared to multiplex ones. These results suggest that de novo mutations and non-inherited CNVs may be significant risk factors for simplex ASD families, presenting a lesser degree in multiplex ASD families (BERNIER et al., 2012; SEBAT et al., 2007). However, similar studies did not confirm these findings, suggesting a low variability of ASD phenotype in multiplex ASD families (PINTO et al., 2010; SPIKER et al., 1994). In addition, a systematic review by Rubenstein and Chawla (2018) quantified the percentage of parents of ASD children who had BAP themselves, presenting a rate of BAP in parents up to 80%, with more prevalent in fathers than mothers.

Several studies suggest a wide range of impairments in infant siblings of ASD children. Qualitative analyses suggest that the overall performance of unaffected high-risk children could be considered at an intermediate level, performing slightly worse than the low-risk children and better than ASD children. Emerging by 24 months, the cognitive differences support the increasing demand for early monitoring of high-risk children to identify risk and promote optimal development (KELLERMAN et al., 2019). Although BAP is not a clinical diagnosis, it does confer risks and challenges, supporting the importance of continuous monitoring in high-risk siblings, even in the absence of a complete ASD diagnosis (GANGI et al., 2021). Some differences in these high-risk siblings are probably due to a later ASD diagnosis. Although, such infant siblings are also at a high risk of developing BAP traits. Details about previous studies that studied the early phenotype of ASD and the BAP traits can be seen at Pisula and Ziegart-Sadowska (2015), Jones et al. (2014), and Yirmiya and Ozonoff (2007).

2.7 Summary

This chapter presented an overview of ASD, especially concerning its classification, prevalence, and etiology. We further explored these three domains because they are essential to understanding the rest of this work. It was mainly dedicated to understanding what autism is, how it is described, the penetration of autism in our society, and the leading causes of the disorder.

We started showing the changes in ASD definition over time given by the two primary and most used diagnostic manuals (DSM and ICD). Earlier defined by several distinct nomenclatures, phenotypic descriptions, and diagnostic manuals, autism is currently recognized as a broad spectrum named Autism Spectrum Disorder. Although ASD screening and diagnosis remain complex in practice, it has a more straightforward set of definitions for its phenotypic manifestations and better diagnostic criteria.

Prevalence studies are essential to understand some characteristics of the disease, such as its causes, the demography of affected individuals (e.g., sex and age), social, racial, and geographical aspects. Prevalence rates are also essential to estimate other disease dimensions, such as heredity patterns. Surveys regarding ASD prevalence showed that autism does not seems related to race, ethnicity, or geographic location, with an average prevalence of $\approx 1.5\%$ (female: $\approx 0.7\%$; male: $\approx 2.55\%$), with a male:female ratio of 3-4:1.

Due to the known genetic nature of ASD, we broadly explored heritability and recurrence rate studies once these two types of studies pursuing to describe such nature of ASD. Although presenting relatively different results concerning ASD heritability, studies that investigated relatively larger sample sizes and employed powerful statistical methods estimate an ASD heritability from $\approx 80\%$ to $\approx 85\%$. Similar outcomes were observed regarding the ASD recurrence rate and the BAP traits among siblings. ASD recurrence researchers that explored relatively larger populations presented overall recurrence rates from $\approx 10\%$ to $\approx 25\%$, again showing differences between sex (from $\approx 14\%$ to $\approx 35\%$ for males; from $\approx 5\%$ to $\approx 20\%$ for females). BAP recurrence researches presented overall recurrence rates generally from $\approx 20\%$ to $\approx 30\%$, including peaks up to 50%. The BAP estimates show small differences between sex, with average of $\approx 32\%$ for males and average of $\approx 18\%$ for females (male:female ratio of 1.8:1) when excluding uncommon cases in which the BAP estimates in females surpassed the BAP estimates in males.

As genetic factors seem to represent a consistently larger role regarding ASD risk than environmental factors, we focused on genetic influences on ASD risk to model our inference methods. Thus, the next chapter introduces the probabilistic graphical models we aim to use.

3 Probabilistic Graphical Models

Any technique that allows the computer to imitate human behaviors is popularly classified as Artificial Intelligence (AI). However, we also have AI techniques based on other biological systems, probability, statistics, and mathematics. Emerging in the 1950s, AI is a relatively new science and engineering field that uses many spheres of human knowledge, such as logic, probability, and mathematics. Its primary goal is a theory of intelligence that explains the behavior of natural intelligent entities and guides the creation of artificial agents capable of smart behaviors. An intelligent agent must be capable of precisely perceive the environment and perform proper actions. An agent is rational if it does the right thing, given its acquired knowledge (RUSSELL; NORVIG, 2020; GENESERETH; NILSSON, 2012).

In the AI context, we can define intelligence as human or rational. Human intelligence is committed to human performance, while rational is a formal and ideal performance measure called rationality. Different techniques are used to pursue these dimensions. Human intelligence approaches usually use empirical science related to psychology, cognitive science, biology, and neuroscience. These approaches involve observations, hypotheses, and the study of how humans behave, how our minds operate, and how human brains process information. Rationalist approaches consist of a combination of formalism from logic, mathematics, statistics, and control theory. These methods aim to create more strict rules for the decision process (RUSSELL; NORVIG, 2020).

3.1 Artificial Intelligence Sub-fields

Russell and Norvig (2020) argue that most Artificial Intelligence comprises the following sub-fields: Machine Learning, Natural Language Processing, Computer Vision, Robotics, and Planning. These subdivisions relate more to the sub-fields' practical goals rather than the technologies employed by each one. For example, Artificial Neural Networks is a Machine Learning technique commonly used in Natural Language Processing, Robotics, Computer Vision, and Planning.

Machine Learning techniques enable machines to learn by observing data and creating models based on such information. Computers use these models as both a hypothesis about the problem and software capable of solving then. There are three main types of learning: 1) supervised learning, in which the computer observes pairs of input and output data to learn a function that maps from inputs to outputs; 2) unsupervised learning, in which the computer uses the inputted data to learn patterns but does not receive explicit feedback; and 3) reinforcement learning, in which the computer learns through a set of reinforcements that can be rewards or punishments.

Natural Language Processing techniques enable machines to communicate successfully in natural languages, such as English or Portuguese. However, as natural languages are different from formal languages, a common problem is the language models, which are models to predict the probability distribution of the language expressions.

Intelligent agents can use several sensors to sense the environment (e.g., images, noises, distances, positions, temperatures). Through this perceptual channel, machines receive stimulus and create a representation of the real world. Computer Vision techniques enable machines to perceive objects from the environment through the use of sensors like cameras. Based on external information acquired by sensors (such as images), Computer Vision agents can build a real-world model, known as reconstruction (e.g., creating geometries), or describe distinctions among the objects they "see" (e.g., labeling objects), known as recognition.

Robotics are techniques that enable machines to move and manipulate the physical world. Such machines (robots) usually are equipped with actuators (e.g., arms, grippers, legs, wheels) designed to produce physical forces on the environment and sensors (e.g., gyroscopes, accelerometers, GPS, cameras, radars, lasers) dedicated to perceiving the environment. This robot-environment interaction can change the state of the robot, the state of the environment, and the state of the people around it.

Planning techniques allow finding a sequence of actions to achieve a goal. Given an initial state, a goal (or a set of them), and a set of possible actions, the planning problem synthesizes a sequence of steps to be executed in the initial state to turn the environment into a goal state. Planning has applications in several areas such as Games, Logistics, Robotics, Manufacturing, etc.

In general, real-world applications still require other abilities from so-called intelligent agents, such as automated reasoning and knowledge representation. Automated reasoning refers to performing reasoning sequences electronically and automatically finding suitable reasoning steps to infer new knowledge from a given data, answer questions, and outline new conclusions. Knowledge representation refers to how to represent real-world events (what machines know, hear, or see) in a pattern that machines can use to reason and solve problems (RUSSELL; NORVIG, 2020).

3.2 Probabilistic Networks

Nowadays, science and technology are omnipresent in our everyday lives, becoming the new basis for belief and, together, bring new ways to improve the quality of life of our entire society (FEENBERG, 2006). Most real-world events are unpredictable, demanding that intelligent applications handle the uncertainty from partial observability, nondeterminism, or any eventuality (RUSSELL; NORVIG, 2020). The uncertainty arises from some information deficiency. So the information may be incomplete, imprecise, incorrect, or contradictory (KLIR, 2006).

Deductive Logic is insufficient for reasoning under uncertain environments once it does not attribute a degree of uncertainty to the premises nor conclusions. Then, the Inductive Logic, supported by the Probability Theory, has emerged as a proper alternative for expressing reasoning (WILLIAMSON, 2002), once the nature of the knowledge from which inferences are produced is uncertain and subjective (PEARL, 1986).

The probability theory provides ways to deal with the uncertainty coming from laziness and ignorance. Laziness is due to the extensive work to consider every possible explanation for given evidence. Ignorance is due to a non-complete knowledge about the domain or uncertainty about a particular situation once we can not evaluate all premises. Address uncertainty with numeric degrees of belief solves the qualification problem, which specifies the impossibility of identifying all preconditions needed to succeed in the desired action (RUSSELL; NORVIG, 2020).

Then emerged the probabilistic graphical models, a graph-based representation for compactly encoding a complex distribution over a high dimensional space. Nodes express variables, and edges denote the interactions between them. Known as probabilistic networks, these models allow inter-causal reasoning, a vital aspect that distinguishes them from other automated inference systems (KOLLER; FRIEDMAN, 2009; KJAERULFF; MADSEN, 2013). In the inter-causal reasoning process, taking evidence about a hypothesis decreases the belief in the competing unobserved hypotheses automatically (KJAERULFF; MADSEN, 2013), which constitutes a safe and complete inference mechanism (PEARL, 1988).

There are two graphical families to represent probability distributions. The Bayesian Networks (BNs) (Section 3.4), and the Markov Models (MMs) (Section 3.5). Both models provide the duality of independencies and factorization. However, they differ regarding the set of independencies they can encode and the factorization of the distribution they induce (KOLLER; FRIEDMAN, 2009).

3.3 Basic Probability Theory

Before introducing the graphical models, this subsection aims to introduce some fundamental probability theories suited to the requirements of the probabilistic networks.

A **sample space** is the set of all possible worlds in a specified domain. The possible worlds are mutually exclusive, once two or more possible worlds can not be the case simultaneously. The possible worlds are also exhaustive because one possible world must be the case. For example, the throw of two distinguishable dices has 36 possible worlds $\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),\cdots,(6,6)\}$. A fully specified sample space associates a probability P (with values between 0 and 1) to each possible world and the total probability of all possible worlds must add up to 1. These associations are called **probability distribution** (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009).

Probability theory names its variables with the first letter in uppercase, and such variables are called aleatory or random variables. An **aleatory variable** is a numerical function defined in a sample space, and it maps from all possible worlds to a set of possible values it can assume. It gives a numerical value X to a phenomenon within the sample space S and is associated with a probability distribution ($P(X)$ in S) such that: Equation 3.1 (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009):

$$\forall x \; P(X{=}x) \geq 0 \; and \; \sum_{x} P(X{=}x) = 1 \tag{3.1}$$

We can take the sum value obtained from two dices throw as an aleatory variable.

Function 3.2 is the numerical function (x and y are the values obtained from each dice), and Table 6 is the probability distribution for each possible sum.

$$F(x,y) = X = x + y \tag{3.2}$$

Table 6 – Probability distribution of Function 3.2

X	$P(X)$	X	$P(X)$
2	$\frac{1}{36}$	8	$\frac{1}{7.2}$
3	$\frac{1}{18}$	9	$\frac{1}{9}$
4	$\frac{1}{12}$	10	$\frac{1}{12}$
5	$\frac{1}{9}$	11	$\frac{1}{18}$
6	$\frac{1}{7.2}$	12	$\frac{1}{36}$
7	$\frac{1}{6}$		

A variable can be discrete or continuous. **Discrete** random variables have a finite set of possible values, usually obtained by counting. **Continuous** random variables take any value in a given interval of real numbers, usually obtained by measuring (KJAERULFF; MADSEN, 2013). From this point on, we deal with the discrete variables aspects, once they are the type of variables used in this study.

Dealing with probability distributions of multiple variables requires a special notation. Suppose a domain with the aleatory variables Weather $W = \{sunny, rain, snow\}$ and Traffic $T = \{jam, normal\}$, with probability distributions $P(W) = \{0.6, 0.25, 0.15\}$ and $P(T) = \{0.42, 0.58\}$, respectively. The probabilities of all combinations of the values of W and T produce a matrix M_{3x2} called the **joint probability distribution** of W and T. **Joint probabilities** measure the likelihood of two or more events co-occurring at the same point in time. It can be represented as the probability of the intersection of the co-occurring events. A **full joint probability distribution** determines the distribution for all aleatory variables completely. This full joint distribution is sufficient as a knowledge base to calculate the probability of any possible event in the model (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009). Table 7 shows the joint probability distribution of W and T.

Table 7 – Full joint probability distribution of W and T

W	***T***	
	jam	*normal*
sunny	0.10	0.50
rain	0.20	0.05
snow	0.12	0.03

A particular subset of possible worlds is called **events**. An example of an event is the list of worlds where two rolled dice add up to 3 $\{(1,2),(2,1)\}$. Probabilistic assertions and queries are usually about pre-defined events. The sum of the probabilities associated with each world of an event defines the **event probability**. For example, the probability that two rolled dice add up to 3 is $P(Sum{=}3) = P((1,2)) + P((2,1)) = 1/36 + 1/36 = 1/18$ (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009). Equation 3.3 shows the probability for any event E.

$$P(E) = \sum_{w \in E} P(w) \tag{3.3}$$

We could calculate the probabilities of some events in Table 7. To do so, we must apply Equation 3.3 to add up the probability values where the desired event is true. For example, the probability of *rain* ($P(W{=}rain) = 0.20 + 0.05 = 0.25$), and the probability of *snow* and traffic *jam* ($P(W{=}snow, T{=}jam) = 0.12$).

Probabilities like $P(W{=}rain)$ are called **unconditional** or **prior probabilities** once probability theory does not require comprehensive knowledge of the sample space. **Prior probabilities** refer to the belief in the events in the absence of any other information (RUSSELL; NORVIG, 2020).

Most of the time, it is necessary to know the probability of an event given that we have some information already revealed, usually called **evidence**. This probability is called **conditional** or **posterior probability** and is written as $P(X|Y)$ (RUSSELL; NORVIG, 2020). Equation 3.4 shows how to compute the conditional probability of event X given an event Y.

$$P(X|Y) = \frac{P(X,Y)}{P(Y)} \tag{3.4}$$

For example, the probability of a traffic jam given that it is raining is:

$$\begin{aligned} P(T{=}jam|W{=}rain) &= \frac{P(T{=}jam, W{=}rain)}{P(W{=}rain)} \\ P(T{=}jam|W{=}rain) &= 0.2/0.25 = 0.8 \end{aligned} \tag{3.5}$$

Given a conditional probability, the joint distribution of X and Y can be written following the product rule (RUSSELL; NORVIG, 2020), as in Equation 3.6.

$$P(X, Y) = P(X|Y)\ P(Y) \tag{3.6}$$

We can apply the product rule to compute the joint probability of n variables by a successive product of conditional and joint probabilities of these same variables. Each subsequent product reduces the joint probability to a conditional probability and a shorter joint probability. Equation 3.7 presents the **chain rule**.

$$\begin{aligned} P(X_1, \cdots, X_n) &= P(X_1|X_2, \cdots, X_n)\ P(X_2, \cdots, X_n) \\ &= P(X_1|X_2, \cdots, X_n)\ P(X_2|X_3, \cdots, X_n)\ P(X_3, \cdots, X_n) \\ &\quad \cdots \\ &= P(X_1|X_2, \cdots, X_n)\ P(X_2|X_3, \cdots, X_n)\ P(X_3, \cdots, X_n) \cdots P(X_{n-1}|X_n)\ P(X_n) \\ &= \left[\prod_{i=1}^{n-1} P(X_i|X_{i+1}, \cdots, X_n)\right] P(X_n) \end{aligned} \tag{3.7}$$

Given observed evidence, to compute the posterior probability for query propositions is known as **probabilistic inference**. We can use the full joint distribution to perform inference. Given the full joint distribution of a model, Equation 3.8 can be used to answer queries.

$$P(X|E{=}e) = \alpha P(X, E{=}e) = \alpha \sum_y P(X, E{=}e, Y{=}y) \tag{3.8}$$

Where:

- $P(X|E{=}e)$ is what we want to know (query);
- e is the list of observed values;
- y is all possible combinations of the values of the remaining unobserved variables;
- α is a normalization constant.

X, E, and Y are the entire domain set of aleatory variables. $P(X|E{=}e, Y{=}y)$ is a subset of the full joint distribution probabilities. The full joint distribution in tabular form does not scale well. However, it is the theoretical foundation to build effective reasoning systems (RUSSELL; NORVIG, 2020).

There is a fundamental property between events known as **independence** (also known as marginal independence or absolute independence). If an event X does not influence an event Y and vice-versa, they are independent events ($X \perp\!\!\!\perp Y$). This independence means that the occurrence of X does not affect the probability of occurrence of Y and vice-versa. The independence between two events (X and Y) can be written as in Equation 3.9.

$$\begin{aligned} P(X,Y) &= P(X)\ P(Y)\ or \\ P(X|Y) &= P(X)\ or \\ P(Y|X) &= P(Y) \end{aligned} \tag{3.9}$$

The knowledge of the domain supports performing assertions over independent events. Suppose we can split the aleatory variables into independent subsets. In that case, we can factor the full joint distribution into separate joint distributions, which reduces the size of the domain representation and the complexity of the inference model. However, it is difficult to identify independent variables in complex domains once independence will fail if a connection, even indirect, exists between two variables (RUSSELL; NORVIG, 2020).

Although independence is a valuable property, it is difficult to identify fully independent events in real-world domains. It is most common to determine the independence of two events, given a third event. Known as **conditional independence**, this relationship of three variables defines the independence of two variables X and Y, given a third variable Z ($X \perp\!\!\!\perp Y|Z$), as in Equation 3.10.

$$\begin{aligned} P(X,Y|Z) &= P(X|Z)\ P(Y|Z)\ or \\ P(X|Y,Z) &= P(X|Z)\ or \\ P(Y|X,Z) &= P(Y|Z) \end{aligned} \tag{3.10}$$

As for absolute independence, conditional independence assertions also allow the decomposition of the full joint distribution. Once conditional independence is more commonly available, it can enable probabilistic systems to scale up. This decomposition of large probabilistic domains into weakly connected subsets makes conditional independence one of the most basic and robust structures of knowledge representation in uncertainty environments (KOLLER; FRIEDMAN, 2009; RUSSELL; NORVIG, 2020).

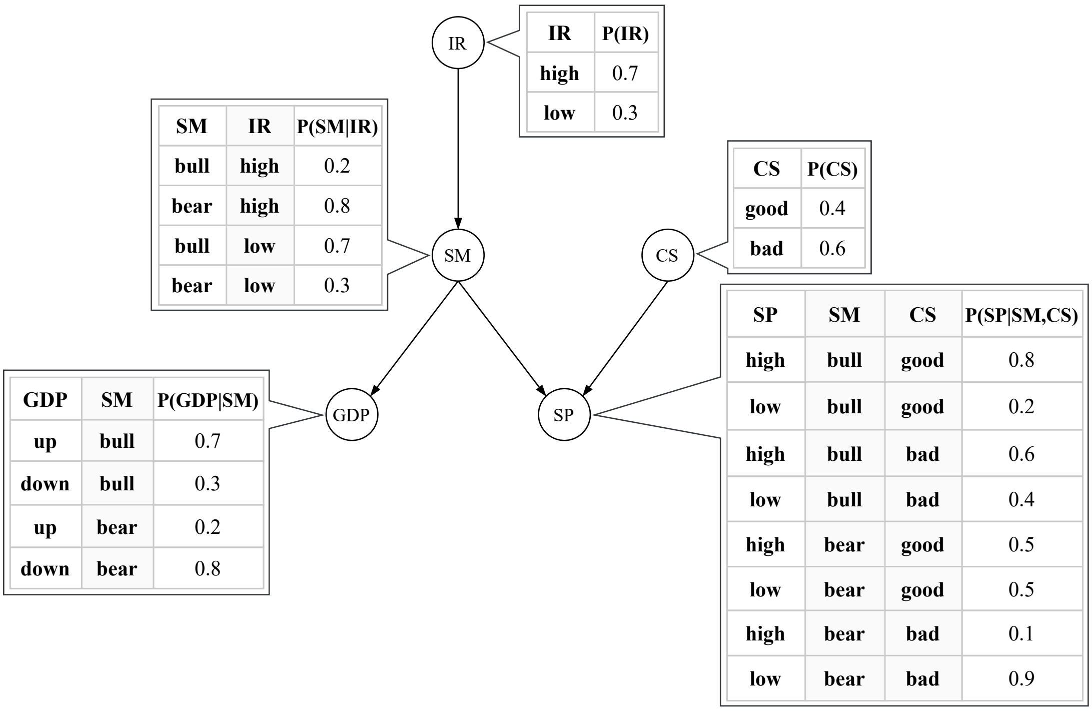
3.4 Bayesian Networks

Bayesian Networks (also known as Causal Networks, Belief Networks, Causal Probabilistic Networks, Probabilistic Cause-Effect Models, Probabilistic Influence Diagrams, and Graphical Probability Networks) are graphical models of causal relationships in a given domain. Describing dependencies among variables, BNs enable solving logical problems that involve probabilistic concepts, expanding the initial models of knowledge representation and manipulation (HOLMES; JAIN, 2008; NEIL; FENTON; NIELSON, 2000).

Essentially, BNs can represent, concisely, any full joint probability distribution. By employing a rigorous and efficient formalism to uncertain knowledge structuring as well as practical algorithms for probabilistic reasoning, BNs support any reasoning with causal variables, such as diagnosis, prediction, or causal explanation (RUSSELL; NORVIG, 2020; WILLIAMSON, 2002).

BNs are models for knowledge representation consisting of two components: a **qualitative component**, representing the network structure as a Directed Acyclic Graph (DAG), and a **quantitative component**, representing the probabilistic element as a set of conditional probabilities. Both components are fundamental to the definition, construction, and underlying inference process (KJAERULFF; MADSEN, 2013; DARWICHE, 2008).

Figure 3 shows the structure and the Conditional Probability Tables (CPTs) of a BN representing part of the stock exchange domain. The IR variable represents the country's interest rate. The interest rate directly impacts the stock market (SM) performance. The stock market performance usually indicates how the country's gross domestic product (GDP) will perform. In addition to internal factors, the state of the stock market (SM) and the performance of the company's economic sector (CS) also impact the stock price (SP) of a particular company.



IR	P(IR)
high	0.7
low	0.3

SM	IR	P(SM|IR)
bull	high	0.2
bear	high	0.8
bull	low	0.7
bear	low	0.3

CS	P(CS)
good	0.4
bad	0.6

GDP	SM	P(GDP|SM)
up	bull	0.7
down	bull	0.3
up	bear	0.2
down	bear	0.8

SP	SM	CS	P(SP|SM,CS)
high	bull	good	0.8
low	bull	good	0.2
high	bull	bad	0.6
low	bull	bad	0.4
high	bear	good	0.5
low	bear	good	0.5
high	bear	bad	0.1
low	bear	bad	0.9

Figure 3 – A BN example over five variables. A CPT is associated with each node containing the conditional probabilities of that node given its parents.

3.4.1 Syntax of Bayesian Networks

BNs represent its **qualitative** aspect using graphs that illustrate their probabilistic distributions. A **graph** $G = (V, E)$ consists of a finite set of distinct **vertices** (or **nodes**) $V = \{v_1, v_2, ..., v_N\}$, and a finite set of **edges** (or **links**) $A = \{a_1, a_2, ..., a_{N^2}\}$ connecting its vertices (ROSEN, 2017).

The connection pattern between nodes delimits some properties of a graph. The notation $v_1 \rightarrow v_2$ indicates a connection from vertice v_1 to vertice v_2 by a directed edge, which means a **directed graph** (or **digraph**). The notation v_1 — v_2 designates a connection from v_1 to v_2 by a not directed edge, which means an **undirected graph**. In a digraph, the edges are unidirectional, indicating that the graph can be traversed only in such directions. On the other hand, in an undirected graph, the edges are bidirectional, indicating that the graph can be traversed in either direction (RUSSELL; NORVIG, 2020; ROSEN, 2017).

A graph is **connected** if there is a path between every pair of its vertices. A directed graph is **acyclic** if any path following the directions of the edges will never produce a closed-loop (cycles). In a **directed multiply connected** graph, there is more than one distinct path between two nodes. There is at most one path between any two nodes in a **directed singly connected** graph (**trees**). In **simple trees**, each node has at most one parent. In **polytrees**, nodes can have more than one parent (RUSSELL; NORVIG, 2020; ROSEN, 2017). Figure 4 illustrates these types of graphs.

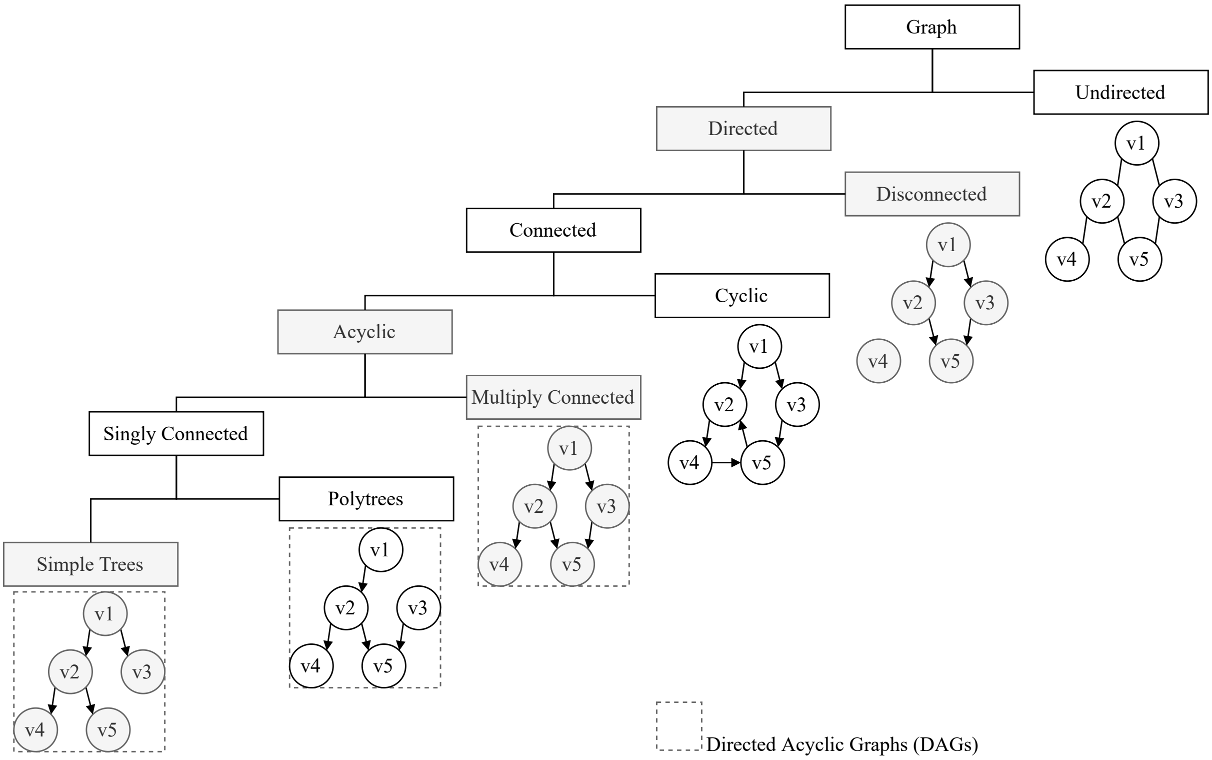


Figure 4 – Types of graphs

Directed Acyclic Graphs (DAGs) represent the qualitative aspect of the BNs graphically. Concerning the DAG that represents a BN, vertices represent the aleatory variables, which correspond to the knowledge domain concepts. The directed edges of a DAG represent, in most cases, a dependency relation between the vertices they connect. Thus, the relation $v_1 \to v_2$ represents a direct dependence of variable v_2 with regard to variable v_1, meaning typically that v_1 has a direct influence on v_2.

Some authors point out that the dependency relation is not necessarily a **cause-effect** relationship and could be just some type of association (SCUTARI; DENIS, 2014). Other authors argue the causal relationship, assuming that the dependency relation is a cause-effect relationship (KJAERULFF; MADSEN, 2013). Based on probabilistic properties, other authors argue in favor of both points of view. They argue that the direction of the edges does not need to have a specific meaning. Although they agree that meaningful BNs express cause-effect relationships, once they correspond to more sparse and natural graphs, resulting in a more transparent and significant interpretation (PEARL, 2009;

KOLLER; FRIEDMAN, 2009). Bayesian models in which the directed edges represent a causal effect are called **causal models**.

3.4.2 Dependencies and Independencies in Graphs

Dependencies and independencies are crucial for understanding BNs behavior and answering queries once the inference model estimates the probability of unobserved variables through other variables whose state has been observed (NIELSEN; JENSEN, 2009).

There is a direct dependency between X and Y if a directed edge exists between X and Y. Thus, X and Y are correlated regardless of evidence about any other variable. Given two not directly linked variables, X and Y, a third variable, Z, in the middle of the undirected path determines conditional independence between X and Y. A vertice Z connecting X and Y specifies an indirect dependency between X and Y (NIELSEN; JENSEN, 2009).

The topology of a BN encodes mainly the conditional independence of the model. Figure 5 illustrates the four cases where the vertice Z connects X and Y: a) indirect causal effect; b) indirect evidential effect; c) common cause; and d) common effect.

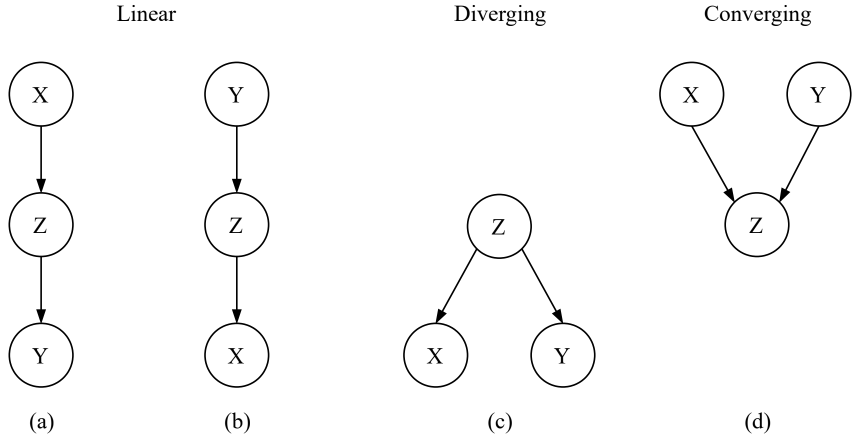


Figure 5 – D-connection types between vertices/variables.

Evidence can be forwarded through the variables of **linear** (serial) connections unless the state of a variable in the middle is known. In the linear connection shown in Figure 5a, if the state of Z is known the cause X can not influence the effect Y. In the linear connection shown in Figure 5b, if the state of Z is known the effect X can not evidence the cause Y (or the effect X can evidence the cause Y only if Z is unknown) (NIELSEN; JENSEN, 2009; CHARNIAK, 1991).

Evidence can pass between all children of a parent variable (vertice) Z in **diverging** connections unless the state of Z is known. In the diverging connection shown in Figure 5c, evidence can pass between X and Y unless the state of Z is known (X is correlated with Y if and only if Z is not known) (NIELSEN; JENSEN, 2009; CHARNIAK, 1991).

It is impossible to infer anything about the parents of a variable Z in **converging** connections unless something is known about Z or its descendants. In the converging connection presented in Figure 5d, if something is known about Z or its descendants, evidence in X can tell us something about Y and vice-versa (NIELSEN; JENSEN, 2009).

In linear and diverging connections, X and Y are independent only if the state of Z is known. Thus X and Y are **d-separated** given Z (**d** connotes "directional"). In converging connections, X and Y are independent and d-separated only if the state of Z or any of its descendants are not known (NIELSEN; JENSEN, 2009; CHARNIAK, 1991).

There are still other general conditional independence properties. As shown in Figure 6, the **Markov condition** states that a variable X is conditionally independent of its non-descendants (ND_1 and ND_2), given its parents (P_1 and P_2).

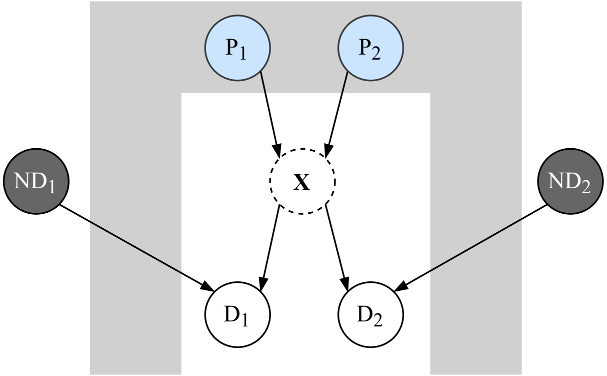


Figure 6 – Conditional independence of non-descendants.

As shown in Figure 7, a variable X is conditionally independent of all other variables in the BN given its **Markov blanket**. The Markov blanket of a variable X is the set composed of its parents, children, and children's parents. Markov blankets follow the d-separation property since the Markov blanket of a variable d-separates it from all other variables.

Grays areas in Figure 6 and 7 represent evidence, these areas "block" probability propagation (RUSSELL; NORVIG, 2020).

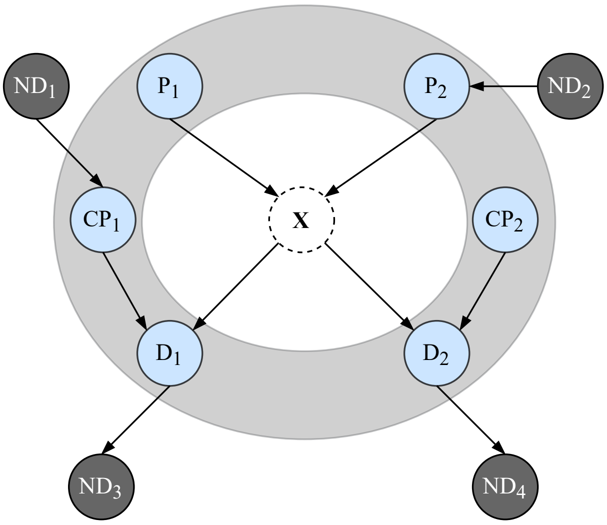


Figure 7 – Conditional independence given the Markov blanket of a variable.

3.4.3 Semantics of Bayesian Networks

Sucintlly, BNs are DAGs in which each vertice corresponds to an aleatory variable. Directed edges connecting pairs of vertices indicate a direct influence of one vertice (parent) over another (child). The **qualitative aspect** of the BNs specifies the correspondence between their syntax with the joint probability distribution over the BN variables. Once the topology of a BN was specified, a conditional distribution must be computed as the local probability for each variable given its parents. As a probabilistic model, each vertice has a CPT that quantifies the effects of its parents on it. The topology and the local probability define the full joint distribution for all variables of a BN (RUSSELL; NORVIG, 2020).

As mentioned before, the full joint probability distribution of a domain will increase as the number of its variables grows. However, given the topology of a BN, only the conditional probabilities for the vertices involved in direct dependencies are required, which means the probability for every node given all possible combinations of its parents. A complete example of a BN (topology and CPTs) can be seen in Figure 3.

The edges in a BN specify the independence assumptions that must hold between the random variables. These assumptions determine what probability information is required to specify the probability distribution among the network's random variables. Each node X_i has an associated probability $P(X_i|parents(X_i))$ that quantifies the effect of its parents on it (NIELSEN; JENSEN, 2009; RUSSELL; NORVIG, 2020). Therefore, the chain rule can be reduced, as shown in Equation 3.11.

$$P(x_i|x_{i+1}, \cdots, x_n) = P(x_i|parents(X_i)) \tag{3.11}$$

Suppose a BN which contains n variables $\{X_1, \cdots, X_n\}$. The product of the relevant elements of the local conditional distributions represents each entry in the joint probability distribution table, as in Equation 3.12. Thus, BNs allow defining the joint distribution based only on their conditional probabilities, reducing the number of probability values needed substantially (NIELSEN; JENSEN, 2009; RUSSELL; NORVIG, 2020).

$$P(X_1{=}x_1, \cdots, X_n{=}x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i)) \tag{3.12}$$

Equation 3.13 presents the **Bayes theorem**, which is the probabilistic basis of the BNs. The Bayes theorem allows computing unknown probabilities from known and stable ones. This simple equation underlies all modern AI approaches for probabilistic inference by helping to simplify the intermediate calculations (RUSSELL; NORVIG, 2020).

$$P(h|D) = \frac{P(D|h) \cdot P(h)}{P(D)} \tag{3.13}$$

Where:

- $P(h)$ is the prior probability of a hypothesis h;
- $P(D)$ is the prior probability of the observed data D;
- $P(D|h)$ is the conditional probability of D given h; and
- $P(h|D)$ is the posterior probability of h given D. It is the belief in the model after seeing the data.

Given the topology and the conditional probabilities of a BN, it is possible to infer the probability of any variable in the network applying basically the Bayes theorem together with some new evidence. Thus, it is possible to take action or search for further evidence to increase the network's confidence (RUSSELL; NORVIG, 2020).

3.4.4 Learning in Bayesian networks

Learning is the task of estimate and select models. Usually, the topology and the probabilities required to define a BN are given by specialists, preview studies, or obtained with experiments and calculus. It is also possible to reach the topology and the statistical information using methods that extract them from the data available (CHARNIAK, 1991; HECKERMAN, 2008; HECKERMAN; GEIGER; CHICKERING, 1995).

There are different learning approaches such as **manual** methods supported by the experience of domain experts, known as supervised learning; **(semi-)automatic** methods that learn from data, known as unsupervised learning; or a combination of both approaches, which combines observed data with experts' experience (KJAERULFF; MADSEN, 2013).

We will adopt the manual construction approach to building our networks due to our data and problem characteristics. Networks created exclusively from the knowledge of experts encode known and expected causal relationships, resulting in the construction of causal models (SCUTARI; DENIS, 2014).

The manual construction of a BN is usually a challenging task. It requires distinct expertise such as model engineering abilities and a comprehensive understanding of the problem domain. The model elicitation process requires: 1) a solid problem definition; 2) a careful identification of the relevant variables; 3) a precise definition of dependences/independences relationships among the chosen variables; and 4) a proper elicitation of many conditional and prior probabilities (KJAERULFF; MADSEN, 2013).

As the parameters of a BN are determined by its structure, creating a BN always proceeds in three consecutive stages. The first step refers to the selection of the variables of interest. The second step refers to identifying the causal, functional or informational relations among the variables to construct the network structure (DAG). The last step refers to estimating the set of conditional and prior probabilities for all network nodes.

3.4.4.1 Variables Identification

The aleatory variables constitute the entire model basis. Selecting variables is one of the most pervasive selection problems in statistical applications. The problem is the uncertainty about which set of variables should establish the relationship between a variable of interest and a subset of potential explanatory or predictor variables (GEORGE, 2000). Domain experts are usually those who perform the selection of the aleatory variables.

There are some fundamental approaches to address this problem, such as the **clarity test** proposed by Kjaerulff and Madsen (2013). According to the clarity test, a variable A must meet three principles to probe whether it has been clearly defined:

- All possible values in the A domain must be exhaustive and mutually exclusive. If the possible values of A are not mutually exclusive, they should be split into several variables;
- Usually, A should represent a unique set of events with no competing variables. That is, the state of `A` should not be given deterministically by the state of another variable;
- A must be clearly defined, leaving no ambiguity concerning its semantics.

Kjaerulff and Madsen (2013) also recommends that it is essential to understand the **types of variables** that may arise. The identification and classification of the variables make it easier to connect them.

- **Problem or hypothesis variables** are the variables of interest from which we may want to calculate the posterior probability given some evidence (information variables). Usually non-observable, these variables relate to the diagnoses or predictions to be made;
- **Information variables** are usually the observable variables that usually have relevant information to the problem-solving. The author separates these variables into the background and symptom variables:
  - **Background variables** usually are among the network root variables and represent the information available before a problem occurs, holding a causal influence over both the problem variables and the symptom variables;
  - **Symptom variables** are the consequence variables usually available after the occurrence of a problem. These variables are children of the problem variables or background variables.

- **Mediating variables** are usually non-observable. Their posterior probability is not of immediate interest, but they help maintain the essential network independence relationships. They tend to be parents of symptom variables and children of problem variables and background variables.

3.4.4.2 Developing the Network Structure

The network structure development defines the dependency relationship between the selected variables. There are two main creational approaches: a **basic approach** based on the natural causal ordering among the previously mentioned types of variables and the **Neil method** proposed by Neil, Fenton and Nielson (2000).

The **basic approach** maintains a causal perspective in the model construction, once this causality is crucial to construct influence diagrams. Such a causal approach may lead to a more suitable representation of the dependence and independence relations and a more reliable estimate of the conditional probabilities (KJAERULFF; MADSEN, 2013).

Thus, the next step in a BN construction process involves identifying and verifying causal links among the selected variables. According to the types of variables, Figure 8 gives an overall view of the causal dependence relations of a BN. The process of eliciting the network structure follows the structure in Figure 8.

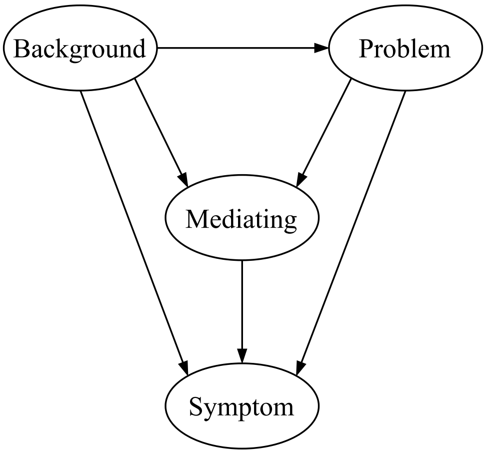


Figure 8 – Typical overall causal structure of a BN. Adapted from Kjaerulff and Madsen (2013).

The **Neil method** creates the network structure based on five commonly occurring substructures. These substructures are known as idioms, and their semantics and syntax represent different methods of uncertain reasoning, covering the vast majority of substructures that can occur in a BN. As described by Neil, Fenton and Nielson (2000), the five idioms are:

- **Definitional or synthesis** integrates many variables into a single variable aiming to organize the BN;
- **Cause/consequence** models cause-effect mechanisms;
- **Measurement** models the uncertainty associated with the accuracy of a measurement instrument;
- **Induction** models inductive reasoning based on populations of similar or exchangeable members;
- **Reconciliation** models the reconciliation of results from competing measurement or prediction systems.

According to the types of variables defined in Section 3.4.4.1, the variables classification depends on their position in the DAG structure. Which of the idioms to chose depends on how we perceive the relationships among the variables. However, the cause/-consequence idiom is the most frequently used substructure. Thus, considering if the relations among the subset of variables are best described using one or more cause/consequence relations is a good starting point (KJAERULFF; MADSEN, 2013). Neil, Fenton and Nielson (2000) present a guide to choosing the proper idiom.

3.4.5 Inference in Bayesian Networks

BNs answer questions concerning the nature of their data through the use of partial **queries**. These queries are performed through techniques known as inference, probabilistic reasoning or belief updating. Given a BN B with n variables $\{X_1, \cdots, X_n\}$, a partial question $Q = \{B, A, E\}$ consists of computing the conditional probability $P(A|E{=}e)$ where:

- A is a target set of non-observed variables;
- The evidence $E{=}e$ is a set of k observed variables $E = (E_1{=}e_1, ..., E_k{=}e_k)$;

- Variables in X not included in A nor E constitute the set of hidden variables H.

Evidence may combine multiple and not always perfect sources of information. Thus, the observation can be uncertain and imprecise, which generates what is known as uncertain evidence. Therefore, there are different types of evidence, such as hard evidence and probabilistic evidence (virtual evidence and soft evidence) (MRAD et al., 2015).

The classic notion of evidence is hard or regular evidence that precisely specifies the state of a random variable. It is an observation that a variable A definitely has a particular value (e.g., A=1) (PEARL, 1988).

Virtual evidence, also known as likelihood evidence, corresponds to the cases where the observation is uncertain. It is usually interpreted as evidence with uncertainty and is commonly represented as a likelihood ratio. A likelihood $P(A)$ represents virtual evidence of a variable A as in Equation 3.14 (PEARL, 1988).

$$P(A) = (P(a_1|a_1), ..., P(a_n|a_n)) \tag{3.14}$$

Where $P(a_i|a_i)$ is the probability of observe A in the state a_i if it really is in the state a_i.

Soft evidence is usually interpreted as evidence of uncertainty and is represented as a probability distribution of one or more variables. There is uncertainty concerning the precise value of a variable A, but certainty regarding its probability distribution $P(A)$. Since $P(A)$ distribution is a certain observation, updating network belief should preserve it (VALTORTA; KIM; VOMLEL, 2002).

This preservation of the local distribution of the evidence variable is the main difference between soft evidence and virtual evidence, once virtual evidence does not require this preservation. Belief in virtual evidence is not fixed and can be modified by further evidence on other variables (MRAD et al., 2015).

There are three main categories of partial queries: Conditional Probability Query (CPQ), Maximum a Posteriori (MAP), and Most Probable Explanation (MPE) or Marginal MAPs (SCUTARI; DENIS, 2014; KOLLER; FRIEDMAN, 2009).

Koller and Friedman (2009) classify **CPQs** as:

- **Causal, deductive or predictive reasoning**: that estimates the probability of a variable given the observation of non-descending variables (from causes to effects). In the BN example in Figure 3, $P(SP|SM, CS)$ represents this type of query;

- **Evidence, abductive or explanation reasoning**: that estimates the probability of a variable given the observation of descending variables (from effects to causes). In the BN example in Figure 3, $P(IR|GDP, SP)$ represents this type of query;

- **Inter-causal reasoning**: that addresses the interaction of causing variables with regard to the same effect variable. It refers to the decrease in the belief of competing hypotheses once observed the occurrence of one or several hypotheses. In the BN example in Figure 3, $P(CS|SM, SP)$ represents this type of query.

MPAs and MPEs consist of identifying the most likely configuration for all variables in A that maximize the posterior probability of E. In MPE, A coincides with all remaining variables in the subset $\{X - E\}$. In MAP, A is a strict subset of “hypothesis” variables in $\{X - E\}$. Thus, MPEs and MAPs calculate the most probable assignment for A (a^*) in a model X given evidence $E{=}e$, as in Equation 3.15 (DARWICHE, 2008).

$$(A|E{=}e) = a^* = \underset{A}{\text{argmax}} P(A|E) \tag{3.15}$$

All these inference problems are complex. The decision version of MPEs, CPQs, and MAPs are known to be NP-complete, PP-complete, and NP^{PP}-complete[1], respectively. There are exact and approximate algorithms for answering these queries. All exact inference algorithms have an exponential complexity regarding the BN treewidth. Approximate inference algorithms usually are not sensitive to the BN treewidth and can be pretty efficient regardless of the BN topology. However, the approximate methods usually present issues regarding the quality of answers they compute, which is commonly related to the amount of time scheduled by the algorithm (DARWICHE, 2008).

The most suitable inference algorithm will depend on the accuracy required and the computational cost. The structure of our BNs will be directed singly connected graph (polytrees). Therefore, we will use exact inference algorithms to perform our queries. Besides generate reliable results, the time and space complexity of exact inference in

[1] NP-, PP-, and NP^{PP}-complete are classifications for the complexity of common problems in computer science. These classifications usually describe the amount of computer time (elementary operations performed) and space an algorithm takes to run (OZTOK; CHOI; DARWICHE, 2016; PAPADIMITRIOU, 1994).

polytrees is linear in the size of the BN (the number of CPT entries) (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009).

Variable elimination is the simplest algorithm for exact inference in PGMs, and is very efficient on models whose DAG representation is a tree. Belief propagation is another algorithm to satisfy CPQs with exact inference when the DAG is a tree (RUSSELL; NORVIG, 2020; PEYRARD et al., 2019).

3.5 Markov Models

It may be necessary to model dynamic systems that allow reasoning about the state of the world as it evolves. These systems states are also represented as a set of aleatory variables, whose values at time t are a snapshot of the relevant system attributes. It is possible to model BNs representing a temporal probability model, known as Dynamic Bayesian Networks (DBNs). DBNs model stochastic processes over time intervals (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009).

DBMs are not the first temporal method of reasoning under uncertainty. Hidden Markov Models have great popularity due to their compact representation, fast learning, and fast inference techniques (RUSSELL; NORVIG, 2020). According to Koller and Friedman (2009), the Hidden Markov Models are the simplest nontrivial type of these **state-observation temporal models**.

In probability theory, a Markov Model is a **Stochastic Process** (SP) which consists of a family of variables that evolve regarding some parameter, usually time. An SP is represented by $\{X_t \mid t \in T\}$, where:

- T is the parametric space, formed by a set of ordered values (e.g., time);
- t is a given value in T; and
- Each X_t is an aleatory variable. The set of its possible values is called the states space, and its specific values at any given time are the process states.

In general, SPs are used to study the evolution of phenomena or systems. Given an initial condition, all system evolution is unknown, having several possible trajectories for its evolution. The SPs analysis determines the probability distributions for each set of aleatory variables, using them to predict future behaviors (states) given past behaviors (states). In contrast with deterministic models, those specified by a set of equations that describe exactly how a system evolves, the evolution of stochastic models is random, and if the process runs several times (realizations of the process), it will not give the same results (JELINEK, 1997; RABINER, 1989).

Let $\{X_0, X_1, \ ... \ X_t, \ ..., \ X_T\}$ be a sequence of stochastic variables, where $(0 \ \leq \ t \ \leq \ T)$ represents a discrete time order, defined for the same discrete and finite state space. If nothing else is considered, the joint probability of these stochastic variables is given by the chain rule (JELINEK, 1997), as shown in Equation 3.16.

$$\begin{aligned} P(X_0, X_1, ..., \ X_T) &= \prod_{t=0}^{T} P(X_t | X_0, X_1, ..., \ X_{t-1}) \\ &= P(X_0)P(X_1|X_0)P(X2|X_0, X_1) \ ... \\ &\quad P(X_T|X_0, X_1, X_2, ..., \ X_{T-1}) \end{aligned} \tag{3.16}$$

An SP is taken as **Markovian** if it satisfies the property shown in Equation 3.17.

$$P(X_t|X_0, X_1, X_2, ..., \ X_{t-1}) = P(X_t|X_{t-1}) \tag{3.17}$$

When dealing with a **markovian process**, the Equation 3.16 can be simplified, as shown in Equation 3.18.

$$\begin{aligned} P(X_0, X_1, ..., \ X_T) &= \prod_{t=0}^{T} P(X_t|X_{t-1}) \\ &= P(X_0)P(X_1|X_0)P(X_2|X_1)P(X_3|X_2) \ ... \\ &\quad P(X_T|X_{T-1}) \end{aligned} \tag{3.18}$$

3.5.1 Markov Chains

Markovian processes in discrete state spaces are known as Markov Chains (MCs). An MC is a memoryless SP whose future state depends only on its current state, disregarding past states. Satisfying what is known as **Markov property**, a MC X_t is a SP where given a value of X_t, the values of X_s $(t \ < \ s)$ are not influenced by the values of X_u $(u \ < \ t)$. Or, more succinctly, successive steps are statistically independent (REICHL, 2016).

Grinstead and Snell (1998) made an interesting description of MCs by defining it as a set of **states** $S = \{s_1, s_2, ..., s_r\}$ in a **process**. The process starts in one of these states and **moves** successively from one state to another. Each move is called a **step**. If the chain is in a current state s_i, then it moves to a state s_j at the next step with a probability denoted by p_{ij}, and this probability does not depend upon which states the chain was before the current state s_i. The probabilities p_{ij} are called **transition probabilities**. The process can remain in the state it is in, and this occurs with probability p_{ii}. An **initial probability distribution**, defined on S, specifies the starting state and is calculated as a vector π that indicates the initial probability of each state.

This probability distribution of the states transitions is typically represented in a **transition matrix**. If a MC has N possible states, its transition matrix will be an NxN matrix, where each entry N_{ij} is the transition probability from state i to state j. The transition matrix must be stochastic, which is a matrix where entries in each row must add up to exactly one ($\sum_{j=1}^{n} P_{ij} = 1$) since each row represents its probability distribution. The transition matrix probabilities can vary over time or be stationary (when its probabilities are time-independent). The π vector and the hidden states (BuM, StM, BeM) in Figure 9 illustrate an MC.

Through the transition matrix it is possible to obtain the absolute probability of the system states after a given number of transitions. The probability of a system composed by: 1) N states $(1, 2 \ldots N)$; 2) a transition matrix A_{NxN}; and 3) an initial state vector π_0, stay in one of its N states after k transitions is seen in equation 3.19.

$$\pi_k = \pi_0 (A_{NxN})^k \tag{3.19}$$

Where:

- Each A_{ixj}^k position is the probability of staying in state j, since it started in state i, after k transitions; and
- π_k has the probabilities of staying in each state after k transitions when considering the initial state vector π_0.

3.5.2 Hidden Markov Models

Most Markovian processes consist of states that can be directly observed. However, HMMs are used to model Markovian processes that generate **indirectly observable**

states through the transitions between the states of the MC that govern the process, but which can not be directly observed. HMMs are double-layered SPs with a nonvisible SP that can be observed through another SP that produces the sequence of observations (RABINER, 1989).

The **hidden process** is a set of states connected by transitions with probabilities (an MC). In contrast, the **observable process** is a set of outputs or visible states emitted by each not observable state according to some output of a probability density function. The challenge is to determine the hidden states from the visible states (RABINER, 1989).

The fundamental difference between HMMs and the rest of the Markovian processes is how the system is observed. HMMs have an indirect observation of the states, carried out by inference since the observable ones are probabilistic functions regarding the states of the chain or regarding the transition between these states. In contrast, the rest of the markovian processes has direct observation, where the observable ones are the states themselves.

Most Neural Networks are probabilistic methods. They work in a discriminative approach to take inputs from a high-dimensional space and map it to a lower-dimensional space. On the other hand, HMMs are statistical methods that work in a generative approach that models conditional dependencies of hidden states. Each state has a probability distribution regarding the observations. An HMM hidden state is the entity's identity that caused each observation, and this hidden cause is translated statistically into the observed data. Through the forward-backward algorithms, it is possible to find the conditional distribution over the hidden states (CAPPÉ; MOULINES; RYDÉN, 2006; RABINER, 1989).

Described for the first time in the late 1960s and early 1970s (BAUM; PETRIE, 1966; BAUM; EAGON, 1967), HMM applications began to be used in word recognition in the middle 1970s (BAKER, 1975). HMMs appear in the literature under various names, such as Hidden Markov Processes, Markov Sources, Hidden Markov Chains, and Probabilistic Functions of Markov Chains. HMM's first applications focused on speech and handwriting recognition and DNA sequencing, reaching, later, great importance in bioinformatics.

3.5.2.1 Hidden Markov Models Structure

An HMM structure is characterized by:

- T: the observation sequence length;
- N: the number of distinct states in the model;
- S: a set of states. Individual states are labeled $\{1, 2, \ ..., N\}$ and the state at time t as Q_t;
- M: the number of distinct observable symbols in the model.
- V: a set of symbols. Individual symbols are denoted as $\{v_1, v_2, \ ..., v_M\}$;
- $A = \{a_{ij}\}$: the transition probability distribution from state a, where: $a_{ij} = \ P[q_{t+1} = j|q_t = i], 1 \leq i, j \leq N$ (a_{ij} can be read as $P(state\ q_j\ at\ t+1|state\ q_i\ at\ t)$);
- B: a NxM probability distribution matrix which relates the states of the set S (rows) to the observable symbols of the set V (columns). $B = \{b_j(k)\}$ defines the observation probability distribution of symbols in the state $j, \{j_1, j_2, ..., j_N\}$, where: $b_j(k) = \ P[O_t = \ v_k|q_t = j], \ 1 \leq k \leq M$. As A, B is stochastic and its probabilities $b_j(k)$ are time independent (($b_j(k)$ can be read as $P(observation\ k\ at\ t|state\ q_j\ at\ t)$);
- $\pi = \{\pi_i\}$: the initial state distribution, where: $\pi_i = \ P[q_1 = i], 1 \leq i \leq N$.

Thus, the HMM specification requires the definition of two model parameters (N and M), a symbol observation specification, and the definition of three sets of probability distribution A, B, and π. The complete set of model parameters is defined as $\lambda = (A, B, \pi)$. This set of parameters defines the measure of probability for O, $P(O|\lambda)$, where O is a set of observed states.

A different graphical notation depicts the HMMs structure. Directed (generally cyclic) graphs represent the HMMs transition/emission model, in which vertices denote the different states and edges indicate the transitions/emissions between states (KOLLER; FRIEDMAN, 2009).

Figure 9 presents the structure of an HMM that represents part of the stock exchange domain. The three hidden variables that form the hidden MC represent the stock market states, Bull Market (BuM), Bear Market (BeM), and Stagnant Market (StM). The edges between these hidden states represent the possible transitions. The values next to each edge indicate the transition probabilities between the hidden states. The two observable symbols represent two critical economic indicators: a high interest rate (HIR) and a growing gross domestic product (GGDP). The dashed edges arriving at the observable states represent the possible emissions. The values next to each dashed edge indicate the emission probabilities from hidden to observable states. This HMM would make it possible to predict the stock market direction by observing the economic indicators.

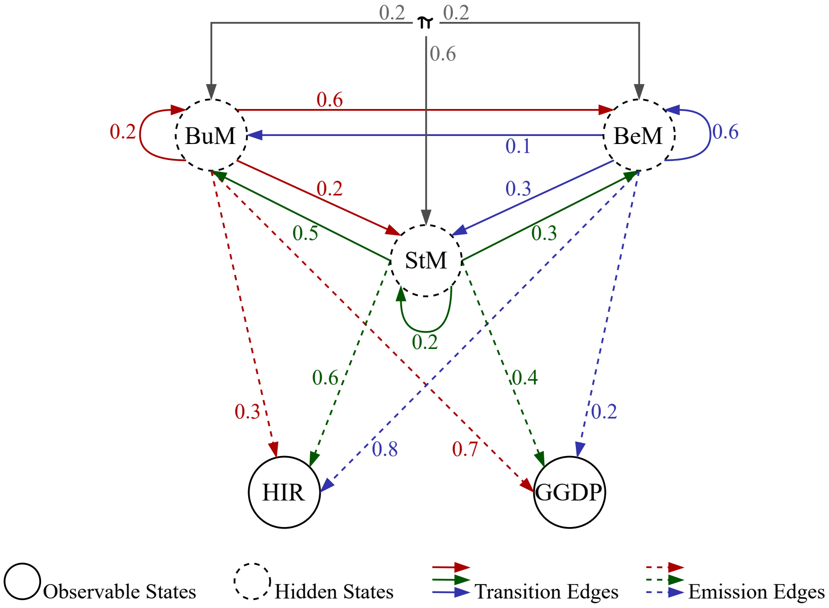


Figure 9 – An HMM example over five states (three hidden/non-observable states and two observable states).

The parameters of the HMM displayed in Figure 9 are listed below:

- $N = 3$; $S = \{\text{BuM, StM, BeM}\}$;
- $M = 2$; $V = \{\text{HIR, GGDP}\}$;
- $A =$

$$\begin{array}{cccc} \text{BuM} & \text{StM} & \text{BeM} & \\ 0.2 & 0.2 & 0.6 & \text{BuM} \\ 0.5 & 0.2 & 0.3 & \text{StM} \\ 0.1 & 0.3 & 0.6 & \text{BeM} \end{array}$$

- B=

$$\begin{array}{cc} \text{HIR} & \text{GGDP} & \\ \left[\begin{array}{cc} 0.3 & 0.7 \\ 0.6 & 0.4 \\ 0.8 & 0.2 \end{array}\right] & \begin{array}{c} \text{BuM} \\ \text{StM} \\ \text{BeM} \end{array} \end{array}$$

- π=

$$\begin{array}{ccc} \text{BuM} & \text{StM} & \text{BeM} \\ \left[\,0.2\right. & 0.6 & \left.0.2\,\right] \end{array}$$

There are three main **problems** we can solve using HMMs:

1. **Evaluation problem**: given an observation sequence O, and a model λ, how to calculate the probability of O be produced by the model ($P(O|\lambda)$);

2. **Best sequence of states**: given an observation sequence O, and a model λ, how to calculate an optimal state sequence Q for a given sequence of observations;

3. **Training**: how to adjust the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$.

These three problems are traditionally solved, respectively, by **Forward-backward**, **Viterbi** and **Baum-Welch** or **K-Means** algorithms (CAPPÉ; MOULINES; RYDÉN, 2006; RABINER, 1989).

3.6 Artificial Intelligence Applications in Medical Researches

Several AI techniques have been applied in medical research. Deep neural architectures are being applied in different biomedical areas, such as public and medical health management, bio and medical imaging, and brain and body machine interface (ZEMOURI; ZERHOUNI; RACOCEANU, 2019; LITJENS et al., 2017; LEE et al., 2017). Current and potential uses of AI in healthcare also include dermatology, ophthalmology, radiology, histopathology, and nuclear medicine (LEE et al., 2019).

Some researches involving the use of intelligent systems applied to autism propose the formulation of diagnostic methods based on magnetic resonance imaging (HEINSFELD et al., 2018; BHAUMIK et al., 2018; KHOSLA et al., 2018; LIAO; LU, 2018; ZHAO et al., 2018; DVORNEK; VENTOLA; DUNCAN, 2018; DEKHIL et al., 2018b; DEKHIL et al., 2018a; HAZLETT et al., 2017; EMERSON et al., 2017; DVORNEK et al., 2017; YAHATA et al., 2016), early prediction approaches from behavioral and developmental measures (BUSSU et al., 2018), the use of robots and other AI techniques applied to the therapy processes of ASD children (ALVES et al., 2020), wearable assistive technologies (BENSSASSI et al., 2018), approaches to predicting autism risk genes (BRUEGGEMAN; KOOMAR; MICHAELSON, 2020; LIN et al., 2018; LE; VAN, 2017), to reveal differences in regional brain structure between ASD and TD people (GÓRRIZ et al., 2019), and to model the diagnostic heterogeneity of ASD (LOMBARDO; LAI; BARON-COHEN, 2019).

Applied in several areas, BNs are among the AI tools that have been most successful in practical applications for medicine (SAHEKI, 2005). The most common approaches are for medical diagnoses, such as diagnosing diseases of the lymph node (HECKERMAN; HORVITZ; NATHWANI, 1992), heart disease diagnosis (SPIEGELHALTER; FRANKLIN; BULL, 2013; SAHEKI, 2005), and computerized tongue diagnosis (ZHANG; ZHANG; ZHANG, 2017). Other human applications include automated language (CHARNIAK; GOLDMAN, 1990) and text understanding (GOLDMAN, 1991), describing the interaction between genes (FRIEDMAN et al., 2000) and control of Computer Vision systems (LEVITT; AGOSTA; BINFORD, 1990).

Regarding mental disorders, Palmer, Lawson and Hohwy (2017) gathered Bayesian approaches to autism within a framework that extends from simple to complex Bayesian inference models. Given that the ASD core features relate to how individuals interact with the world around them, they propose that ASD is characterized by a greater weighting of sensory information in updating probabilistic representations of the environment. Thus, ASD may relate to finer mechanisms involved in the adjustment of sensory perception, and the hypotheses regarding atypical sensory weighting in ASD have direct implications for behavior regulation. They base their work on a theory called predictive processing, in which top-down and bottom-up messages passing across the cerebral cortex implement hierarchical probabilistic inference on the sensory stimulation causes. The hypothesis regarding ASD is that the incoming sensory signals are weighted more highly when integrated with the brain's existing model of the environment, such that neural processes like perception are dictated to a greater extent by the present sensory data rather than prior or contextual information.

HMMs also have been used for modeling several different problems in medical researches (KROGH; MIAN; HAUSSLER, 1994; MEYER; DURBIN, 2002; TESTA et al., 2015), including approaches to diagnose cancer (MANOGARAN et al., 2018), for genotype imputation (BROWNING; BROWNING, 2009; LI et al., 2010; MARCHINI et al., 2007; HOWIE; DONNELLY; MARCHINI, 2009; MARCHINI; HOWIE, 2010), and to investigate heart abnormalities (FAHAD et al., 2018; DWIVEDI; IMTIAZ; RODRIGUEZ-VILLEGAS, 2018; SARAÇOĞLU, 2012; CHAUHAN et al., 2008; WANG et al., 2007; UĞUZ; ARSLAN; TÜRKOĞLU, 2007).

Regarding mental disorders, HMMs have been applied to evaluate the pronunciation quality and acquisition of language skills (SCHIPOR; PENTIUC; SCHIPOR, 2012; SAZ et al., 2009), to forecast a possible future ASD diagnosis from infants with a high risk of ASD (ALIE et al., 2011), to diagnose emotion-related mental diseases (GUO et al., 2017), and to recognize the stereotyped gestures which are typical of ASD people (CAMADA; CERQUEIRA; LIMA, 2017).

3.7 Summary

This chapter presented an overview of the AI sub-fields, with emphasis on the probabilistic graphical models. We further explored these approaches because they are widely used for inference in environments of uncertainty. Moreover, an overview concerning probability theory also was necessary due to its importance to the probabilistic models. We dedicated special attention to understanding the models' fundamentals, how they work, and what they can do.

We started showing the vast dimension of the AI field by succinctly defining it and describing its main sub-fields. Then, we presented the probabilistic networks, which allow inter-causal reasoning to build inference models. We also introduced both the probability theory basics and the graphs fundamentals since these techniques underlie the development of the graphical models and the inference process.

Both Bayesian and Markovian approaches seem the most suitable methods to model the complex genetic nature of ASD etiology. The statistical characteristics of the ASD heritability and recurrence rates also contributed to the adoption of these methods. Thus, BNs and HMMs were sufficiently explored once these two types of methodologies can infer unknown states given a piece of evidence.

HMMs were applied to model the causal relationship between diagnosed/non-diagnosed parents regarding the ASD risk in their future descendants. Details in Chapter 4.

We will employ BNs to model an ASD causal network capable of estimating, for example:

- The probability of a female individual being autistic, given that she has a paternal male cousin, son of an aunt, diagnosed as autistic; or
- The probability of a male individual being autistic, given that he has one diagnosed older brother and one typical older sister; or
- The risk of ASD in parents, given that they have one or more diagnosed children.

Details in Chapters 5, 6 and 7.

4 Hidden Markov Models to Estimate the Risk of Having Autistic Children

Genetic factors have been pointed out as the primary root associated with the risk of autism. Recent works indicate that $\approx 80\%$ of autistic people have inherited the condition from their parents (see Section 2.4). However, there are no estimates that indicate the likelihood of an autistic parent having an autistic child. Using HMMs and the data of autism heritability, we developed a model to investigate the likelihood of autistic parents generating autistic children. Our model was built and validated using statistical data from the association of gender with recurrence of autism among siblings and statistical data from the association of genetic factors with autism.

Based on the possible observation of some characteristics of the parents, our approach was to estimate the probability variation of generating ASD children, a not observable condition before birth. Given the statistical nature of ASD heritability and recurrence data available in the literature, HMMs seemed to be the most straightforward, appropriate, and transparent strategy to start this investigation. This adequacy is mainly due to the HMMs generative approach, which, based on prior probabilities of each state, allows to infer a distribution probability over the possible values of the hidden states.

For this, we used two sets of statistical data. A set of statistical information about ASD recurrence among siblings was used to model the hidden states transition probabilities, and a set of statistical data about ASD heritability was used to model the observable states emission probabilities. We did not use direct individual observations to train our HMMs parameters because, to the best of our acknowledgment, there is no such kind of public data available. Thus, we used relevant statistical data in the literature for the adjustment (training) of the proposed model parameters. The use of these known statistical relationships does not mean that our models are static or deterministic. Such data may change when arising either new data to use as training data or new relevant statistical data about ASD heritability.

The remaining sections explain our assumptions about the probabilities used to model our HMMs. We used such HMMs to estimate the likelihood of ASD parents generating ASD children. We created six variations of the chains, each one according to the children's gender and transition matrices. Thus, it was possible to estimate the probabilities for ASD girls and ASD boys separately, which is essential given the difference in the ASD prevalence between genders. The following sections describe our methodology stages.

4.1 Hidden Markov Models States

It was necessary to define the hidden and observable states before creating our HMMs.

The hidden chain was composed of two states ($N = 2$):

- **TD**: meaning a Typical Girl/Boy;
- **ASD**: meaning an ASD Girl/Boy.

The observable chain was composed of two states ($M = 2$):

- **TP**: meaning Typical Parents (father AND mother without ASD diagnosis);
- **AP**: meaning ASD Parents (father OR mother with ASD diagnosis).

A clinical investigation or genetic characteristics recognized as possible causes of ASD may characterize an ASD diagnosis. In addition, specific autistic traits may indicate an ASD diagnosis once there is a genetic sharing between ASD and some autistic traits (e.g., childhood behavior, rigidity, and attention to detail) (BRALTEN et al., 2018; LUNDSTRÖM et al., 2012; ROBINSON et al., 2011).

We used the hidden and observable states defined in this section for modeling all of our HMMs.

4.2 Initial State Distribution (π)

Although there are different and important studies related to ASD prevalence (Table 1), we used an ASD prevalence among children calculated from the ASD diagnosis data presented by Palmer et al. (2017). From these diagnosis data, we calculated three important probabilities: 1) the children general ASD prevalence, regardless of gender ($P(A) = 0.0125$); 2) the ASD prevalence among girls ($P(AG) = 0.005$); and 3) the ASD prevalence among boys ($P(AB) = 0.0197$). One of the most important researches on ASD prevalence indicates that 1/54 children were diagnosed inside the spectrum (MAENNER et al., 2020). This same research shows that for each ASD girl (prevalence of 20%), there are four ASD boys (prevalence of 80%). Although the prevalence calculated from Palmer et al. (2017) shows a lower ASD prevalence, both general and by gender, it corroborates the relation of four ASD boys to each ASD girl.

We chose to use the calculated prevalence because the research of Palmer et al. (2017) was conducted among pairs of siblings, recording the probability of younger siblings being autistic concerning the older sibling's condition. Thus, this pattern was essential to use that information to develop our transition data and validate our prediction model accurately by simulating the population of Palmer et al. (2017).

From the probabilities and the hidden states previously defined, it was possible to determine our initial state distribution vectors (π) for both girls (πG) and boys (πB).

$$\pi = \begin{array}{cc} \text{TD} & \text{ASD} \\ \left[1 - P(ASD) \right. & \left. P(ASD) \right] \end{array}$$

$$\pi G = \begin{array}{cc} \text{TG} & \text{AG} \\ \left[0.995 \right. & \left. 0.005 \right] \end{array}$$

$$\pi B = \begin{array}{cc} \text{TB} & \text{AB} \\ \left[0.9803 \right. & \left. 0.0197 \right] \end{array}$$

These initial state distribution vectors were used for modeling all of our HMMs according to the respective children's gender.

4.3 Transition Matrix (A)

Several works have studied the ASD recurrence rates among siblings (Table 4). Palmer et al. (2017) made a more uniform estimate of ASD sex-specific recurrence rates among sibling pairs, with a similar amount of siblings pairs according to siblings gender (395 220 male:male pairs, 400 249 female:male pairs, 359 679 female:female pairs, and 405 389 male:female pairs).

According to Palmer et al. (2017), when the older male sibling had the ASD diagnosis, ASD was diagnosed in 4.2% of female siblings and 12.9% of male siblings. When the older female sibling had the ASD diagnosis, ASD was diagnosed in 7.6% of female siblings and 16.8% of male siblings. These statistics clearly show us the increased likelihood of a younger sibling being diagnosed as autistic when he/she has an older sibling already diagnosed. Alternatively, when the older male sibling did not have the ASD diagnosis, ASD was diagnosed in 0.4% of female siblings and 1.5% of male siblings. When the older female sibling did not have the ASD diagnosis, ASD was diagnosed in 0.4% of female siblings and 1.8% of male siblings. These statistics clearly show us the decreased likelihood of a younger sibling being diagnosed as autistic when he/she has an older sibling not diagnosed.

Although an ASD older sibling suggests an increase in the likelihood of ASD in a younger sibling, the older sibling condition is not the determining genetic factor. Instead, the determining genetic factor is what they have in common, their parents. Therefore, as we aim to estimate the risk of ASD children based on the parents' characteristics, we used the data of ASD recurrence among siblings to calculate the transition probabilities among our HMMs states.

We created three transition matrices for each gender since the probabilities significantly changed according to the older sibling gender. Two of them, according to the older sibling gender, and the other one disregarding the older sibling gender. We calculated all transition probabilities presented in the following subsections from the diagnostic data of the population studied by Palmer et al. (2017).

4.3.1 Transition Matrices for the Birth of Females

To simulate female births, given that there is an older brother, we calculated the following conditional probabilities: 1) a girl being autistic, given that she has an autistic older brother ($P(AG|AB) = 0.0422$); 2) a girl being autistic, given that she has a typical older brother ($P(AG|TB) = 0.0038$); 3) a girl being typical, given that she has an autistic older brother ($P(TG|AB) = 1 - P(AG|AB)$); and 4) a girl being typical, given that she has a typical older brother ($P(TG|TB) = 1 - P(AG|TB)$). These conditional probabilities constitute the transition matrix $A(MF)$.

$$A(MF) = \begin{array}{c c} \begin{array}{cc} \text{TG} & \text{AG} \end{array} & \\ \begin{bmatrix} P(TG|TB) & P(AG|TB) \\ P(TG|AB) & P(AG|AB) \end{bmatrix} & \begin{array}{c} \text{TB} \\ \text{AB} \end{array} \end{array}$$

$$A(MF) = \begin{array}{c c} \begin{array}{cc} \text{TG} & \text{AG} \end{array} & \\ \begin{bmatrix} 0.9962 & 0.0038 \\ 0.9578 & 0.0422 \end{bmatrix} & \begin{array}{c} \text{TB} \\ \text{AB} \end{array} \end{array}$$

For clarification purposes, position $\{A(MF)_{0,1} = 0.0038\}$ is the conditional probability value of $P(AG|TB)$, which is the transition probability from the state TB to the state AG. In other words, it means the probability of a TD older brother having an ASD younger sister.

To simulate female births, given that there is an older sister, we calculated the following conditional probabilities: 1) a girl being autistic, given that she has an autistic older sister ($P(AG|AG) = 0.0759$); 2) a girl being autistic, given that she has a typical older sister ($P(AG|TG) = 0.0045$); 3) a girl being typical, given that she has an autistic older sister ($P(TG|AG) = 1 - P(AG|AG)$); and 4) a girl being typical, given that she has a typical older sister ($P(TG|TG) = 1 - P(AG|TG)$). These conditional probabilities constitute the transition matrix $A(FF)$.

$$A(FF) = \begin{array}{c c} \begin{array}{cc} \text{TG} & \text{AG} \end{array} & \\ \begin{bmatrix} P(TG|TG) & P(AG|TG) \\ P(TG|AG) & P(AG|AG) \end{bmatrix} & \begin{array}{c} \text{TG} \\ \text{AG} \end{array} \end{array}$$

$$A(FF) = \begin{array}{cc} \text{TG} & \text{AG} \\ \left[\begin{array}{cc} 0.9955 & 0.0045 \\ 0.9241 & 0.0759 \end{array}\right] \end{array} \begin{array}{c} \\ \text{TG} \\ \text{AG} \end{array}$$

To simulate female births, regardless the older sibling gender, we calculated the following conditional probabilities: 1) a girl being autistic, given that she has an autistic older sibling ($P(AG|ASD) = 0.0486$); 2) a girl being autistic, given that she has a typical older sibling ($P(AG|TD) = 0.0041$); 3) a girl being typical, given that she has an autistic older sibling ($P(TG|ASD) = 1 - P(AG|ASD)$); and 4) a girl being typical, given that she has a typical older sibling ($P(TG|TD) = 1 - P(AG|TD)$). These conditional probabilities constitute the transition matrix $A(XF)$.

$$A(XF) = \begin{array}{cc} \text{TG} & \text{AG} \\ \left[\begin{array}{cc} P(TG|TD) & P(AG|TD) \\ P(TG|ASD) & P(AG|ASD) \end{array}\right] \end{array} \begin{array}{c} \\ \text{TD} \\ \text{ASD} \end{array}$$

$$A(XF) = \begin{array}{cc} \text{TG} & \text{AG} \\ \left[\begin{array}{cc} 0.9959 & 0.0041 \\ 0.9514 & 0.0486 \end{array}\right] \end{array} \begin{array}{c} \\ \text{TD} \\ \text{ASD} \end{array}$$

4.3.2 Transition Matrices for the Birth of Males

To simulate male births, given that there is an older brother, we calculated the following conditional probabilities: 1) a boy being autistic, given that he has an autistic older brother ($P(AB|AB) = 0.1293$); 2) a boy being autistic, given that he has a typical older brother ($P(AB|TB) = 0.0154$); 3) a boy being typical, given that he has an autistic older brother ($P(TB|AB) = 1 - P(AB|AB)$); and 4) a boy being typical, given that he has a typical older brother ($P(TB|TB) = 1 - P(AB|TB)$). These conditional probabilities constitute the transition matrix $A(MM)$.

$$\begin{array}{cccc} & \text{TB} & \text{AB} & \\ A(MM) = & \left[\begin{array}{c} P(TB|TB) \\ P(TB|AB) \end{array}\right. & \left.\begin{array}{c} P(AB|TB) \\ P(AB|AB) \end{array}\right] & \begin{array}{c} \text{TB} \\ \text{AB} \end{array} \end{array}$$

$$\begin{array}{cccc} & \text{TB} & \text{AB} & \\ A(MM) = & \left[\begin{array}{c} 0.9846 \\ 0.8707 \end{array}\right. & \left.\begin{array}{c} 0.0154 \\ 0.1293 \end{array}\right] & \begin{array}{c} \text{TB} \\ \text{AB} \end{array} \end{array}$$

To simulate male births, given that there is an older sister, we calculated the following conditional probabilities: 1) a boy being autistic, given that he has an autistic older sister ($P(AB|AG) = 0.1681$); 2) a boy being autistic, given that he has a typical older sister ($P(AB|TG) = 0.0180$); 3) a boy being typical, given that he has an autistic older sister ($P(TB|AG) = 1 - P(AB|AG)$); and 4) a boy being typical, given that he has a typical older sister ($P(TB|TG) = 1 - P(AB|TG)$). These conditional probabilities constitute the transition matrix $A(FM)$.

$$\begin{array}{cccc} & \text{TB} & \text{AB} & \\ A(FM) = & \left[\begin{array}{c} P(TB|TG) \\ P(TB|AG) \end{array}\right. & \left.\begin{array}{c} P(AB|TG) \\ P(AB|AG) \end{array}\right] & \begin{array}{c} \text{TG} \\ \text{AG} \end{array} \end{array}$$

$$\begin{array}{cccc} & \text{TB} & \text{AB} & \\ A(FM) = & \left[\begin{array}{c} 0.9820 \\ 0.8319 \end{array}\right. & \left.\begin{array}{c} 0.0180 \\ 0.1681 \end{array}\right] & \begin{array}{c} \text{TG} \\ \text{AG} \end{array} \end{array}$$

To simulate male births, regardless the older sibling gender, we calculated the following conditional probabilities: 1) a boy being autistic, given that he has an autistic older sibling ($P(AB|ASD) = 0.1368$); 2) a boy being autistic, given that he has a typical older sibling ($P(AB|TD) = 0.0167$); 3) a boy being typical, given that he has an autistic older sibling ($P(TB|ASD) = 1 - P(AB|ASD)$); and 4) a boy being typical, given that he has a typical older sibling ($P(TB|TD) = 1 - P(AB|TD)$). These conditional probabilities constitute the transition matrix $A(XM)$.

$$A(XM) = \begin{array}{cc} \text{TB} & \text{AB} & \\ \left[\begin{array}{c} P(TB|TD) \\ P(TB|ASD) \end{array}\right. & \left.\begin{array}{c} P(AB|TD) \\ P(AB|ASD) \end{array}\right] & \begin{array}{c} \text{TD} \\ \text{ASD} \end{array} \end{array}$$

$$A(XM) = \begin{array}{cc} \text{TD} & \text{ASD} & \\ \left[\begin{array}{c} 0.9833 \\ 0.8632 \end{array}\right. & \left.\begin{array}{c} 0.0167 \\ 0.1368 \end{array}\right] & \begin{array}{c} \text{TD} \\ \text{ASD} \end{array} \end{array}$$

4.4 Emission Data (B)

Because genetic factors may point to an ASD risk increase, we took ASD diagnosis/genes in parents as the observable characteristic, which may predict the probability of generating ASD children. We assumed that parents' characteristics (genetics or clinical diagnosis) could be observed before having children.

4.4.1 Autistic Parents Given They Have Autistic Children

A study conducted among siblings has identified 14 516 children diagnosed with ASD. Such work studied 37 570 twin pairs; 2 642 064 full sibling pairs; and 432 281 maternal and 445 531 paternal half-sibling pairs. LTMs were fitted using monozygotic or dizygotic twins, full siblings, and paternal and maternal half-siblings to decompose the variance into four factors: 1) additive genetic effect (inherited); 2) non-additive genetic factors; 3) shared environmental factors; and 4) non-shared environmental factors. This data was used to determine concordant and discordant sibling pairs, which allowed them to calculate ASD heritability. The best model was the one that used additive genetic and non-shared environmental parameters. The ASD heritability estimated was $\approx 83\%$ (SANDIN et al., 2017).

Another recent multinational cohort study with more than two million people also used additive genetic factors and nonshared environmental to estimate the ASD heritability. They estimated that the ASD heritability is $\approx 80\%$, with possible modest differences in the sources of ASD risk replicated across countries (BAI et al., 2019).

No specific estimates indicate the likelihood of a couple of parents being autistic (either one of them or both), given that they have an ASD child. Some of the best ASD heritability estimates are the genetic factors calculated by Sandin et al. (2017) and Bai et al. (2019). These estimates suggest that more than 80% of ASD people have inherited the condition directly from their parents. Thus, we have assumed that given an ASD child, there is a likelihood of 83% that its parents are also autistic ($P(AP|ASD) = 0.83$). Once the ASD heritability estimates of Sandin et al. (2017) do not take the children's gender into account, we used the conditional probability ($P(AP|ASD)$) as the emission data for ASD children of both genders.

4.4.2 Autistic Parents Given They Have Typical Children

Similarly, no known estimates indicate the likelihood of a couple of parents being autistic (either one of them or both), given that they have a TD child. Therefore, the ASD diagnosis data presented by Palmer et al. (2017) also were used for estimating the probability of ASD parents, given that they have a TD child. Such estimates were calculated as follows.

Firstly, we calculated the percentage of parents having both one ASD child and one TD child. Let's call this group of parents as $PwAT$. According to the ASD heritability data, having an ASD child suggests that $PwAT$ have a higher likelihood to be autistic, although they also have a TD child. The fact that $PwAT$ also have a TD child is the starting point to estimate the probability of TD children having ASD parents. Through a statistical analysis over the data of Sandin et al. (2017), the percentage of $PwAT$ is 2.26% ($P(PwAT) = 0.0226$).

Secondly, we determined the percentage of children with $PwAT$ according to their gender. The percentage of TD boys with $PwAT$ is 46.82% ($P(TTB|PwAT) = 0.4682$). The percentage of TD girls with $PwAT$ is 53.18% ($P(TTG|PwAT) = 0.5318$). These two conditional probabilities will allow us to estimate the likelihood of a $PwAT$ occurrence according to the children's gender.

Thirdly, we calculated the prevalence of TD children in the population studied by Palmer et al. (2017). Regardless of the children condition, such population sex-ratio is 51.1% of boys ($P(B) = 0.511$), and 48.9% of girls ($P(G) = 0.489$). Given that, the total percentage of TD boys ($P(TTB)$) and TD girls ($P(TTG)$) was obtained as follows.

$$\begin{aligned} P(TTB) &= (1 - P(AB)) \cdot P(B) \\ &= 0.9803 \cdot 0.511 \\ &= 0.501 \end{aligned} \tag{4.1}$$

$$\begin{aligned} P(TTG) &= (1 - P(AG)) \cdot P(G) \\ &= 0.995 \cdot 0.489 \\ &= 0.487 \end{aligned} \tag{4.2}$$

Finally, we calculated the likelihood of a $PwAT$ occurrence with regard to the children's gender. We calculated two probabilities: the probability of $PwAT$, given that they have a TD boy ($P(PwAT|TTB)$); and the probability of $PwAT$, given that they have a TD girl ($P(PwAT|TTG)$). These two conditional probabilities were calculated using Bayes' theorem.

$$\begin{aligned} P(PwAT|TTB) &= \frac{P(TTB|PwAT) \cdot P(PwAT)}{P(TTB)} \\ &= \frac{0.4682 \cdot 0.0226}{0.501} \\ &= 0.0211 \end{aligned} \tag{4.3}$$

$$\begin{aligned} P(PwAT|TTG) &= \frac{P(TTG|PwAT) \cdot P(PwAT)}{P(TTG)} \\ &= \frac{0.5318 \cdot 0.0226}{0.487} \\ &= 0.0247 \end{aligned} \tag{4.4}$$

As mentioned before, $PwAT$ are those with both one ASD child and one TD child. Thus, $P(PwAT|TTB)$ and $P(PwAT|TTG)$ represent, in fact, the probability of a TD child having an ASD sibling. However, according to Sandin et al. (2017), ASD people have inherited the disorder from their parents in $\approx 83\%$ of the cases. This heritability suggests that $PwAT$ have the likelihood of 83% to be autistic once they also have one ASD child. Taking the ASD heritability into account, we calculated the probability of at least one or both parents (mother OR father) being autistic concerning the children's gender.

Equation 4.5 estimates the probability of APs given that they have a TD boy.

$$\begin{aligned} P(AP|TB) &= P(PwAT|TTB) \cdot P(AP|ASD) \\ &= 0.0211 \cdot 0.83 \\ &= 0.0175 \end{aligned} \tag{4.5}$$

Equation 4.6 estimates the probability of APs given that they have a TD girl.

$$\begin{aligned} P(AP|TG) &= P(PwAT|TTG) \cdot P(AP|ASD) \\ &= 0.0247 \cdot 0.83 \\ &= 0.0205 \end{aligned} \tag{4.6}$$

We used similar reasoning for estimating the overall prevalence of ASD parents in the population studied by Palmer et al. (2017). In such population, 2.38% of the parents had at least one ASD child ($P(PwA) = 0.0238$). Taking the ASD heritability into account, we calculated the overall prevalence of ASD parents as in Equation 4.7.

$$\begin{aligned} P(AP) &= P(PwA) \cdot P(AP|ASD) \\ &= 0.0238 \cdot 0.83 \\ &= 0.0197 \end{aligned} \tag{4.7}$$

We used this estimate of ASD parents to validate our model results by estimating the potential prevalence of ASD in their offspring. We compared the ASD estimated prevalence with the real ASD prevalence of Palmer et al. (2017).

4.4.3 Final Emission Matrices

The emission matrices were defined from the genetic probabilities and the prevalence data referred to above (Subsections 4.4.1 and 4.4.2). Two emission matrices were defined, one for boys ($B(B)$) and one for girls ($B(G)$). These two matrices were used for modeling all of our HMMs according to the children's gender.

$$B(B) = \begin{array}{cc} \text{TP} & \text{AP} & \\ \left[\begin{array}{cc} 1 - P(AP|TB) & P(AP|TB) \\ 1 - P(AP|ASD) & P(AP|ASD) \end{array}\right] & \begin{array}{c} \text{TB} \\ \text{AB} \end{array} \end{array}$$

$$B(B) = \begin{array}{cc} \text{TP} \quad \text{AP} & \\ \begin{bmatrix} 0.9825 & 0.0175 \\ 0.1700 & 0.8300 \end{bmatrix} & \begin{array}{c} \text{TB} \\ \text{AB} \end{array} \end{array}$$

$$B(G) = \begin{array}{cc} \text{TP} \quad \text{AP} & \\ \begin{bmatrix} 1 - P(AP|TG) & P(AP|TG) \\ 1 - P(AP|ASD) & P(AP|ASD) \end{bmatrix} & \begin{array}{c} \text{TG} \\ \text{AG} \end{array} \end{array}$$

$$B(G) = \begin{array}{cc} \text{TP} \quad \text{AP} & \\ \begin{bmatrix} 0.9795 & 0.0205 \\ 0.1700 & 0.8300 \end{bmatrix} & \begin{array}{c} \text{TG} \\ \text{AG} \end{array} \end{array}$$

4.5 Hidden Markov Models Structures and Probabilities

This section intends to simplify the visualization of the proposed methodology. Figure 10 and Figure 11 show our resulting HMMs structures and probabilities for female and male births, respectively. There are HMMs groups divided by gender because of the difference in statistical data regarding the prevalence and genetic inheritance of autism between male and female children.

The initial state distribution vectors of each group of HMMs have the same values for the three chains belonging to the same group. They were fitted according to the ASD prevalence data for each gender, Section 4.2.

Inside each group, the three distinct chains vary basically by the difference between the probabilities distribution between the transition states. This probability distribution variation is related to the gender of the older sibling, as presented in Section 4.3.

The emission data were computed according to the statistical data about autism heritability. Statistics on genetically inherited autism did not take gender into account, case of TD/ASD parents potentially generating ASD children with equal probabilities for both genders, Subsection 4.4.1. Our calculated statistics on non-genetically inherited autism take the gender into account, case of TD/ASD parents potentially generating TD children with distinct probabilities for each gender, Subsection 4.4.2.

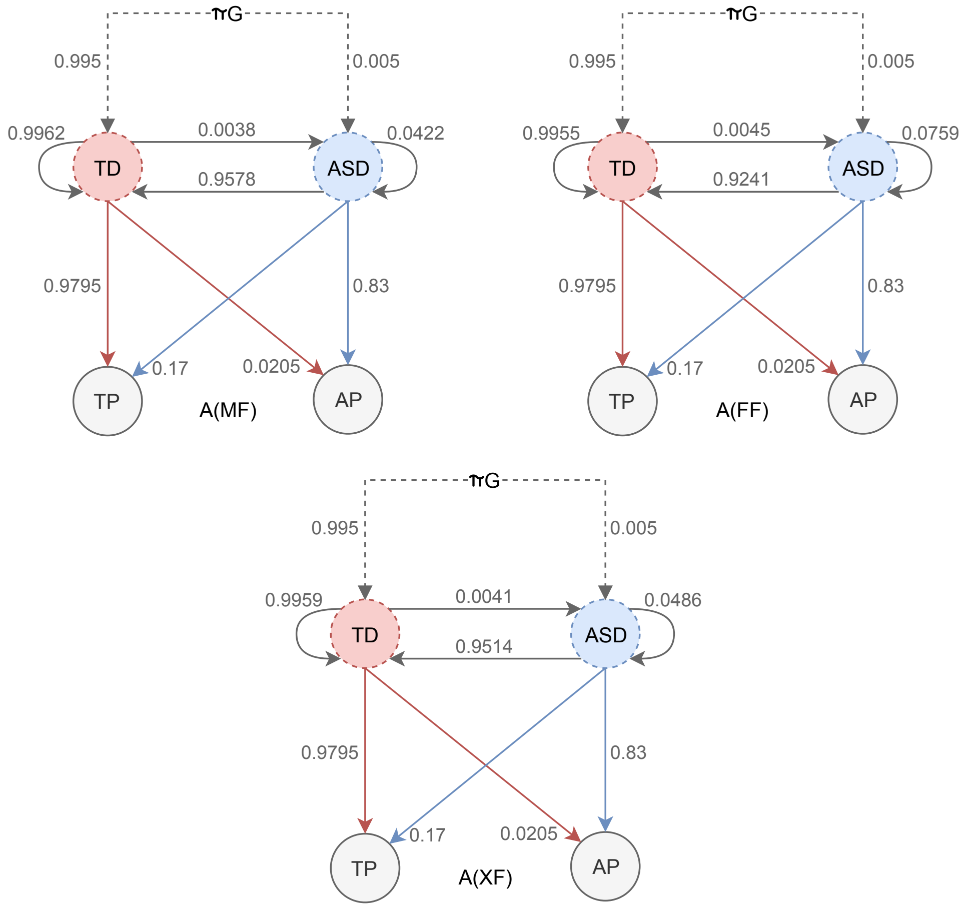


Figure 10 – HMMs for predicting the probability of having ASD girls. A(MF): transition data given that the older sibling is a boy; A(FF): transition data given that the older sibling is a girl; A(XF): transition data regardless the older sibling gender.

4.6 The risk of ASD

There is no consensus regarding the concept of risk, having several different ways of understanding the risk definition. There are definitions based on probability, undesirable events or danger, and others based on uncertainties. In addition, there are subjective and epistemic definitions of risk, dependent on the available knowledge, besides definitions that grant risk an ontological status independent of the assessors (AVEN, 2012).

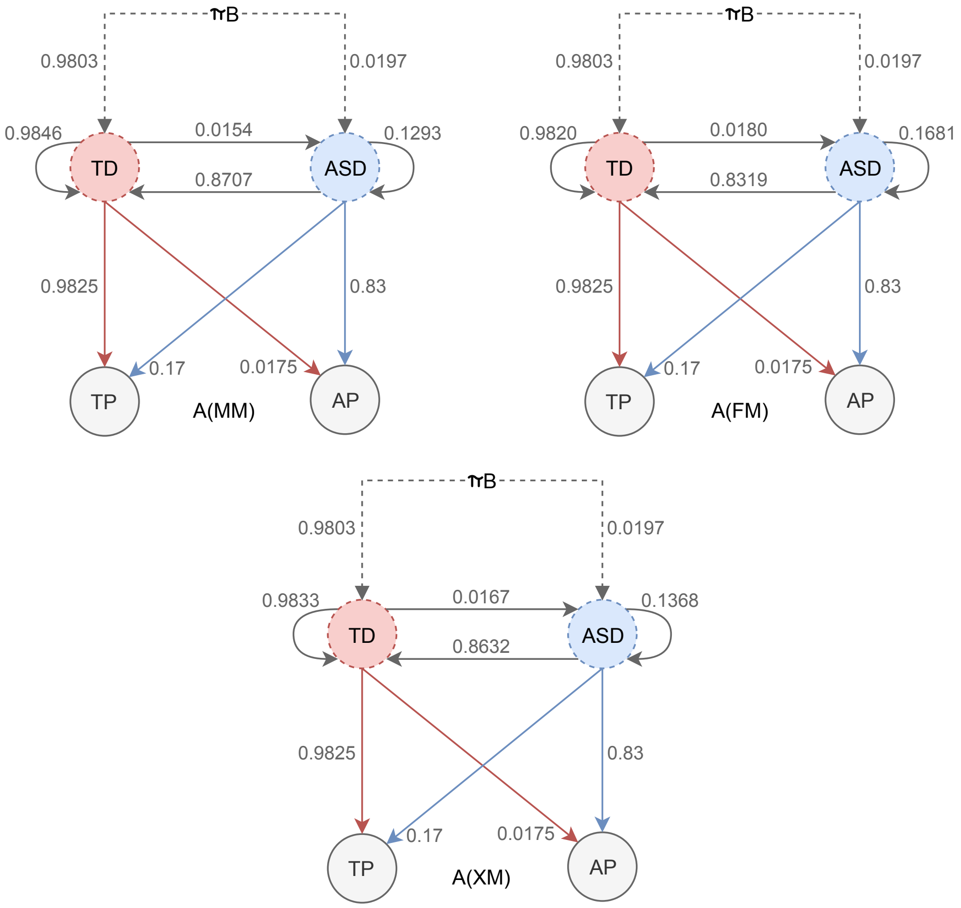


Figure 11 – HMMs for predicting the probability of having ASD boys. A(MM): transition data, given that the older sibling is a boy; A(FM): transition data, given that the older sibling is a girl; A(XM): transition data regardless of the older sibling gender.

The daily use suggests that the term risk can be considered positive or negative, and it could be both a noun (taking risks) and a verb (to risk losses). In this context, there are three main perspectives to the risk definition: 1) a situation involving a possibility of loss, damage, or other unwelcome circumstance; 2) A hazardous journey, undertaking, or course of action; and 3) something with the potential to produce a good or bad outcome in a particular domain (AVEN, 2012).

Even in the scientific context, many risk definitions can be found and classified in different risks categories, such as risk as a quantitative measure, qualitative concepts, or undesirable consequences (AVEN, 2012). However, in the context of our work, the risk refers to the quantitative product of the probability of some future event, without implying any qualitative judgment. Thus, the risk of ASD should be understood exclusively as the probability of a positive ASD diagnosis.

4.7 Implementation

The *hmmlearn*[1] library (version 0.2.1) was used for developing our HMMs. We used *MultinomialHMM* models with multinomial (discrete) emissions. *Hmmlearn* is an open-source set of algorithms and models usually used for modeling HMMs in the Python language. Built on *scikit-learn*[2], *NumPy*[3], and *SciPy*[4], *hmmlearn* adapts and uses these tools to sequence data.

We used the *predict_proba* function for estimating the probabilities of the HMMs' hidden states. This function computes the posterior probability for each state in the model. *Viterbi* was employed as the *predict_proba* decoder algorithm. In an attempt to decrease accuracy errors, our probabilities were rounded to the 15th decimal place. Our HMMs implementation code can be seen at our code repository[5].

4.8 Simulations

We simulated the generation of children from two different parents' profiles. Such profiles were the parents' states defined in Section 4.1 and used for modeling our observable states.

For each HMM set (boys and girls) and parents states/profiles (TP and AP), we simulated the birth of two children, maintaining the pattern of two children per couple presented by Palmer et al. (2017).

We observed the probabilities of the generated children being in one of the states defined in Section 4.1, those states which were used for modeling our HMMs' hidden chains.

4.9 Results

We organized our results according to the parents' profiles and the children's gender. The key findings of this study are shown on the graph displayed in Figure 12, which

[1] <https://hmmlearn.readthedocs.io>
[2] <http://scikit-learn.org>
[3] <http://www.numpy.org>
[4] <https://www.scipy.org/>
[5] <https://github.com/emerson-prof-carvalho/hmm>

summarizes the probabilities of TD/ASD parents generating TD/ASD children. For comparison purposes, the overall ASD prevalence calculated from the data of Palmer et al. (2017) was also plotted.

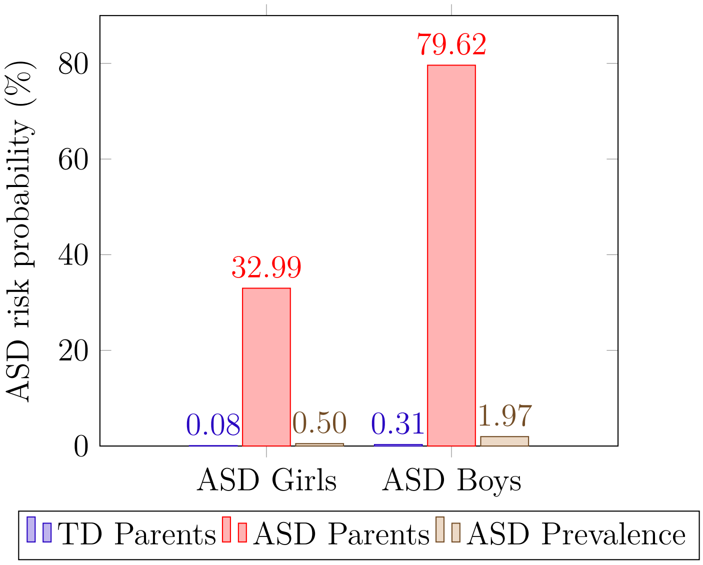


Figure 12 – ASD risk probability according to the parents' profile (TD, ASD); in addition to the overall ASD Prevalence from Palmer et al. (2017).

The following subsections detail our results.

4.9.1 Typical Parents

We displayed our results in tables. Each table row shows results concerning the corresponding transition matrix. Results from $A(FF)$ matrix is 7.3% greater than those from $A(MF)$ matrix (Table 8). This difference was already expected due to the greater likelihood of ASD when there is an older ASD sister. The $A(XF)$ matrix shows results close to the mean of the three matrices results. The mean probability of an ASD girl is ($P(AG|TP) = 0.078\%$). This probability is ≈ 6.5 times lower than the overall probability of ASD girls ($P(AG) = 0.5\%$), Section 4.2.

		States	
		TG(%)	**AG(%)**
Transition Matrix	***A(MF)***	99.9250	0.0750
	A(FF)	99.9195	0.0805
	A(XF)	99.9221	0.0779

Table 8 – Probabilities of TD parents generating TD/ASD girls.

For boys, the ASD probability increases ≈ 4 times with regard to ASD girls (Table 9). Results from $A(FM)$ matrix is 6.6% greater than those from $A(MM)$ matrix. Taking the transition matrix $A(XM)$ into account, the mean probability of an ASD boy is $(P(AB|TP) = 0.306\%)$. This probability is also ≈ 6.5 times lower compared to the overall probability of ASD boys $(P(AB) = 1.97\%)$, Section 4.2.

		States	
		TB(%)	**AB(%)**
Transition Matrix	***A(MM)***	99.7051	0.2949
	A(FM)	99.6855	0.3145
	A(XM)	99.6938	0.3062

Table 9 – Probabilities of TD parents generating TD/ASD boys.

Our experiments suggest that it is unlikely that TD parents could generate an ASD child when genetic inheritance is taken into account. Although, according to the genetic factors presented by Palmer et al. (2017) and Bai et al. (2019), from ≈ 17% to 19% of ASD children are generated by TD parents, with no evident hereditary genetic causes. This percentage of TD parents generating ASD children is due mainly to genetic mutations (not inherited) or gestational environment issues.

4.9.2 Autistic Parents

Results from $A(FF)$ matrix is 32% greater than those from $A(MF)$ matrix (Table 10). Taking the transition matrix $A(XF)$ into account, the probability of an ASD girl is $(P(AG|AP) \approx 33\%)$. This probability is ≈ 65 times higher compared to the overall probability of ASD girls $(P(AG) = 0.5\%)$, Section 4.2.

		States	
		TG(%)	**AG(%)**
Transition Matrix	***A(MF)***	69.2311	30.7689
	A(FF)	59.3439	40.6561
	A(XF)	67.0081	32.9919

Table 10 – Probabilities of ASD parents generating TD/ASD girls.

For boys, the ASD probability increases ≈ 2.5 times with regard to ASD girls (Table 11). Results from $A(FM)$ matrix is 4.3% greater than those from $A(MM)$ matrix.

Taking the transition matrix $A(XM)$ into account, the mean probability of an ASD boy is $(P(AB|AP) = 79.6\%)$. This probability is ≈ 40 times higher compared to the overall probability of ASD boys $(P(AB) = 1.97\%)$, Section 4.2.

		States	
		TB(%)	**AB(%)**
Transition Matrix	***A(MM)***	21.0140	78.9860
	A(FM)	17.6446	82.3554
	A(XM)	20.3818	79.6182

Table 11 – Probabilities of ASD parents generating TD/ASD boys.

4.10 Discussions

The overall direction of our results showed compelling evidence that could help learn about the ASD genetic risk among TD/ASD parents. Our data suggest that the ASD risk significantly increases from 40 to 65 times in parents with ASD diagnosis/risk genes.

Although many authors have investigated AI approaches to predict ASD, to the best of our knowledge, this is the first attempt to use genetic statistics related to the parents' condition to infer the risk of autism in their children. Most works used some data from the subject itself to predict ASD diagnosis. Even studies aimed at predicting autism in newborns used samples of materials from the individuals themselves (BAHADO-SINGH et al., 2019; SKAFIDAS et al., 2014). Therefore, we validate and compare our model results against some well-known statistical data about the heritability and prevalence of autism.

We simulated the population studied by Palmer et al. (2017) to validate our models' results (Section 4.2). Such population contains 3 121 074 children (1 596 078 males and 1 524 996 females), with two children per parents. The parents' states (observable ones) were randomly defined as TPs (98.03%) and APs (1.97%). These percentages of parents' states follow the overall APs prevalence estimated at the end of Subsection 4.4.2. In addition to the gender distinction, we also considered the number of children as first or second descendants.

Using the probabilities for generating ASD children obtained from our HMMs (Section 4.9), our strategy was to estimate the ASD prevalence in that known population. Our estimated ASD prevalence was 0.7% among females. Concerning males, the estimated ASD prevalence was 1.9%. The overall estimated ASD prevalence was 1.3%. As expected, it appears our estimated ASD prevalence is close to the ASD prevalence of the real population. The highest ASD prevalence variation was among females, which suggests that the calculated probabilities of ASD parents generating ASD girls would be a maximum probability.

On the other hand, our estimated prevalence is lower than or equal to the prevalence found by Baio et al. (2018). Their estimates were 0.7% among girls, 2.6% among boys, and 1.7% for the overall ASD prevalence. These statistics corroborate that our probabilities of ASD parents generating ASD girls would be at their maximum and that the probabilities of ASD parents generating ASD boys could be even higher. Thus, it seems that our ASD risk probabilities for TD/ASD parents lead to ASD prevalence estimates close to the real ASD prevalence nowadays.

For ASD parents, our study indicates that the ASD risk for boys ($\approx$ 80%) is approximately 2.5 times higher than it is for girls ($\approx$ 33%). The current literature indicates that the ASD prevalence is three to four times higher in boys, suggesting that the ASD risk difference should be more significant between boys and girls. However, the difference between boys and girls concerning ASD prevalence is reduced when the genetic factors is considered. Messinger et al. (2015) obtained a 3.2:1 male:female ASD ratio among a large sample of high-risk siblings for ASD. In an analysis of the population studied by Palmer et al. (2017), we observed that the difference in ASD prevalence between genders decreases when there is more evidence about the presence of the disorder in the family. Having two children with ASD increases the likelihood of ASD inheritance in a family. Taking only families in the population of Palmer et al. (2017) with both siblings with ASD diagnosis, the ASD sex ratio is approximately 2.8 ASD boys (74%) for each ASD girl (26%), approaching our results concerning the risk of ASD parents having ASD children.

There are pieces of evidence that ASD parents are likely to have more ASD boys than ASD girls (SANDIN et al., 2017; BAI et al., 2019; BAIO et al., 2018; PALMER et al., 2017). Quite distinct probabilities for ASD parents generating ASD children could be obtained if there were more accurate data about the likelihood of a(n) TD/ASD boy/girl having TD/ASD parents. This indicates that we may have in the future a more appropriate emission matrix (B) data, whether by having more assertive data about the genetic factors (currently tending to be close to 80% for both genders) or by having more assertive data to distinguish between TD boys/TD girls concerning the probability of having ASD parents (assumed in this work as 1.75% for boys and 2.05% for girls). A larger sample to create the transition data also would be able to make more accurate predictions.

Some improvements that could lead to different results would be to consider more than one level in the family ancestry chart (HANSEN et al., 2019), and take the parents' age into account (SANDIN et al., 2016). Since only the previous state influences the current state in the Markovian models, there is no reference, for example, about the genetic influence of the grandparents. Such analysis could require other types of statistics about genetic factors in autism and the use of different AI techniques, such as Bayesian Networks.

4.11 Summary

Using our HMMs models, we estimated that ASD parents could generate ASD children with probabilities of $\approx 33\%$ for girls and $\approx 80\%$ for boys. As no previous work has evaluated the ASD risk from parents' characteristics, by quantifying the risk of ASD parents having ASD children, we gave a first look at how much the ASD risk increases for ASD parents (≈ 40 to ≈ 65 times), as well as how much the ASD risk decreases for TD parents (≈ 6.5 times). We also highlighted the decrease between the rate of ASD girls and ASD boys when genetic factors are taken into account (≈ 2.5 boys for each girl). This decrease suggests that genetically inherited autism may affect girls more than other causes of autism. Another key point was the estimation of the (emission) probabilities of ASD parents, even though they have TD children ($P(AP|TB)$ and $P(AP|TG)$). Most ASD cases tend to cluster in families. Thus, our findings support and quantify past evidence that this clustering is due to genetic factors.

Although it is too early to draw statistically significant conclusions, the possibility of contributing to estimating the ASD risk according to the parents' condition is a fascinating proposition. We believe we provide an initial model that can be applied and improved as long as new and potentially more accurate ASD statistical data emerge. For people who intend to have children and have autistic characteristics/diagnoses, our estimates could help clarify the ASD risk and alert them in planning the process of early investigation on their children.

By having more accurate statistical data about the genetic factors in autism, future works could accurately estimate the potential risk of ASD parents generating ASD children. However, these causal probabilities regarding the likelihood of TD/ ASD parents generating TD/ASD children can be used as a basis for building causal models (e.g., BNs) capable of inferring over a family tree, as can be seen in the following three chapters.

5 Common Characteristics Among all Bayesian Networks

Chapter 4 presented some estimates of autistic parents having autistic children. Based on such estimates, we developed a set of BNs to investigate the likelihood of ASD in several family members, given some evidence. Our models' basic structures were built and validated using statistical data from the association of gender with recurrence of autism among siblings and statistical data of the association of genetic factors with autism.

Our approach was to predict future conditions among relatives based on the possible observation of some family individuals' characteristics (e.g., an ASD diagnosis). Given the causal nature of ASD heredity, BNs seemed to be the most straightforward, appropriate, and transparent strategy to start this investigation. This adequacy is due mainly to the BNs causal approach, which allows building an inference system over causal models.

We did not use direct individual observations to train our BNs parameters because, to the best of our acknowledgment, there is no such kind of public data available. Thus, we used the causal probabilities estimates from Carvalho et al. (2020) for the adjustment ("training") of the models' parameters.

The remainder of this chapter explains the design process, the primary structure, and the assumptions about the probabilities used to model our BNs.

5.1 Bayesian Networks Development Process

The manual construction of a BN requires a well-defined problem to solve, a careful identification and selection of the relevant variables, a precise description of dependences/independences relationships among the selected variables, and a proper elicitation of the required prior and conditional probabilities (KJAERULFF; MADSEN, 2013).

We defined each problem to be solved essentially by describing the posterior probabilities that we would like to infer. Thus, such probabilities guided the selection of the smallest possible set of aleatory variables of interest, and the other necessary ones, given the problem specification.

Constructing a suitable graph structure is a prerequisite to correctly representing the dependencies among the model variables and accurately eliciting the required probabilities. However, to specify an adequate graph structure as a practical model of the types of reasoning expected is one of the main barriers to building BNs efficiently (NEIL; FENTON; NIELSON, 2000).

As we adopted the manual construction approach to develop our networks, for all proposed BNs we sought to define their structures following the natural configuration of a family tree, managing the dependence between variables as a cause and effect relationship. This causal modeling performs the substructures of our networks mainly as the cause-consequence idiom proposed by Neil, Fenton and Nielson (2000). Given the BN structure, we defined the type of its random variables as suggested by Kjaerulff and Madsen (2013).

The other classic barrier for creating valuable BNs occurs when eliciting the conditional probability values from a specific domain (NEIL; FENTON; NIELSON, 2000). Following a family tree structure allow both to represent each variable as a unique set of events with no competing ones and clearly define their semantics, leaving no ambiguity. The clarity test proposed by Kjaerulff and Madsen (2013) requires these two principles to probe whether a variable was clearly defined.

This natural modeling for the BNs structures as a family tree and the causal nature of the ASD heritability aimed to eliminate the two classic barriers when building effective BNs identified by Neil, Fenton and Nielson (2000). Therefore, we ensured that the directions of the edges do not conflate cause to effect directions with the directions intended by the inferences we might want to perform.

5.2 The Bayesian Networks Variables

For each BNs designed, we defined the smallest possible number of variables given the problem to be solved. However, we described the entire set of random variables in this section since some appear in more than one BN. We will present the variables by levels once we investigate the risk of ASD over three generations. In general, we adopted two children per couple, except for the variables `F` and `M` in the BN in Section 6.2, where we adopted six children (three of each gender) to estimate the effect of different combinations among siblings. Thus, the children of the couple composed of `F` and `M` represents a central reference point from which we will make several inferences.

The first level variables include the grandparents, paternal and maternal:

- Paternal grandparents
  - `GFF`: representing the Grandfather from Father side;
  - `GMF`: representing the Grandmother from Father side;
- Maternal grandparents
  - `GFM`: representing the Grandfather from Mother side;
  - `GMM`: representing the Grandmother from Mother side;

The Second level variables include parents, maternal and paternal uncles and aunts, and their partners:

- Parents
  - `F`: representing the Father;
  - `M`: representing the Mother;
- Half-siblings parents
  - `FP`: representing another Father Partner (the mother of half-siblings from Father side);
  - `MP`: representing another Mother Partner (the father of half-siblings from Mother side);
- Paternal uncles and aunts.
  - `UF`: representing an Uncle from Father Side;
  - `UFP`: representing the Partner of the Uncle from Father Side;
  - `AF`: representing an Aunt from Father Side;
  - `AFP`: representing the Partner of the Aunt from Father Side;
- Maternal uncles and aunts.

  - `AMP`: representing the Partner of the Aunt from Mother Side;
  - `AM`: representing an Aunt from Mother Side;
  - `UMP`: representing the Partner of the Uncle from Mother Side;
  - `UM`: representing an Uncle from Mother Side;

The third level variables include full-siblings, half-siblings, and cousins:

- Full-siblings
  - `FB`: representing a First Boy (a first male child);
  - `SB`: representing a Second Boy (a second male child);
  - `TB`: representing a Third Boy (a third male child);
  - `FG`: representing a First Girl (a first female child);
  - `SG`: representing a Second Girl (a second female child);
  - `TG`: representing a Third Girl (a third female child);
- Half-siblings
  - `BFP`: representing a Boy (half-brother) from Father Partner;
  - `GFP`: representing a Girl (half-sister) from Father Partner;
  - `BMP`: representing a Boy (half-brother) from Mother Partner;
  - `GMP`: representing a Girl (half-sister) from Mother Partner;
- Paternal cousins
  - `BUF`: representing a Boy (male cousin) son of an Uncle from Father Side;
  - `GUF`: representing a Girl (female cousin) daughter of an Uncle from Father Side;
  - `BAF`: representing a Boy (male cousin) son of an Aunt from Father Side;
  - `GAF`: representing a Girl (female cousin) daughter of an Aunt from Father Side;
- Maternal cousins
  - `BAM`: representing a Boy (male cousin) son of an Aunt from Mother Side;
  - `GAM`: representing a Girl (female cousin) daughter of an Aunt from Mother Side;

  - BUM: representing a Boy (male cousin) son of an Uncle from Mother Side;
  - GUM: representing a Girl (female cousin) daughter of an Uncle from Mother Side;

5.3 Domain of the Random Variables (States)

Our main objective is to estimate the family bias to autism, which means to infer the probability of an individual being typical or presenting some autistic traits. Thus, each node included in the BNs represents a family member and shares the same domain. As will be seen in the following chapters, the states representing the family members can assume one of the following domain values:

- **asd**: reveal/indicates a person with ASD traits;
- **ang**: reveal/indicates a person with ASD traits, with non-hereditary factors (environmental factors or *de novo* mutations) as the most likely cause;
- **td**: reveal/indicates a person with a typical development.

We included the ang domain value for two purposes: 1) to allow using virtual evidence in the inference process, especially concerning explanation queries. For example, when evaluating the impact of particular evidence on non-descendant nodes (e.g., parents and grandparents), it is necessary to consider that ASD may have a non-hereditary factor as a cause; and 2) to consider in causal inferences the ASD cases caused by non-hereditary factors. It would not be correct to assume that all ASD people have genes associated with the syndrome since some of these cases are caused by environmental factors that do not necessarily change the individual's genotype.

This domain sharing enabled all our defined variables concerning the last clarity test principle proposed by Kjaerulff and Madsen (2013), which requires that all possible values in the variable domain must be exhaustive and mutually exclusive.

5.4 Model Probabilities

Among other essential points, the manual construction of a BN requires a proper elicitation of many prior and conditional probabilities. We created the BNs according to the family structure we wanted to investigate. Although we modeled different BNs, we used a standard set of prior and conditional probabilities. We separated these probabilities by gender due to the already known male-female sex ratio difference regarding the ASD prevalence/risk.

5.4.1 Prior Probabilities

Supposing there is no evidence about an individual concerning an ASD diagnosis, we assumed that the best information to represent this prior probability is the ASD prevalence data. Therefore, we used as prior probabilities the ASD prevalence data gathered in Section 2.2, whose central tendencies we summarized in Table 2.

Tables 12 and 13 present the prior probabilities for root nodes (those nodes without a parent node). We calculated the `ang` values using an additive **genetic factor** of ≈ 81%. This percentage is the mean value attributed to additive genetics calculated from the results of Bai et al. (2019), Yip et al. (2018), Sandin et al. (2017) and Tick et al. (2016). Consequently, the value attributed to **non-genetic factors** is ≈ 19%.

Table 12 – Prior probabilities used for any root node that represents a male individual.

Male	`asd`	0.0222
	`ang`	0.0043
	`td`	0.9778

Table 13 – Prior probabilities used for any root node that represents a female individual.

Female	`asd`	0.0065
	`ang`	0.0012
	`td`	0.9935

Figure 13 displays that the ASD prevalence median values are located at the center of the quartiles limits. The histogram shown in Figure 14 displays that the peak frequency of ASD prevalences is also pretty close to the means. The data in these graphs indicate that using the ASD prevalence mean values as prior probabilities seem adequate for this investigation.

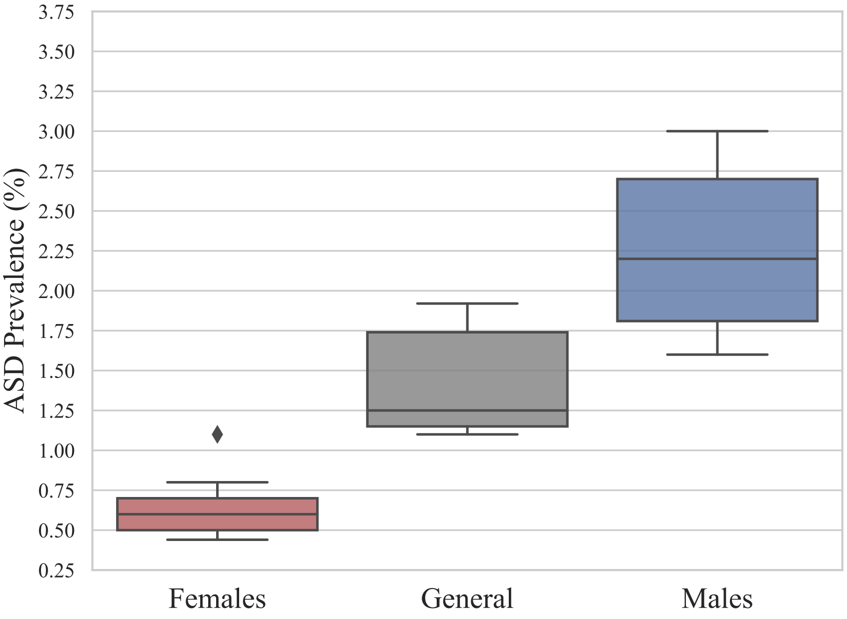


Figure 13 – Dispersion and skewness of the ASD prevalence data from Table 1.

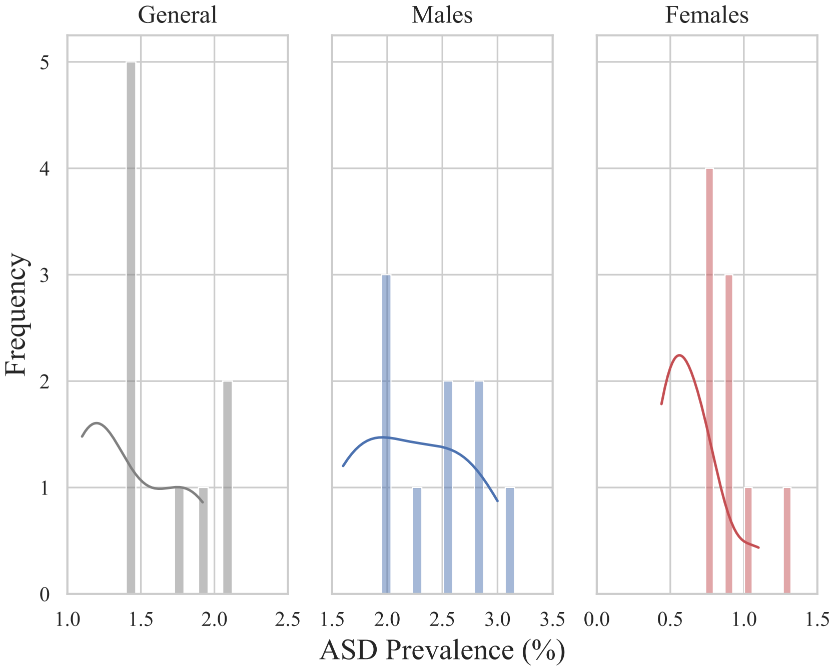


Figure 14 – Distribution of of the ASD prevalence data from Table 1.

5.4.2 Conditional Probabilities

We used as conditional probabilities the causal ASD data gathered in Section 4.9. All conditional probabilities that involve ASD parents are equal, regardless of the parents' gender. We expected such equality once we estimated the probability of autistic parents (father OR mother) generating autistic children. We expected the probability `P(Male=asd|Father=asd,Mother=td)` to be equal to the probability `P(Male=asd|Father =td,Mother=asd)` as well as the probability `P(Female=asd|Father=asd,Mother=td)` to be equal to the probability `P(Female=asd|Father=td,Mother=asd)` because there is no evidence that gender determines how much risk an ASD person represents to the offspring. Even the so-called maternal effect does not represent a greater risk of ASD associated with an ASD mother than an ASD father (BAI et al., 2019; YIP et al., 2018).

However, it is explicit the increased genetic risk for those children who have both parents diagnosed with autism, despite being a scarce case. To have a more realistic conditional probability, we estimated the genetic risk for children with ASD parents (father AND mother) using the probability rule of addition. In our case, it means the probability that an individual will inherit autism from his/her father, or inherit autism from his/her mother, or even inherit autism due to factors coming from both parents. Equation 5.1 presents the rule of addition.

$$P(A \cup B) = P(A) + P(B) - P(A, B) \tag{5.1}$$

Consequently, we estimated `P(Male=asd|Father=asd,Mother=asd)` and `P(Female =asd|Father=asd,Mother=asd)` according to Equations 5.2 and 5.3, respectively.

$$\begin{aligned} P(Male{=}asd|Father{=}asd, Mother{=}asd) = {} & P(Male{=}asd|Father{=}asd, Mother{=}td) + \\ & P(Male{=}asd|Father{=}td, Mother{=}asd) \\ & - \\ & P(Male{=}asd|Father{=}asd, Mother{=}td) \cdot \\ & P(Male{=}asd|Father{=}td, Mother{=}asd) \\ = {} & 0.958458 \end{aligned} \tag{5.2}$$

$$\begin{aligned} P(Female{=}asd|Father{=}asd, Mother{=}asd) &= P(Female{=}asd|Father{=}asd, Mother{=}td)+ \\ &\quad P(Female{=}asd|Father{=}td, Mother{=}asd) \\ &\quad - \\ &\quad P(Female{=}asd|Father{=}asd, Mother{=}td)\cdot \\ &\quad P(Female{=}asd|Father{=}td, Mother{=}asd) \\ &= 0.550991 \end{aligned} \tag{5.3}$$

Results from Equations 5.2 and 5.3 imply that given both parents with ASD, they would generate ASD boys and ASD girls with probabilities of ≈ 96% and ≈ 55%, respectively. Thus, we defined the conditional probabilities for the intermediate and leaf nodes according to Tables 14 (males) and 15 (females). ASD cases due to environmental factors or *de novo* mutations do not depend on the parents' genetics. Thus, the conditional probability for `ang` has the same values as the prior probabilities.

Table 14 – CPT for a male individual given his parents.

Father		asd	asd	asd	ang	ang	ang	td	td	td
Mother		asd	ang	td	asd	ang	td	asd	ang	td
Male	asd	0.9585	0.7962	0.7962	0.7962	0.0031	0.0031	0.7962	0.0031	0.0031
	ang	0.0043	0.0043	0.0043	0.0043	0.0043	0.0043	0.0043	0.0043	0.0043
	td	0.0415	0.2038	0.2038	0.2038	0.9969	0.9969	0.2038	0.9969	0.9969

For clarification purposes, the probability value 0.0415 in Table 14 is `P(Male=td|Father=asd,Mother=asd)`. We followed this pattern for all CPTs presented in the rest of this thesis, where the horizontal headlines are the evidence, and the vertical headlines are the variables of interest.

Table 15 – CPT for a female individual given her parents.

Father		asd	asd	asd	ang	ang	ang	td	td	td
Mother		asd	ang	td	asd	ang	td	asd	ang	td
Female	asd	0.5510	0.3299	0.3299	0.3299	0.0008	0.0008	0.3299	0.0008	0.0008
	ang	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012
	td	0.4490	0.6701	0.6701	0.6701	0.9992	0.9992	0.6701	0.9992	0.9992

Based on Tables 14 and 15, the male:female ASD risk rate decreases to 1.75:1 if we assume both parents with ASD. This value emphasizes the decreasing tendency of this rate as more extensive and intense genetic factors are associated with parents, as discussed in Sections 2.3.2.3 and 4.10, suggesting that inheriting more genetic factors related to autism would imply a higher ASD prevalence for females.

5.5 Implementation Tools

We used *Pgmpy*[1], an open-source *Python* library that provides an easy-to-use Application Programming Interface (API) for working with PGMs. It allows creating graphical models and answer inference or map queries over them once it has implemented many inference algorithms (ANKAN; PANDA, 2015), including the Variable Elimination algorithm used in our queries. *Pgmpy* is one of the most suitable BNs software packages, with a modular architecture designed for extensibility and an active software community that usually provides new updates (MICHIELS; LARRAÑAGA; BIELZA, 2021).

Other *Python* libraries used include: *Pandas*[2], an open-source data analysis and manipulation tool designed to work with structured and time-series data. We used *Pandas* as the data source for producing our graphics; *NumPy*[3], an open-source *Python* library fundamental for scientific computing, with an extensive set of mathematical functions to operate on large multi-dimensional arrays; and *Seaborn*[4], an open-source *Python* library for data visualization. Based on *Matplotlib*[5], *Seaborn* provides a high-level API for drawing statistical graphics.

Our environment used *Python* version 3.9.4 and *Pgmpy* version 0.1.15. Our BNs implementation code can be seen at our code repository[6].

5.6 Summary

This chapter introduced some fundamental aspects regarding our BNs developing process. Such aspects involve the process of structuring the BNs' topology, the definition of their variables, including its domain, the set of prior and conditional probabilities, and the tools we used.

We seek to evaluate the adequacy of the random variables concerning the clarity test and define the type of each one of them. The BNs topologies follow the cause-consequence idiom, attempting to understand the role of each variable given the BN structure.

[1] <https://pgmpy.org>
[2] <https://pandas.pydata.org/>
[3] <https://numpy.org/>
[4] <https://seaborn.pydata.org/>
[5] <https://matplotlib.org/>
[6] <https://github.com/emerson-prof-carvalho/bns>

To the best of our knowledge, the set of prior and conditional probabilities used are the state-of-the-art literature concerning the ASD prevalence, recurrence, and heritability data. Such studies calculated these estimates based on the clinical diagnosis of ASD, which seeks to diagnose phenotypic manifestations of the disorder. Thus, the ASD risk estimates that our BNs will evaluate aim to assess the risk of a clinical diagnosis of ASD traits, which is different from estimating the presence or absence of ASD genetic variants.

The subsequent chapters use the guidelines specified in this chapter to create BNs that will investigate specific problem definitions.

6 Bayesian Network to Estimate the Risk of ASD in Siblings, Parents, and Grandparents

The ASD heritability and recurrence among siblings are well-explored subjects, performed mainly by studies that evaluated sibling pairs. Therefore, we aim to assess the risk of ASD in parents, grandparents, and siblings of ASD individuals so that we can explore different and distinct pieces of evidence.

6.1 Problem Definition

We aimed to estimate conditional probabilities such as:

- `P(Father|Boy)`, `P(Father|Girl)`, `P(Father|Boy, Girl, ...)`;
- `P(Mother|Boy)`, `P(Mother|Girl)`, `P(Mother|Boy, Girl, ...)`;
- `P(Boy|Boy)`, `P(Boy|Girl)`, `P(Boy|Boy, Girl)`, `P(Boy|Boy, Boy, ...)`;
- `P(Girl|Boy)`, `P(Girl|Girl)`, `P(Girl|Boy, Girl)`, `P(Girl|Girl, Girl, ...)`;
- `P(Grandfather|Boy)`, `P(Grandfather|Girl)`, `P(Grandfather|Boy, Girl, ...)`;
- `P(Grandmother|Boy)`, `P(Grandmother|Girl)`, `P(Grandmother|Boy, Girl, ...)`;

Notations like `P(Father|Boy)` summarizes the following conditional probabilities `P(Father=asd|Boy=asd)`, `P(Father=asd|Boy=td)`, `P(Father=td|Boy=asd)`, and `P(Father=td|Boy=td)`.

6.2 Bayesian Network Structure

Using a portion of the variables previously defined in Section 5.2, we created a BN to estimate the risk of ASD in siblings, parents and grandparents. Figure 15 presents the BN structure.

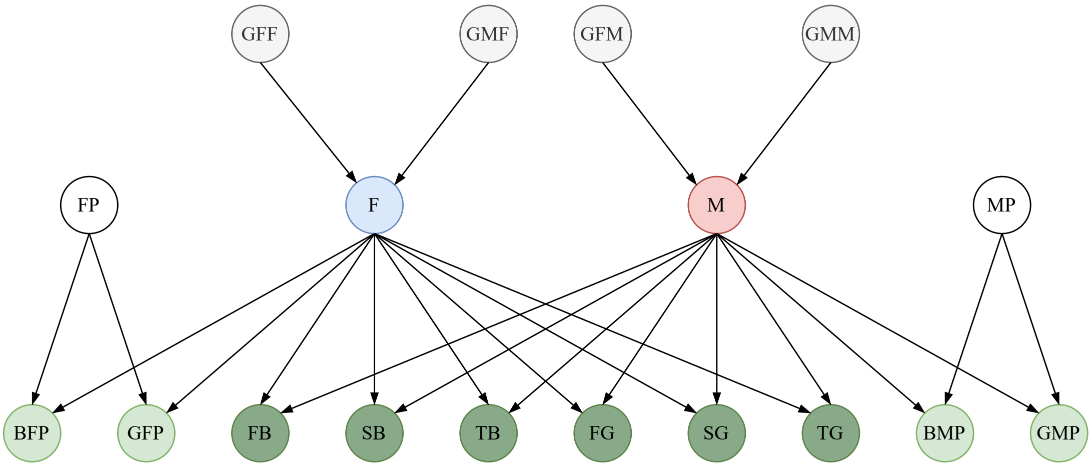


Figure 15 – The structure of a BN to estimate the risk of ASD in siblings, parents, and grandparents.

According to the types of variables defined in Section 3.4.4.1, the nodes representing grandparents can be classified as background or hypothesis variables. They work as background variables because they have an indirect causal relationship regarding the symptom variables. They act as hypothesis variables because they are among the variables of interest given the problem definition (to estimate the risk of ASD in grandparents given evidence in their grandchildren).

The nodes representing parents can be classified as hypothesis variables. They work as hypothesis variables because they have a direct causal relationship with the symptom variables and are among the variables of interest given the problem definition (to estimate parents as the probable explanation given evidence in their children).

The nodes representing children can be classified as symptom or hypothesis variables. They act as symptom variables while we take them as evidence. They serve as hypothesis variables when we make inferences about them, given some evidence at the variables in the same level (to estimate the recurrence among siblings).

We used the prior probabilities defined in Tables 12 and 13 as the discrete probability distributions for root nodes (those representing grandparents, father partner and mother partner). Nodes representing parents (`F` and `M`) and children have their CPTs produced from the conditional probabilities presented in Tables 14 and 15. We emphasize that the probabilities correspond to the gender of the family member represented by each node.

6.3 Marginal Probabilities without Evidence

First, we investigated the probabilities estimated by the BN when no evidence is given. This analysis aimed to evaluate the propagation of uncertainty across the network levels. Table 16 shows the estimated probabilities.

Table 16 – Marginal probabilities of the BN nodes in Figure 15 if no evidence is given.

	Males									**Females**								
	GFF	GFM	MP	F	BFP	BMP	FB	SB	TB	GMF	GMM	FP	M	GFP	GMP	FG	SG	TG
asd	0.022			0.026	0.028		0.031			0.007			0.010	0.011		0.012		
td	0.974			0.970	0.968		0.965			0.992			0.989	0.988		0.987		

$^{\text{GFF}}$Grandfather from Father side; $^{\text{GFM}}$Grandfather from Mother side; $^{\text{MP}}$Mother Partner; $^{\text{F}}$Father; $^{\text{BFP}}$Boy (half-brother) from Father Partner; $^{\text{BMP}}$Boy (half-brother) from Mother Partner; $^{\text{FB}}$First Boy; $^{\text{SB}}$Second Boy; $^{\text{TB}}$Third Boy; $^{\text{GMF}}$Grandmother from Father side; $^{\text{GMM}}$Grandmother from Mother side; $^{\text{FP}}$Father Partner; $^{\text{M}}$Mother; $^{\text{GFP}}$Girl (half-sister) from Father Partner; $^{\text{GMP}}$Girl (half-sister) from Mother Partner; $^{\text{FG}}$First Girl; $^{\text{SG}}$Second Girl; $^{\text{TG}}$Third Girl.

Marginal probabilities range from $\approx$ 2.2% (root nodes) to 3.1% (leaf nodes) for nodes representing males. Regarding nodes representing females, marginal probabilities range from 0.7% to 1.2%. These results show an uncertainty propagation through the BN, especially in leaf nodes. However, these probabilities range are corroborated by the variation also noticed in the ASD prevalence data, which suggests a suitable fit to start our investigation.

We used these estimates to evaluate the BN results when applying the evidence sets we aim to investigate.

6.4 Estimating the Risk of ASD in Parents

We performed conditional probability queries (explanation reasoning queries) to compute the risk of ASD in parents given evidence regarding their descendants. In all inference cases, the non-explicit definition for a variable state means no evidence. We started the investigation by analyzing one child per couple. Following this, we increased the number of children to explore the impacts on the causal explanation. We performed inferences and presented results in such a way we could analyze genders separately.

6.4.1 Risk of ASD in Parents Given One Child

Tables 17 presents the inference results for parents given the evidence regarding a first child. Values in parentheses indicate results when virtual evidence was used.

Table 17 – Inference results for fathers and mothers given one child.

		FB		FG	
		asd	td	asd	td
F	asd	(0.6333) 0.6529	0.0053	(0.6636) 0.6786	0.0173
	td	(0.3651) 0.3456	0.9903	(0.3349) 0.3200	0.9784
M	asd	(0.2528) 0.2606	0.0021	(0.2668) 0.2728	0.0068
	td	(0.7463) 0.7385	0.9967	(0.7323) 0.7263	0.9919

$^{\text{FB}}$First Boy; $^{\text{FG}}$First Girl; $^{\text{F}}$Father; $^{\text{M}}$Mother.

Given one ASD child, the risk of ASD for fathers ranges from $\approx 63\%$ (ASD boy) to $\approx 68\%$ (ASD girl), a 24−26-fold increase in the probability risk of ASD. Meanwhile, the risk of ASD for mothers ranges from $\approx 25\%$ (ASD boy) to $\approx 27\%$ (ASD girl), a 25−27-fold increase in the probability risk of ASD. As expected, if the child with ASD is a girl, the causality associated with the parents' genetics is slightly higher. Figure 16 summarizes the average risk of ASD in parents given the evidence regarding a first child.

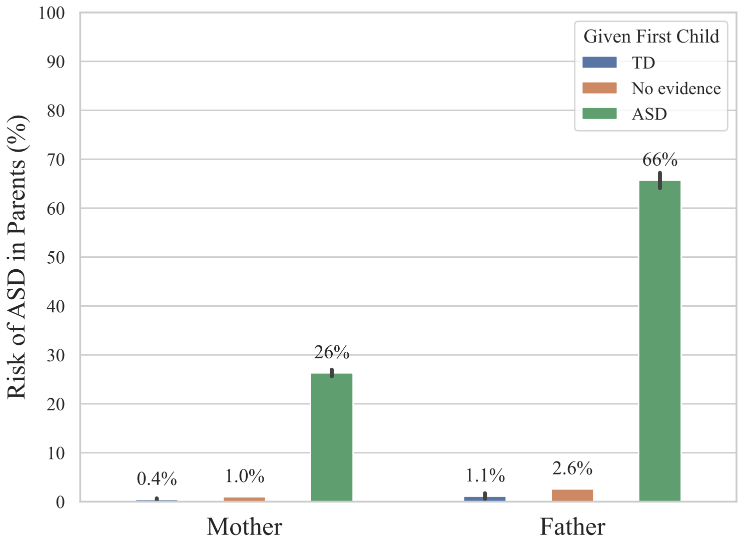


Figure 16 – Risk of ASD in parents given evidence regarding one child.

6.4.2 Risk of ASD in Parents Given Two Children

Table 18 presents the inference results for parents given the evidence regarding two children. Given two ASD children, the risk of ASD for fathers is ≈ 72%, a 28-fold increase in the probability risk of ASD. In comparison, the risk of ASD for mothers is ≈ 29−30%, a 29-fold increase in the probability risk of ASD.

Table 18 – Inference results for parents given two children.

		FB			FB				FG		
		asd	asd td	td	asd	asd	td	td	asd	asd td	td
		SB			SG				SG		
		asd	td asd	td	asd	td	asd	td	asd	td asd	td
F	asd	(0.7210) 0.7212	(0.4274) 0.4748	0.0011	(0.7223) 0.7225	(0.5966) 0.6229	(0.5016) 0.5467	0.0036	(0.7240) 0.7241	(0.6370) 0.6579	0.0117
	td	(0.2778) 0.2776	(0.5701) 0.5229	0.9946	(0.2765) 0.2763	(0.4016) 0.3754	(0.4963) 0.4513	0.9921	(0.2748) 0.2747	(0.3614) 0.3406	0.9840
M	asd	(0.2889) 0.2889	(0.1681) 0.1867	0.0004	(0.2918) 0.2919	(0.2367) 0.2471	(0.1975) 0.2153	0.0014	(0.2959) 0.2959	(0.2539) 0.2623	0.0046
	td	(0.7103) 0.7102	(0.8309) 0.8123	0.9983	(0.7073) 0.7072	(0.7624) 0.7519	(0.8015) 0.7837	0.9973	(0.7033) 0.7032	(0.7451) 0.7368	0.9942

^{FB}First Boy; ^{FG}First Girl; ^{SB}Second Boy; ^{SG}Second Girl; ^{F}Father; ^{M}Mother;

On the other hand, there is a 2−25-fold decrease in the probability risk of ASD for parents given the evidence of two TD children. The risk of ASD for parents decreases 25-fold if the two TD children are boys and 7-fold if at least one TD child is a boy. In other words, having two TD sons reduces the chances of ASD in parents substantially. However, the risk rate reduction given evidence regarding two TD daughters is considerably shorter (2-fold), which is corroborated by studies that point to females being more immune to ASD genetic load. Therefore, we expected that having two TD daughters should not reduce the risk of ASD in parents by the same dimension as if they had two TD sons.

For groupings including one ASD child and one TD child, the risk of ASD for fathers ranges from ≈ 43% to ≈ 66%, while the risk of ASD for mothers ranges from ≈ 17% to ≈ 26%. For parents who have one ASD child, having one TD boy as a second child is a factor that mitigates parental causality by ≈ 15−35% for mothers and ≈ 17−35% for fathers. Having one TD girl as a second child also mitigates the parental causality, although in a modest proportion of at most ≈ 8−9%. As in cases with a single ASD child, if the children with ASD are girls, the parental causality is slightly higher given the same remaining set of evidence.

Figure 17 summarizes the average risk of ASD in parents given the evidence regarding two children.

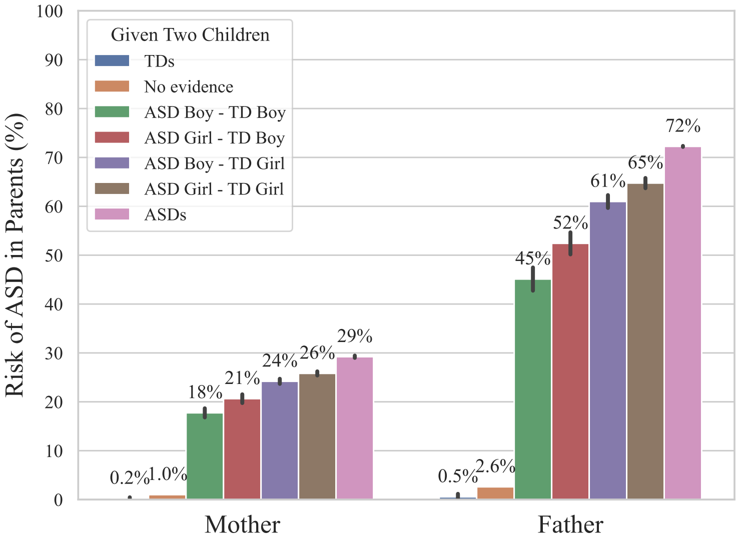


Figure 17 – Risk of ASD in parents given evidence regarding two children.

6.4.3 Risk of ASD in Parents Given Three Children

Tables 19-22 present the inference results for fathers given the evidence regarding three children. Fathers with three ASD children have ≈ 72−73% of ASD risk, with no significant increase than two children with ASD. Thus, this set of evidence shows that one TD child after two ASD cases does not decrease the risk of ASD for fathers meaningfully. However, given the evidence of three TD children, the risk of ASD for fathers tends to zero if at least one child is a boy.

Table 19 – Inference results for fathers given three boys.

		FB							
		asd	asd	asd	asd	td	td	td	td
		SB				SB			
		asd	asd	td	td	asd	asd	td	td
		TB		TB		TB		TB	
		asd	td	asd	td	asd	td	asd	td
F	asd	(0.7221) 0.7221	(0.7167) 0.7177		(0.1655) 0.2042	(0.7167) 0.7177	(0.1655) 0.2042		0.0002
	td	(0.2767) 0.2767	(0.2821) 0.2810		(0.8309) 0.7923	(0.2821) 0.2810	(0.8309) 0.7923		0.9954

[FB]First Boy; [SB]Second Boy; [TB]Third Boy; [F]Father.

Figure 18 summarizes the risk of ASD in fathers with three children. Given only one ASD case among three children, the causality associated with fathers varies substantially,

Table 20 – Inference results for fathers given two boys and one girl.

		FB							
		asd	asd	asd	asd	td	td	td	td
		SB				SB			
		asd	asd	td	td	asd	asd	td	td
		TG		TG		TG		TG	
		asd	td	asd	td	asd	td	asd	td
F	asd	(0.7235) 0.7235	(0.7198) 0.7201	(0.7178) 0.7185	(0.3562) 0.4067	(0.7178) 0.7185	(0.3562) 0.4067	(0.2298) 0.2822	0.0007
	td	(0.2753) 0.2753	(0.2790) 0.2787	(0.2810) 0.2803	(0.6410) 0.5907	(0.2810) 0.2803	(0.6410) 0.5907	(0.7668) 0.7147	0.9949

^{FB}First Boy; ^{SB}Second Boy; ^{TG}Third Girl; ^{F}Father.

Table 21 – Inference results for fathers given one boy and two girls.

		FB							
		asd	asd	asd	asd	td	td	td	td
		SG				SG			
		asd	asd	td	td	asd	asd	td	td
		TG		TG		TG		TG	
		asd	td	asd	td	asd	td	asd	td
F	asd	(0.7253) 0.7254	(0.7208) 0.7211		(0.5496) 0.5835	(0.7187) 0.7192	(0.4364) 0.4888		0.0024
	td	(0.2735) 0.2735	(0.2780) 0.2777		(0.4485) 0.4147	(0.2801) 0.2796	(0.5612) 0.5090		0.9933

^{FB}First Boy; ^{SG}Second Girl; ^{TG}Third Girl; ^{F}Father.

Table 22 – Inference results for fathers given three girls.

		FG							
		asd	asd	asd	asd	td	td	td	td
		SG				SG			
		asd	asd	td	td	asd	asd	td	td
		TG		TG		TG		TG	
		asd	td	asd	td	asd	td	asd	td
F	asd	(0.7279) 0.7279	(0.7221) 0.7222		(0.6017) 0.6300	(0.7221) 0.7222	(0.6017) 0.6300		0.0078
	td	(0.2709) 0.2709	(0.2767) 0.2766		(0.3966) 0.3684	(0.2767) 0.2766	(0.3966) 0.3684		0.9878

^{FG}First Girl; ^{SG}Second Girl; ^{TG}Third Girl; ^{F}Father.

according to the evidence regarding the other two children. The risk of ASD for fathers decreases to:

- $\approx$ 17–28% if the two TD children are boys, which represents a decrease of $\approx$ 58–74% in comparison to only one child with ASD (Section 6.4.1);
- $\approx$ 36–49% if exactly one of the TD children is a boy, a decrease of $\approx$ 26–45% in comparison to only one child with ASD; and
- $\approx$ 55–63% if the two TD children are girls, only $\approx$ 5–17% lower than cases where there is no evidence of TD children.

Tables 23-26 present the inference results for mothers given the evidence regarding three children. Mothers with three ASD children have $\approx$ 29–31% risk of ASD, presenting

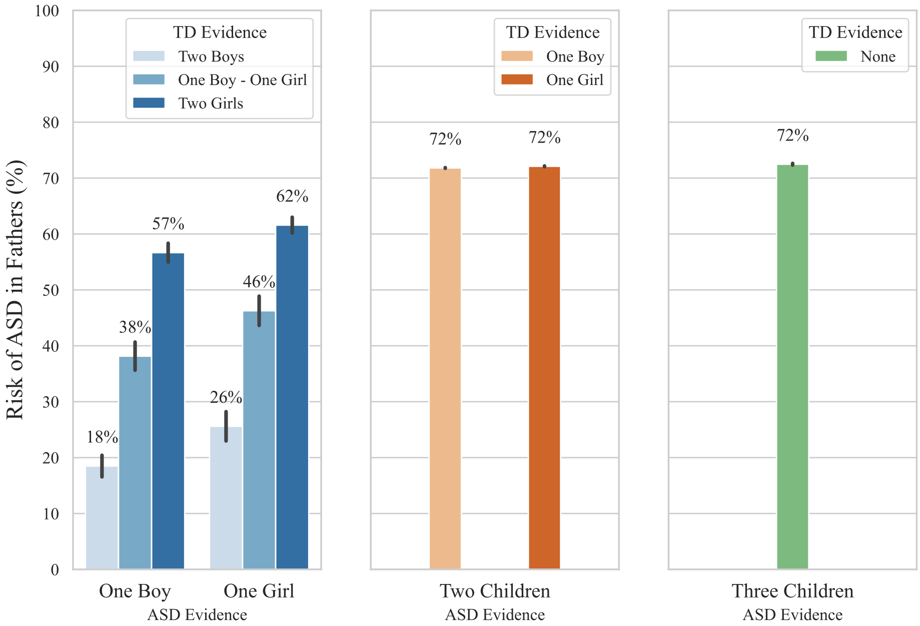


Figure 18 – Risk of ASD in fathers given evidence regarding three children.

no significant increase than in cases when they have two ASD children. As for fathers, this set of evidence shows that one TD child after two ASD cases does not decrease the risk of ASD for mothers significantly. Again, given the evidence of three TD children, the risk of ASD for mothers tends to zero if at least one child is a boy.

Table 23 – Inference results for mothers given three boys.

		FB							
		asd	asd	asd	asd	td	td	td	td
		SB				SB			
		asd	asd	td	td	asd	asd	td	td
		TB		TB		TB		TB	
		asd	td	asd	td	asd	td	asd	td
M	asd	(0.2906) 0.2906	(0.2820) 0.2825	(0.0649) 0.0800	(0.2820) 0.2825	(0.0649) 0.0800	0.0001		
	td	(0.7085) 0.7085	(0.7171) 0.7166	(0.9340) 0.9188	(0.7171) 0.7166	(0.9340) 0.9188	0.9987		

^{FB}First Boy; ^{SB}Second Boy; ^{TB}Third Boy; ^{M}Mother.

Table 24 – Inference results for mothers given two boys and one girl.

		FB							
		asd	asd	asd	asd	td	td	td	td
		SB				SB			
		asd	asd	td	td	asd	asd	td	td
		TG		TG		TG		TG	
		asd	td	asd	td	asd	td	asd	td
M	asd	(0.2941) 0.2941	(0.2863) 0.2864	(0.2830) 0.2833	(0.1399) 0.1597	(0.2830) 0.2833	(0.1399) 0.1597	(0.0901) 0.1107	0.0003
	td	(0.7051) 0.7050	(0.7128) 0.7127	(0.7161) 0.7158	(0.8590) 0.8392	(0.7161) 0.7158	(0.8590) 0.8392	(0.9087) 0.8882	0.9985

^{FB}First Boy; ^{SB}Second Boy; ^{TG}Third Girl; ^{M}Mother.

Table 25 – Inference results for mothers given one boy and two girls.

		FB							
		asd	asd	asd	asd	td	td	td	td
		SG				SG			
		asd	asd	td	td	asd	asd	td	td
		TG		TG		TG		TG	
		asd	td	asd	td	asd	td	asd	td
M	asd	(0.2988) 0.2988	(0.2883) 0.2884		(0.2171) 0.2305	(0.2841) 0.2842	(0.1715) 0.1922		0.0009
	td	(0.7003) 0.7003	(0.7108) 0.7107		(0.7819) 0.7685	(0.7150) 0.7149	(0.8274) 0.8068		0.9978

$^{\mathrm{FB}}$First Boy; $^{\mathrm{SG}}$Second Girl; $^{\mathrm{TG}}$Third Girl; $^{\mathrm{M}}$Mother.

Table 26 – Inference results for mothers given three girls.

		FG							
		asd	asd	asd	asd	td	td	td	td
		SG				SG			
		asd	asd	td	td	asd	asd	td	td
		TG		TG		TG		TG	
		asd	td	asd	td	asd	td	asd	td
M	asd	(0.3053) 0.3053	(0.2911) 0.2912		(0.2385) 0.2497	(0.2911) 0.2912	(0.2385) 0.2497		0.0031
	td	(0.6938) 0.6938	(0.7080) 0.7079		(0.7606) 0.7494	(0.7080) 0.7079	(0.7606) 0.7494		0.9957

$^{\mathrm{FG}}$First Girl; $^{\mathrm{SG}}$Second Girl; $^{\mathrm{TG}}$Third Girl; $^{\mathrm{M}}$Mother.

Figure 19 summarizes the risk of ASD for mothers with three children. Given only one ASD case among three children, the causality associated with mothers also varies substantially according to the evidence regarding the other two children. The risk of ASD for mothers decreases to:

- $\approx$ 6–11% if the two TD children are boys, a decrease of $\approx$ 58–77% in comparison to the case of only one child with ASD (Section 6.4.1);
- $\approx$ 14–19% if exactly one of the TD children is a boy, a decrease of $\approx$ 27–46% in comparison to the case of only one child with ASD; and
- $\approx$ 22–25% if the two TD children are girls, only $\approx$ 4–15% lower than cases where there is no evidence of TD children.

6.4.4 Risk of ASD in Parents Given Six Children

We also estimated six children per couple to evaluate the limits of the risk of ASD for parents and the emerging hypothesis that having two children with ASD seems sufficient to assume a high risk of ASD for parents, despite evidence regarding other children. Tables 27 and 28 present the inference results for fathers and mothers, respectively, given

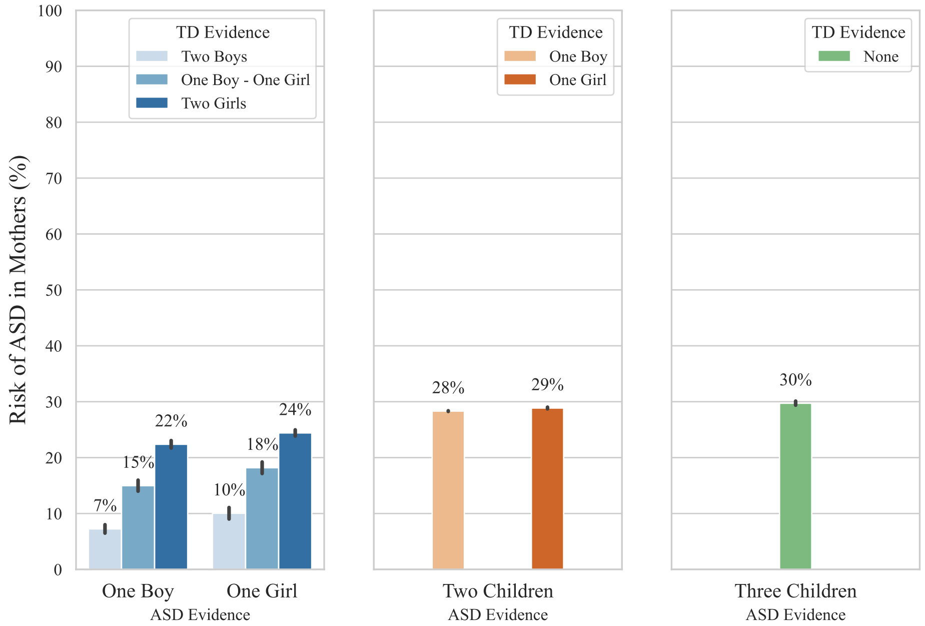


Figure 19 – Risk of ASD in mothers given evidence regarding three children.

the evidence regarding six children (three children of each gender in this case). These estimates allowed us to achieve two significant findings.

Table 27 – Inference results for fathers given six children.

FB		asd	asd	asd	asd	asd	asd	asd	asd	asd	asd	asd	asd	td	td	td	td
SB		asd	asd	asd	asd	asd	asd	asd	asd	td	td	td	td	td	td	td	td
TB		asd	asd	asd	asd	td	td	td	td	td	td	td	td	td	td	td	td
FG		asd	asd	asd	td	asd	asd	asd	td	asd	asd	asd	td	asd	asd	asd	td
SG		asd	asd	td	td	asd	asd	td	td	asd	asd	td	td	asd	asd	td	td
TG		asd	td	td	td	asd	td	td	td	asd	td	td	td	asd	td	td	td
F	asd	(0.7344) 0.7345	(0.7252) 0.7252	(0.7212) 0.7212	(0.7196) 0.7196	(0.7214) 0.7214	(0.7197) 0.7197	(0.7190) 0.7190	(0.7105) 0.7140	(0.7190) 0.7190	(0.7187) 0.7187	(0.7017) 0.7092	(0.0595) 0.0768	(0.7185) 0.7186	(0.6841) 0.6999	(0.0298) 0.0403	0.0001
	td	(0.2644) 0.2644	(0.2737) 0.2736	(0.2776) 0.2776	(0.2792) 0.2792	(0.2774) 0.2774	(0.2791) 0.2791	(0.2798) 0.2798	(0.2882) 0.2847	(0.2798) 0.2798	(0.2801) 0.2801	(0.2970) 0.2895	(0.9365) 0.9192	(0.2802) 0.2802	(0.3145) 0.2988	(0.9660) 0.9555	0.9956

FB First Boy; SB Second Boy; TB Third Boy; FG First Girl; SG Second Girl; TG Third Girl; F Father.

Table 28 – Inference results for mothers given six children.

FB		asd	asd	asd	asd	asd	asd	asd	asd	asd	asd	asd	asd	td	td	td	td
SB		asd	asd	asd	asd	asd	asd	asd	asd	td	td	td	td	td	td	td	td
TB		asd	asd	asd	asd	td	td	td	td	td	td	td	td	td	td	td	td
FG		asd	asd	asd	td	asd	asd	asd	td	asd	asd	asd	td	asd	asd	asd	td
SG		asd	asd	td	td	asd	asd	td	td	asd	asd	td	td	asd	asd	td	td
TG		asd	td	td	td	asd	td	td	td	asd	td	td	td	asd	td	td	td
M	asd	(0.3221) 0.3221	(0.2983) 0.2983	(0.2883) 0.2883	(0.2842) 0.2842	(0.2887) 0.2887	(0.2844) 0.2844	(0.2826) 0.2826	(0.2787) 0.2801	(0.2827) 0.2827	(0.2819) 0.2819	(0.2750) 0.2780	(0.0233) 0.0301	(0.2816) 0.2817	(0.2680) 0.2742	(0.0117) 0.0158	0.0000
	td	(0.6771) 0.6770	(0.7008) 0.7008	(0.7108) 0.7108	(0.7149) 0.7149	(0.7104) 0.7104	(0.7147) 0.7147	(0.7165) 0.7165	(0.7204) 0.7190	(0.7164) 0.7164	(0.7172) 0.7172	(0.7241) 0.7211	(0.9755) 0.9687	(0.7175) 0.7174	(0.7311) 0.7249	(0.9871) 0.9830	0.9987

FB First Boy; SB Second Boy; TB Third Boy; FG First Girl; SG Second Girl; TG Third Girl; M Mother.

First, it was possible to estimate the maximum risk of ASD for parents. Such limits are ≈ 32% for mothers and ≈ 73% for fathers. We believe in these limits because increasing the number of evidence (e.g., the number of ASD children from two to six) did not significantly increase parental causality.

Second, given the evidence of two children with ASD, we may attribute these ASD occurrences to genetics regardless of evidence about other children. We believe this because the risk of ASD for parents with only two children, both autistic, is similar to the risk of ASD for parents with six children (with two of them autistic). The closeness of these two estimates led us to believe that after two ASD children, evidence of TD children does not seem to decrease the parental causality significantly. Figure 20 summarizes the risk of ASD for parents with six children.

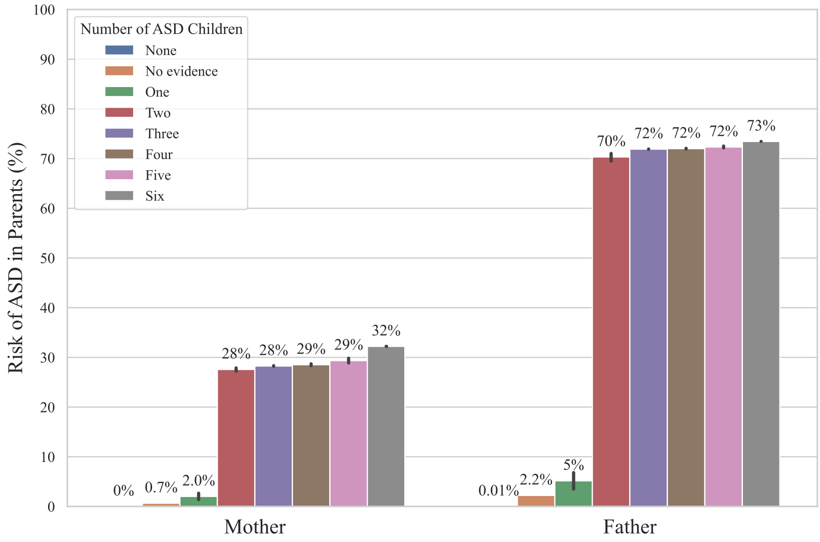


Figure 20 – Risk of ASD in parents given evidence regarding six children.

6.4.5 Results Accordance with ASD Heritability

This section aims to assess our inference results regarding the risk of ASD for parents contrasted to the well-known data regarding ASD heritability and prevalence.

As described in Section 2.4, the ASD heritability ranges from ≈ 80% to ≈ 93%. These probabilities indicate cases where an ASD child inherited the condition from at least one of his/her parents.

We calculated the joint probability for fathers and mothers given evidence of ASD children to estimate genetics as the most likely explanation for ASD cases. The inferences performed by our BN attribute to genetics from ≈ 88% to ≈ 94% if one ASD child, and nearly 100% if two or more ASD children. Observing values regarding only one individual, the mean probability attributed to genetics is ≈ 91%, which is supported by the ASD heritability estimates of Bai et al. (2019), Yip et al. (2018), and Sandin et al. (2017). In families with at least two ASD children, both parents present BAP features in ≈ 59% of cases. In ≈ 92% of these families, BAP features are present in at least one parent (LOSH et al., 2008).

The risks of ASD attributed to fathers were ≈ 2.5-fold greater than the risks attributed to mothers. If we take the ASD prevalence rate between males and females into account, this father:mother ASD risk ratio (2.5:1) is supported by the male:female ASD prevalence rate when there is evidence of genetic factors, as discussed in Sections 2.3.2.3 and 4.10.

6.5 Estimating the Risk of ASD in Grandparents

We evaluated the risk of ASD in grandparents by investigating ASD evidence in their grandchildren. We estimated the risk of ASD given they have up to two grandchildren, once having two children with ASD seems sufficient to assume a high risk of ASD for parents.

6.5.1 The Risk of ASD in Grandparents given One Grandchild

Table 29 presents the inference results for grandparents given the evidence regarding one grandchild. Given one ASD grandchild, the risk of ASD ranges from ≈ 6% for maternal grandmothers to ≈ 13−14% for paternal grandmothers. Regarding grandfathers, the risk ranges from ≈ 19−21% (maternal) to ≈ 44−47% (paternal). These results represent a ≈ 9−21-fold increase in the risk of ASD for grandparents given one grandchild with ASD.

Figure 21 summarizes the risk of ASD in grandparents given one grandchild.

Table 29 – Inference results for grandparents given one grandchild.

		FB		FG	
		asd	td	asd	td
GFF	asd	(0.4371) 0.4505	0.0083	(0.4579) 0.4681	0.0165
	td	(0.5604) 0.5471	0.9874	(0.5398) 0.5296	0.9793
GMF	asd	(0.1284) 0.1323	0.0024	(0.1345) 0.1375	0.0048
	td	(0.8705) 0.8666	0.9963	(0.8645) 0.8615	0.9939
GFM	asd	(0.1933) 0.1988	0.0164	(0.2031) 0.2074	0.0198
	td	(0.8032) 0.7978	0.9793	(0.7934) 0.7892	0.9760
GMM	asd	(0.0571) 0.0587	0.0048	(0.0600) 0.0613	0.0058
	td	(0.9417) 0.9401	0.9940	(0.9388) 0.9375	0.9930

^{FB}First Boy; ^{FG}First Girl; ^{GFF}Grandfather from Father side; ^{GMF}Grandmother from Father side; ^{GFM}Grandfather from Mother side; ^{GMM}Grandmother from Mother side.

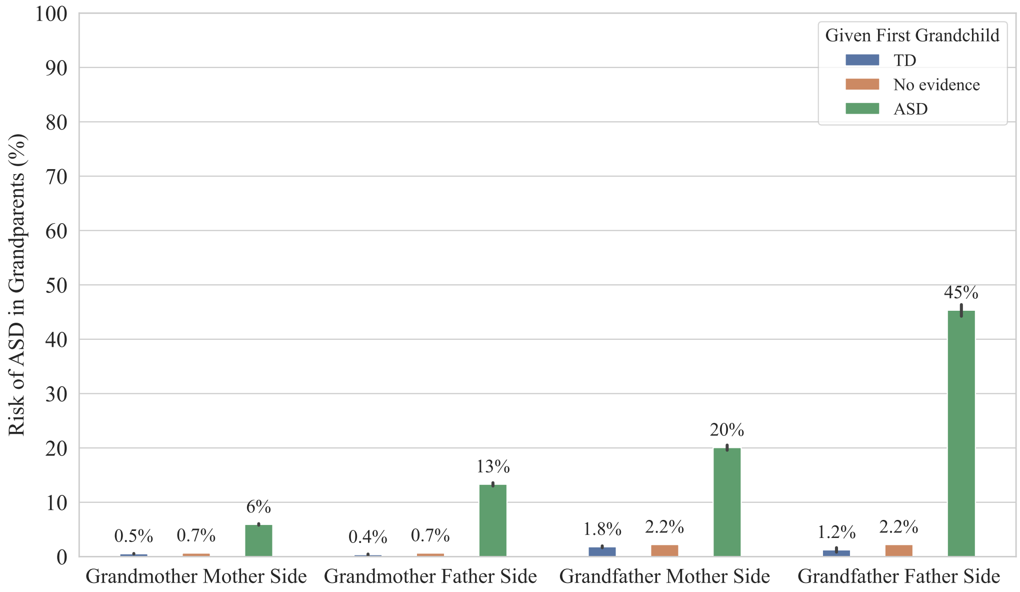


Figure 21 – Risk of ASD in grandparents given evidence regarding one grandchild.

6.5.2 The Risk of ASD in Grandparents given Two Grandchildren

Table 30 presents the inference results for grandparents given the evidence regarding two grandchildren. Given two ASD grandchildren, the risk of ASD ranges from ≈ 6−7% for maternal grandmothers to ≈ 15% for paternal grandmothers. Regarding grandfathers, the risk ranges from ≈ 22% (maternal) to ≈ 50% (paternal).

For groupings including one ASD grandchild and one TD grandchild, the risk of ASD for grandparents decreases ≈ 33−38% when the TD grandchild is a boy, and ≈ 13−17% when the TD grandchild is a girl.

Table 30 – Inference results for grandparents given two grandchildren.

		FB				FB				FG			
		asd	asd	td	td	asd	asd	td	td	asd	asd	td	td
		SB				SG				SG			
		asd	td	asd	td	asd	td	asd	td	asd	td	asd	td
GFF	asd	(0.4970) 0.4972	(0.2965) 0.3289		0.0054	(0.4979) 0.4981	(0.4121) 0.4301	(0.3472) 0.3780	0.0071	(0.4991) 0.4992	(0.4397) 0.4540		0.0126
	td	(0.5008) 0.5006	(0.7004) 0.6682		0.9903	(0.4999) 0.4998	(0.5854) 0.5675	(0.6500) 0.6193	0.9886	(0.4987) 0.4987	(0.5579) 0.5437		0.9831
GMF	asd	(0.1460) 0.1460	(0.0871) 0.0966		0.0016	(0.1462) 0.1463	(0.1210) 0.1263	(0.1019) 0.1110	0.0021	(0.1466) 0.1466	(0.1291) 0.1333		0.0037
	td	(0.8530) 0.8529	(0.9118) 0.9023		0.9972	(0.8527) 0.8527	(0.8779) 0.8726	(0.8969) 0.8879	0.9967	(0.8524) 0.8524	(0.8698) 0.8656		0.9951
GFM	asd	(0.2187) 0.2188	(0.1335) 0.1467		0.0153	(0.2208) 0.2209	(0.1819) 0.1893	(0.1543) 0.1668	0.0160	(0.2237) 0.2237	(0.1941) 0.2000		0.0182
	td	(0.7779) 0.7778	(0.8627) 0.8496		0.9805	(0.7758) 0.7758	(0.8145) 0.8072	(0.8420) 0.8295	0.9798	(0.7730) 0.7729	(0.8024) 0.7966		0.9775
GMM	asd	(0.0646) 0.0647	(0.0394) 0.0433		0.0044	(0.0653) 0.0653	(0.0538) 0.0559	(0.0456) 0.0493	0.0047	(0.0661) 0.0661	(0.0574) 0.0591		0.0053
	td	(0.9342) 0.9342	(0.9594) 0.9555		0.9943	(0.9336) 0.9336	(0.9451) 0.9429	(0.9532) 0.9495	0.9941	(0.9327) 0.9327	(0.9415) 0.9397		0.9934

$^{\text{FB}}$First Boy; $^{\text{FG}}$First Girl; $^{\text{SB}}$Second Boy; $^{\text{SG}}$Second Girl; $^{\text{GFF}}$Grandfather from Father side; $^{\text{GMF}}$Grandmother from Father side; $^{\text{GFM}}$Grandfather from Mother side; $^{\text{GMM}}$Grandmother from Mother side.

Figure 22 summarizes the risk of ASD for grandparents given two grandchild.

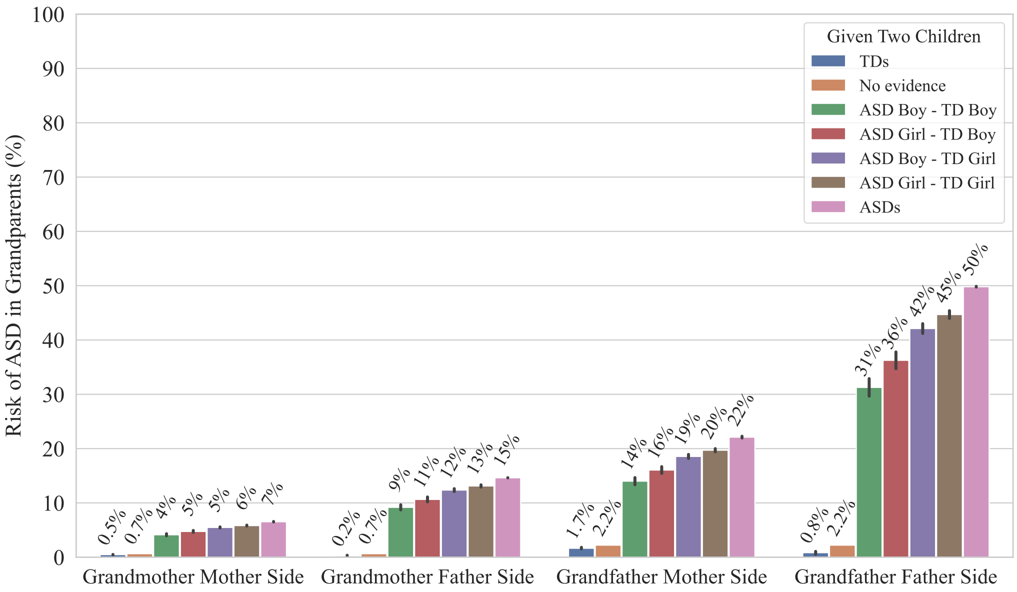


Figure 22 – Risk of ASD in grandparents given evidence regarding two grandchild.

Given two grandchildren with ASD, the risk of ASD is up to ≈ 77% for grandfathers and ≈ 23% for grandmothers if the two grandchildren with ASD are not siblings. That is, they are grandchildren of different children (simulated using the BN defined in

Chapter 7).

6.5.3 Results Accordance with Mental and Neurological Disorders Among Grandparents

This section aims to assess our inference results regarding the risk of ASD for grandparents contrasted to the literature concerning family history of mental or neurological disorders with shared symptomatic and possibly etiologic overlap with ASD.

A study conducted by Xie et al. (2019) pointed that ≈ 81% of ASD people have a grandparent with some mental or neurological disorder, reaching up to 90% for some autistic subgroups. Given one ASD grandchild, our results suggest that the combined risk of ASD in grandparents ranges from ≈ 82% to ≈ 88%, ≈ 26% for maternal grandparents, and ≈ 58% for paternal grandparents. The combined risk of ASD in grandparents can be up to ≈ 94% given two ASD grandchildren, ≈ 29% for maternal grandparents, and ≈ 65% for paternal grandparents. Once few works quantified the association between family history of mental or neurological disorders with ASD beyond first-degree relatives, we believe this is a reasonable comparison.

6.6 Estimating the Risk of ASD Among Siblings

We evaluated the risk of ASD in some individuals by investigating ASD evidence in their siblings. We estimated the risk of ASD in individuals with one older sibling with ASD since most of the works that studied ASD heritability/recurrence evaluated pairs of siblings. Nevertheless, we also estimated the risk of ASD in the case of two affected older siblings to assess the ASD risk thresholds in the younger siblings.

6.6.1 The Risk of ASD Given One Affected Sibling

Table 31 presents the inference results for full- and half-siblings given evidence regarding one older sibling.

Given one older ASD sibling, the risk of ASD for younger full brothers ranges from ≈ 70% to ≈ 75%, approximately 23-fold increase in the risk of ASD. Regarding younger half brothers, the risk of ASD ranges from ≈ 51% to ≈ 54% for paternal half brothers

Table 31 – Inference results for second siblings given one older sibling.

		FB		FG	
		asd	td	asd	td
SB	asd	(0.6980) 0.7194	0.0089	(0.7308) 0.7472	0.0220
	td	(0.2978) 0.2763	0.9869	(0.2649) 0.2485	0.9737
BFP	asd	(0.5057) 0.5211	0.0124	(0.5296) 0.5413	0.0217
	td	(0.4900) 0.4747	0.9834	(0.4662) 0.4544	0.9740
BMP	asd	(0.2166) 0.2227	0.0222	(0.2275) 0.2322	0.0258
	td	(0.7791) 0.7731	0.9736	(0.7683) 0.7636	0.9699
SG	asd	(0.2912) 0.3002	0.0032	(0.3054) 0.3123	0.0087
	td	(0.7075) 0.6986	0.9955	(0.6933) 0.6865	0.9901
GFP	asd	(0.2106) 0.2171	0.0047	(0.2206) 0.2255	0.0086
	td	(0.7881) 0.7817	0.9941	(0.7782) 0.7732	0.9902
GMP	asd	(0.0905) 0.0931	0.0087	(0.0951) 0.0971	0.0103
	td	(0.9082) 0.9057	0.9900	(0.9036) 0.9017	0.9885

^{FB}First Boy; ^{FG}First Girl; ^{SB}Second Boy; ^{BFP}Boy (half-brother) from Father Partner; ^{BMP}Boy (half-brother) from Mother Partner; ^{SG}Second Girl; ^{GFP}Girl (half-sister) from Father Partner; ^{GMP}Girl (half-sister) from Mother Partner.

and ≈ 22% to ≈ 23% for maternal half brothers, nearly 17 and 7-fold increase in the risk of ASD, respectively.

The risk of ASD for younger full sisters ranges from ≈ 29% to ≈ 31% given one older ASD sibling, approximately 25-fold increase in the risk of ASD. Regarding younger half sisters, the risk of ASD ranges from ≈ 21% to ≈ 23% for paternal half sisters and ≈ 9% to ≈ 10% for maternal half sisters, nearly 18 and 8-fold increase in the risk of ASD, respectively.

For all second siblings cases, the higher risk occurs when the older sibling is a girl. Figures 23 and 24 summarize the risk of ASD in second brothers and sisters, respectively.

6.6.2 The Risk of ASD Given Two Affected Siblings

Table 32 presents the inference results for younger siblings given the evidence regarding two older siblings.

Given two older ASD siblings, the risk of ASD for younger full brothers is ≈ 79%, approximately 25-fold increase in the risk of ASD. Regarding younger half brothers, the risk of ASD is ≈ 57% for paternal half brothers and ≈ 25% for maternal half brothers,

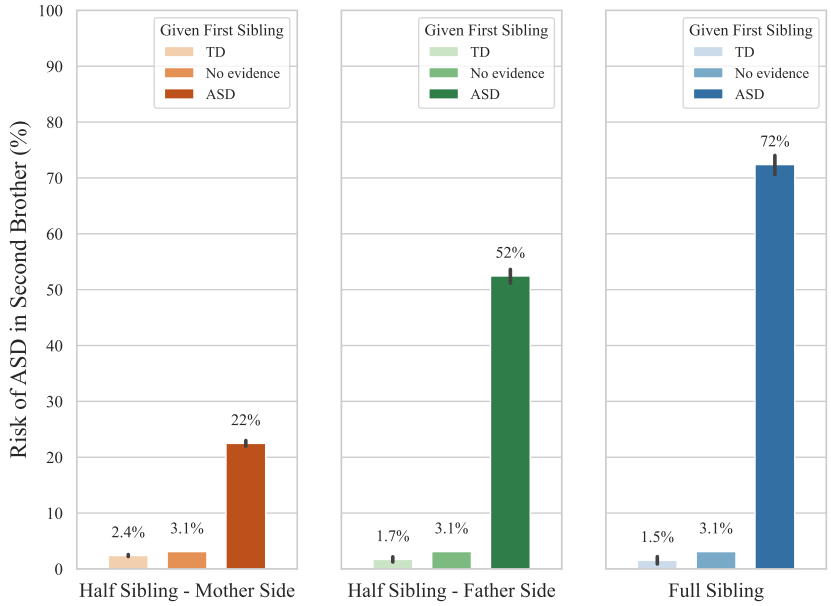


Figure 23 – Risk of ASD in second brother given evidence regarding one older sibling.

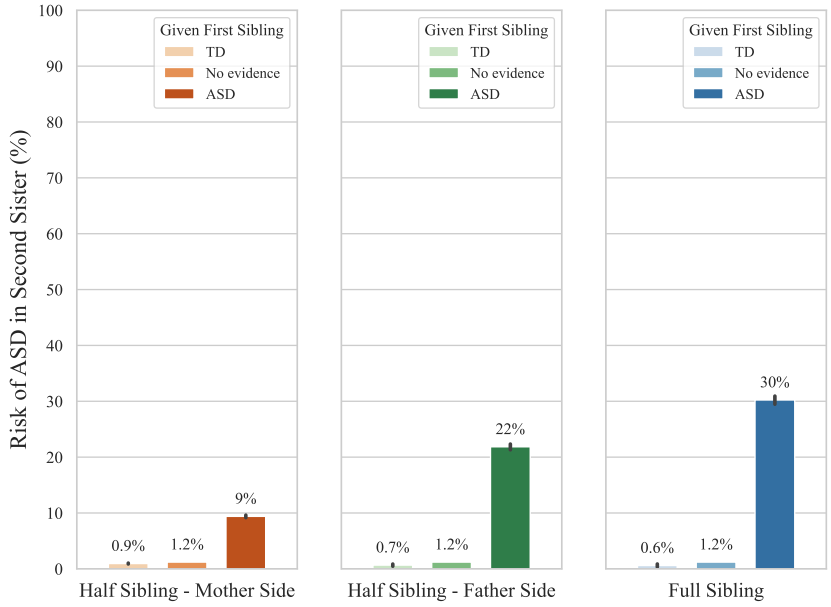


Figure 24 – Risk of ASD in second sister given evidence regarding one older sibling.

nearly 18 and 8-fold increase in the risk of ASD, respectively. These risks estimates do not increase significantly given three older siblings with ASD.

The risk of ASD for younger full sisters is $\approx 33\%$ given two older ASD siblings,

Table 32 – Inference results for third siblings given two older siblings.

		FB			FB				FG		
		asd	asd td	td	asd	asd	td	td	asd	asd td	td
		SB			SG				SG		
		asd	td asd	td	asd	td	asd	td	asd	td asd	td
TB	asd	(0.7940) 0.7942	(0.4727) 0.5248	0.0043	(0.7948) 0.7950	(0.6581) 0.6870	(0.5541) 0.6037	0.0070	(0.7958) 0.7960	(0.7022) 0.7251	0.0159
	td	(0.2018) 0.2016	(0.5231) 0.4710	0.9915	(0.2009) 0.2008	(0.3377) 0.3088	(0.4417) 0.3921	0.9888	(0.1999) 0.1998	(0.2936) 0.2707	0.9799
BFP	asd	(0.5747) 0.5748	(0.3440) 0.3812	0.0090	(0.5757) 0.5758	(0.4769) 0.4976	(0.4022) 0.4377	0.0110	(0.5770) 0.5771	(0.5087) 0.5251	0.0173
	td	(0.4211) 0.4209	(0.6518) 0.6145	0.9867	(0.4201) 0.4200	(0.5188) 0.4982	(0.5935) 0.5580	0.9848	(0.4187) 0.4187	(0.4871) 0.4707	0.9784
BMP	asd	(0.2446) 0.2447	(0.1509) 0.1654	0.0209	(0.2470) 0.2470	(0.2042) 0.2123	(0.1738) 0.1876	0.0216	(0.2501) 0.2501	(0.2175) 0.2240	0.0241
	td	(0.7511) 0.7511	(0.8448) 0.8304	0.9749	(0.7488) 0.7488	(0.7916) 0.7835	(0.8220) 0.8082	0.9741	(0.7457) 0.7456	(0.7782) 0.7718	0.9717
TG	asd	(0.3316) 0.3317	(0.1964) 0.2181	0.0013	(0.3326) 0.3327	(0.2742) 0.2862	(0.2304) 0.2511	0.0024	(0.3339) 0.3339	(0.2929) 0.3025	0.0061
	td	(0.6671) 0.6670	(0.8023) 0.7806	0.9975	(0.6662) 0.6661	(0.7246) 0.7125	(0.7684) 0.7477	0.9963	(0.6649) 0.6648	(0.7059) 0.6963	0.9926
GFP	asd	(0.2394) 0.2395	(0.1431) 0.1587	0.0033	(0.2399) 0.2399	(0.1986) 0.2073	(0.1674) 0.1823	0.0041	(0.2404) 0.2404	(0.2119) 0.2187	0.0067
	td	(0.7593) 0.7593	(0.8556) 0.8401	0.9955	(0.7589) 0.7589	(0.8001) 0.7915	(0.8313) 0.8165	0.9947	(0.7584) 0.7583	(0.7869) 0.7800	0.9920
GMP	asd	(0.1023) 0.1024	(0.0629) 0.0690	0.0082	(0.1033) 0.1033	(0.0853) 0.0887	(0.0725) 0.0783	0.0085	(0.1046) 0.1046	(0.0909) 0.0936	0.0096
	td	(0.8964) 0.8964	(0.9359) 0.9298	0.9906	(0.8955) 0.8954	(0.9135) 0.9101	(0.9262) 0.9204	0.9902	(0.8942) 0.8941	(0.9078) 0.9051	0.9892

^{FB}First Boy; ^{FG}First Girl; ^{SB}Second Boy; ^{SG}Second Girl; ^{TB}Third Boy; ^{BFP}Boy (half-brother) from Father Partner; ^{BMP}Boy (half-brother) from Mother Partner; ^{TG}Third Girl; ^{GFP}Girl (half-sister) from Father Partner; ^{GMP}Girl (half-sister) from Mother Partner.

approximately 28-fold increase in the risk of ASD. Regarding younger half sisters, the risk of ASD is ≈ 24% for paternal half sisters and ≈ 10% for maternal half sisters, nearly 20 and 8-fold increase in the risk of ASD, respectively. These risk of ASD estimates do not increase significantly given three older siblings with ASD.

Figures 25 and 26 summarize the risk of ASD in third brothers and sisters, respectively.

Some research present a recurrence risk increases up to 2.5-fold for third-born children with two younger siblings with ASD than for second-born children with one younger sibling with ASD (RISCH et al., 2014). However, our results suggest a more modest increase in the recurrence risk for third-born children, approximately a 1.1-fold increase compared with second-born children.

The ASD etiology is primarily due to inherited genetics. Thus, we believe in the accuracy of our estimates once which determines the risk is the parents' genetics, not

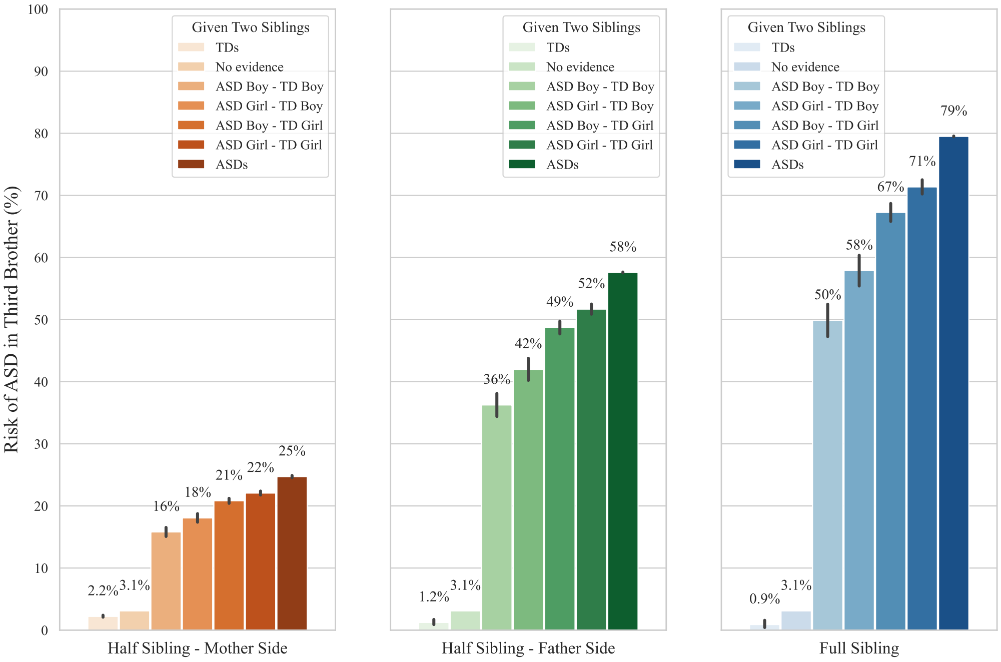


Figure 25 – Risk of ASD in third brother given evidence regarding two older siblings.

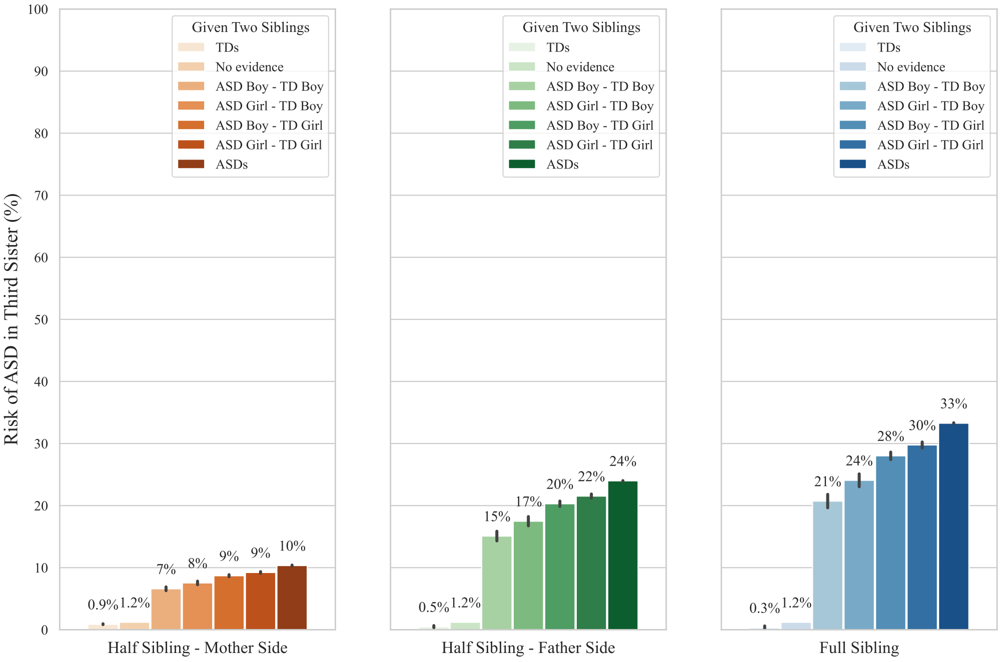


Figure 26 – Risk of ASD in third sister given evidence regarding two older siblings.

the number of affected children. An aspect that can explain this difference is the curtailment of reproduction, known as reproductive stoppage. The stoppage is a phenomenon that parents tend to stop having children after the birth of an affected one, which has implications for recurrence risk estimation and genetics, once it can bias recurrence risk estimates negatively if not addressed accurately (HANSEN et al., 2019; WOOD et al., 2015; HOFFMANN et al., 2014).

6.6.3 Results Accordance with ASD and BAP Recurrence Among Siblings

This section assesses our inference results concerning the risk of ASD in siblings contrasted to the well-known data regarding ASD and BAP recurrence among siblings.

As previously described in Section 2.6, overall estimates of ASD plus BAP for high-risk siblings are up to 77% (average: 50%), up to 52% for females (average: 32%), and up to 88% for males (average: 60%). These probabilities indicate the risk of ASD/BAP for younger full siblings given an older sibling with ASD.

Our results suggest an ASD risk of $\approx$ 72% for males and $\approx$ 30% for females given one older full sibling with ASD. In cases with two older siblings with ASD, the risk of ASD in younger full siblings is $\approx$ 79% for males and $\approx$ 33% for females. These estimates are similar to the ASD plus BAP recurrences estimated by Hudry et al. (2014), Chawarska et al. (2014), Ozonoff et al. (2014), Hoffmann et al. (2014), Wan et al. (2013), Macari et al. (2012), Schwichtenberg et al. (2010) and Christensen et al. (2010).

Compared to full-siblings, our results suggest that maternal half-siblings have $\approx$ 3-fold decrease in the risk of ASD, while paternal half-siblings have $\approx$ 1.4-fold decrease. Studies on ASD recurrence among siblings have reported risk for full-siblings up to 3-fold that of half-siblings (HANSEN et al., 2019; WOOD et al., 2015; SANDIN et al., 2014; RISCH et al., 2014; GRØNBORG; SCHENDEL; PARNER, 2013).

However, the estimates made by Risch et al. (2014) presented a recurrence for maternal half-siblings twice as high as paternal half-siblings. Similarly, the estimates of Hansen et al. (2019), and Grønborg, Schendel and Parner (2013) showed an ASD recurrence for maternal half-siblings slightly larger than for paternal half-siblings, although not significantly different. Factors associated with pregnancy and the maternal intrauterine environment may support these higher recurrence risks in maternal half-siblings (GRØN-

BORG; SCHENDEL; PARNER, 2013).

It suggests that the genotype of the disorder comes in similar amounts from both parents, although the ASD phenotype in fathers is more common. Thus, ASD recurrence between half-siblings must be interpreted carefully, as the genetic inheritance may have come from either parent, possibly even both. Consequently, it is suitable to assume the risk of ASD for half-siblings to be half the risk of full-siblings.

6.7 Interpreting Virtual Evidences

Cases with uncertainty about the evidence can be better evaluated using virtual evidence, as they allow addressing such situations. The use of virtual evidence allowed treating the possible cause of ASD in individuals diagnosed according to the disorder's etiology. Thus, it was possible to attribute the cause to hereditary factors and non-hereditary factors.

The use of virtual evidence regarding one ASD individual reduced the risk of ASD by only $\approx 2{-}3\%$ for parents, grandparents, and siblings. The extremely low prevalence of ASD due to non-hereditary factors, $\approx 0.1\%$ for females and $\approx 0.4\%$ for males can explain this slight decrease.

There was no decrease in the risk of ASD for cases with virtual evidence of two or more ASD individuals. Results from Messinger et al. (2015), Gerdts et al. (2013), and Losh et al. (2008) suggest that genetic causes of ASD may vary between simplex and multiplex families. ASD and BAP traits are higher among members of multiplex families, and such individuals are more vulnerable to ASD symptoms. Thus, we have further evidence that two cases of autism in a family are sufficient to attribute the ASD cause to inherited genetic factors, even using virtual evidence.

Therefore, estimates obtained from virtual evidence are helpful for cases of a single affected individual. In those cases, where some uncertainty is added to the only evidence, the risk of ASD is slightly reduced. Although statistically insignificant, this information decreases the range of ASD risk and can be important as directives for different decision processes.

6.8 Summary

This chapter proposed a BN to assess the risk of ASD in parents, grandparents, and siblings. Our results produced the following contributions:

- Well reasoned estimates regarding the risk of ASD for parents, grandparents, and siblings given evidence of ASD individuals in the family;
- The estimate of reliable limits to the risk of ASD for parents, grandparents, and siblings;
- An estimate regarding the number of ASD cases in a family sufficient to attribute the ASD occurrences to the genetic inheritance;
- Having an ASD child indicates a higher risk for ASD in parents, even higher if the ASD child is a girl. Having a TD child implies a lower risk of ASD in parents, even lower if the ASD child is a boy. Our results corroborate and quantify these implications.

Besides these suggested contributions, when it was reasonable to compare, some of our estimates have substantial similarities to the ASD heritability and recurrence data in the literature, which validates the adequacy of our proposed BN model as a tool to estimate the risk of ASD in family members.

The next chapter presents a BN proposed to estimate the risk of ASD in males and females given additional evidence sets still unexplored.

7 Bayesian Network to Estimate the Risk of ASD Given Evidence in Others Relatives

The current literature suggests that the ASD recurrence risk dimension among family members depends significantly on relatedness (genetic similarity) and likely on their genders. The more closely the affected family member, the higher the risk of ASD (XIE et al., 2019; HANSEN et al., 2019).

This chapter aims to assess the risk of ASD in male and female individuals given evidence regarding some of their second- and third-degree relatives. Thus, we will estimate the risk of ASD given evidence in grandparents, aunts, uncles, nieces, nephews, and cousins.

7.1 Problem Definition

We aimed to estimate conditional probabilities such as:

- `P(Individual|Paternal Grandparents)`;
- `P(Individual|Maternal Grandparents)`;
- `P(Individual|Paternal Uncle)`, `P(Individual|Maternal Uncle)`;
- `P(Individual|Paternal Aunt)`, `P(Individual|Maternal Aunt)`;
- `P(Individual|Nieces)`, `P(Individual|Nephews)`; and
- `P(Individual|Paternal Cousins)`, `P(Individual|Maternal Cousins)`.

Notations like `P(Individual|Paternal Grandparents)` summarizes the following conditional probabilities `P(Male=asd|GFF=asd)`, `P(Male=asd|GMF=asd)`, `P(Male=asd |GFM=asd)`, `P(Male=asd|GMM=asd)`, `P(Female=asd|GFF=asd)`, `P(Female=asd|GMF=asd)`, `P(Female=asd|GFM=asd)`, `P(Female=asd|GMM=asd)`. We also explored combinations with two individuals as evidence, such as `P(Male=asd|GFF=asd,GMF=asd)`, `P(Female=asd|GFF =asd,GMM=asd)`.

7.2 Bayesian Network Structure

Using a portion of the variables previously defined in Section 5.2, we created a BN to estimate the risk of ASD in second- and third-degree relatives. Figure 27 presents the structure of the proposed BN.

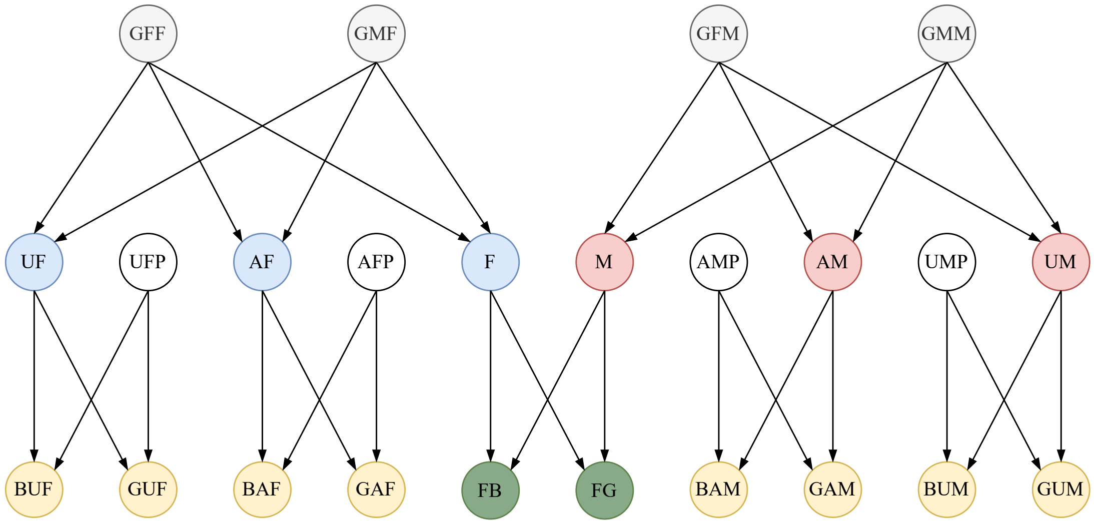


Figure 27 – The structure of a BN to estimate the risk of ASD in second- and third-degree relatives.

According to the types of variables defined in Section 3.4.4.1, the nodes representing grandparents can be classified as background variables. They work as background variables because they will be observed and have an indirect causal relationship regarding the symptom variables.

The nodes representing parents, uncles, and aunts can be classified as background or problem variables. The nodes representing relatives in law (white nodes) work as background variables because they directly relate to the symptom variables. The nodes representing blood relatives (blue and red nodes) work as problem variables because they are parents of symptom variables and children of background variables. Furthermore, in a few cases, some of these variables are among the variables of interest given the problem definition (to estimate the risk of ASD given evidence regarding nephews and nieces).

The nodes representing children and cousins can be classified as symptom or hypothesis variables. The variables representing cousins (yellow nodes) work as symptom variables while we take them as evidence. The variables representing children (green nodes) act as hypothesis variables when we make inferences about them given the problem definition (to estimate the risk of ASD given evidence in grandparents, uncles, aunts, and cousins).

We used the prior probabilities defined in Tables 12 and 13 as the discrete probability distributions for root nodes, those representing grandparents (gray nodes), uncles in law, and aunts in law (white nodes). Nodes representing parents, blood uncles, blood aunts (blue and red nodes), children (green nodes), and cousins (yellow nodes) have their CPTs produced from the conditional probabilities presented in Tables 14 and 15. We emphasize that the probabilities correspond to the gender of the family member represented by each node.

Given no evidence, the marginal probabilities produced by this BN are similar to those in Section 6.3, considering variables at the same level and relationship type. Our inferences were performed using both hard and virtual evidence, which allowed us to consider the uncertainty of the evidence.

7.3 Risk of ASD Given Evidence Regarding Grandparents

Given the ASD diagnosis in one grandparent, the risk of ASD for male individuals ranges from $\approx$ 27% (if maternal grandparent with ASD) to $\approx$ 64% (if paternal grandparent with ASD). The risk of ASD for female individuals ranges from $\approx$ 11% (if maternal grandparent with ASD) to $\approx$ 27% (if paternal grandparent with ASD). In all cases, the risk of ASD is slightly higher when there is evidence of ASD in grandmothers. Figure 28 summarizes the average risk of ASD given the evidence regarding grandparents.

The cup model assumes that females have a higher cup representing ASD risk than males. Larger cups mean a higher tolerance threshold to develop ASD in women (HOANG; CYTRYNBAUM; SCHERER, 2018). For example, a couple of children from the same parents share their genetics. Suppose both children received the same ASD-related genetic variants. In this case, the necessary variants for boys to develop ASD may not be enough for girls to develop ASD, despite ASD genetic variants being present in these unaffected girls.

Since ASD-related genetic variants may be present in unaffected females, these females may transfer such variants onto their offspring, likewise affected males. Therefore, we can conclude that the ASD diagnosis in males reflects the ASD genotype more

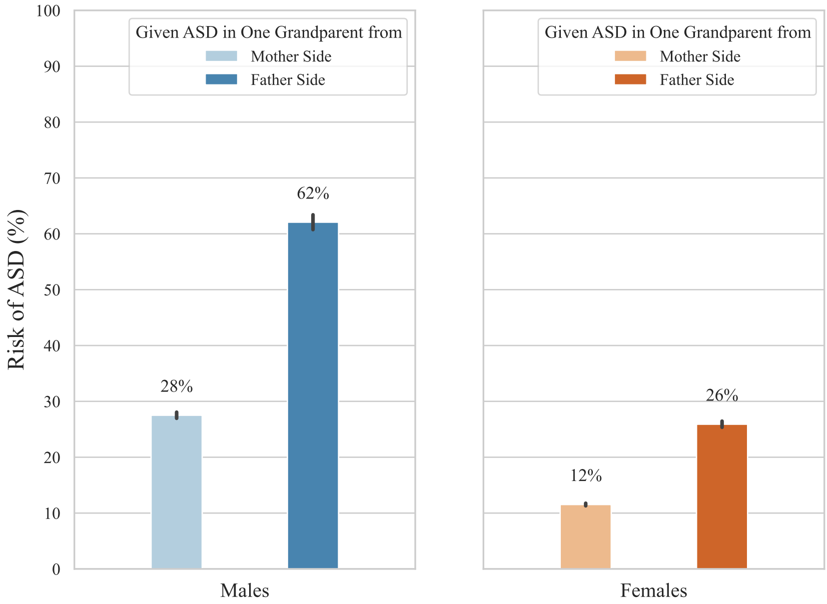


Figure 28 – Risk of ASD given evidence regarding one grandparent.

efficiently. Once this premise is supported by ASD prevalence, recurrence, and genetic data, hereafter, we will take as truth the proposition that the ASD phenotype in males better reflects the ASD genotype.

As the probabilities of our BNs are based on data related to the ASD phenotype (clinical diagnosis), it is most accurate to interpret the results of causal queries taking the paths involving male individuals into account.

Thus, taking results from the path "`paternal grandparents→father`", the risk of ASD tends to be:

- ≈ 26% for females and ≈ 62% for males given one grandparent with ASD, regardless of being paternal or maternal (Figure 29a);
- ≈ 31% for females and ≈ 75% for males given two grandparents with ASD, the paternal grandparents or the maternal grandparents (Figure 29b); and
- ≈ 45% for females and ≈ 85% for males given two grandparents with ASD, one paternal grandparent and one maternal grandparent (estimated using the probability addition rule and the conditional probability of males in the node representing the mother, Figure 29c).

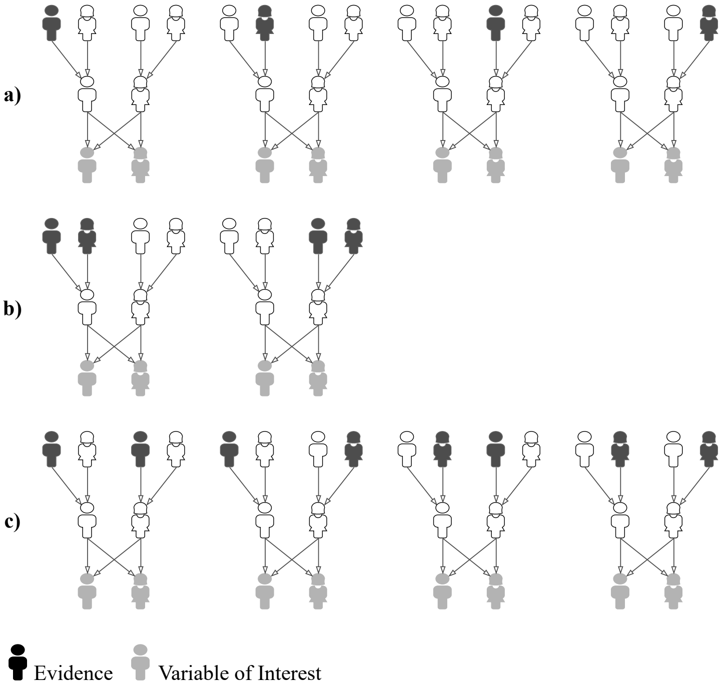


Figure 29 – Evidence setup for the risk of ASD given evidence regarding grandparents.

7.4 Risk of ASD Given Evidence Regarding Uncles and Aunts

Given the ASD diagnosis in one uncle-or-aunt, the risk of ASD for male individuals ranges from ≈ 24% (if one uncle from the mother side with ASD) to ≈ 59% (if one aunt from the father side with ASD). The risk of ASD for female individuals ranges from ≈ 10% (if one uncle from the mother side with ASD) to ≈ 25% (if one aunt from the father side with ASD). In all cases, the risk of ASD is slightly higher when there is evidence of ASD in aunts. Figure 30 summarizes the average risk of ASD given the evidence regarding uncles and aunts.

Considering males reflects the ASD genotype more efficiently, and taking results from the path “`uncle-or-aunt`→`paternal grandparents`→`father`”, the risk of ASD tends to be:

- ≈ 24% for females and ≈ 57% for males given one uncle-or-aunt with ASD, regardless of being paternal or maternal (Figure 31a);
- ≈ 26% for females and ≈ 63% for males given one uncle and one aunt with ASD, both from the father side or both from the mother side (Figure 31b); and
- ≈ 42% for females and ≈ 80% for males given one paternal uncle-or-aunt with ASD

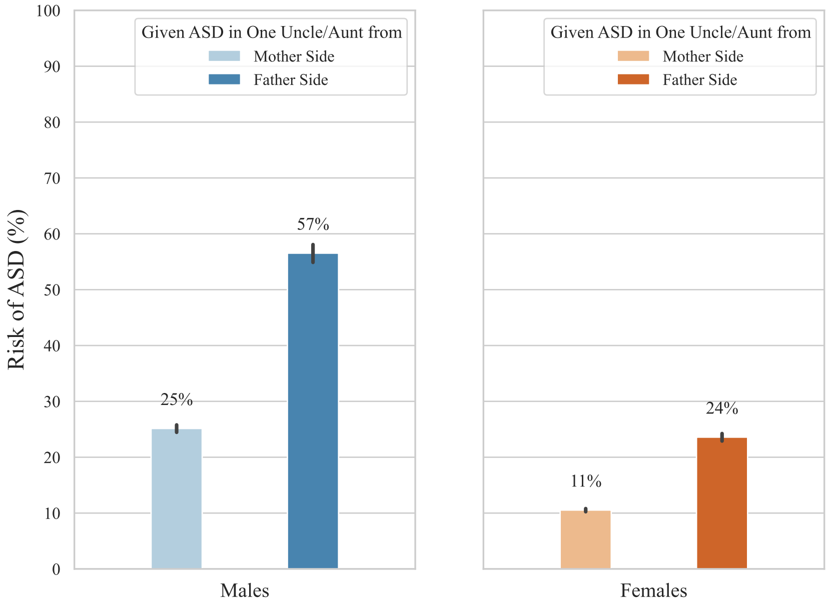


Figure 30 – Risk of ASD given evidence regarding one uncle-or-aunt.

and one maternal uncle-or-aunt with ASD (estimated using the probability addition rule and the conditional probability of males in the node representing the mother, Figure 31c).

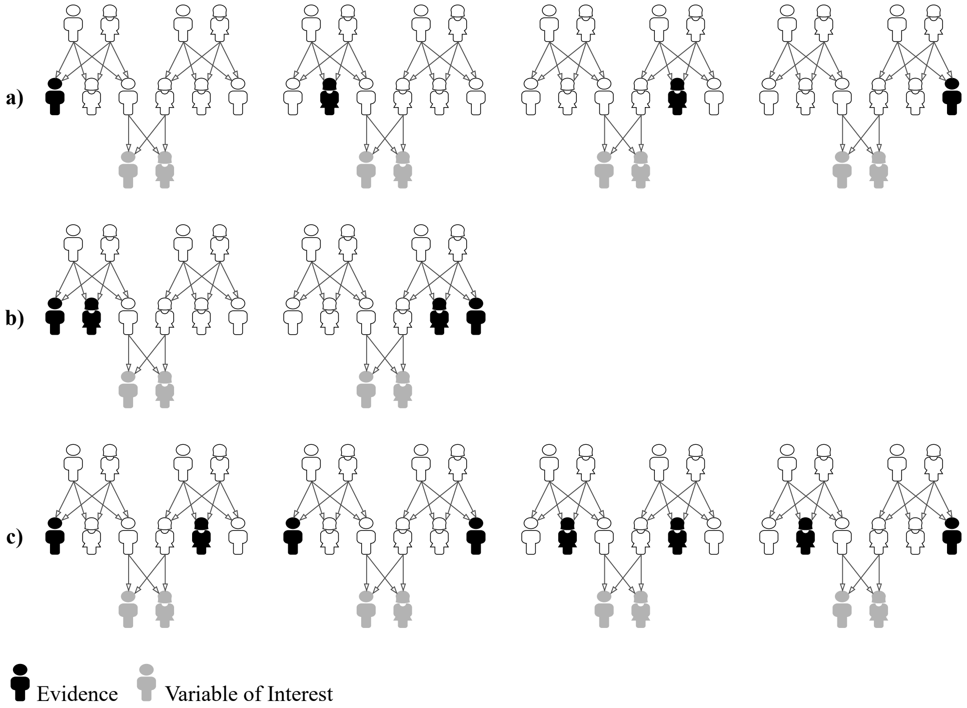


Figure 31 – Evidence setup for the risk of ASD given evidence regarding uncles and aunts.

7.5 Risk of ASD Given Evidence Regarding Nephews and Nieces

Given the ASD diagnosis in one nephew-or-niece, the risk of ASD for male individuals ranges from ≈ 21% (if ASD in one nephew, son of a sister) to ≈ 53% (if ASD in one niece, daughter of a brother). The risk of ASD for female individuals ranges from ≈ 9% (if ASD in one nephew, son of a sister) to ≈ 22% (if ASD in one niece, daughter of a brother). In all cases, the risk of ASD is slightly higher when there is evidence of ASD in nieces. Figure 32 summarizes the average risk of ASD given the evidence regarding nephews and nieces.

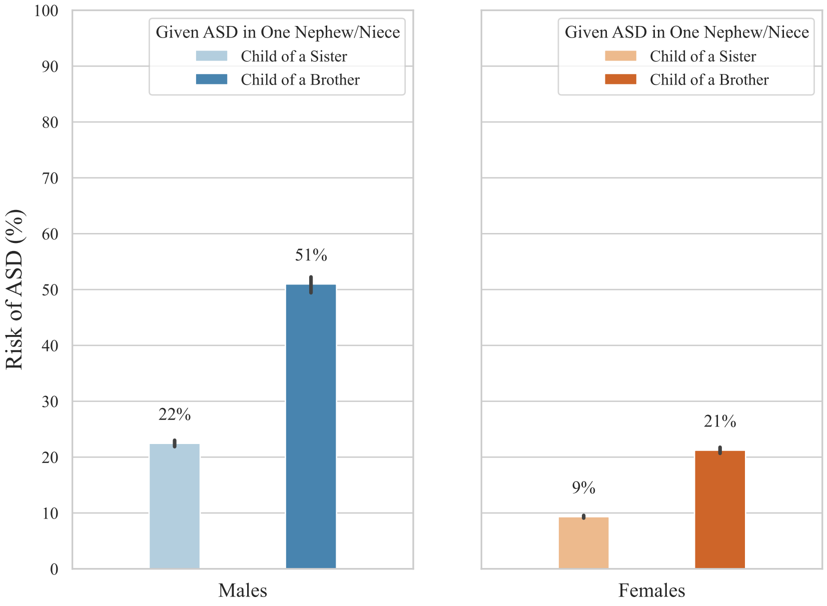


Figure 32 – Risk of ASD given evidence regarding one nephew-or-niece.

Considering males reflects the ASD genotype more efficiently, and taking results from the path "`nephew-or-niece`→`brother`→`parents`", the risk of ASD tends to be:

- ≈ 21% for females and ≈ 51% for males given one nephew-or-niece with ASD, regardless of being a child of a sister or a child of a brother (Figure 33a);
- ≈ 24% for females and ≈ 57% for males given one nephew and one niece with ASD, children of the same brother or sister (Figure 33b); and
- ≈ 33% for females and ≈ 79% for males given one nephew-or-niece with ASD, child of a brother-or-sister, and another nephew-or-niece with ASD, child of a distinct

brother-or-sister (Figure 33c).

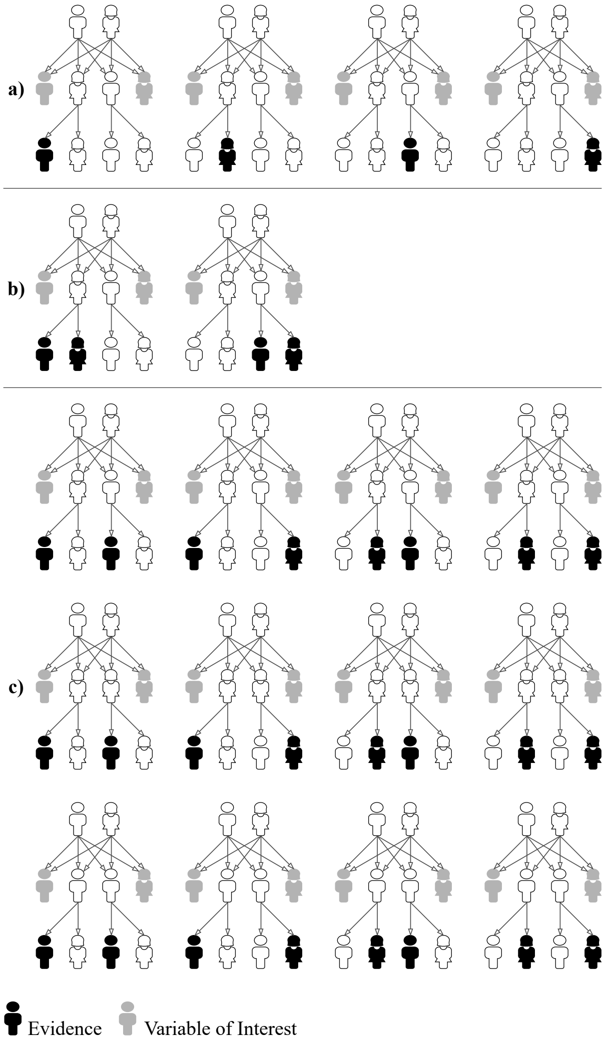


Figure 33 – Evidence setup for the risk of ASD given evidence regarding nephews and nieces.

7.6 Risk of ASD Given Evidence Regarding Cousins

Given the ASD diagnosis in one cousin, the risk of ASD for male individuals ranges from ≈ 9% (if ASD in one male cousin, son of an aunt from the mother side) to ≈ 42% (if ASD in one female cousin, daughter of an uncle from the father side). The risk of ASD for female individuals ranges from ≈ 4% (if ASD in one cousin, child of an aunt from the mother side) to ≈ 18% (if ASD in one female cousin, daughter of an uncle from the father side). In all cases, the risk of ASD is slightly higher when there is evidence of ASD in female cousins. Figure 34 summarizes the average risk of ASD given the evidence regarding cousins.

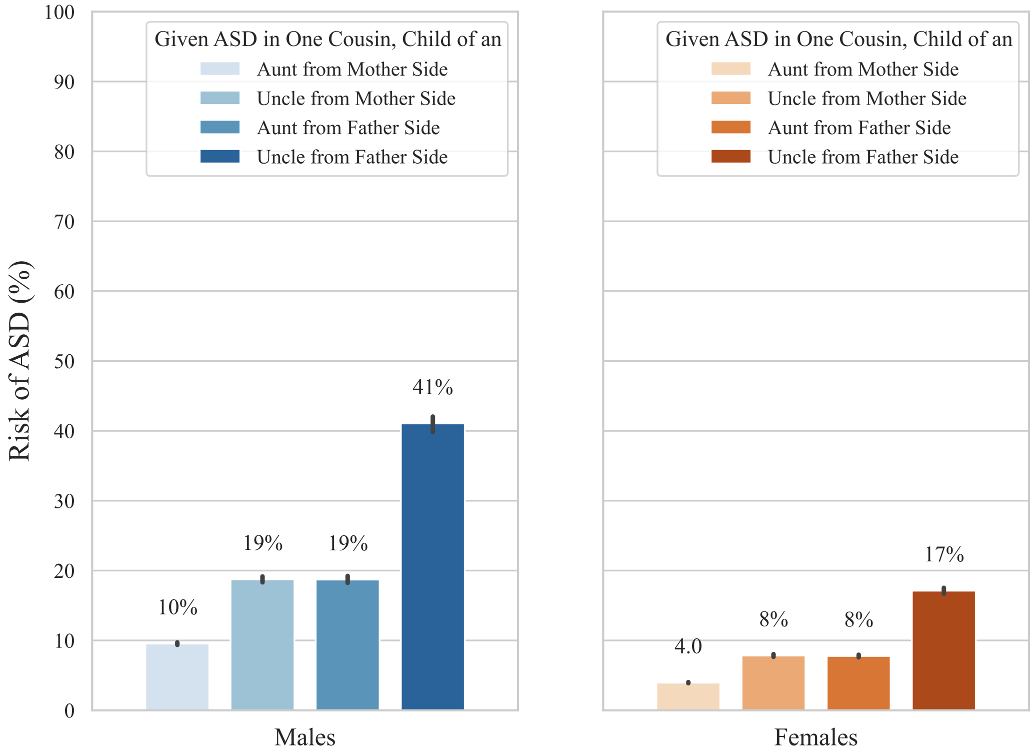


Figure 34 – Risk of ASD given evidence regarding one cousin.

Considering males reflects the ASD genotype more efficiently, and taking results from the path "`cousin`→`paternal uncle`→`paternal grandparents`→`father`", the risk of ASD tends to be:

- ≈ 17% for females and ≈ 41% for males given one cousin with ASD, regardless of being paternal or maternal (Figure 35a);

- ≈ 19% for females and ≈ 46% for males given two cousins with ASD, children of the same uncle-or-aunt (Figure 35b);

- ≈ 26% for females and ≈ 61% for males given two cousins with ASD, one cousin child of an uncle and one cousin child of an aunt, both from the father side or both from the mother side (Figure 35c); and

- ≈ 31% for females and ≈ 65% for males given two cousins with ASD, one cousin from the father side and one cousin from the mother side (estimated using the probability addition rule and the conditional probability of males in the node representing the mother - Figure 36).

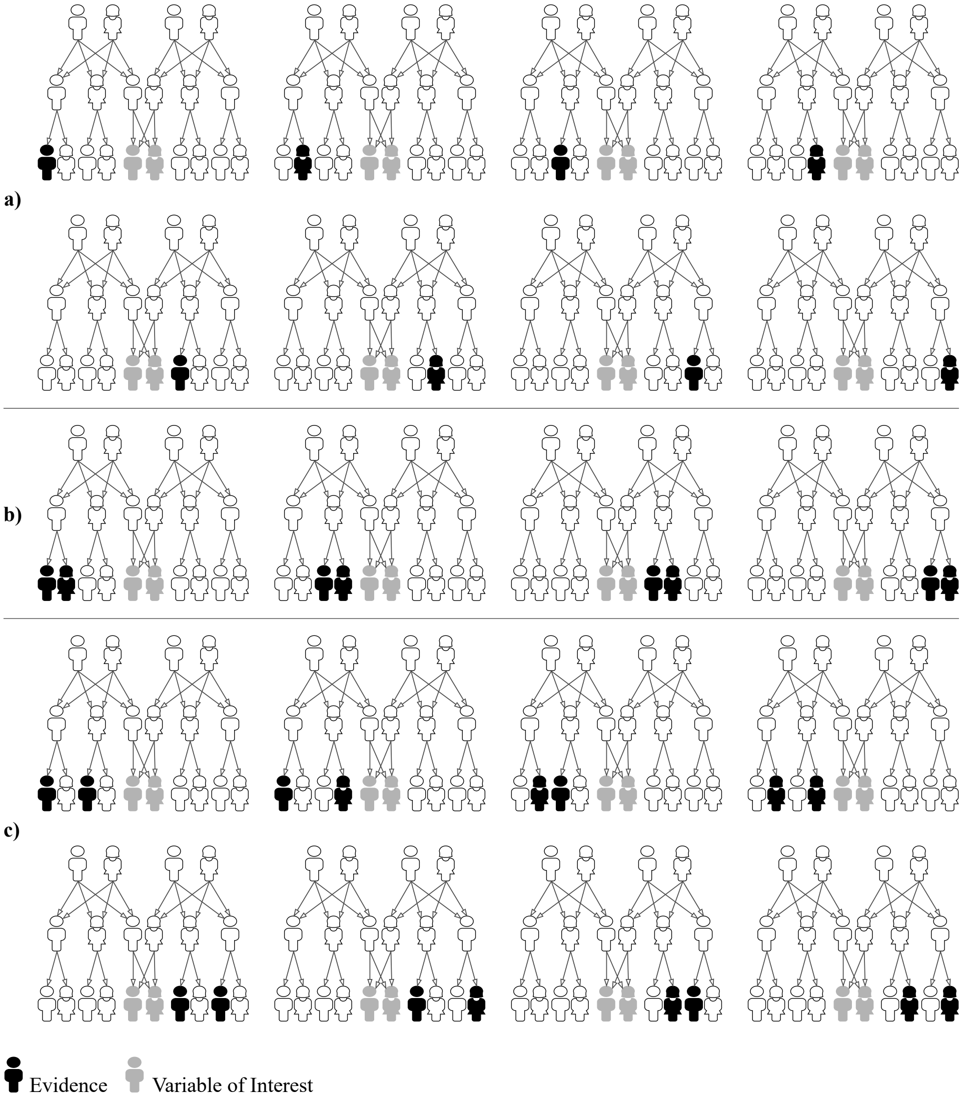


Figure 35 – Evidence setup for the risk of ASD given evidence regarding two cousins (both maternal or paternal).

7.7 Risk of ASD and the Genetic Similarity

Genetic similarity is a measure of the genetic relatedness among individuals. Human beings have 23 pairs of chromosomes, of which 22 pairs are autosomes. An autosome is any numbered chromosome that does not differ between the sexes. Autosomal DNA is inherited equally from both parents and describes DNA inherited from the autosomal chromosomes. Therefore, the statistics regarding autosomal DNA represent the amount of autosomal DNA shared between two people, describing the connection between the

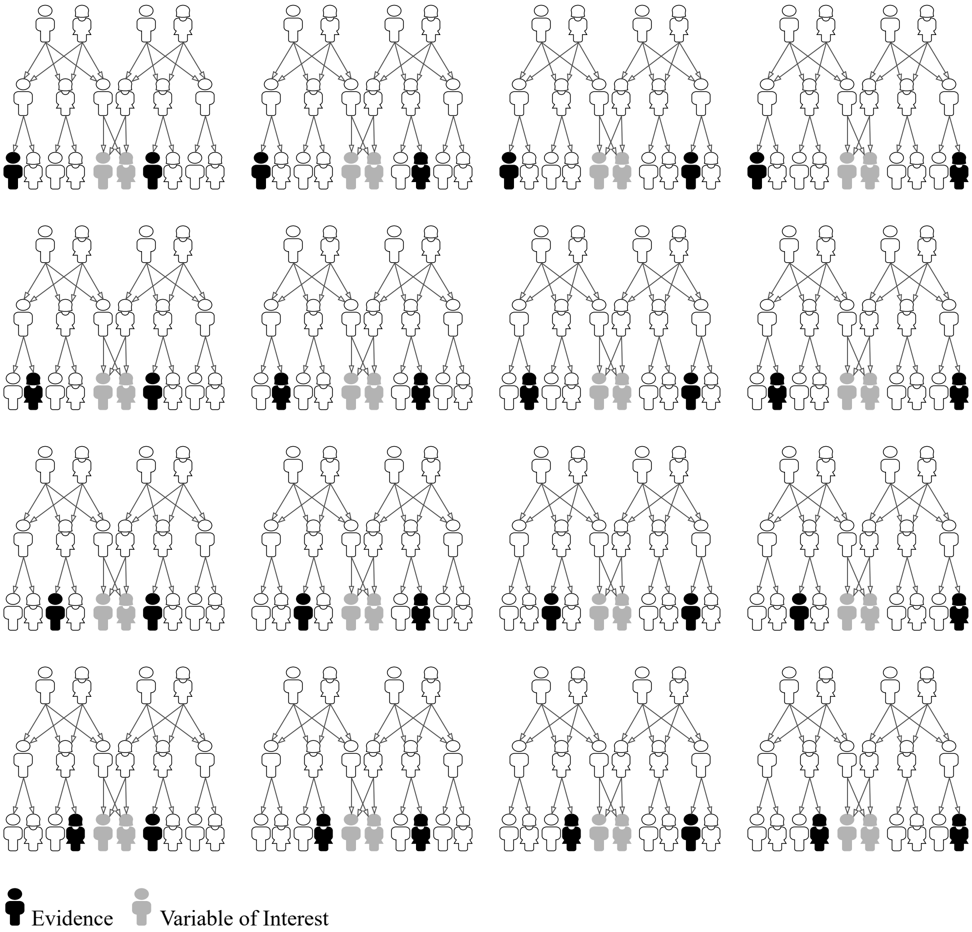


Figure 36 – Evidence setup for the risk of ASD given evidence regarding two cousins (one maternal and one paternal).

genealogical relationship. There are mathematical methods to calculate the autosomal DNA shared by two individuals. Such methods show the average amount of DNA shared by pairs of relatives in percentages (ISOGG, 2022; LEWIS, 2018).

Parents, children, and full-siblings have 50% of genetic similarity. Given one of these relatives with ASD, the risk of ASD ranges from ≈ 26% to ≈ 33% for females and ≈ 66% to ≈ 80% for males.

Grandparents, grandchildren, aunts-or-uncles, nieces-or-nephews, and half-siblings have 25% of genetic similarity. Given one of these relatives with ASD, the risk of ASD ranges from ≈ 21% to ≈ 26% for females and ≈ 51% to ≈ 62% for males.

Cousins have 12.5% of genetic similarity. Given one cousin with ASD, the risk of ASD is ≈ 17−18% for females and ≈ 41−42% for males. Once double cousins have 25% of genetic similarity, we can assume that the risk of ASD for this type of cousins is higher, approaching the estimated risk for relatives with equivalent genetic similarity.

Figure 37 summarizes the risk of ASD by genetic similarity given one relative with ASD.

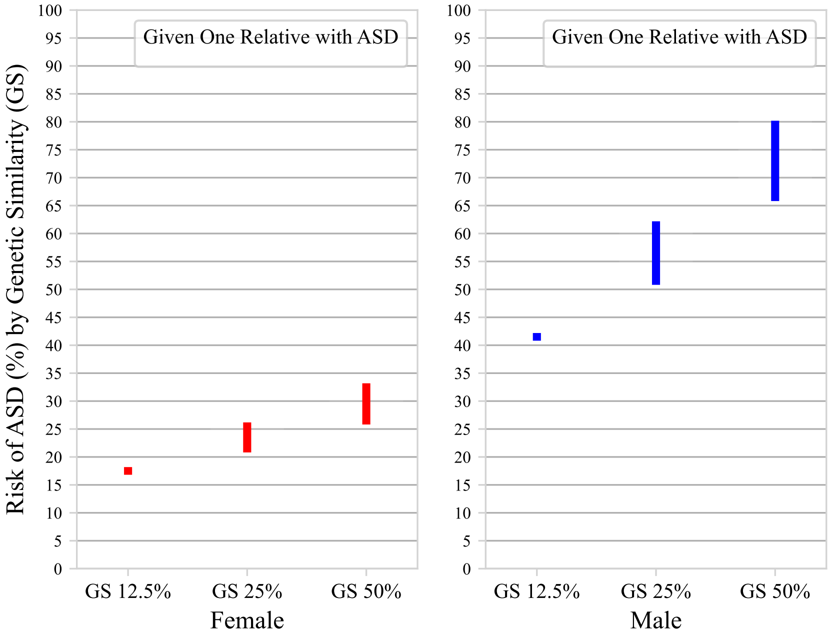


Figure 37 – Risk of ASD by genetic similarity given one relative with ASD.

These estimates increase as new evidence of relatives with ASD is exposed. Figure 38 summarizes the risk of ASD by genetic similarity given two relatives with ASD.

7.8 Results Accordance with Mental and Neurological Disorders Among Relatives

This section aims to assess our inference results regarding the risk of ASD for second-degree relatives contrasted to the literature concerning family history of mental or neurological disorders. The work carried out by Xie et al. (2019) investigated the following mental disorders: alcohol misuse, drug misuse, nonaffective psychotic disorders, bipolar disorder, depression, anxiety disorders, obsessive-compulsive disorder, stress-related disorders, other neurotic disorders, eating disorders, personality disorder, intellectual disability, attention-deficit/hyperactivity disorder, and other childhood disorders. In addition, they also examined the following neurological disorders: cerebral palsy, epilepsy, multiple sclerosis, migraine, dementia, stroke, and Parkinson's disease. These disorders shared symptomatic and possibly etiologic overlap with ASD.

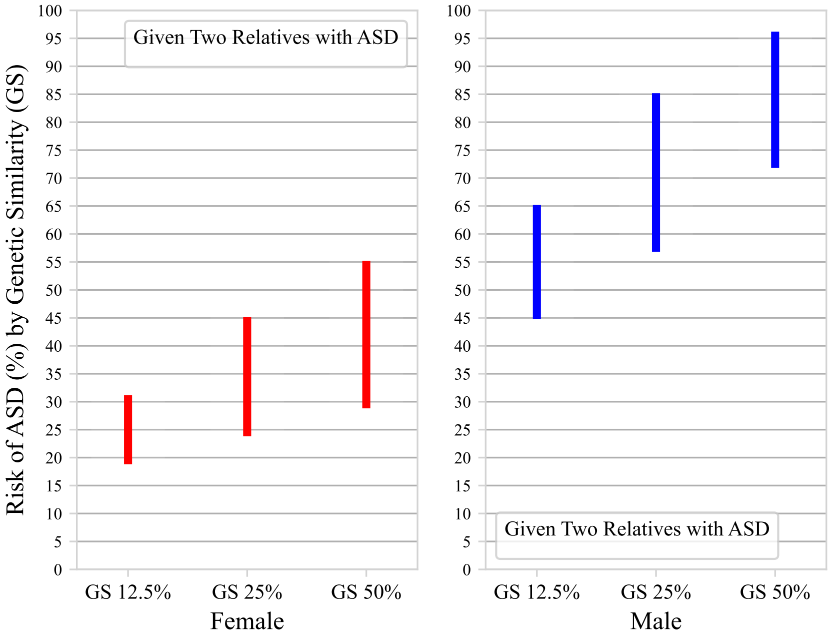


Figure 38 – Risk of ASD by genetic similarity given two relatives with ASD.

The outcomes achieved by Xie et al. (2019) in an autistic sample (71% males) show that from ≈ 80% to ≈ 90% of people with ASD have at least one grandparent with a mental or neurological condition. Let us consider this association between grandchildren and grandparents in terms of recurrence. Maintaining this same proportion of males, our results suggest an ASD recurrence of ≈ 53% given one grandparent with ASD, ≈ 62% given two grandparents with ASD (both paternal or maternal), and up to ≈ 74% given two grandparents with ASD (one paternal and one maternal).

Regarding uncles or aunts, Xie et al. (2019) point that from ≈ 49% to ≈ 61% of people with ASD have at least one uncle-or-aunt with a mental or neurological condition. Maintaining the same proportion of males, our results suggest an ASD recurrence of ≈ 44% given one nephew or one niece with ASD, ≈ 47% given one nephew and one niece with ASD (children of the same brother or sister), and up to ≈ 63% given one nephew-or-niece with ASD (child of a brother) and another nephew-or-niece with ASD (child of a sister). Our estimates concerning uncles and aunts have similar values to nephews and nieces, which we expected due to genetic similarity (25%). Therefore, our results suggest an ASD recurrence of ≈ 49% given one uncle-or-aunt with ASD, ≈ 52% given one uncle and one aunt with ASD (both from the father side or both from the mother side), and up to ≈ 70% given one paternal uncle-or-aunt with ASD and one maternal uncle-or-aunt with ASD.

Finally, Xie et al. (2019) point that from ≈ 44% to ≈ 61% of people with ASD have at least one first cousin with a mental or neurological condition. Maintaining the same proportion of males, our results suggest an ASD recurrence of ≈ 35% given one cousin with ASD, ≈ 38% given two cousins with ASD (children of the same uncle-or-aunt), ≈ 51% given two cousins with ASD, one cousin child of an uncle and one cousin child of an aunt (both from the father side or both from the mother side), and up to ≈ 55% given two cousins with ASD (one cousin from the father side and one cousin from the mother side).

The work performed by Trevis et al. (2020) assessed two large multiplex ASD families. When two or more siblings were identified with ASD/BAP traits in these families, the recurrence of ASD/BAP is widespread among their children, nephews and nieces, grandchildren, parents, uncles and aunts, and cousins, as can be noticed in Figure 39.

7.9 Summary

This chapter proposed a BN to assess the risk of ASD given evidence regarding second- and third-degree relatives. Our results produced the following contributions:

- Well reasoned estimates regarding the risk of ASD/BAP traits in second- and third-degree relatives, assuming one and two affected individuals as evidence;
- Well reasoned estimates regarding the risk of ASD/BAP traits by genetic similarity, suggesting a risk range according to such similarity.

Besides these suggested contributions, although it was not a simple and explicit comparison, some of our estimates have substantial similarities regarding the risk of ASD and the family history of mental and neurological disorders, which also supports validating the adequacy of our proposed BN model as a tool to estimate the risk of ASD in family members.

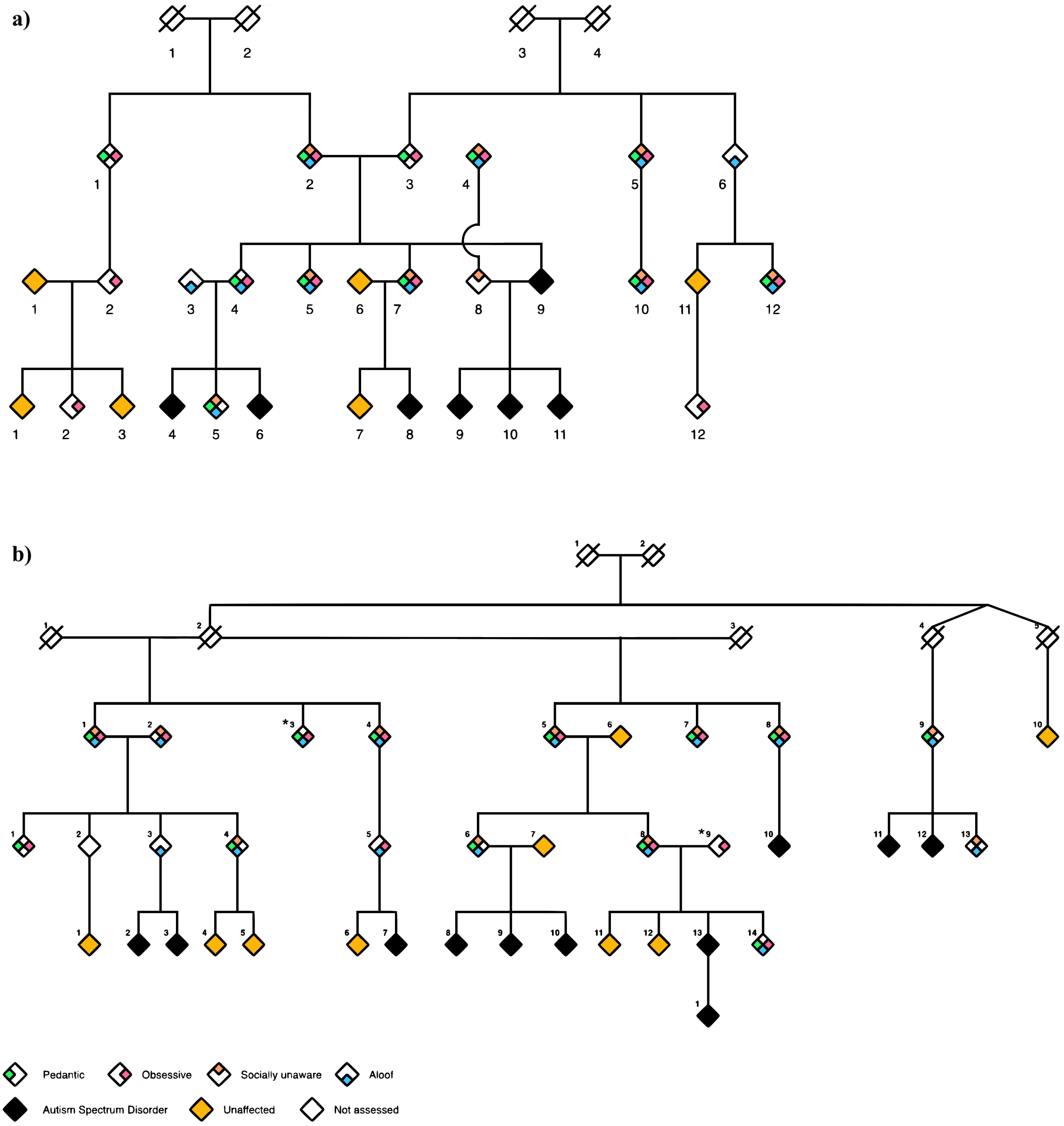


Figure 39 – ASD and BAP recurrence in two multiplex ASD families (**a** and **b**). Pedantic, obsessive, socially unaware, and aloof are BAP endophenotypes. All individuals with at least one of these endophenotypes had a BAP diagnosis. Adapted from Trevis et al. (2020).

8 Final Thoughts

The previous chapters described some essential aspects of autism concerning our work aims, including its classification, etiology, and recurrence rates. Also, we introduced the AI approaches we used to estimate the family bias to autism. Moreover, we build an HMM-based model to infer the likelihood of ASD parents generating ASD children. Our HMM results have shed light on fundamental causal probabilities to propose BN models with the potential to infer the ASD risk over a given family structure. We then proposed the manual construction of BN models representing the family tree. Finally, we used these BN models to perform several inferences regarding the risk of ASD given a set of evidence.

Our outcomes demonstrated that our BNs seem promising towards our primary objective once the central structure and the set of probabilities we used to build the models were validated by reputable statistical data regarding ASD in the literature. This chapter intends to summarize the achievements of our work, discussing its strengths, limitations, and future works.

8.1 Project Achievements

In Chapter 1, we defined some critical steps to achieve our primary objective. In this section, we discuss the achievement of these specific objectives.

As probabilistic models, the quality of Bayesian and Markov models fundamentally depends on an accurately defined set of probabilities. Thus, we needed to investigate state-of-the-art statistical information regarding ASD prevalence, heritability, and recurrence rates. In addition to supporting the models' construction, these data also helped validate our results.

The ASD prevalence data were used as probabilities for the HMMs initial states and the BNs root nodes. Regarding prevalence, we believe to be achieved adequate accuracy in Section 2.2. We used the heritability data to model the emission probabilities of our HMMs. Regarding heritability, we believe to be achieved adequate accuracy in Section 2.4.

Such HMMs were used to calculate the conditional probabilities for the BN models, specifically regarding the likelihood of autistic parents generating autistic children, which we believe to be achieved adequate accuracy in our estimates, as can be seen in Chapter 4. We also investigate and summarize statistical information regarding ASD and BAP recurrence rates among siblings, which we believe to be achieved adequate accuracy in Section 2.5.

Then, we aimed a causal probabilistic model to infer the probabilities of ASD in family members, given some evidence of ASD in the family genealogical tree. We believe we have achieved this objective in Section 4.9 once our HMMs model gave us the causal probabilities of ASD/TD parents generating ASD/TD children and in Sections 6.4, 6.5, and 6.6 in which we estimated the risk of ASD in parents, grandparents, and siblings. We evaluate the proposed BN models with the ASD heritability, prevalence, and recurrence data existing in the state-of-the-art literature, as seen in Sections 6.4.5, 6.5.3, and 6.6.3, where we discussed our results accordance.

The last objective was to introduce some ASD probabilities estimates from predefined family compositions to infer the likelihood of ASD in other family members, both below and above in the family tree. We achieved this goal in Chapter 7, in which we estimated the risk of ASD in males and females given evidence in grandparents, aunts-or-uncles, nieces-or-nephews, and cousins. Along with the results obtained in Chapter 6, these results culminated in estimating ASD risk ranges by genetic similarity, an essential measurement for risk analysis.

Thus, we believe that we achieved our main objective once:

- Our BNs were created respecting the fundamental rules required to ensure the quality of the models;
- Our BNs used state-of-the-art probabilities we obtained through in-depth literature reviews; and
- Mainly because our BN models could estimate the risk of ASD reliably.

Therefore, the answer to our research question is that *probabilistic networks are a suitable and promising approach to model the family bias to autism.* Although they provide highly relevant information, it is noteworthy that our estimates do not intend to be assertive. Such estimates must always be taken as directive information once non-deterministic models provide them.

8.2 Strengths and Limitations

Among the diverse strengths of this work, it is noteworthy that:

- The PGMs typical structure allowed to represent family structures naturally. Furthermore, this similarity enabled the assembly of the networks' syntax following best practices;
- In-depth literature reviews produced the set of probabilities that comprised the semantics of the networks;
- The adequacy of our inference results regarding de risk of ASD if compared to the well-known probabilities in the literature; and
- The option and flexibility to estimate the risk of ASD by evaluating different sets of evidence.

Among possible limitations of this work, it is noteworthy that:

- Although the probabilities employed resulted from in-depth literature reviews, these probabilities reflect the ASD phenotype. Therefore, genotype-based probabilities could lead to more accurate estimates, especially when females are involved;
- Since females require additional genetic variants to manifest the disorder in general, it is rational to assume that females with ASD diagnosis impose a higher risk of ASD to their offspring than males. However, our models do not handle this specificity once the ASD heritability data in the literature do not measure this potential difference concerning the causality; and
- The female protective effect decreases the risk of ASD in females, which is a known fact. However, given our BNs structures, such reduction in the ASD risk consequently reduces the risk of ASD in their descendants, which seems inaccurate, especially when the offspring are males. Such limitation also exists when explanation queries with females in the path are considered.

8.3 Future Works

Future works may assess the use of genotype-related probabilities, although there are no sufficient probabilistic data of this nature, if any exists. Genotype-related data could minimize the impact of the female protective effect over the estimates and deal with the causality of the disorder differently according to gender.

A dynamic system for building the network's syntax and semantics and defining the set of evidence and variables of interest could simplify the investigation of on-demand real-world scenarios, in addition to providing a straightforward process to fit the networks when new probabilities become available.

Finally, variables representing other ASD risks could be added to the networks, such as parental age or environmental factors. However, inferences using such causes depend on a robust set of conditional probabilities, which are not entirely known or available yet.

References

ALIE, D. et al. Analysis of eye gaze pattern of infants at risk of autism spectrum disorder using markov models. In: IEEE. *2011 IEEE Workshop on Applications of Computer Vision (WACV)*. [S.l.], 2011. p. 282–287. 84

ALMANDIL, N. B. et al. Environmental and genetic factors in autism spectrum disorders: Special emphasis on data from arabian studies. *International journal of environmental research and public health*, Multidisciplinary Digital Publishing Institute, v. 16, n. 4, p. 658, 2019. 21, 42

ALSHABAN, F. et al. Prevalence and correlates of autism spectrum disorder in qatar: a national study. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 60, n. 12, p. 1254–1268, 2019. 32

ALVES, F. J. et al. Applied behavior analysis for the treatment of autism: A systematic review of assistive technologies. *IEEE Access*, IEEE, v. 8, p. 118664–118672, 2020. 21, 83, 183

ALVES, F. J. et al. Robôs como suporte às intervenções baseadas em aba para o transtorno do espectro autista: uma revisão narrativa. In: FRANÇA, G.; PINHO, K. R. (Ed.). *Autismo: Tecnologias e formação de professores para a escola pública*. Palmas, TO: Nagô Editora, 2021. cap. 9, p. 125–135. 183

AMARAL, J. N. et al. About computing science research methodology. Technical report, Alberta University, 2011. 25

AMENDAH, D. et al. The economic costs of autism: A review. *Autism spectrum disorders*, Oxford University Press Oxford, UK, p. 1347–1360, 2011. 19

ANDERSEN, S. K. et al. Hugin-a shell for building bayesian belief universes for expert systems. In: *IJCAI*. [S.l.: s.n.], 1989. v. 89, n. August, p. 1080–1085. 22

ANKAN, A.; PANDA, A. pgmpy: Probabilistic graphical models using python. In: CITESEER. *Proceedings of the 14th Python in Science Conference (SCIPY 2015). Citeseer*. [S.l.], 2015. v. 10. 116

APA, A. P. A. *Diagnostic and Statistical Manual of Mental Disorders (DSM-V)*. 5. ed. Arlington VA: American Psychiatric Association Publishing, 2013. 18, 29

ASSOCIATION, A. P. *Diagnostic and Statistical Manual of Mental Disorders-DSM*. 1. ed. [S.l.]: American Psychiatric Pub, 1952. 28

ASSOCIATION, A. P. *Diagnostic and statistical manual of mental disorder-DSM*. 2. ed. [S.l.]: American Psychiatric Pub, 1968. 28

ASSOCIATION, A. P. *Diagnostic and statistical manual of mental disorder-DSM*. 3. ed. [S.l.]: American Psychiatric Pub, 1980. 28

ASSOCIATION, A. P. *Diagnostic and statistical manual of mental disorder-DSM*. 4. ed. [S.l.]: American Psychiatric Pub, 1994. 29

ATSEM, S. et al. Paternal age effects on sperm foxk1 and kcna7 methylation and transmission into the next generation. *Human molecular genetics*, Oxford University Press, v. 25, n. 22, p. 4996–5005, 2016. 39

AVEN, T. The risk concept—historical and recent development trends. *Reliability Engineering & System Safety*, Elsevier, v. 99, p. 33–44, 2012. 98, 99

BAHADO-SINGH, R. O. et al. Artificial intelligence analysis of newborn leucocyte epigenomic markers for the prediction of autism. *Brain research*, Elsevier, v. 1724, p. 146457, 2019. 103

BAI, D. et al. Association of genetic and environmental factors with autism in a 5-country cohort. *JAMA psychiatry*, 2019. 21, 32, 36, 42, 43, 45, 93, 94, 102, 104, 112, 114, 129

BAILEY, A. et al. Autism as a strongly genetic disorder: evidence from a british twin study. *Psychological medicine*, Cambridge University Press, v. 25, n. 1, p. 63–77, 1995. 42

BAIO, J. et al. Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2014. *MMWR Surveillance Summaries*, Centers for Disease Control and Prevention, v. 67, n. 6, p. 1, 2018. 19, 104

BAKER, J. K. *Stochastic modeling as a means of automatic speech recognition.* [S.l.], 1975. 79

BARON-COHEN, S. et al. Prevalence of autism-spectrum conditions: Uk school-based population study. *The British Journal of Psychiatry*, Cambridge University Press, v. 194, n. 6, p. 500–509, 2009. 19, 32

BASELMANS, B. M. et al. Risk in relatives, heritability, snp-based heritability and genetic correlations in psychiatric disorders: a review. *Biological Psychiatry*, Elsevier, v. 89, n. 1, p. 11–19, 2020. 37, 44, 45

BAUM, L. E.; EAGON, J. A. An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, v. 73, n. 3, p. 360–363, 1967. 79

BAUM, L. E.; PETRIE, T. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, JSTOR, v. 37, n. 6, p. 1554–1563, 1966. 79

BAUMAN, M. L.; KEMPER, T. L. Neuroanatomic observations of the brain in autism: a review and future directions. *International journal of developmental neuroscience*, Elsevier, v. 23, n. 2-3, p. 183–187, 2005. 35

BAXTER, A. J. et al. The epidemiology and global burden of autism spectrum disorders. *Psychological medicine*, Cambridge University Press, v. 45, n. 3, p. 601–613, 2015. 31, 32, 34

BECK, R. G. Estimativa do número de casos de transtorno do espectro autista no sul do brasil. *Programa de Pós-Graduação em Ciência da Saúde - Universidade do sul de Santa Catarina*, 2017. 34

BEDFORD, R. et al. Precursors to social and communication difficulties in infants at-risk for autism: gaze following and attentional engagement. *Journal of autism and developmental disorders*, Springer, v. 42, n. 10, p. 2208–2218, 2012. 50

BENSSASSI, E. M. et al. Wearable assistive technologies for autism: opportunities and challenges. *IEEE Pervasive Computing*, IEEE, v. 17, n. 2, p. 11–21, 2018. 83

BERG, J. et al. Action recognition in assembly for human-robot-cooperation using hidden markov models. *Procedia CIRP*, Elsevier, v. 76, p. 205–210, 2018. 23

BERNIER, R. et al. Evidence for broader autism phenotype characteristics in parents from multiple-incidence autism families. *Autism Research*, Wiley Online Library, v. 5, n. 1, p. 13–20, 2012. 52

BHAUMIK, R. et al. Predicting autism spectrum disorder using domain-adaptive cross-site evaluation. *Neuroinformatics*, Springer, v. 16, n. 2, p. 197–205, 2018. 21, 82, 83

BISWAS, S.; SINGH, A.; REDDY, C. *Block-5 Human Genetics*. [S.l.]: IGNOU, 2017. 37

BÖLTE, S.; GIRDLER, S.; MARSCHIK, P. B. The contribution of environmental exposure to the etiology of autism spectrum disorder. *Cellular and Molecular Life Sciences*, Springer, v. 76, n. 7, p. 1275–1297, 2019. 35, 36

BONIS, S. Stress and parents of children with autism: a review of literature. *Issues in mental health nursing*, Taylor & Francis, v. 37, n. 3, p. 153–163, 2016. 18, 20

BONNET, A. *Heritability Estimation in High-dimensional Mixed Models: Theory and Applications*. Tese (Doutorado) — Université Paris-Saclay, 2016. 44

BONNET, A. et al. Heritability estimation in high dimensional sparse linear mixed models. *Electronic Journal of Statistics*, The Institute of Mathematical Statistics and the Bernoulli Society, v. 9, n. 2, p. 2099–2129, 2015. 45

BOOMSMA, D.; BUSJAHN, A.; PELTONEN, L. Classical twin studies and beyond. *Nature reviews genetics*, Nature publishing group, v. 3, n. 11, p. 872–882, 2002. 44

BOURGERON, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews Neuroscience*, Nature Publishing Group, v. 16, n. 9, p. 551–563, 2015. 38

BOYLE, E. A.; LI, Y. I.; PRITCHARD, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell*, Elsevier, v. 169, n. 7, p. 1177–1186, 2017. 40, 44

BRALTEN, J. et al. Autism spectrum disorders and autistic traits share genetics and biology. *Molecular psychiatry*, Nature Publishing Group, v. 23, n. 5, p. 1205–1212, 2018. 87

Brazil's Ministry of Health. *Diretrizes de Atenção à Reabilitação da Pessoa com Transtorno do Espectro Autista (TEA)*. [S.l.]: Ministério da Saúde Brasília, 2014. 20

BREESE, J. S.; HECKERMAN, D. Decision-theoretic troubleshooting: A framework for repair and experiment. *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, Cornell University, p. 124–132, 1996. 22

BROWNING, B. L.; BROWNING, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *The American Journal of Human Genetics*, Elsevier, v. 84, n. 2, p. 210–223, 2009. 84

BRUEGGEMAN, L.; KOOMAR, T.; MICHAELSON, J. J. Forecasting risk gene discovery in autism with machine learning and genome-scale data. *Scientific reports*, Nature Publishing Group, v. 10, n. 1, p. 1–11, 2020. 83

BUESCHER, A. V. et al. Costs of autism spectrum disorders in the united kingdom and the united states. *JAMA pediatrics*, American Medical Association, v. 168, n. 8, p. 721–728, 2014. 19

BUSSU, G. et al. Prediction of autism at 3 years from behavioural and developmental measures in high-risk infants: A longitudinal cross-domain classifier analysis. *Journal of autism and developmental disorders*, Springer, p. 1–16, 2018. 21, 83

CAMADA, M. Y.; CERQUEIRA, J. J.; LIMA, A. M. N. Stereotyped gesture recognition: An analysis between hmm and svm. In: IEEE. *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*. [S.l.], 2017. p. 328–333. 84

CAMELI, C. et al. An increased burden of rare exonic variants in nrxn1 microdeletion carriers is likely to enhance the penetrance for autism spectrum disorder. *Journal of cellular and molecular medicine*, Wiley Online Library, v. 25, n. 5, p. 2459–2470, 2021. 41

CAPPÉ, O.; MOULINES, E.; RYDÉN, T. *Inference in hidden Markov models*. New York, NY: Springer Science & Business Media, 2006. 79, 82

CARVALHO, E. A. et al. Hidden markov models to estimate the probability of having autistic children. *IEEE Access*, IEEE, v. 8, p. 99540–99551, 2020. 107, 183

CASTRO, R. et al. Network tomography: Recent developments. *Statistical science*, JSTOR, p. 499–517, 2004. 22

CHARMAN, T. et al. Non-asd outcomes at 36 months in siblings at familial risk for autism spectrum disorder (asd): A baby siblings research consortium (bsrc) study. *Autism Research*, Wiley Online Library, v. 10, n. 1, p. 169–178, 2017. 49, 50

CHARNIAK, E. Bayesian networks without tears. *AI magazine*, v. 12, n. 4, p. 50–50, 1991. 66, 67, 70

CHARNIAK, E.; GOLDMAN, R. Plan recognition in stories and in life. In: *Machine Intelligence and Pattern Recognition*. [S.l.]: Elsevier, 1990. v. 10, p. 343–351. 83

CHAUHAN, S. et al. A computer-aided mfcc-based hmm system for automatic auscultation. *Computers in biology and medicine*, Elsevier, v. 38, n. 2, p. 221–233, 2008. 84

CHAWARSKA, K. et al. 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: a baby siblings research consortium study. *Journal of the American Academy of Child & Adolescent Psychiatry*, Elsevier, v. 53, n. 12, p. 1317–1327, 2014. 48, 50, 137

CHIAROTTI, F.; VENEROSI, A. Epidemiology of autism spectrum disorders: A review of worldwide prevalence estimates since 2014. *Brain Sciences*, Multidisciplinary Digital Publishing Institute, v. 10, n. 5, p. 274, 2020. 34

CHRISTENSEN, L. et al. Play and developmental outcomes in infant siblings of children with autism. *Journal of autism and developmental disorders*, Springer, v. 40, n. 8, p. 946–957, 2010. 50, 137

COLVERT, E. et al. Heritability of autism spectrum disorder in a uk population-based twin sample. *JAMA psychiatry*, American Medical Association, v. 72, n. 5, p. 415–423, 2015. 45

CONRAD, D. F. et al. Origins and functional impact of copy number variation in the human genome. *Nature*, Nature Publishing Group, v. 464, n. 7289, p. 704–712, 2010. 38

CONSTANTINO, J. N. et al. Sibling recurrence and the genetic epidemiology of autism. *American Journal of Psychiatry*, Am Psychiatric Assoc, v. 167, n. 11, p. 1349–1356, 2010. 46

CORNEW, L. et al. Atypical social referencing in infant siblings of children with autism spectrum disorders. *Journal of autism and developmental disorders*, Springer, v. 42, n. 12, p. 2611–2621, 2012. 50

COUTEUR, A. L. et al. Autism diagnostic interview: a standardized investigator-based instrument. *Journal of autism and developmental disorders*, Springer, v. 19, n. 3, p. 363–387, 1989. 49

DAMIANO, C. R. et al. What do repetitive and stereotyped movements mean for infant siblings of children with autism spectrum disorders? *Journal of autism and developmental disorders*, Springer, v. 43, n. 6, p. 1326–1335, 2013. 50

DARWICHE, A. Bayesian networks. In: *Handbook of knowledge representation*. 1. ed. United Kingdom: Elsevier, 2008. p. 467–499. 63, 75

DAWSON, G. et al. Defining the broader phenotype of autism: Genetic, brain, and behavioral perspectives. *Development and psychopathology*, Cambridge University Press, v. 14, n. 3, p. 581–611, 2002. 47

DEFRIES, J. C.; FULKER, D. W. Multiple regression analysis of twin data. *Behavior genetics*, Springer, v. 15, n. 5, p. 467–473, 1985. 43

DEKHIL, O. et al. Identifying personalized autism related impairments using resting functional mri and ados reports. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland, 2018. p. 240–248. 21, 82, 83

DEKHIL, O. et al. Using resting state functional mri to build a personalized autism diagnosis system. In: IEEE. *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. [S.l.], 2018. p. 1381–1385. 21, 82, 83

DENG, W. et al. The relationship among genetic heritability, environmental effects, and autism spectrum disorders: 37 pairs of ascertained twin study. *Journal of Child Neurology*, SAGE Publications Sage CA: Los Angeles, CA, v. 30, n. 13, p. 1794–1799, 2015. 43, 45

DICKER, R. C. et al. Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics. 2006. 31

DIETZ, P. M. et al. National and state estimates of adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, Springer, p. 1–9, 2020. 31, 32

DUDBRIDGE, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*, Public Library of Science, v. 9, n. 3, p. e1003348, 2013. 40

DURKIN, M. S. et al. Autism spectrum disorder among us children (2002–2010): socioeconomic, racial, and ethnic disparities. *American Journal of Public Health*, American Public Health Association, v. 107, n. 11, p. 1818–1826, 2017. 34

DURKIN, M. S.; WOLFE, B. L. Trends in autism prevalence in the us: A lagging economic indicator? *Journal of Autism and Developmental Disorders*, Springer, v. 50, n. 3, p. 1095–1096, 2020. 34

DVORNEK, N. C.; VENTOLA, P.; DUNCAN, J. S. Combining phenotypic and resting-state fmri data for autism classification with recurrent neural networks. In: IEEE. *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. [S.l.], 2018. p. 725–728. 21, 82, 83

DVORNEK, N. C. et al. Identifying autism from resting-state fmri using long short-term memory networks. In: SPRINGER. *International Workshop on Machine Learning in Medical Imaging*. Cham, Switzerland, 2017. p. 362–370. 21, 82, 83

DWIVEDI, A. K.; IMTIAZ, S. A.; RODRIGUEZ-VILLEGAS, E. Algorithms for automatic analysis and classification of heart sounds: a systematic review. *IEEE Access*, IEEE, v. 7, p. 8316–8345, 2018. 84

D'ABATE, L. et al. Predictive impact of rare genomic copy number variations in siblings of individuals with autism spectrum disorders. *Nature communications*, Nature Publishing Group, v. 10, n. 1, p. 1–9, 2019. 50, 51

ELSABBAGH, M. et al. Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, Wiley Online Library, v. 5, n. 3, p. 160–179, 2012. 19, 34

ELSABBAGH, M. et al. Epidemiology and clinical characterization of autism and other pervasive developmental disorders across the world: evidence, opportunities, and challenges. *Int J Epidemiol*, 2010. 34

ELSABBAGH, M.; JOHNSON, M. H. Getting answers from babies about autism. *Trends in cognitive sciences*, Elsevier, v. 14, n. 2, p. 81–87, 2010. 19

EMERSON, R. W. et al. Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. *Science translational medicine*, American Association for the Advancement of Science, v. 9, n. 393, p. eaag2882, 2017. 20, 21, 82, 83

FAHAD, H. et al. Microscopic abnormality classification of cardiac murmurs using anfis and hmm. *Microscopy research and technique*, Wiley Online Library, v. 81, n. 5, p. 449–457, 2018. 84

FEENBERG, A. What is philosophy of technology? In: *Defining technological literacy*. [S.l.]: Springer, 2006. p. 5–16. 21, 57

FOLSTEIN, S.; RUTTER, M. Infantile autism: a genetic study of 21 twin pairs. *Journal of Child psychology and Psychiatry*, Wiley Online Library, v. 18, n. 4, p. 297–321, 1977. 42

FOMBONNE, E. The rising prevalence of autism. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 59, n. 7, p. 717–720, 2018. 31

FRAZIER, T. W. et al. A twin study of heritable and shared environmental contributions to autism. *Journal of autism and developmental disorders*, Springer, v. 44, n. 8, p. 2013–2025, 2014. 36, 43, 45

FRIEDMAN, N. et al. Using bayesian networks to analyze expression data. *Journal of computational biology*, Mary Ann Liebert, Inc., v. 7, n. 3-4, p. 601–620, 2000. 83

FU, W. et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*, Nature Publishing Group, v. 493, n. 7431, p. 216–220, 2013. 39

GANGI, D. N. et al. Measuring social-communication difficulties in school-age siblings of children with autism spectrum disorder: Standardized versus naturalistic assessment. *Autism Research*, Wiley Online Library, 2021. 47, 48, 52

GARDENER, H.; SPIEGELMAN, D.; BUKA, S. L. Perinatal and neonatal risk factors for autism: a comprehensive meta-analysis. *Pediatrics*, Am Acad Pediatrics, v. 128, n. 2, p. 344–355, 2011. 35

GAUGLER, T. et al. Most genetic risk for autism resides with common variation. *Nature genetics*, Nature Publishing Group, v. 46, n. 8, p. 881–885, 2014. 35, 39, 42, 43, 45

GELLMAN, M. D.; TURNER, J. R. et al. *Encyclopedia of behavioral medicine*. New York, NY: Springer New York, 2013. 35

GENESERETH, M. R.; NILSSON, N. J. *Logical foundations of artificial intelligence*. [S.l.]: Morgan Kaufmann, 2012. 21, 55

GEORGE, E. I. The variable selection problem. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 95, n. 452, p. 1304–1308, 2000. 71

GERDTS, J. A. et al. The broader autism phenotype in simplex and multiplex families. *Journal of autism and developmental disorders*, Springer, v. 43, n. 7, p. 1597–1605, 2013. 51, 138

GHAHRAMANI, Z. An introduction to hidden markov models and bayesian networks. In: *Hidden Markov models: applications in computer vision*. Singapore: World Scientific Publishing, 2001. p. 9–41. 23

GIRAULT, J. B. et al. Quantitative trait variation in asd probands and toddler sibling outcomes at 24 months. *Journal of neurodevelopmental disorders*, Springer New York, v. 12, n. 1, 2020. 46, 47

GIRIRAJAN, S.; CAMPBELL, C. D.; EICHLER, E. E. Human copy number variation and complex genetic disease. *Annual review of genetics*, Annual Reviews, v. 45, p. 203–226, 2011. 38

GLIGA, T. et al. Spontaneous belief attribution in younger siblings of children on the autism spectrum. *Developmental psychology*, American Psychological Association, v. 50, n. 3, p. 903, 2014. 50, 51

GOLDMAN, R. *A Probabilistic Approach to Language Understanding," Department of Computer Science.* [S.l.], 1991. 83

GOMES, P. T. et al. Autism in brazil: a systematic review of family challenges and coping strategies. *Jornal de Pediatria (Versão em Português)*, Elsevier, v. 91, n. 2, p. 111–121, 2015. 20

GOODMAN, R. et al. The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, Cambridge University Press, v. 41, n. 5, p. 645–655, 2000. 50

GÓRRIZ, J. M. et al. A machine learning approach to reveal the neurophenotypes of autisms. *International journal of neural systems*, World Scientific, v. 29, n. 07, p. 1850058, 2019. 83

GRINSTEAD, C. M.; SNELL, J. L. *Introduction to probability.* 2. ed. Rhode Island, USA: American Mathematical Society, 1998. 77

GRØNBORG, T. K.; SCHENDEL, D. E.; PARNER, E. T. Recurrence of autism spectrum disorders in full- and half-siblings and trends over time: a population-based cohort study. *JAMA pediatrics*, American Medical Association, v. 167, n. 10, p. 947–953, 2013. 31, 32, 46, 47, 137, 138

GROVE, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics*, Nature Publishing Group, v. 51, n. 3, p. 431–444, 2019. 20, 21

GUO, K. et al. Eeg-based emotion classification using innovative features and combined svm and hmm classifier. In: IEEE. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* [S.l.], 2017. p. 489–492. 84

HALLMAYER, J. et al. Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of general psychiatry*, American Medical Association, v. 68, n. 11, p. 1095–1102, 2011. 20, 21, 35, 36, 42, 43, 45

HANSEN, S. N. et al. Recurrence risk of autism in siblings and cousins: A multinational, population-based study. *Journal of the American Academy of Child & Adolescent Psychiatry*, Elsevier, v. 58, n. 9, p. 866–875, 2019. 19, 32, 46, 47, 105, 137, 140

HANSEN, S. N.; SCHENDEL, D. E.; PARNER, E. T. Explaining the increase in the prevalence of autism spectrum disorders: the proportion attributable to changes in reporting practices. *JAMA pediatrics*, American Medical Association, v. 169, n. 1, p. 56–62, 2015. 34

HAZLETT, H. C. et al. Early brain development in infants at high risk for autism spectrum disorder. *Nature*, Nature Publishing Group, v. 542, n. 7641, p. 348, 2017. 20, 21, 82, 83

HECKERMAN, D. Probabilistic similarity networks. *Networks*, Wiley Online Library, v. 20, n. 5, p. 607–636, 1990. 22

HECKERMAN, D. A tutorial on learning with bayesian networks. *Innovations in Bayesian networks*, Springer, p. 33–82, 2008. 70

HECKERMAN, D.; GEIGER, D.; CHICKERING, D. M. Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, Springer, v. 20, n. 3, p. 197–243, 1995. 70

HECKERMAN, D.; HORVITZ, E.; NATHWANI, B. *Toward Normative Expert Systems: Part I, the Pathfinder Project. Knowledge Systems Laboratory, Medical Computer Science.* [S.l.]: Stanford University, 1992. 83

HEINSFELD, A. S. et al. Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: Clinical*, Elsevier, v. 17, p. 16–23, 2018. 21, 82, 83

HERTZ-PICCIOTTO, I.; DELWICHE, L. The rise in autism and the role of age at diagnosis. *Epidemiology (Cambridge, Mass.)*, NIH Public Access, v. 20, n. 1, p. 84, 2009. 34

HILKER, R. et al. Heritability of schizophrenia and schizophrenia spectrum based on the nationwide danish twin register. *Biological psychiatry*, Elsevier, v. 83, n. 6, p. 492–498, 2018. 45

HINBEST, C.; CHMILIAR, L. Autism as a global challenge: Examining the increased childhood prevalence of autism. *Journal of Student Research*, v. 10, n. 1, 2021. 34

HOANG, N.; CYTRYNBAUM, C.; SCHERER, S. W. Communicating complex genomic information: A counselling approach derived from research experience with autism spectrum disorder. *Patient education and counseling*, Elsevier, v. 101, n. 2, p. 352–361, 2018. 41, 142

HODGES, H.; FEALKO, C.; SOARES, N. Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. *Translational Pediatrics*, AME Publications, v. 9, n. Suppl 1, p. S55, 2020. 36

HOEKSTRA, R. A. et al. Heritability of autistic traits in the general population. *Archives of pediatrics & adolescent medicine*, American Medical Association, v. 161, n. 4, p. 372–377, 2007. 43, 45

HOFFMANN, T. J. et al. Evidence of reproductive stoppage in families with autism spectrum disorder: a large, population-based cohort study. *JAMA psychiatry*, American Medical Association, v. 71, n. 8, p. 943–951, 2014. 46, 47, 137

HOLMES, D. E.; JAIN, L. C. Introduction to bayesian networks. In: *Innovations in Bayesian Networks*. [S.l.]: Springer, 2008. p. 1–5. 22, 63

HOPPER, J. Variance components for statistical genetics: applications in medical research to characteristics related to human diseases and health. *Statistical Methods in Medical Research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 2, n. 3, p. 199–223, 1993. 44

HORVITZ, E. J.; BARRY, M. Display of information for time-critical decision making. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Cornell University, p. 296–305, 1995. 22

HORVITZ, E. J. et al. The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Cornell University, p. 256–265, 1998. 22

HOWIE, B. N.; DONNELLY, P.; MARCHINI, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics*, Public Library of Science, v. 5, n. 6, p. e1000529, 2009. 84

HUDRY, K. et al. Early language profiles in infants at high-risk for autism spectrum disorders. *Journal of autism and developmental disorders*, Springer, v. 44, n. 1, p. 154–167, 2014. 50, 137

HUGUET, G.; BOURGERON, T. Genetic causes of autism spectrum disorders. In: *Neuronal and synaptic dysfunction in autism spectrum disorder and intellectual disability.* [S.l.]: Elsevier, 2016. p. 13–24. 39, 40

HUTMAN, T. et al. Selective visual attention at twelve months: Signs of autism in early social interactions. *Journal of autism and developmental disorders*, Springer, v. 42, n. 4, p. 487–498, 2012. 50, 51

IAKOUCHEVA, L. M.; MUOTRI, A. R.; SEBAT, J. Getting to the cores of autism. *Cell*, Elsevier, v. 178, n. 6, p. 1287–1298, 2019. 20, 21, 38, 40

IBGE, I. B. de Geografia e E. *População Brasileira Estimada*. 2020. Accessed: February 12, 2021. Disponível em: <https://www.ibge.gov.br/cidades-e-estados.html?view=municipio>. 35

IDRING, S. et al. Autism spectrum disorders in the stockholm youth cohort: design, prevalence and validity. *PloS one*, Public Library of Science, v. 7, n. 7, 2012. 32

IOSSIFOV, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, Nature Publishing Group, v. 515, n. 7526, p. 216–221, 2014. 39

ISOGG, I. S. of G. G. *Autosomal DNA statistics*. 2022. Accessed: February 12, 2022. Disponível em: <https://isogg.org/wiki/Autosomal_DNA_statistics>. 150

JACQUEMONT, S. et al. A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. *The American Journal of Human Genetics*, Elsevier, v. 94, n. 3, p. 415–425, 2014. 40

JAMIL, R.; GRAGG, M. N.; DEPAPE, A.-M. The broad autism phenotype: Implications for empathy and friendships in emerging adults. *Personality and Individual Differences*, Elsevier, v. 111, p. 199–204, 2017. 48

JANVIER, Y. M. et al. Screening for autism spectrum disorder in underserved communities: Early childcare providers as reporters. *Autism*, SAGE Publications Sage UK: London, England, v. 20, n. 3, p. 364–373, 2016. 33

JELINEK, F. *Statistical methods for speech recognition*. [S.l.]: MIT press, 1997. 77

JONES, E. J. et al. Developmental pathways to autism: a review of prospective studies of infants at risk. *Neuroscience & Biobehavioral Reviews*, Elsevier, v. 39, p. 1–33, 2014. 48, 49, 52

KANNER, L. et al. Autistic disturbances of affective contact. *Nervous child*, v. 2, n. 3, p. 217–250, 1943. 18, 28

KELLERMAN, A. et al. Dyadic interactions in children exhibiting the broader autism phenotype: Is the broader autism phenotype distinguishable from typical development? *Autism Research*, Wiley Online Library, v. 12, n. 3, p. 469–481, 2019. 48, 50, 52

KHOSLA, M. et al. 3d convolutional neural networks for classification of functional connectomes. *arXiv preprint arXiv:1806.04209*, 2018. 21, 82, 83

KIM, Y. S. et al. Prevalence of autism spectrum disorders in a total population sample. *American Journal of Psychiatry*, Am Psychiatric Assoc, v. 168, n. 9, p. 904–912, 2011. 19, 32

KING, M.; BEARMAN, P. Diagnostic change and the increased prevalence of autism. *International journal of epidemiology*, Oxford University Press, v. 38, n. 5, p. 1224–1234, 2009. 34

KJAERULFF, U. B.; MADSEN, A. L. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. 2. ed. New York NY: Springer, 2013. 9, 22, 58, 59, 63, 65, 70, 71, 72, 73, 107, 108, 111

KLIR, G. J. Uncertainty and information: foundations of generalized information theory. *Kybernetes*, Emerald Group Publishing Limited, 2006. 21, 57

KOGAN, M. D. et al. A national profile of the health care experiences and family impact of autism spectrum disorder among children in the united states, 2005–2006. *Pediatrics*, Am Acad Pediatrics, v. 122, n. 6, p. e1149–e1158, 2008. 19

KOGAN, M. D. et al. The prevalence of parent-reported autism spectrum disorder among us children. *Pediatrics*, Am Acad Pediatrics, v. 142, n. 6, 2018. 19, 32

KOLLER, D.; FRIEDMAN, N. *Probabilistic graphical models: principles and techniques*. Cambridge, MA: The MIT press, 2009. 58, 59, 60, 63, 65, 66, 74, 76, 80

KONG, A. et al. Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, Nature Publishing Group, v. 488, n. 7412, p. 471–475, 2012. 39

KROGH, A.; MIAN, I. S.; HAUSSLER, D. A hidden markov model that finds genes in e.coli dna. *Nucleic Acids Research*, Oxford University Press, v. 22, n. 22, p. 4768–4778, 1994. 23, 84

KRONCKE, A. P.; WILLARD, M.; HUCKABEE, H. The causes of autism. In: *Assessment of Autism Spectrum Disorder*. [S.l.]: Springer, 2016. p. 11–21. 21, 42

LANDA, R. J. Efficacy of early interventions for infants and young children with, and at risk for, autism spectrum disorders. *International Review of Psychiatry*, Taylor & Francis, v. 30, n. 1, p. 25–39, 2018. 20

LANDA, R. J. et al. Latent class analysis of early developmental trajectory in baby siblings of children with autism. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 53, n. 9, p. 986–996, 2012. 50

LE, D.-H.; VAN, N. T. Meta-analysis of whole-transcriptome data for prediction of novel genes associated with autism spectrum disorder. In: *Proceedings of the 8th International Conference on Computational Systems-Biology and Bioinformatics*. [S.l.: s.n.], 2017. p. 56–61. 83

LEE, J.-G. et al. Deep learning in medical imaging: general overview. *Korean journal of radiology*, v. 18, n. 4, p. 570–584, 2017. 82

LEE, L. I. et al. The current state of artificial intelligence in medical imaging and nuclear medicine. *BJR Open*, The British Institute of Radiology., v. 1, p. 20190037, 2019. 82

LEEKAM, S. R. et al. The diagnostic interview for social and communication disorders: algorithms for icd-10 childhood autism and wing and gould autistic spectrum disorder. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 43, n. 3, p. 327–342, 2002. 29

LEIGH, J. P.; DU, J. Brief report: Forecasting the economic burden of autism in 2015 and 2025 in the united states. *Journal of autism and developmental disorders*, Springer, v. 45, n. 12, p. 4135–4139, 2015. 19

LEVITT, T. S.; AGOSTA, J. M.; BINFORD, T. O. Model-based influence diagrams for machine vision. In: *Machine Intelligence and Pattern Recognition*. [S.l.]: Elsevier, 1990. v. 10, p. 371–388. 83

LEVY, D. et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*, Elsevier, v. 70, n. 5, p. 886–897, 2011. 40

LEWIS, D. R. *Human genetics: concepts and applications*. [S.l.]: McGraw-Hill Education, 2018. ISBN 9781259700934. 37, 38, 39, 150

LI, Y. et al. Mach: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology*, Wiley Online Library, v. 34, n. 8, p. 816–834, 2010. 84

LIAO, D.; LU, H. Classify autism and control based on deep learning and community structure on resting-state fmri. In: IEEE. *Advanced Computational Intelligence (ICACI), 2018 Tenth International Conference on*. [S.l.], 2018. p. 289–294. 21, 82, 83

LIN, H.-C. et al. Developmental and mental health risks among siblings of patients with autism spectrum disorder: a nationwide study. *European Child & Adolescent Psychiatry*, Springer, p. 1–6, 2021. 52

LIN, Y. et al. A machine learning approach to predicting autism risk genes: Validation of known genes and discovery of new candidates. *bioRxiv*, Cold Spring Harbor Laboratory, p. 463547, 2018. 83

LINDSTRÖM, L. S. et al. Etiology of familial aggregation in melanoma and squamous cell carcinoma of the skin. *Cancer Epidemiology and Prevention Biomarkers*, AACR, v. 16, n. 8, p. 1639–1643, 2007. 44

LITJENS, G. et al. A survey on deep learning in medical image analysis. *Medical image analysis*, Elsevier, v. 42, p. 60–88, 2017. 82

LITTLE, T. D. *The Oxford handbook of quantitative methods*. [S.l.]: Oxford University Press, USA, 2014. v. 2. 44

LOMBARDO, M. V.; LAI, M.-C.; BARON-COHEN, S. Big data approaches to decomposing heterogeneity across the autism spectrum. *Molecular psychiatry*, Nature Publishing Group, v. 24, n. 10, p. 1435–1450, 2019. 83

LOOMES, R.; HULL, L.; MANDY, W. P. L. What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, Elsevier, v. 56, n. 6, p. 466–474, 2017. 19

LORD, C. et al. The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, Springer, v. 30, n. 3, p. 205–223, 2000. 51

LORD, C. et al. Autism diagnostic observation schedule. *Western Psychological Services*, v. 12031, p. 90025–1251, 1989. 51

LOSH, M. et al. Neuropsychological profile of autism and the broad autism phenotype. *Archives of general psychiatry*, American Medical Association, v. 66, n. 5, p. 518–526, 2009. 47

LOSH, M. et al. Defining key features of the broad autism phenotype: A comparison across parents of multiple-and single-incidence autism families. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, Wiley Online Library, v. 147, n. 4, p. 424–433, 2008. 129, 138

LUCAS, H. M. et al. Parental perceptions of genetic testing for children with autism spectrum disorders. *American Journal of Medical Genetics Part A*, Wiley Online Library, 2021. 41

LUNDSTRÖM, S. et al. Autism spectrum disorders and autisticlike traits: similar etiology in the extreme end and the normal variation. *Archives of general psychiatry*, American Medical Association, v. 69, n. 1, p. 46–52, 2012. 87

LVOVS, D.; FAVOROVA, O.; FAVOROV, A. A polygenic approach to the study of polygenic diseases. *Acta Naturae*, v. 4, n. 3, 2012. 39

LYALL, K. et al. The changing epidemiology of autism spectrum disorders. *Annual review of public health*, Annual Reviews, v. 38, p. 81–102, 2017. 20, 35, 36

MACARI, S. L. et al. Predicting developmental status from 12 to 24 months in infants at risk for autism spectrum disorder: A preliminary report. *Journal of autism and developmental disorders*, Springer, v. 42, n. 12, p. 2636–2647, 2012. 50, 51, 137

MAENNER, M. J. et al. Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2016. *MMWR Surveillance Summaries*, Centers for Disease Control and Prevention, v. 69, n. 4, p. 1, 2020. 18, 32, 38, 88

MAENNER, M. J. et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2018. *MMWR Surveillance Summaries*, Centers for Disease Control and Prevention, v. 70, n. 11, p. 1, 2021. 18, 34

MANOGARAN, G. et al. Machine learning based big data processing framework for cancer diagnosis using hidden markov model and gm clustering. *Wireless personal communications*, Springer, v. 102, n. 3, p. 2099–2116, 2018. 84

MARCHINI, J.; HOWIE, B. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, Nature Publishing Group, v. 11, n. 7, p. 499, 2010. 84

MARCHINI, J. et al. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics*, Nature Publishing Group, v. 39, n. 7, p. 906, 2007. 84

MASI, A. et al. An overview of autism spectrum disorder, heterogeneity and treatment options. *Neuroscience bulletin*, Springer, v. 33, n. 2, p. 183–193, 2017. 20

MATSON, J. L.; KOZLOWSKI, A. M. The increasing prevalence of autism spectrum disorders. *Research in Autism Spectrum Disorders*, Elsevier, v. 5, n. 1, p. 418–425, 2011. 34

MATTILA, M.-L. et al. An epidemiological and diagnostic study of asperger syndrome according to four sets of diagnostic criteria. *Journal of the American Academy of Child & Adolescent Psychiatry*, Elsevier, v. 46, n. 5, p. 636–646, 2007. 34

MESSINGER, D. et al. Beyond autism: a baby siblings research consortium study of high-risk children at three years of age. *Journal of the American Academy of Child & Adolescent Psychiatry*, Elsevier, v. 52, n. 3, p. 300–308, 2013. 49, 50

MESSINGER, D. S. et al. Early sex differences are not autism-specific: A baby siblings research consortium (bsrc) study. *Molecular autism*, Springer, v. 6, n. 1, p. 1–12, 2015. 46, 104, 138

MEYER, I. M.; DURBIN, R. Comparative ab initio prediction of gene structures using pair hmms. *Bioinformatics*, Oxford University Press, v. 18, n. 10, p. 1309–1318, 2002. 23, 84

MICHAELSON, J. J. et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell*, Elsevier, v. 151, n. 7, p. 1431–1442, 2012. 39

MICHIELS, M.; LARRAÑAGA, P.; BIELZA, C. Bayesuites: An open web framework for massive bayesian networks focused on neuroscience. *Neurocomputing*, Elsevier, v. 428, p. 166–181, 2021. 116

MILLER, M. et al. Early pragmatic language difficulties in siblings of children with autism: Implications for dsm-5 social communication disorder? *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 56, n. 7, p. 774–781, 2015. 49, 50

MORJARIA, M.; SANTOSA, F. Monitoring complex systems with causal networks. *IEEE Computational Science and Engineering*, IEEE, v. 3, n. 4, p. 9–10, 1996. 22

MRAD, A. B. et al. An explication of uncertain evidence in bayesian networks: likelihood evidence and probabilistic evidence. *Applied Intelligence*, SPRINGER VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS, v. 43, n. 4, p. 802–824, 2015. 74

MULLEN, E. M. et al. *Mullen Scales of Early Learning*. [S.l.]: AGS Circle Pines, MN, 1995. 51

MUSTAFA, M. K.; ALLEN, T.; APPIAH, K. A comparative review of dynamic neural networks and hidden markov model methods for mobile on-device speech recognition. *Neural Computing and Applications*, Springer, v. 31, n. 2, p. 891–899, 2019. 23

NEALE, B. Liability threshold models. *Encyclopedia of Statistics in Behavioral Science*, Wiley Online Library, 2005. 44

NEALE, M.; CARDON, L. R. *Methodology for genetic studies of twins and families*. [S.l.]: Springer Science & Business Media, 2013. v. 67. 42

NEIL, M.; FENTON, N.; NIELSON, L. Building large-scale bayesian networks. *The Knowledge Engineering Review*, Cambridge University Press, v. 15, n. 3, p. 257–284, 2000. 22, 63, 72, 73, 108

NEVISON, C. D. A comparison of temporal trends in united states autism prevalence to trends in suspected environmental factors. *Environmental Health*, BioMed Central, v. 13, n. 1, p. 73, 2014. 34

NICHOLS, C. M. et al. Social smiling and its components in high-risk infant siblings without later asd symptomatology. *Journal of autism and developmental disorders*, Springer, v. 44, n. 4, p. 894–902, 2014. 50

NIELSEN, T. D.; JENSEN, F. V. *Bayesian networks and decision graphs*. 2. ed. New York, NY: Springer Science & Business Media, 2009. 66, 67, 68, 69

NOH, M. et al. Multicomponent variance estimation for binary traits in family-based studies. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, Wiley Online Library, v. 30, n. 1, p. 37–47, 2006. 44

OFNER, M. et al. *Autism spectrum disorder among children and youth in Canada 2018*. [S.l.]: Public Health Agency of Canada Ottawa, 2018. 31, 32

ONAOLAPO, A.; ONAOLAPO, O. Global data on autism spectrum disorders prevalence: A review of facts, fallacies and limitations. *Universal Journal of Clinical Medicine*, v. 5, n. 2, p. 14–23, 2017. 33

ORGANIZATION, W. H. et al. International statistical classification of diseases and related health problems: 10th revision (icd-10). *http://www.who.int/classifications/apps/icd/icd*, 1992. 29

ORGANIZATION, W. H. et al. Icd-11 for mortality and morbidity statistics. *Retrieved June*, v. 22, p. 2018, 2018. 30

ORGANIZATION, W. W. H. *Autism spectrum disorders*. 2019. Accessed: February 12, 2021. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>. 35

OUSLEY, O.; CERMAK, T. Autism spectrum disorder: defining dimensions and subgroups. *Current developmental disorders reports*, Springer, v. 1, n. 1, p. 20–28, 2014. 29

ÖZERK, K. The issue of prevalence of autism/asd. *International Electronic Journal of Elementary Education*, ERIC, v. 9, n. 2, p. 263–306, 2016. 18, 33, 34

OZONOFF, S. et al. The broader autism phenotype in infancy: when does it emerge? *Journal of the American Academy of Child & Adolescent Psychiatry*, Elsevier, v. 53, n. 4, p. 398–407, 2014. 49, 50, 137

OZONOFF, S. et al. Recurrence risk for autism spectrum disorders: a baby siblings research consortium study. *Pediatrics*, Am Acad Pediatrics, v. 128, n. 3, p. e488–e495, 2011. 46

OZTOK, U.; CHOI, A.; DARWICHE, A. Solving pppp-complete problems using knowledge compilation. In: *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning*. [S.l.: s.n.], 2016. p. 94–103. 75

O'NEILL, D. Language use inventory. *Ontario, Canada: Knowledge in Development*, 2009. 51

PALMER, C. J.; LAWSON, R. P.; HOHWY, J. Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological bulletin*, American Psychological Association, v. 143, n. 5, p. 521, 2017. 21, 83

PALMER, N. et al. Association of sex with recurrence of autism spectrum disorder among siblings. *JAMA pediatrics*, American Medical Association, v. 171, n. 11, p. 1107–1112, 2017. 9, 32, 38, 46, 47, 88, 89, 94, 96, 100, 101, 102, 103, 104

PAPADIMITRIOU, C. H. *Computational Complexity*. [S.l.]: Addison-Wesley, 1994. 75

PAUL, R. et al. Out of the mouths of babes: Vocal production in infant siblings of children with asd. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 52, n. 5, p. 588–598, 2011. 50

PAULA, C. S. et al. Brief report: prevalence of pervasive developmental disorder in brazil: a pilot study. *Journal of autism and developmental disorders*, Springer, v. 41, n. 12, p. 1738–1742, 2011. 34

PAWITAN, Y. et al. Estimation of genetic and environmental factors for binary traits using family data. *Statistics in medicine*, Wiley Online Library, v. 23, n. 3, p. 449–465, 2004. 45

PEARL, J. Fusion, propagation, and structuring in belief networks. *Artificial intelligence*, Elsevier, v. 29, n. 3, p. 241–288, 1986. 57

PEARL, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1. ed. San Francisco CA: Morgan Kaufmann, 1988. 22, 58, 74

PEARL, J. *Causality: Models, Reasoning, and Inference.* 2. ed. New York, NY: Cambridge University Press, 2009. 65, 66

PEARSON, K.; LEE, A. Mathematical contributions to the theory of evolution. viii. on the inheritance of characters not capable of exact quantitative measurement. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, v. 195, p. 79–150, 1900. 44

PEYRARD, N. et al. Exact or approximate inference in graphical models: why the choice is dictated by the treewidth, and how variable elimination can be exploited. *Australian & New Zealand Journal of Statistics*, Wiley Online Library, v. 61, n. 2, p. 89–133, 2019. 76

PINTO, D. et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, Nature Publishing Group, v. 466, n. 7304, p. 368–372, 2010. 52

PISULA, E.; ZIEGART-SADOWSKA, K. Broader autism phenotype in siblings of children with asd — a review. *International journal of molecular sciences*, Multidisciplinary Digital Publishing Institute, v. 16, n. 6, p. 13217–13258, 2015. 48, 49, 51, 52

POURRET, O.; NAÏM, P.; MARCOT, B. *Bayesian networks: a practical guide to applications.* [S.l.]: John Wiley & Sons, 2008. 22

QUINLAN, C. A. et al. Parental age and autism spectrum disorders among new york city children 0–36 months of age. *Maternal and child health journal*, Springer, v. 19, n. 8, p. 1783–1790, 2015. 36

RABINER, L. R. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, Ieee, v. 77, n. 2, p. 257–286, 1989. 22, 77, 79, 82

RAHUL, M. et al. An efficient technique for facial expression recognition using multistage hidden markov model. In: *Soft Computing: Theories and Applications.* Singapore: Springer, 2019. p. 33–43. 23

RANKIN, J. A.; TOMENY, T. S. Screening of broader autism phenotype symptoms in siblings: Support for a distinct model of symptomatology. *Journal of autism and developmental disorders*, Springer, v. 49, n. 11, p. 4686–4690, 2019. 48

RAPIN, I.; TUCHMAN, R. F. Autism: definition, neurobiology, screening, diagnosis. *Pediatric Clinics*, Elsevier, v. 55, n. 5, p. 1129–1146, 2008. 18

REICHL, L. E. *A modern course in statistical physics.* [S.l.]: John Wiley & Sons, 2016. 77

RISCH, N. et al. Familial recurrence of autism spectrum disorder: evaluating genetic and environmental contributions. *American Journal of Psychiatry*, Am Psychiatric Assoc, v. 171, n. 11, p. 1206–1213, 2014. 46, 47, 135, 137

ROBINSON, E. B. et al. Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Archives of general psychiatry*, American Medical Association, v. 68, n. 11, p. 1113–1121, 2011. 87

ROBINSON, E. B. et al. Examining and interpreting the female protective effect against autistic behavior. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 110, n. 13, p. 5258–5262, 2013. 40

ROELFSEMA, M. T. et al. Are autism spectrum conditions more prevalent in an information-technology region? a school-based study of three regions in the netherlands. *Journal of autism and developmental disorders*, Springer, v. 42, n. 5, p. 734–739, 2012. 32

ROGGE, N.; JANSSEN, J. The economic costs of autism spectrum disorder: a literature review. *Journal of autism and developmental disorders*, Springer, v. 49, n. 7, p. 2873–2900, 2019. 19

ROSEN, K. H. *Handbook of discrete and combinatorial mathematics*. 2. ed. Boca Raton, FL: CRC press, 2017. 64, 65

ROSTI, R. O. et al. The genetic landscape of autism spectrum disorders. *Developmental Medicine & Child Neurology*, Wiley Online Library, v. 56, n. 1, p. 12–18, 2014. 35

ROTHOLZ, D. A. et al. Improving early identification and intervention for children at risk for autism spectrum disorder. *Pediatrics*, Am Acad Pediatrics, v. 139, n. 2, p. e20161061, 2017. 33

RUBEIS, S. D. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, Nature Publishing Group, v. 515, n. 7526, p. 209–215, 2014. 38

RUBENSTEIN, E.; CHAWLA, D. Broader autism phenotype in parents of children with autism: a systematic review of percentage estimates. *Journal of child and family studies*, Springer, v. 27, n. 6, p. 1705–1720, 2018. 48, 49, 52

RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. [S.l.]: Pearson Education, 2020. v. 4. 21, 22, 55, 57, 58, 59, 60, 61, 62, 63, 64, 65, 67, 68, 69, 76

RUTTER, M.; BAILEY, A.; LORD, C. The social communication questionnaire. manual. los angeles, ca.: Western. *Psychological Services*, 2003. 49

RUTTER, M. et al. Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services*, v. 29, n. 2003, p. 30, 2003. 49

SAHEKI, A. H. *Construção de uma rede Bayesiana aplicada ao diagnóstico de doenças cardíacas*. Tese (Doutorado) — Universidade de São Paulo, 2005. 22, 83

SANDERS, S. J. et al. Multiple recurrent de novo cnvs, including duplications of the 7q11. 23 williams syndrome region, are strongly associated with autism. *Neuron*, Elsevier, v. 70, n. 5, p. 863–885, 2011. 38

SANDERS, S. J. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron*, Elsevier, v. 87, n. 6, p. 1215–1233, 2015. 39

SANDIN, S. et al. The familial risk of autism. *Jama*, American Medical Association, v. 311, n. 17, p. 1770–1777, 2014. 20, 21, 36, 42, 43, 137

SANDIN, S. et al. The heritability of autism spectrum disorder. *Jama*, American Medical Association, v. 318, n. 12, p. 1182–1184, 2017. 21, 42, 43, 45, 93, 94, 95, 104, 112, 129

SANDIN, S. et al. Autism risk associated with parental age and with increasing difference in age between the parents. *Molecular Psychiatry*, Nature Publishing Group, v. 21, n. 5, p. 693, 2016. 20, 21, 105

SARAÇOĞLU, R. Hidden markov model-based classification of heart valve disease with pca for dimension reduction. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 25, n. 7, p. 1523–1528, 2012. 84

SATTERSTROM, F. K. et al. Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk. *bioRxiv*, Cold Spring Harbor Laboratory, p. 484113, 2018. 39

SATTERSTROM, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, Elsevier, v. 180, n. 3, p. 568–584, 2020. 21

SAZ, O. et al. Tools and technologies for computer-aided speech and language therapy. *Speech Communication*, Elsevier, v. 51, n. 10, p. 948–967, 2009. 84

SCHAAF, C. P. et al. A framework for an evidence-based gene list relevant to autism spectrum disorder. *Nature Reviews Genetics*, Nature Publishing Group, v. 21, n. 6, p. 367–376, 2020. 38

SCHAEFER, G. B.; MENDELSOHN, N. J. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. *Genetics in Medicine*, Nature Publishing Group, v. 15, n. 5, p. 399–407, 2013. 46

SCHIEVE, L. A. et al. Have secular changes in perinatal risk factors contributed to the recent autism prevalence increase? development and application of a mathematical assessment model. *Annals of epidemiology*, Elsevier, v. 21, n. 12, p. 930–945, 2011. 36

SCHIPOR, O.; PENTIUC, S.; SCHIPOR, M. Automatic assessment of pronunciation quality of children within assisted speech therapy. *Elektronika ir Elektrotechnika*, v. 122, n. 6, p. 15–18, 2012. 84

SCHWICHTENBERG, A. et al. Can family affectedness inform infant sibling outcomes of autism spectrum disorders? *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 51, n. 9, p. 1021–1030, 2010. 50, 51, 137

SCUTARI, M.; DENIS, J.-B. *Bayesian networks: with examples in R*. 1. ed. Boca Raton, FL: Chapman and Hall/CRC, 2014. 65, 70, 74

SEALEY, L. et al. Environmental factors in the development of autism spectrum disorders. *Environment international*, Elsevier, v. 88, p. 288–298, 2016. 36

SEBAT, J. et al. Strong association of de novo copy number mutations with autism. *Science*, American Association for the Advancement of Science, v. 316, n. 5823, p. 445–449, 2007. 52

SELTZER, M. M. et al. Maternal cortisol levels and behavior problems in adolescents and adults with asd. *Journal of autism and developmental disorders*, Springer, v. 40, n. 4, p. 457–469, 2010. 20

SHIMABUKURO, T. T.; GROSSE, S. D.; RICE, C. Medical expenditures for children with an autism spectrum disorder in a privately insured population. *Journal of Autism and Developmental Disorders*, Springer, v. 38, n. 3, p. 546–552, 2008. 19

SILBERSTEIN, M. et al. A system for exact and approximate genetic linkage analysis of snp data in large pedigrees. *Bioinformatics*, Oxford University Press, v. 29, n. 2, p. 197–205, 2013. 22

SIMONOFF, E. Genetic counseling in autism and pervasive developmental disorders. *Journal of autism and developmental disorders*, Springer, v. 28, n. 5, p. 447–456, 1998. 46

SKAFIDAS, E. et al. Predicting the diagnosis of autism spectrum disorder using gene pathway analysis. *Molecular psychiatry*, Nature Publishing Group, v. 19, n. 4, p. 504–510, 2014. 103

SOCIETY, A. *What is autism?* 2020. Disponível em: <https://www.autism-society.org/what-is/>. 28

SPARROW, S. S.; CICCHETTI, D. V. *The Vineland Adaptive Behavior Scales.* [S.l.]: Allyn & Bacon, 1989. 51

SPIEGELHALTER, D. J.; FRANKLIN, R. C.; BULL, K. Assessment, criticism and improvement of imprecise subjective probabilities for a medical expert system. *arXiv preprint arXiv:1304.1529*, 2013. 83

SPIKER, D. et al. Genetics of autism: characteristics of affected and unaffected children from 37 multiplex families. *American Journal of Medical Genetics*, Wiley Online Library, v. 54, n. 1, p. 27–35, 1994. 52

STANKIEWICZ, P.; LUPSKI, J. R. Structural variation in the human genome and its role in disease. *Annual review of medicine*, Annual Reviews, v. 61, p. 437–455, 2010. 38

STONE, W.; OUSLEY, O. Screening tool for autism in two-year-olds (stat). *Nashville: Vanderbilt University*, 2004. 51

SUCKSMITH, E.; ROTH, I.; HOEKSTRA, R. Autistic traits below the clinical threshold: re-examining the broader autism phenotype in the 21st century. *Neuropsychology review*, Springer, v. 21, n. 4, p. 360–389, 2011. 48

SZATMARI, P. The classification of autism, asperger's syndrome, and pervasive developmental disorder. *The Canadian Journal of Psychiatry*, SAGE Publications Sage CA: Los Angeles, CA, v. 45, n. 8, p. 731–738, 2000. 18, 28

SZATMARI, P. et al. The familial aggregation of the lesser variant in biological and nonbiological relatives of pdd probands: a family history study. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, Cambridge University Press, v. 41, n. 5, p. 579–586, 2000. 47

TAMPOSIS, I. A. et al. Semi-supervised learning of hidden markov models for biological sequence analysis. *Bioinformatics*, Oxford University Press, v. 35, n. 13, p. 2208–2215, 2019. 23

TANIAI, H. et al. Genetic influences on the broad spectrum of autism: Study of proband-ascertained twins. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, Wiley Online Library, v. 147, n. 6, p. 844–849, 2008. 43, 45

TAYLOR, B.; JICK, H.; MACLAUGHLIN, D. Prevalence and incidence rates of autism in the uk: time trend from 2004–2010 in children aged 8 years. *BMJ open*, British Medical Journal Publishing Group, v. 3, n. 10, 2013. 34

TAYLOR, M. J. et al. Etiology of autism spectrum disorders and autistic traits over time. *JAMA psychiatry*, 2020. 36

TENESA, A.; HALEY, C. S. The heritability of human disease: estimation, uses and abuses. *Nature Reviews Genetics*, Nature Publishing Group, v. 14, n. 2, p. 139–149, 2013. 44, 45

TESTA, A. C. et al. Codingquarry: highly accurate hidden markov model gene prediction in fungal genomes using rna-seq transcripts. *BMC genomics*, BioMed Central, v. 16, n. 1, p. 170, 2015. 23, 84

TICK, B. et al. Heritability of autism spectrum disorders: a meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 57, n. 5, p. 585–595, 2016. 20, 21, 36, 42, 43, 45, 112

TORRE-UBIETA, L. de la et al. Advancing the understanding of autism disease mechanisms through genetics. *Nature medicine*, Nature Publishing Group, v. 22, n. 4, p. 345–361, 2016. 39

TREVIS, K. J. et al. Tracing autism traits in large multiplex families to identify endophenotypes of the broader autism phenotype. *International journal of molecular sciences*, Multidisciplinary Digital Publishing Institute, v. 21, n. 21, p. 7965, 2020. 10, 153, 154

TSAI, L.; STEWART, M. A.; AUGUST, G. Implication of sex differences in the familial transmission of infantile autism. *Journal of Autism and Developmental Disorders*, Springer, v. 11, n. 2, p. 165–173, 1981. 40

TURNER, L. M.; STONE, W. L. Variability in outcome for children with an asd diagnosis at age 2. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 48, n. 8, p. 793–802, 2007. 49

UĞUZ, H.; ARSLAN, A.; TÜRKOĞLU, İ. A biomedical system based on hidden markov model for diagnosis of the heart valve diseases. *Pattern recognition letters*, Elsevier, v. 28, n. 4, p. 395–404, 2007. 84

VALTORTA, M.; KIM, Y.-G.; VOMLEL, J. Soft evidential update for probabilistic multiagent systems. *International Journal of Approximate Reasoning*, Elsevier, v. 29, n. 1, p. 71–106, 2002. 74

VERHULST, B.; NEALE, M. C. Best practices for binary and ordinal data analyses. *Behavior Genetics*, Springer, p. 1–11, 2021. 44

VISSCHER, P. M.; HILL, W. G.; WRAY, N. R. Heritability in the genomics era—concepts and misconceptions. *Nature reviews genetics*, Nature Publishing Group, v. 9, n. 4, p. 255–266, 2008. 45

VOS, T. et al. Gbd 2016 disease and injury incidence and prevalence collaborators. global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the global burden of disease study 2016. *Lancet*, v. 390, n. 10100, p. 1211–59, 2017. 31

WAN, M. W. et al. Quality of interaction between at-risk infants and caregiver at 12–15 months is associated with 3-year autism outcome. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 54, n. 7, p. 763–771, 2013. 50, 137

WANG, C. et al. Prenatal, perinatal, and postnatal factors associated with autism: a meta-analysis. *Medicine*, Wolters Kluwer Health, v. 96, n. 18, 2017. 20, 21

WANG, J. et al. Sparse multiview task-centralized ensemble learning for asd diagnosis based on age-and sex-related functional connectivity patterns. *IEEE transactions on cybernetics*, IEEE, n. 99, p. 1–14, 2018. 18

WANG, P. et al. Phonocardiographic signal analysis method using a modified hidden markov model. *Annals of Biomedical Engineering*, Springer, v. 35, n. 3, p. 367–374, 2007. 84

WAZANA, A.; BRESNAHAN, M.; KLINE, J. The autism epidemic: fact or artifact? *Journal of the American Academy of Child & Adolescent Psychiatry*, Elsevier, v. 46, n. 6, p. 721–730, 2007. 34

WEINER, D. J. et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nature genetics*, Nature Publishing Group, v. 49, n. 7, p. 978–985, 2017. 40

WERLING, D. M.; GESCHWIND, D. H. Recurrence rates provide evidence for sex-differential, familial genetic liability for autism spectrum disorders in multiplex families and twins. *Molecular autism*, BioMed Central, v. 6, n. 1, p. 1–14, 2015. 40

WILLIAMSON, J. *Probability logic. Handbook of the Logic of Argument and Inference.* [S.l.]: North-Holland, 2002. 22, 57, 63

WOOD, C. L. et al. Evidence for asd recurrence rates and reproductive stoppage from large uk asd research family databases. *Autism Research*, Wiley Online Library, v. 8, n. 1, p. 73–81, 2015. 46, 47, 137

XIAO, F.; AI, Q. Data-driven multi-hidden markov model-based power quality disturbance prediction that incorporates weather conditions. *IEEE Transactions on Power Systems*, IEEE, 2018. 23

XIE, F.; PELTIER, M.; GETAHUN, D. Is the risk of autism in younger siblings of affected children moderated by sex, race/ethnicity, or gestational age? *Journal of Developmental & Behavioral Pediatrics*, v. 37, n. 8, p. 603–609, 2016. 46, 47

XIE, S. et al. Family history of mental and neurological disorders and risk of autism. *JAMA network open*, American Medical Association, v. 2, n. 3, p. e190154–e190154, 2019. 19, 32, 132, 140, 151, 152, 153

XU, B. et al. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nature genetics*, Nature Publishing Group, v. 40, n. 7, p. 880–885, 2008. 38

XU, G. et al. Prevalence of autism spectrum disorder among us children and adolescents, 2014-2016. *Jama*, American Medical Association, v. 319, n. 1, p. 81–82, 2018. 32

YAHATA, N. et al. A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nature communications*, Nature Publishing Group, v. 7, p. 11254, 2016. 21, 82, 83

YANG, J. et al. Gcta: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, Elsevier, v. 88, n. 1, p. 76–82, 2011. 43, 45

YIP, B. H. K. et al. Heritable variation, with little or no maternal effect, accounts for recurrence risk to autism spectrum disorder in sweden. *Biological psychiatry*, Elsevier, v. 83, n. 7, p. 589–597, 2018. 42, 43, 45, 112, 114, 129

YIRMIYA, N.; OZONOFF, S. The very early autism phenotype. *Journal of Autism and Developmental Disorders*, Springer, v. 37, n. 1, p. 1–11, 2007. 52

YODER, P. et al. Predicting social impairment and asd diagnosis in younger siblings of children with autism spectrum disorder. *Journal of autism and developmental disorders*, Springer, v. 39, n. 10, p. 1381–1391, 2009. 50

ZABLOTSKY, B. et al. Estimated prevalence of autism and other developmental disabilities following questionnaire changes in the 2014 national health interview survey. 2015. 34

ZABLOTSKY, B. et al. Prevalence and trends of developmental disabilities among children in the united states: 2009–2017. *Pediatrics*, Am Acad Pediatrics, v. 144, n. 4, p. e20190811, 2019. 32

ZARREI, M. et al. A copy number variation map of the human genome. *Nature reviews genetics*, Nature Publishing Group, v. 16, n. 3, p. 172–183, 2015. 38

ZEMOURI, R.; ZERHOUNI, N.; RACOCEANU, D. Deep learning in the biomedical applications: Recent and future status. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 9, n. 8, p. 1526, 2019. 82

ZHANG, D.; ZHANG, H.; ZHANG, B. Computerized tongue diagnosis based on bayesian networks. In: *Tongue Image Analysis*. [S.l.]: Springer, 2017. p. 265–280. 83

ZHAO, Y. et al. 3d deep convolutional neural network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland, 2018. p. 172–180. 21, 82, 83

ZWAIGENBAUM, L. et al. Early intervention for children with autism spectrum disorder under 3 years of age: recommendations for practice and research. *Pediatrics*, Am Acad Pediatrics, v. 136, n. Supplement 1, p. S60–S81, 2015. 20

ZWAIGENBAUM, L. et al. Sex differences in children with autism spectrum disorder identified within a high-risk infant cohort. *Journal of autism and developmental disorders*, Springer, v. 42, n. 12, p. 2585–2596, 2012. 46, 47

Appendix

APPENDIX A – Publications

This appendix presents our published and (to be) submitted works. To date, the results include two published papers, one published book chapter, one book chapter under review, two papers submitted for acceptance, and two papers to be submitted.

Published papers:

- *Hidden Markov Models to Estimate the Probability of Having Autistic Children* (CARVALHO et al., 2020); and
- *Applied Behavior Analysis for the Treatment of Autism: A Systematic Review of Assistive Technologies* (ALVES et al., 2020).

Published book chapter:

- Chapter *Robôs como suporte às intervenções baseadas em aba para o Transtorno do Espectro Autista: uma revisão narrativa* in the book *Autismo: Tecnologias e formação de professores para a escola pública* (ALVES et al., 2021).

Submitted book chapter:

- Chapter *Tecnologias Assistivas Aplicadas ao TEA: Aspectos Filosóficos, Éticos e Legais* in the book *Autismo: Tecnologias Educacionais e Práticas nas Escolas* (to be published).

Submitted papers:

- *Identifying Potential Brain Regions for Autism Severity Diagnosis using Machine Learning and fMRI*; and
- *rs-fMRI and Machine Learning for ASD Diagnosis - A Systematic Review and Meta-analysis.*

Papers to be submitted:

- *Autism Spectrum Disorder: A Literature Review on Prevalence and Etiology Measures*, which refers to Chapter 2 of this thesis;
- *Estimating the Family Bias to Autism: A Bayesian Approach*, which refers to Chapters 5, 6, and 7 of this thesis.