

Article

Probabilistic Causal Modeling of Barriers to Accessibility for Persons with Disabilities in Canada

Mouri Zakir ^{1,*}, Gregor Wolbring ^{2,†} and Svetlana Yanushkevich ^{1,*}¹ Biometric Technologies Laboratory, Department of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada² Community Rehabilitation and Disability Studies, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 1N4, Canada; gwolbrin@ucalgary.ca

* Correspondence: mouri.zakir@ucalgary.ca (M.Z.); syanshk@ucalgary.ca (S.Y.)

† Current address: Department of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada.

Highlights:

This research investigates the causal relationship between accessibility barriers and demographic and other traits of people with disabilities based on the Canadian Disability Survey 2022. We apply a unique computational intelligence tool—probabilistic causal models such as structural equation modeling and Bayesian networks—to discover those relationships and model scenarios to assess the risks and barriers in accessing services, transportation and communications in the context of inclusive smart cities. In addition, we apply both generative and causal models to produce synthetic data in order to test the accuracy of the causal models' structure learning.

What are the main findings?

- Probabilistic models are a feasible tool to derive reasoning from data about vulnerable groups.
- A causal relation between accessibility experience, demographics and disabilities is confirmed.
- The synthetic data generated using probabilistic models preserve the structure of real data.

What is the implication of the main finding?

- Insights into accessibility for diverse groups will be essential in designing inclusive smart cities.
- Synthetic data can help overcome the insufficiency of real data while maintaining confidentiality.

Abstract: This paper utilizes a methodological two-step process incorporating statistical and causal probabilistic modeling techniques to investigate factors affecting the accessibility experiences of persons with disabilities in Canada. We deploy a network-based approach using empirical data to perform a holistic assessment of the relations between various demographic features (e.g., age, gender and type of disability) and accessibility barriers. A statistical measurement method is applied that utilizes structural equation modeling supported by exploratory factor analysis. For causal probabilistic modeling, Bayesian networks are employed as a straightforward and compact way to interpret knowledge representation. This causal reasoning approach analyzes the nature and frequency of encountering barriers based on data to understand the risk factors contributing to pressing accessibility issues. Furthermore, to evaluate network performance and overcome any data limitations, synthetic data generation techniques are applied to create and validate artificial data built on real-world knowledge. The proposed framework strives to provide reasoning to understand the prevalence of physical, social, communication or technological barriers encountered by persons with disabilities in their daily lives. This study contributes to the identification of areas for prioritization in facilitating accessibility regulation and practices to realize an inclusive society.

Keywords: accessibility barriers; persons with disabilities; inclusive smart city; causality; exploratory factor analysis; structural equation modeling; regression; machine reasoning; Bayesian network; synthetic data generation; statistical hypotheses testing



Academic Editor: Pierluigi Siano

Received: 8 October 2024

Revised: 16 December 2024

Accepted: 19 December 2024

Published: 24 December 2024

Citation: Zakir, M.; Wolbring, G.; Yanushkevich, S. Probabilistic Causal Modeling of Barriers to Accessibility for Persons with Disabilities in Canada. *Smart Cities* **2025**, *8*, 4. <https://doi.org/10.3390/smartcities8010004>

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A growing emphasis has been placed on providing equal opportunities for all people, particularly people with disabilities, to support their participation in society without discrimination [1]. In recent years, the world has witnessed increased smart city initiatives that employ emerging data and technologies to tackle complex urban challenges [2]. This life-enhancing process is promoted by inclusion and social sustainability to build a community for people to thrive with equal opportunities [3,4].

The global “smart city” movement ideates the application of Information and Communication Technologies (ICT) as a crucial means of addressing the negative consequences of urban development and improving the quality of life of residents [2]. In this context, inclusion is understood in many ways, such as social cohesion and sense of belonging, equity, participation, quality of life, and human rights [2]. As a tool to discover whether equity and inclusion have been achieved, accessibility is investigated here using data on accessibility barriers for the case of Canada. An overview of the motivations for this research is presented in Figure 1 and discussed in this section.

Cities are complex systems that face a growing number of societal challenges that city governments are struggling to address [5]. An individual’s ability to integrate into a city, navigate the environment and comprehend information affects their well-being and overall quality of life [5,6]. Vulnerable groups are particularly exposed to exclusion from social activities and full enjoyment of their rights and duties as citizens in urban environments [5].

Impacted by the rapidly increasing global population, cities face an increase in the number of people with various disabilities, whether sensory or physical. Recent findings from the 2022 Canadian Survey on Disability (CSD) [7] show that 27% of Canadians aged 15 years and older, or 8.0 million people, have one or more disabilities that limit them in their daily activities. This indicates a 5% increase since 2017 [8], which can be partially attributed to factors such as demographics and the type of disability. Despite representing a significant portion of the population and the diversity of conditions, people with disabilities face common barriers that prevent their full participation in society [9].

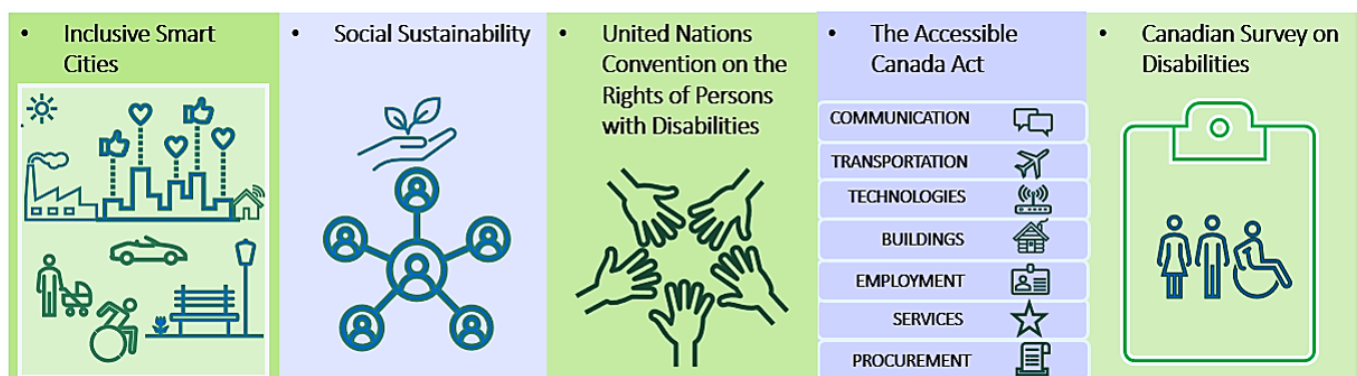


Figure 1. Motivations for this investigation on accessibility in Canada: Inclusive smart cities, social sustainability, the United Nations Convention on the Rights of Persons with Disabilities, the Accessible Canada Act and the Canadian Survey on Disabilities [2,3,6,7,9].

Smart cities are promising communities that leverage intelligent technologies to connect citizens, which is especially crucial for citizens with disabilities, who face substantial challenges in urban living [10]. But this falls short of a comprehensive urban development paradigm, which needs the critical analysis of the extent to which smart cities also promote social sustainability of urban areas [11]. Social sustainability is a dynamic concept that promotes infrastructure to support social needs and concerns equally for all. While smart cities primarily aim to enhance performance through the innovative use of data and

technology, a social sustainability perspective stresses critical interconnections between people and places [11].

An inclusive city should accommodate everyone, regardless of their economic means, gender, race, ethnicity, and religion, enabling and empowering people to fully participate in the social, economic and political opportunities that cities have to offer [5]. For instance, in the city of Toronto, a barrier-free, well-run, modern, and connected community enabling technology is envisioned to improve the quality of life of its residents as a long term vision of the city administration's [12].

However, research has shown that the idea of an inclusive city is mostly perceived in general terms of poverty and social inclusion, while "disability-inclusion" is neglected [5]. Despite some disability-specific political, legislative and standardization frameworks in place, such as the UN Convention on the Rights of Persons with Disabilities [13], there is a lack of applicable approaches and tools that can help relevant stakeholders with translating these frameworks into more pragmatic terms to better understand the overall complexity of their requirements [5].

The UN Convention on the Rights of Persons with Disabilities (CRPD) aims to recognize the inherent dignity and equal rights of all members of the human family [13]. The CRPD was ratified and signed by the Government of Canada (GC) in 2010 [9] and is fulfilled for federal organizations by the Accessible Canada Act (ACA) [9]. The ACA is landmark federal legislation that aims to realize a barrier-free Canada in the area of communications, Information and Communication Technologies (ICT) and physical/transportation aspects by 2040 [9].

The findings discussed above align with the goals of the Accessible Canada Act [9] and serve as the motivation for our study on the application of smart tools such as causal models to investigate the challenges faced by persons with disabilities when accessing services, physical spaces, communication, transportation and other areas.

1.1. Objective

Smart cities are increasingly targeted to create inclusive environments that cater to the needs of all residents, including those facing limitations and hindrances in the use of "everyday" technologies [6]. This objective can be achieved by ensuring accessibility for vulnerable groups such as persons with disabilities in any environment so that all individuals can be independent and safe enough to reach their full potential [14].

The technological solutions driving smart cities do not counter the fundamental barriers to exclusion, especially for vulnerable populations, as there are no one-size-fits-all solutions for marginalization [2]. To improve the living conditions of persons with disabilities within smart city initiatives, it is essential to improve the accessibility of federal services in terms of physical spaces, communications, transportation, and other areas for population with disabilities. This requires the investigation of the nature of any barriers to accessibility and the influence of demographic traits on the extent the experience of such barriers utilizing advanced frameworks beyond simple statistical/correlation analysis.

In addition, data collection is largely dependent on self-identification, which people with disabilities might be reluctant to participate in. This may result in insufficient data, which can be addressed by generating synthetic data using Artificial Intelligence (AI) techniques, which would also aid in protecting sensitive information.

The objective of our paper is to confirm that probabilistic causal reasoning models are feasible tools for understanding the effects of barriers encountered by persons with disabilities depending on their demographic attributes when accessing federal services and facilities. We deploy a modeling tool to discover the cause-and-effect relationships between variables to investigate the nature of accessibility barriers and the influence of

demographic traits on the extent of those barriers. This approach presents the discovery of the underlying factors that drive changes in accessibility experiences.

1.2. Prior Work, Novelty and Outline

This investigation lays the groundwork for the prediction of trends in social development research using causal reasoning algorithms and first-hand accounts of user experience from surveys.

1.2.1. Prior work

Our preliminary research on causal analysis reported in [15] involved accessibility experiences of persons with disabilities based on the nature of their reported long-term conditions based on the CSD 2017 survey [16]. The prevalence of accessibility barriers related to communications, information and communication technology (ICT) and transportation was assessed based on the type of disabilities reported.

1.2.2. Novelty

We utilize probabilistic causal models to provide a comprehensive assessment of the relationships between various demographic features (e.g., age, gender and type of disability) and accessibility barriers in the public and private sectors. Additionally, synthetic data generation and statistical testing are applied using the created Bayesian probabilistic model to handle any limitation that may occur due to occasional and, in many cases, unbalanced data, especially concerning persons with disabilities.

Based on our research and to the best of our knowledge, that application of causal modeling is a new approach to the qualitative assessment of accessibility experience. This study aims to help stakeholders and policy makers in strategic planning and execution with respect to accessible environments for persons with disabilities.

1.2.3. Outline

This paper is structured as follow: Section 2 outlines the problem formulation for this investigation and literature review of related work. Section 3 elaborates on the proposed approach for the methodology, including a brief description of data pre-processing. Sections 4–7 comprehensively discuss the definitions, hypothesis formulation and implementation of Exploratory Factor Analysis (EFA), Structural Equation Modeling (SEM) and Bayesian Networks (BNs), respectively, followed by the generation of Synthetic Data (SD), along with analyses of the experimental results for each method. The discussion and conclusion are presented in Sections 8 and 9, respectively.

Collectively, these sections provide a comprehensive narrative of the application of AI modeling techniques for the analysis of accessibility for persons with disabilities, unraveling the intricacies of their effects on decision making, engaging the smart city discourse and ultimately envisioning a future where cities are not just digitally enabled but dynamically understood and optimized through the lens of inclusion.

2. Background and Literature Review

This section includes the background of the problem addressed by this study and a discussion of the various techniques adopted by other investigators in the analysis of underlying factors in various social development studies.

2.1. Problem Formulation

Cities increasingly rely on digital analytics and smart solutions to address key challenges associated with the recent proliferation of big data driving smart city transformations [17]. Big data refer to large and complex sets of data that generally represent digital

traces of human activities characterized by the scale or volume, analysis method or effect on organizations [18]. Cities around the world collect large quantities of data related to urban living to produce useful content for various stakeholders, including citizens, patrons, local governments and companies [19]. For instance, in San Francisco, crime records were analyzed to improve public security services; in Seoul, patterns and demands for the usage of the city buses were identified to improve public bus services; and in Rio de Janeiro, data from cameras and sensors were used to address safety concerns [19].

Despite the emergence of such cases, the understanding of accessibility in the public and private sectors for smart cities remains limited in the literature [19]. Smart cities' potential to improve the lives of people with disabilities also remains underexplored [20]. By leveraging data-driven insights, smart cities can create more inclusive environments that enhance the overall quality of life of people with disabilities [18,20]. This paper discusses the multitude of challenges encountered by individuals with disabilities reflected in the CSD data and proposes techniques for utilizing AI-driven smart tools to effectively identify the patterns in these challenges.

The CSD is premised on the social model of disability, which perceives disability as the result of the interaction between a person's functional limitations and barriers in the surrounding environment [7,21]. This includes social and physical barriers that make it difficult to function day to day. Thus, disability is a social disadvantage that an unfavorable environment imposes on top of an individual's functional limitation [7]. Disabilities can encompass significant activity limitations that result from the association of a long-term condition with contextual factors [22], directly affecting more than one billion people worldwide [23]. Hence, the experience is unique for each individual living with a disability, shaped by their physiological condition, along with social and physical environments [22].

Disabling barriers stem from inadequate policies or standards, negative attitudes and prejudices, service deficiencies, inaccessible built environments and a lack of evidence and data to inform effective policies and programs [23]. Moreover, people affected by disabilities are at an increased risk of poor health outcomes, lower education attainment, reduced employment and earning potential, living in poverty and higher dependency on others [23].

Accessibility is about removing barriers and ensuring equal participation in everyday activities. A variety of societal barriers still prevalent in Canada contribute to inadequate accessibility faced by people with disabilities and often lead to undesirable consequences for people with disabilities, such as social exclusion, negative stereotypes and perceptions and challenges that often lead to disability burnout [24].

Although disabilities affect people of all ages, genders and geographical regions, some groups may be more likely to develop disabilities, while individuals and groups may differ in their ability to manage adverse health conditions [22]. Hence, it is crucial to identify the obstacles and risk factors to improve accessibility in federal services and private facilities to ensure equal participation in society for persons with disabilities.

Our study seeks to leverage probabilistic causal models to determine the likelihood of experiencing specific barriers given a particular type of disability, age or gender. These projections have the potential to infer implications for entities under the federal government in providing barrier-free services and further predict trends in accessibility experiences a decade or more ahead of time to help with evaluation and preparedness. Subsequently, we strive to provide a comprehensive understanding of the correlation between data on multiple biological traits to decipher the risk factors and pathways contributing to varying accessibility experiences, ultimately providing recommendations for policy-making decisions and resource utilization.

2.2. Related Work

This section discusses published work regarding a range of accessibility issues experienced by individuals with disabilities, as well as related methodologies addressing real-world user-experience scenarios for quality assessment.

- General Considerations for Accessibility

The inherent problems associated with measuring accessibility are discussed in [25] for people with disabilities in terms of surfaces, barriers and travel modes. The proposed method includes the absolute measure currently used in ADA compliance [26] to map accessibility, as well as to help select the mitigation or renovation projects that yield the greatest increases in accessibility. The author of [27] discussed limitations of AI personalization algorithms driven by data in creating inclusive technology such as device accessibility settings or driving guides for autonomous vehicles. The study concluded with suggestions for assistive technologies developed with each unique disability in mind rather than for a broad category of disability for optimum accessibility experience.

- Probabilistic Modeling

The authors of [28] combined knowledge with available demographic and health survey data to construct a BN and investigate the factors associated with childhood health conditions. They investigated the effects of missing values, sample size and knowledge-based constraints and assessed their accuracy with multiple scoring functions.

- Statistical Modeling

Statistical modeling in the context of causal analysis is explored in a tribute to researchers who have significantly contributed to the advancement of SEM [29]. An overview of SEM including its beginnings and usefulness in social sciences, is presented here. Another paper [30] focused on SEM and a test confirmatory factor analysis system to better explain how social networking systems could be utilized by students for educational purposes, outlining the relations among different types of independent and dependent variables.

- Factor Analysis with Statistical Modeling

The research described in [31] developed a self-aware computational model to assess student satisfaction in a university engineering faculty depending on different aspects of academic experience using FA and SEM. This model provides the basis for probabilistic inference using a reasoning mechanism.

- Probabilistic Analysis and Statistical Modeling

In [32], the techniques of BN and SEM were combined to model service quality in the rail transit service of Seville, Spain. The proposed approach initiates with a BN to find certain hypotheses about the relationships between the dimensions that define service quality, followed by the use of SEM to validate the overall model. Moreover, SEM helped with validation of the relationships of the service quality framework extracted from the BN. Hence, this approach allowed researchers to examine more than one relationship and hypotheses with a large amount of information at the same time.

An approach combining deep learning with BN techniques was employed in [33] for the analysis of mass migration patterns. Migration crises were addressed in the context of crisis management by using a combination of modeling and imbalance assessment tools. By using deep learning for forecasting and causal reasoning via Bayesian networks, this methodology enabled the evaluation of risks and imbalances in the socio-technological landscape. The causal analysis revealed significant impacts of economic conditions, crises and government policies on migration patterns.

The study reported in [34] proposed linking a BN to SEM, providing an advantage in testing causal relationships in order to overcome the limitation of BN in differentiating causal from spurious relationships among decision factors. This study applied the integrated approach to decision support for customer retention in a virtual community, showing a pathway for Bayesian modeling at the latent variable level and extending SEM for managerial prediction and diagnosis.

Another study [35] assessed the risk of a stroke given stroke patient data through the application and validation of causal graphical models. The risk factors identified by a stroke expert were modeled with a BN, while SEM was later applied to provide a quantitative assessment of causality among the variables in the BN. The authors of [36] explore assistive technologies for people with disabilities in smart cities and emergency assistance provided by first responders. Technologies designed for interactions between responders and the assisted persons were discussed based on AI formalism.

- **Generative AI**

This literature reported in [37] examined the interplay between urban form and social behavior by identifying the key indicators of urban morphology and presenting state-of-the-art methodologies for data collection. The computational capability of foundation models and generative AI models to simulate interactions between individuals and urban built environments in a diverse group across the globe was also discussed. Through this exploration, the potential of advanced AI systems to assist urban scientists in understanding cities was demonstrated for the optimal application of generative AI in urban development research and policy making.

The literature review revealed that while quantitative data have been analyzed using probabilistic modeling in many references, causal models have not been applied, as far as we know, to investigate accessibility barriers faced by people with disabilities.

3. Proposed Approach and Materials

This section describes the proposed approach for developing a computational tool for causal modeling, along with an overview of the primary dataset [38] used in this study.

3.1. Methodology

Smart cities, which involve citizens leveraging advanced data analytics, predictive models and digital twin techniques, can offer a transformative model for socially sustainable and inclusive urban development [17]. The probabilistic framework describes how to represent and manipulate uncertainty about models and predictions, which has a central role in scientific data analysis and AI [39]. On the other hand, predictive analytics is critical to proactive planning, enabling cities to cultivate urban environments that are centered around the needs of citizens [17].

This paper proposes an approach constituting the above-mentioned methods, promoting the development of such urban public spaces catering to a diverse community, with a focus on people with disabilities. Our proposed approach comprises a multivariate statistical analysis framework and probabilistic reasoning techniques, as illustrated in Figure 2, to identify causal relations between different combinations of demographic traits and barriers to accessibility for persons with disabilities. Statistical analysis is applied to empirical data to provide a quantitative assessment of causal relationships among the variables used in the Bayesian network, thus statistically validating expert knowledge.

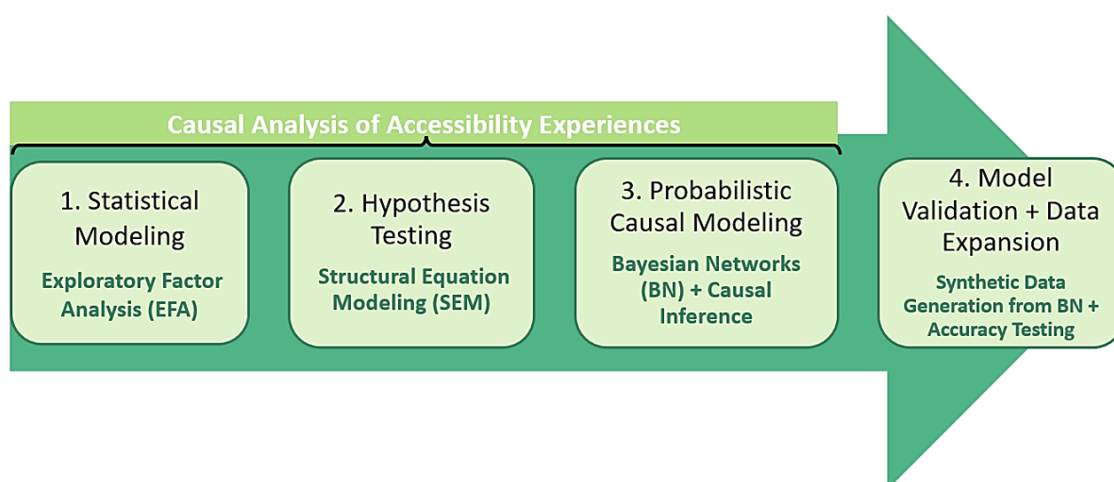


Figure 2. Proposed methodology: A probabilistic causal reasoning framework is employed to identify the underlying causes of barriers for varying demographics with disabilities. Here, EFA, SEM and BN techniques are combined for causal analysis, followed by SD generation using the probabilistic model to overcome limitations of the real data.

We use the causal probabilistic method to identify challenges and considerations of the use of synthetic data for smart cities. In this direction, diagnostic analysis and predictive analysis are keystones for identifying reasons behind specific outcomes and preparing for future accessibility concerns. Leveraging historical data, these tools identify potential challenges and predict trends for hypothetical scenarios, empowering cities to adopt a proactive stance in addressing issues [17]. The integration of predictive analytics into smart city frameworks allows decision makers to make informed choices regarding infrastructure development, resource distribution and policy implementation to thrive in an ever-evolving urban landscape.

For statistical modeling, we use EFA, an exploratory data analysis method, to search factors or latent variables from a set of observed variables. It helps with data interpretations by reducing the number of variables. Next, we implement a linear model framework called SEM to represent the various aspects of accessibility thought to be causally connected.

To support the confirmatory analysis of SEM, we then apply probabilistic reasoning based on a causal model such as a BN. Several scenarios are demonstrated for risk assessment using the inference mechanism of the network. Finally, the created BN is used for data generation to explore the possibility of using Synthetic Data (SD) for future investigations in case of inconclusive or limited real-world data.

Additionally, the BN is used as a template to replicate the structure and variables in the dataset to generate a new set of data by applying parameter learning techniques. This synthetic set of data is then compared with the original data through statistical hypothesis testing to evaluate the accuracy of the former. Furthermore, using the structure learning method, the BN is recreated in an automated manner to explore the precision of this learning algorithm with the synthetic data. Adjustment is applied to observe improvement in the results in the form of conditional probability.

3.2. Dataset

The primary data (PD) used for this paper were obtained from the Statistics Canada data hub and published by the Government of Canada in May 2024. This section contains provides an overview, in addition to addressing accuracy, limitations and data processing steps leading to statistical and causal probabilistic modeling later.

The dataset [38] is based on the Canadian Survey on Disability (CSD) 2022, a national survey that collects information about Canadians aged 15 and over whose everyday ac-

tivities are limited because of a long-term condition or health-related problem [7]. As demonstrated in (Figure 3), a range of data is provided on 10 different disability types leading to activity limitations related to *seeing, hearing, vision, mobility, flexibility, dexterity, pain, learning, memory, developmental and mental health*. The total sample size for the CSD is approximately 54,000 individuals, with an overall collection response rate of 61.1% [7]. The data includes barrier scenarios categorized under four groups: (1) communications, (2) attitude, (3) information and communication technologies (ICT) and (4) physical/transportation estimated by both number and percentage.

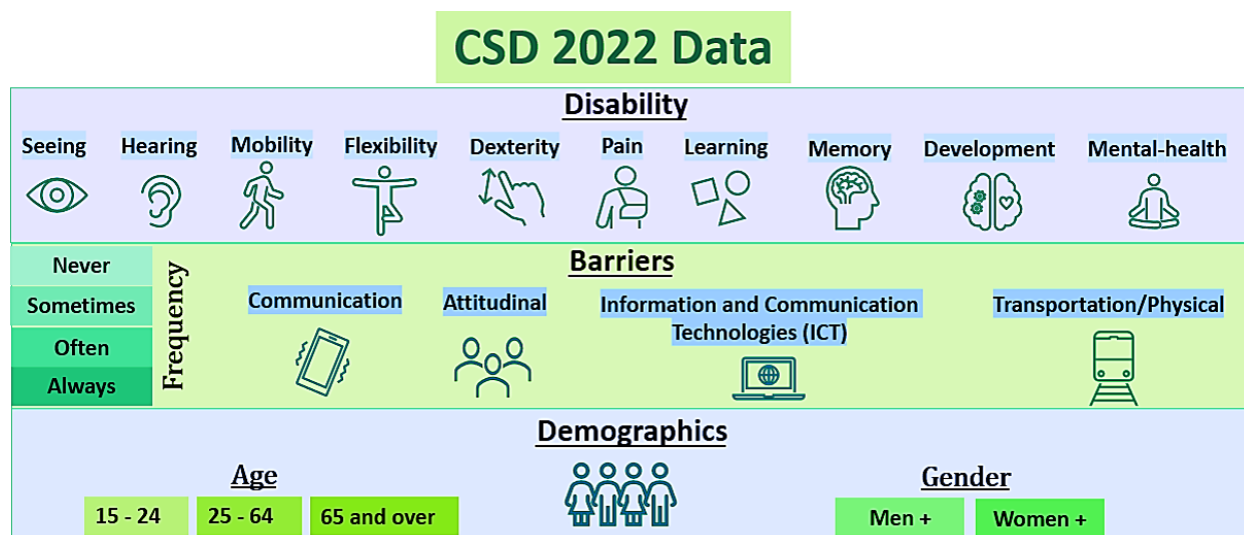


Figure 3. Primary dataset (PD)—Categorical variables and their subcategories presented in the CSD 2022 accessibility barrier data. The “+” symbol next to each class under **Gender** denotes “non-binary persons” [38].

In the dataset, there are two types of variables: numerical and categorical. The categorical variables and their subcategories are presented in Figure 3. Observing the categorical features, there two kinds of categorical data: ordinal and nominal. An ordinal variable has values defined by an order relation between the different categories, whereas the categories of nominal data are purely descriptive and do not possess any quantitative or numeric value [40].

For the disability types, the dataset focuses on activity limitations related to *hearing, vision, mobility, flexibility, dexterity, pain, learning, mental health, memory and developmental* disabilities. Brief descriptions of these categories, along with the corresponding recommended necessary accessibility measures, are included in Table A2. The data for demographic with disabilities are categorized in terms of barrier scenarios related to *Communications, attitude, information and communication technologies (ICT) and physical/transportation*. The collected data are estimated by both number and percentage in the CSD dataset.

To prepare the data to implement the modeling algorithms, we performed some data normalization steps. The data cleaning steps were followed by splitting the ‘number’ and ‘percentage’ data. One-hot encoding was performed on nominal features; however, ordinal data were mapped into numerical data so that the inner order could be preserved.

Since the source is a sample survey, all estimates are subject to both sampling and non-sampling errors. Estimates and confidence interval that do not meet Statistics Canada’s quality standards are suppressed [7].

4. Exploratory Factor Analysis (EFA) for Identification of Data Structure

With the accessibility barrier data, the goal was to investigate the causality between demographic attributes and accessibility experiences. To establish a causal connection between variables, it was assumed that there are underlying latent constructs in the data. However, with no prior knowledge, dimension reduction and structure identification were needed before proceeding with statistical modeling.

4.1. EFA Overview

Factor analysis is a widely used unsupervised machine learning algorithm in dimension reduction. It creates factors from the observed variables to represent the common variance, i.e., variance due to correlation among the observed variables [31,41]. In experimental design, the term “factor” refers to a categorical explanatory variable and is the reason the observable variables have the correlations they do [42]. Factors are also known as latent or hypothetical variables.

EFA aims to explain the covariances or correlations among the variables and to understand what constructs underlie the data [43]. Hence, EFA was selected as a statistical framework to identify latent constructs or factors that explain the patterns of correlations within the observed variables. It is often used as a first analytic step to determine the most appropriate method of conceptualizing scores in a given dataset without any modeling limitations [44].

With EFA, the goal is to describe and summarize a dataset by explaining a set of observed variables in terms of a smaller number of latent variables or factors [42]. The equation below represents features in terms of factors (β) independently for $i = 1, \dots, n$. Given a variable (x), factor F , factor loading (l) and error term (ϵ) (weight of the factor for the corresponding variable), the number of factors is equal to the number of variables [41]:

$$x_i - \mu_i = l_1 F_1 + \dots + l_n F_n + \epsilon_i$$

To determine the most effective structure with an appropriate number of factors for statistical modeling with SEM later, factor validity is assessed here using EFA.

4.2. Application of the EFA Method

The pre-processed dataset (Section 3.2) was used for EFA as illustrated in Figure 4, employing the following steps. To check the suitability of data for EFA, correlation tests were performed. The co-linearity variable obtained from these tests, as included in Table 1, indicates how strongly a single variable is correlated with other variables.

4.2.1. Bartlett’s Test of Sphericity

Bartlett’s test checks whether correlation is present in the given data. It tests the null hypothesis that the correlation matrix is an identical matrix [41]. The identical matrix consists of all diagonal elements, represented as 1, so, the null hypothesis assumes that no correlation is present among the variables.

We needed to reject this null hypothesis because factor analysis aims to explain the common variance, i.e., the variation due to correlation among the variables. If the p-test statistic value is less than 0.05 [32,34], we can decide that the correlation is not an identical matrix, i.e., correlation is present among the variables with a 95% confidence level.

Table 1. Bartlett’s sphericity and KMO test parameters to check variable correlations for EFA.

Bartlett Sphericity Test			KMO Test
Component	Chi-squared value: 5890.8469	p value: 0.0	KMO test statistic: 0.6489

Since the p -test statistic is less than 0.05, as shown in Table 1, we can conclude that correlation is present among the variables, which allows us to apply factor analysis.

4.2.2. Kaiser–Meyer–Olkin (KMO) Test

The KMO test measures the proportion of variance that might be common among the variables [41]. Larger proportions are expected to show several correlations present among the variables, thereby ensuring the application of dimension reduction techniques such as factor analysis [42]. The KMO score is always between 0 and 1, and values more than 0.6 are highly desirable as a measure of how well-suited our data are for factor analysis [45].

Table 1 includes the co-linearity variable obtained from these tests, which indicates how strongly a single variable is correlated with the other variables. There are eight observed variables in the pre-processed dataset. It can be observed that age, disability, barriers and the four levels of the frequency of encountering barriers can all influence the accessibility experiences of persons with disability. Since these variables are correlated, we can form factors to explain the common variance, i.e., variance due to correlation among these features [41].

4.3. Interpretation of the EFA Results

In the EFA method, every factor explains a certain variance in observed variables. The factors with the lowest variance were dropped to obtain optimal number of factors, i.e., four in our case.

4.3.1. Description of the EFA Components

Here, we interpret the factors influencing variable correlation by making use of components such as loadings, variance, eigenvalues and communalities, as included in Table 2.

- **Factor loading:** is a matrix that shows the relationship of each variable to the underlying factor. It shows the correlation coefficient for an observed variable and factor, along with variance explained by the observed variables [41].
- **Eigenvalues:** represent variance explained by each factor against the total variance, also known as characteristic roots. We selected a number of factors whose eigenvalues are greater than 1 [41]. In a standard normal distribution with a mean of 0 and standard deviation of 1, the variance is 1 [42]. Since we standard-scaled the data, the variance of a feature is 1. This is the reason for selecting factors whose eigenvalues (variance) are greater than 1, i.e., the factors that explain more variance than a single observed variable.
- **Communalities:** are the sum of the squared loadings for each variable. They represent the common variance [41,42], ranging from 0 to 1, with a value close to 1 representing more variance.

Table 2. EFA parameters for the frequency of encountering barriers.

Variable	Factor Loadings				Communalities
	Factor 1	Factor 2	Factor 3	Factor 4	
Age	0.8270	0.0228	0.1227	0.0258	0.7002
Gender	−0.0067	−0.0018	0.0423	−0.2223	0.0512
Disability	0.0022	0.0746	−0.0438	0.0277	0.0083
Barriers	0.0094	0.0052	0.0423	−0.0024	0.0019
Never	0.4332	0.5239	0.0901	0.6612	0.9074
Sometimes	0.1653	0.9587	0.2186	−0.0425	0.9961
Often	−0.5055	−0.1514	0.7928	−0.1716	0.9365
Always	−0.5543	−0.3019	0.2974	−0.1847	0.5209

4.3.2. Determining the Optimum Number of Factors

The factors and their eigenvalues, which represent the amount of variance the factor explains, are plotted in Figure 4. We selected the number of factors whose eigenvalues are greater than 1, since in a standard normal distribution with a mean of 0 and standard deviation of 1, the variance is 1.

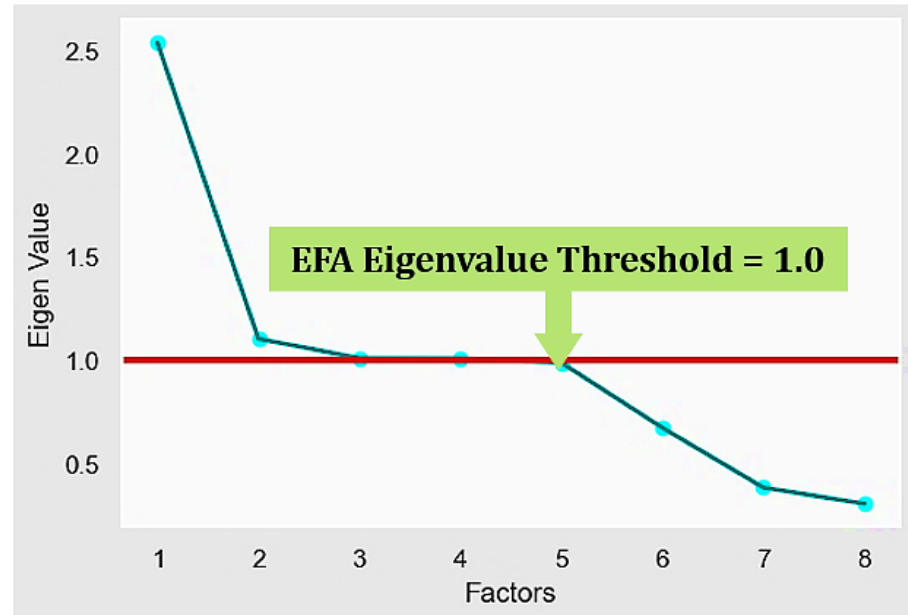


Figure 4. Exploratory factor analysis: The scree plot shows the effect of eigenvalues with an increase in the number of factors.

The number of factors is equal to the number of variables in our dataset. Not all of the factors provide a significant amount of useful information about the common variance among the variables, so we decided the number of factors on the basis of the common variance the factors explain [41].

For our data, the eigenvalues are greater than 1 for four factors (or unobserved variables), given the eight observed variables. This matches the number of factors that we have in our model, as shown in Figure 5. For example, factor 4 has an eigenvalue of 2.5k = m whereas the eigenvalue for factor 5 is 0.99, which is less than 1.

Considering only the original eigenvalues from the graph in Figure 4, we can see that the eigenvalues drop below 1 starting for the fifth factor for dimension reduction. Therefore, four latent variables were selected for SEM in Figure 5 based on these EFA results.

The EFA method examined the data to find possible underlying factors given a set of observed variables. This helped with data interpretations by reducing the number of unobserved variables or factors. After extracting the maximum common variance from all variables, there were four optimum latent factors discovered. This simplification made the data more manageable and easier to interpret to define the structure of our SEM model.

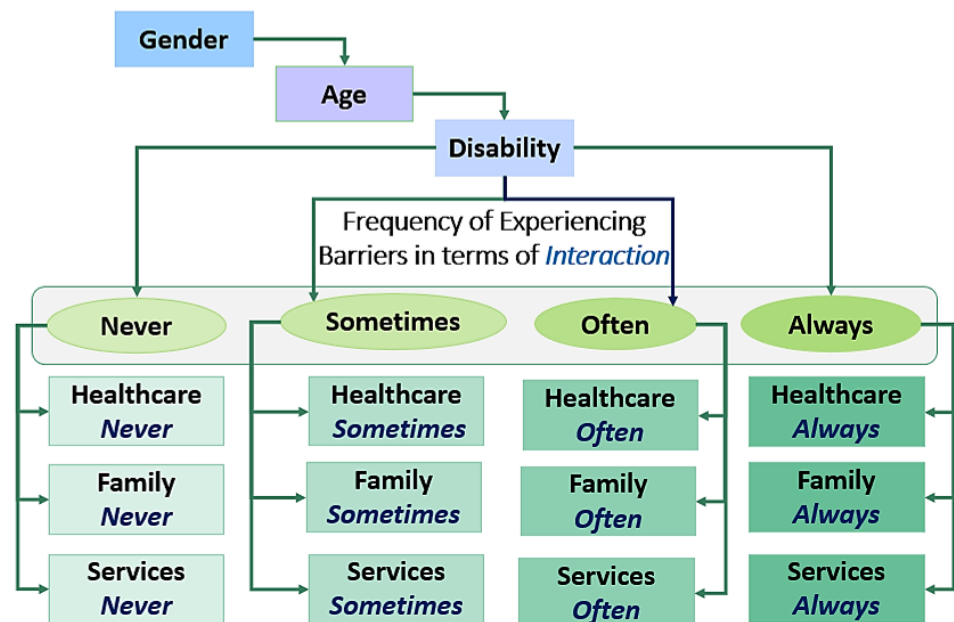


Figure 5. SEM structure 1: SEM application for measuring the correlation between accessibility experience and overall user attributes for *interaction*-related barriers. Each rectangle is an observed variables with multiple subcategories, whereas the oval shapes represent the latent variable of *accessibility experience* attributed to *gender*, *age*, *disability*, *barriers* and all four responses for *frequency*.

5. SEM Architecture for Statistical Analysis

SEM is a multivariate analysis technique widely applied in the field of social sciences [32,46]. Its factor analysis element allows for assessment of the relationships among different observed variables (available data) through causal networks. These relationships allow for understanding of the unobserved or latent variables. The multiple regression element estimates the hypothesized causal relationships among variables to be tested [47].

5.1. Application of the SEM Method

In order to define the structure in Figure 5 to model the dataset, we used the Semopy tool in Python [48], which facilitates parameter estimation through a restricted Maximum Likelihood Estimation (MLE) approach while simultaneously obtaining unbiased estimates to analyze structural relationships among different variables in the data.

5.1.1. Measurement Model

The definition of the SEM structure typically consists of left-sided values and right-sided values separated by an operator character that constitutes the nature of the relationship between them [30,48].

The model was outlined by determining the number and relationships of measured and latent variables. Only the variables that are valid indicators of the constructs of disability and accessibility were selected carefully for SEM. Measured or observed variables are presented in the dataset with values, whereas latent variables are usually inherent in the data. The latent variables were chosen to quantify factors that can only be understood as a concept but cannot be measured. Next, a path diagram depicting the structural and measurement models was generated for SEM analysis.

5.1.2. Defining the Variables

To define the structure, the data used for factor analysis were considered for SEM. The purpose of the SEM analysis was to assess the holistic causal connections among the given

personal attributes. Each barrier scenario includes four separate data points based on the frequency of encountering them in the PD.

The extent of the barriers lies in the frequency with which they were experienced, i.e., if 80% of a barrier had been reported as **never** being experienced, then it would be less significant than one that had been reported as **always** being experienced in a majority of responses. Hence, our goal is to assess the barriers in terms of the frequency of encountering them, and as shown in Figure 5, we used frequency to center the SEM. The latent variables for this model are *never*, *sometimes*, *often* and *always*, measured by the observed variables for each barrier category as described in Figure 3.

The goal was to explore all possible structures in order to determine the most efficient one reflecting the true nature of the variable dynamics while minimizing the number of insignificant relationships. Therefore, another structure that we regarded is shown in Figure A1 with latent variables of *adjustment*, *understanding* and *assistance* to reflect issues people with disabilities can experience while interacting in person with *healthcare*, *family* and *services*, respectively.

5.2. Analysis of the SEM Results

SEM models determine the relationship between the measured and latent variables [30,46–48]. In the SEM definition, we attempted statistical modeling with two structures (Figures 5 and A1). Since the results from of EFA suggest four possible latent variables, we selected the first structure for our analysis.

5.2.1. Description of the SEM Parameters

P-values and regression estimates collectively determine which relationships in the model are statistically significant and their nature [29,30,47]. To investigate the effects of the independent variable on the dependent ones, parameter analysis is presented in Table 3 for all barriers based on the model in Figure 5.

- **Regression Estimate (RE):** The primary result of an SEM analysis is a set of estimates of the regression coefficients. The sign of the estimate tells us whether there is a positive or negative correlation between variables [29,30,46]. Equation (1) expresses the formula for calculating the estimate (β), given an independent variable (X), a dependent variable (Y) and their means ($\bar{X}$ and $\bar{Y}$, respectively) [49].

$$\beta = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} \quad (1)$$

- **Standardized Regression Estimates (SRE):** Since the regular estimates use different scales for different variables in a model, they cannot be compared to identify their level of importance. However, standardized estimates all use the same scale and can be compared to determine which independent variables are crucial in the model [46–48]. The standardized estimate (β^*) is calculated by multiplying the estimate (β) by the standard deviation of X and the standard deviation of Y [29,49] as shown in Equation (2).

$$\beta^* = \beta \times \frac{\sigma_X}{\sigma_Y} \quad (2)$$

- ***P*-value:** The *p*-value indicates whether relationships between variables are statistically significant or not. It measures how likely it is that any observed association between variables is due to chance [29,47]. A common threshold for the *p*-value is 0.05, stating that 5% of the time, variable correlation might have been falsely concluded.

The first variable in the measurement model for each barrier category does not produce any *p*-value, as seen in Table 3. A *p*-value > 0.05 indicates an insignificant relationship [50].

Table 3. SEM Parameter analysis for the frequency of encountering barriers.

Barrier Scenario	Regression Analysis for Frequency of Encounter							
	Never		Sometimes		Often		Always	
	SRE	p-Value	SRE	p-Value	SRE	p-Value	SRE	p-Value
Attitude (Family)	0.9923	-	0.9355	-	0.7295	-	0.7086	-
Attitude (Healthcare)	0.9987	0.0001	0.9605	0.0025	0.7630	0.0002	0.7530	0.0005
Attitude (Business)	0.9993	0.0002	0.9443	0.0012	0.7065	0.0003	0.4539	0.0066
Alarms	0.9996	0.0011	0.7371	0.0019	0.6466	0.0004	0.8754	0.0017
Lighting/ Sound	0.9997	0.0003	0.9395	0.0013	0.8613	0.0007	0.9030	0.0004
Videoconferencing	0.9996	0.0021	0.9230	0.0003	0.7604	0.0005	0.5729	0.0048
Automated message	0.9973	0.0023	0.9175	0.0001	0.8475	0.0006	0.5274	0.0344
Interaction (Family)	0.9938	0.0014	0.9547	0.0009	0.7543	0.0013	0.9178	0.0045
Interaction (Services)	0.9967	0.0001	0.9823	0.0011	0.8688	0.0001	0.8806	0.0002
Interaction (Healthcare)	0.9966	0.0001	0.9708	0.0002	0.8663	0.0001	0.9124	0.0003
Signs	0.9988	0.0001	0.8289	0.0009	0.5889	0.0106	0.7729	0.0001
Traffic signals	0.9977	0.0010	0.2630	0.0017	0.2268	0.0071	0.6647	0.0244
Online (Business)	0.9986	0.0023	0.4603	0.0120	0.1098	0.1933	0.6428	0.0198
Online (Booking)	0.9986	0.0021	0.5110	0.0383	0.2354	0.0051	0.8007	0.0001
Online (Services)	0.9986	0.0002	0.6653	0.0010	0.4839	0.0062	0.9017	0.0029
Phone (health care)	0.9989	0.0011	0.9876	0.0001	0.8989	0.0007	0.8217	0.0007
Phone (Family)	0.9974	0.0003	0.9774	0.0013	0.9036	0.0001	0.8701	0.0002
Phone (Services)	0.9980	0.0004	0.9842	0.0001	0.9005	0.0012	0.7396	0.0002
Doorway	0.9898	0.0001	0.7763	0.0002	0.7069	0.0001	0.6429	0.0019
Sidewalks	0.9882	0.0003	0.8821	0.0002	0.7557	0.0001	0.5785	0.0057
Sidewalks (snow)	0.9790	0.0001	0.9194	0.0005	0.6211	0.0051	0.1154	0.1702
Washrooms	0.9972	0.0013	0.7480	0.0001	0.6924	0.0002	0.8274	0.0002
Queues	0.9904	0.0001	0.9352	0.0006	0.7060	0.0001	0.5674	0.0143
Self-serve	0.9970	0.0019	0.8059	0.0103	0.7447	0.0001	0.7377	0.0002
Social media	0.9996	0.0210	0.7963	0.0001	0.4823	0.0069	0.8881	0.0011

SEM parameter values below the threshold are highlighted in red in the table.

5.2.2. SEM Parameter Analysis

The SEM parameters include the standardized regression coefficients or simply *co-efficient* and *p*-values derived from the analysis of data on experiencing barriers: *never*, *sometimes*, *often* and *always* for all observed barrier scenarios. The threshold for significance is greater than 0.50 for *SRE* and 0.05 for the *p*-value. The values that are not significant are highlighted in red.

Communication Barriers: Accessibility in communication via *interaction*, *phone*, *automated message* and *videoconferencing*, as shown in Table 3, implies that all relationships are significant, with *p*-values being less than 0.05. All barrier scenarios with *SRE* values close to 1 imply strong connections between the disability type and barriers while accessing services for persons with disabilities.

ICT Barriers: Most *p*-values for ICT-related barriers encountered through online services using are below 0.0200 (Table 3). However, the frequency category of *sometimes* for *online (business)* has a *p*-value of 0.1933, which is above the threshold, rendering this connection insignificant.

Transportation Barriers: The parameters tabulated in Table 3 show that disability type significantly relates to transportation, where *physical*, *navigation* and *announcement*-related barriers can be experienced. The majority of *p*-values are below 0.03, with a *p*-value of only 0.1702 for *sidewalks (snow):always*.

Observing the statistical modeling analysis, the majority of *p*-values are less than 0.05, indicating that the associations between disabilities and barriers are rationally well-

grounded. The SEM confirms the assumed strong correlation between the frequency of experiencing accessibility barriers and the type of disability. The analysis confirms the hypotheses about the valid connection between specific accessibility needs that comes with the varying nature of disabilities and the nature of barriers experienced due to a lack of accommodation in federal sector facilities and services.

6. Probabilistic Modeling with Causal Bayesian Network

This section presents the application of Bayesian networks (BNs) and experimental results of causal Bayesian inference to evaluate the probabilities of various accessibility scenarios. The objective of this application is to examine how a BN can serve as an effective system for decision making in social development, including with respect to accessibility in federal sector organizations.

6.1. BN Overview

Although SEM techniques have been utilized to learn causal relationships from data, they do not entail the predictive and diagnostic capabilities of BNs when handling non-linear relationships [34,46]. Thus, we apply BNs for reasoning under uncertainty conditions, which, in our dataset, may exist due to limited knowledge about systems, an ambiguous understanding of disabilities or randomness in the demographic attributes driving accessibility experiences [28,34,35].

The network was modeled after the SEM structure shown in Figure 5. For demonstration, the BN for barriers related to in-person *Interactions* is illustrated in (Figure 6). Here, the graph represents connections between each node, both causal and non-causal. The causality between nodes such as *Disability* and the nature/ frequency of *Barriers* would be assumed during inference later in this process.

6.1.1. Fundamental Concepts

A BN represents conditional probability and causality relationships between variables expressed by two key concepts in probability: conditional probability and Bayes' theorem.

- **Conditional probability** is a measure of the change in likelihood of an event in relation to information about another event, as expressed by the following equation [51].

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (3)$$

This states that the probability of event A occurring given that event B has occurred is equal to the probability of both events divided by the probability of event B [51]. The calculation for conditional probability involves the joint probability of both events and the probability of the given event.

- **Bayes' theorem** propagates probability through the model, which relates the conditional and marginal probabilities of random events. It is an equation that allows for the manipulation of conditional probabilities in a BN [52]. The theorem is expressed below and is used to update the probabilities of hypotheses based on new evidence:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (4)$$

Here, A and B are events and $P(B) \neq 0$. $P(A|B)$ and $P(B|A)$ are conditional probabilities that indicate the probability of an event occurring given another event, also called posterior probability. $P(A)$ and $P(B)$ are the probabilities of observing A and B, respectively, without any given conditions, also known as the prior and marginal probability respectively [53].

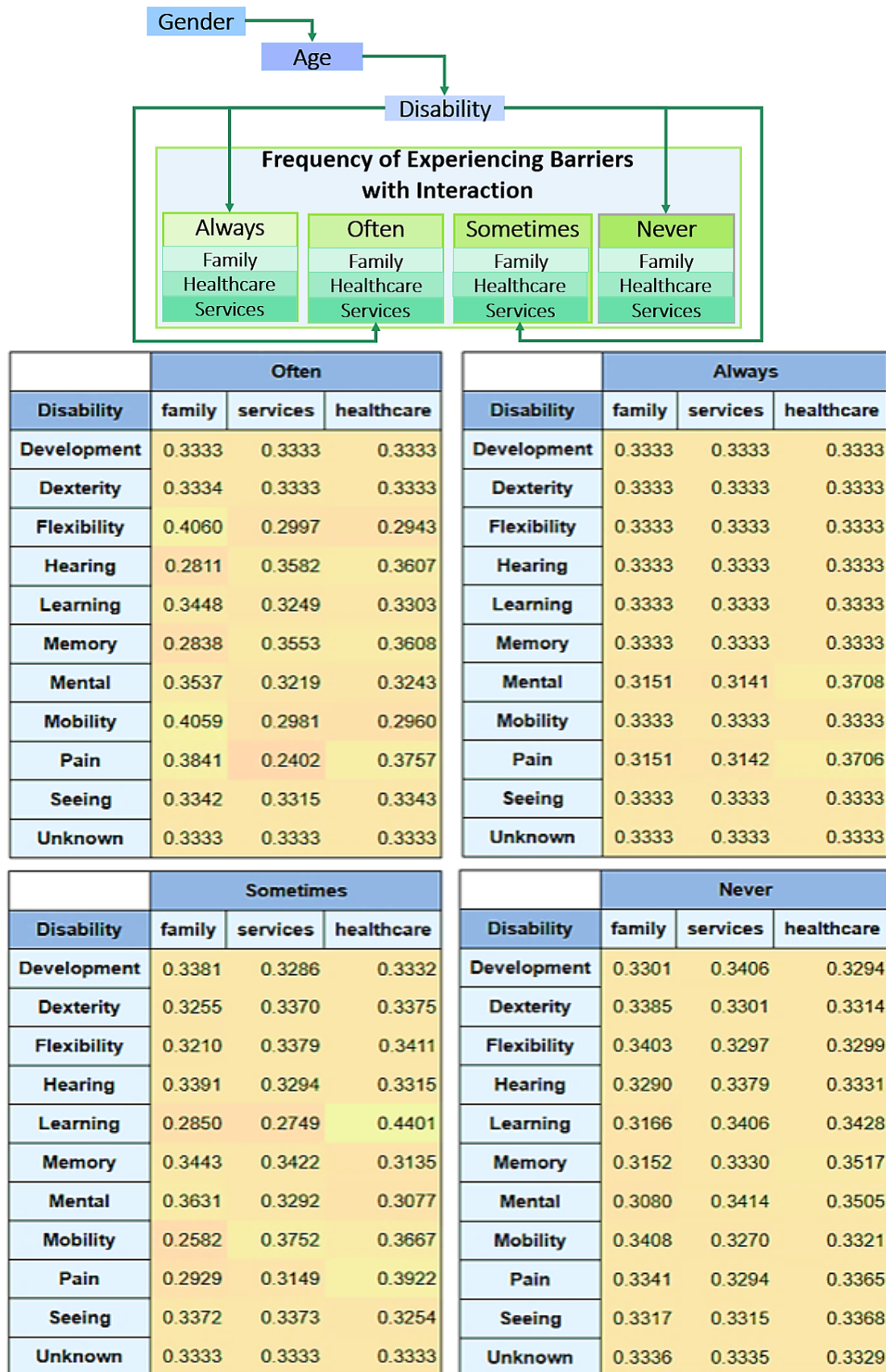


Figure 6. BN probabilistic graph model designed according to the SEM structure in Figure 5, showing dependencies among the data variables. The Conditional Probability tables (CPT) contain probabilities (color-coded values between 0 and 1) of barriers under 4 nodes associated with the frequency of experiencing barriers (*never*, *sometimes*, *often* and *always*) under the parent node of *disability*. Here, shades closer to red is the lower and closer to green represent higher probabilities.

6.1.2. Graphical Model

The BN model consists of a qualitative part to represent the dependencies and a quantitative part to specify the probabilistic relationships.

- **Graph Structure:** A BN is a type of causal probabilistic model that is described qualitatively by a directed acyclic graph as illustrated in Figure 6 for in-person *Interaction*. The vertices of the graph representing variables are called nodes. The nodes are represented as circles containing the variable name. The connections between the nodes are called arcs or edges, which are drawn as arrows between the nodes representing dependence between the variables. Therefore, for any pair of nodes indicating that one node is the parent of the other, there is no independence assumptions, which are implied in a BN by the absence of a link. Moreover, a conditional probability table (CPT) is associated with each node containing the probability distribution of the node given its parent nodes [54].
- **Conditional Probability Tables (CPTs):** Once the network topology is constructed, the CPTs are initialized. The values in a CPT represents prior probabilities used in the inference, which is the calculation of posterior probabilities using Bayes' formula [28]. In Bayesian statistics, a posterior probability is the revised or updated probability of an event occurring after factoring in new information [55]. To draw conclusions about the causal dependencies, the posterior probabilities are deduced using prior knowledge with Bayesian inference [54].

6.2. BN Experimental Results from Causal Inference

By representing conditional dependence, probabilistic graphical models such as BNs allow for the identification of causality from data via inference for the probability computation [56] of frequency of experiencing barriers depending on disability type. Here, the BN application is analyzed in two primary ways for prediction and diagnosis [57]:

1. **Diagnostic reasoning** involves inspecting causality by reasoning backwards from the observed effects of an event.
2. **Predictive reasoning** involves obtaining the expected probabilities of certain outcomes given some observed evidence.

6.2.1. Diagnostic Reasoning with Evidence of *Barriers*

The diagnosis of accessibility experiences is carried out with BN inference to learn about the likelihood of a demographic trait being present in individuals experiencing specific barriers. The probability of the cause (*disability*) is continuously shifting either upward or downward, depending on the information gathered about the effect (*barriers*) during the process [58].

The purpose of diagnostic testing is to move from the probability of disability before considering the causality with barriers (prior probability) to the probability after considering the causality with barriers based on the report/data (posterior probability) [58] using Bayes' theorem (Equation (4)). This aims to determine the most and least considerable effects of the selected barriers on disabilities.

- **Demographic traits correlated with *interaction barriers*:**

The inference of *disability* is illustrated in Figure 7, assuming the evidence of *interaction barriers* to estimate posterior probabilities for all 11 types of *disability* (*development, dexterity, flexibility, hearing, learning, memory, mental health, mobility, pain, seeing and unknown*).

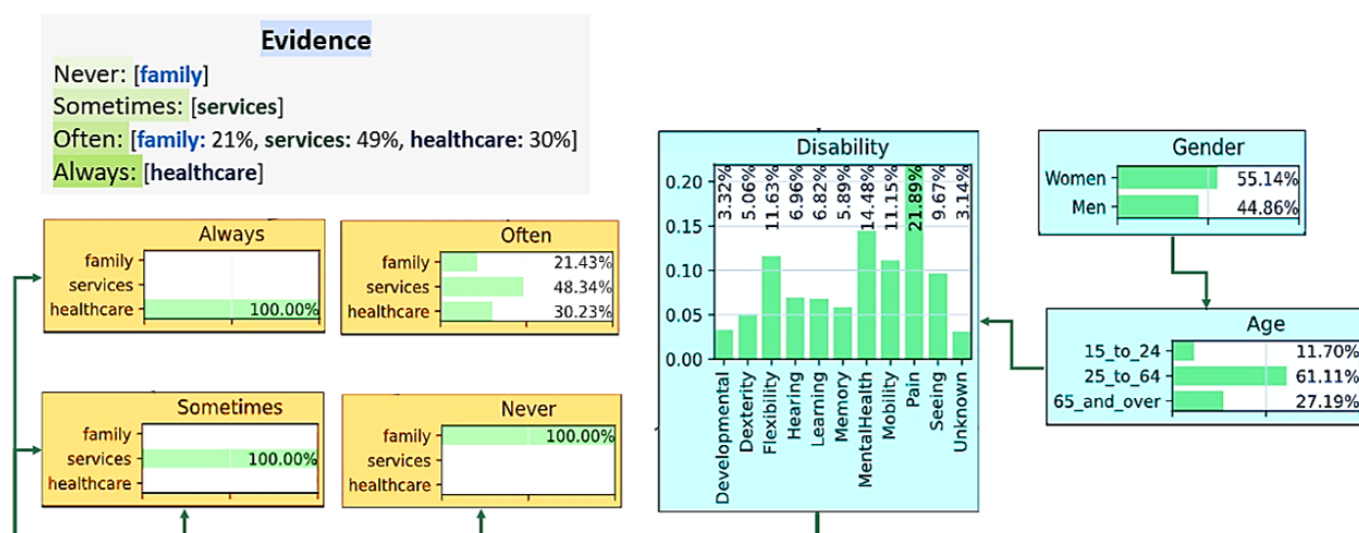


Figure 7. BN inference: likelihood of disability for persons encountering communication barriers during *interaction* with family, healthcare and other services. Evidence of the nature and frequency of barriers (rectangles highlighted in orange) was randomly selected to demonstrate the diagnostic reasoning.

To update the probabilities, *interaction barriers* were considered to deduce their effects on demographics with disabilities. For example, $P(\text{Disability} | \text{Never} = \text{family})$ is the expression for the evidence of *never* in the BN definition to predict the chances of a person having various disability types when no barrier is experienced interacting with family.

In Figure 7, although the partial evidence is defined initially as whole numbers for *often*, such as 20% for *family*, the algorithm adjusts it to a decimal value, such as 21.43%, to match real-world conditions.

While individuals with *pain-related disability* (21.89%) are most likely to have this experience, *mental health* (14.48%) and *mobility* (11.63%) also seem to have a substantial likelihood. Also *Women* between the ages of 25 and 64 years also appear to be most vulnerable when it comes to facing challenges with in-person interactions.

- **Likelihood of Disability types for each barrier scenario:**

The holistic process of updating the posterior probabilities of the BN *disability* parent node is executed by providing evidence of barriers experienced in specific scenarios. The overall inference results are included in Table 4 for each type of disability when barriers are encountered *always*.

The observations from these updated posterior probabilities indicate that for attitudinal barriers (*attitude (business)*, *attitude (family)* and *attitude (healthcare)*), an individual is most likely to have *pain-related disabilities*. Communications scenarios seem to mostly affect persons with the following disabilities: *videoconferencing—learning* (20.80%); *interaction with family—mental* (33.82%), *services—mobility* (20.59%) and *healthcare—pain* (22.73%); *automated message: mental* (25.62%). All issues related to *phone with family, services and healthcare* show the greatest likelihood for *pain-related disabilities*.

Additionally, ICT barriers appear to impact persons with *pain-related disability* (*online (booking)*: 34.49%; *social media*: 32.69%; *online (services)*: 18.49%). For *self-serve technology*, *flexibility disability* has the highest probability (26.11%).

Furthermore, Physical and transportation-related issues yield the greatest probability for individuals with *pain-related disability* (*traffic signals*: 37.15%; *wait lines or queues*: 28.67%; *level of lighting/sound indoors*: 24.94%). Substantial effects can be observed for *flexibility dis-*

ability (doorways: 23.50%; sidewalks: 19.35%). Mental health issues seem to be greatly affected by barriers pertinent to announcements or *alarms* (28.59%) and *signs* or direction (21.94%).

Table 4. Probability of an individual having a disability given *always* encountering a specific barrier.

Barrier	Development	Dexterity	Flexibility	Hearing	Learning	Memory	Mental	Mobility	Pain	Seeing	Unknown
Attitude (business)	0.0371	0.0539	0.0005	0.2261	0.0587	0.1584	0.0139	0.1946	0.2399	0.0106	0.0063
Attitude (family)	0.0553	0.0404	0.0168	0.0988	0.0081	0.0129	0.0655	0.2552	0.2900	0.1567	0.0002
Attitude (healthcare)	0.0250	0.0200	0.1071	0.0408	0.1593	0.0718	0.1484	0.0657	0.2341	0.1047	0.0232
Alarms	0.0566	0.0641	0.1485	0.0822	0.0134	0.1113	0.2859	0.0918	0.0261	0.0560	0.0640
Lighting/ sound	0.0069	0.0760	0.2066	0.0341	0.1154	0.0768	0.0858	0.1305	0.2494	0.0033	0.0153
Video conferencing	0.0370	0.0964	0.0772	0.0949	0.2080	0.0888	0.0738	0.0829	0.1740	0.1443	0.0228
Automated message	0.0428	0.0734	0.1634	0.1259	0.0109	0.0900	0.2562	0.0192	0.0582	0.1298	0.0301
Interaction (family)	0.0064	0.0461	0.1663	0.0062	0.0976	0.0069	0.3382	0.1067	0.0074	0.1450	0.0732
Interaction (services)	0.0573	0.0891	0.0183	0.0142	0.0229	0.0913	0.1627	0.2059	0.1458	0.0801	0.1123
Interaction (healthcare)	0.0282	0.0477	0.1208	0.0073	0.1228	0.0563	0.2230	0.0529	0.2273	0.0996	0.0142
Signs	0.0297	0.0652	0.1701	0.1246	0.1000	0.0226	0.2194	0.0460	0.0884	0.1276	0.0064
Traffic signals	0.0574	0.0506	0.0173	0.0036	0.1042	0.0689	0.0777	0.0666	0.3715	0.1198	0.0624
Online (booking)	0.0261	0.0170	0.0750	0.0694	0.0982	0.0695	0.1188	0.0440	0.3449	0.1011	0.0359
Online (services)	0.0341	0.0347	0.0669	0.1295	0.0412	0.0932	0.1599	0.0955	0.1849	0.0925	0.0675
Doorways	0.0712	0.0117	0.2350	0.0392	0.0394	0.0568	0.0985	0.0874	0.1802	0.0939	0.0868
Sidewalks	0.0813	0.0070	0.1935	0.0796	0.1017	0.0145	0.0752	0.1587	0.1216	0.1538	0.0131
Washrooms	0.0083	0.0718	0.1580	0.0785	0.1133	0.0953	0.0779	0.1356	0.1704	0.0324	0.0584
Queues	0.0371	0.0155	0.0782	0.0359	0.0477	0.0260	0.2211	0.1527	0.2867	0.0908	0.0083
Phone (healthcare)	0.0602	0.0368	0.1450	0.0596	0.0676	0.0034	0.1032	0.1774	0.3101	0.0368	0.0001
Phone (family)	0.0360	0.0382	0.0627	0.0678	0.0869	0.0853	0.1584	0.1026	0.2438	0.1021	0.0162
Phone (services)	0.0558	0.0551	0.0435	0.0539	0.0673	0.0761	0.1407	0.1181	0.2515	0.1106	0.0274
Self-serve tech	0.0819	0.1038	0.2611	0.1701	0.0771	0.0534	0.0443	0.0965	0.0420	0.0374	0.0324
Social media	0.0150	0.0270	0.1427	0.0382	0.0643	0.0083	0.2294	0.0253	0.3269	0.1095	0.0134

For each barrier, the sum of the probabilities across all disabilities is 1. The highest probability for each barrier scenario is highlighted in green.

6.2.2. Predictive Reasoning with the Evidence of Disability

This section describes BN inference for the nature and frequency of experiencing barriers given the presence of an attribute such as *disability* (Figure 8 and Table 5). To measure the likelihood of an event given that another event has occurred, conditional probability (Equation (3) and Table 5) is employed. The evidence used here is one or a combination of multiple attributes, including hard (100% certainty) and soft (with some level of uncertainty) evidence.

- **Interaction barriers for particular demographics:** For the inference in Figure 8, we use hard evidences for each attribute node, using the BN for *interaction* barriers in Figure 6. The algorithm estimates probabilities of encountering various barriers based on specified attributes i.e., *women* with *mobility* disabilities aged *15 to 24* years. The results show that this group is most likely to face barriers from interaction with *family* (**sometimes** (48%); **often** (72%); **always** (57%)) and *healthcare* (**never** (39%)).
- **Interaction barriers for persons with mobility disability:** Based on the Bayesian analysis through inference, the probability of an individual with *mobility* disabilities experiencing each type of barrier is illustrated in Table 5. In the table, each column represents a barrier scenario except for 'Frequency'. The evidence is randomly assumed here, pertaining to the parent *disability* parent node with subtypes to predict the

probabilities for child nodes (frequency) *never*, *sometimes*, *often* and *always* subject to those types.

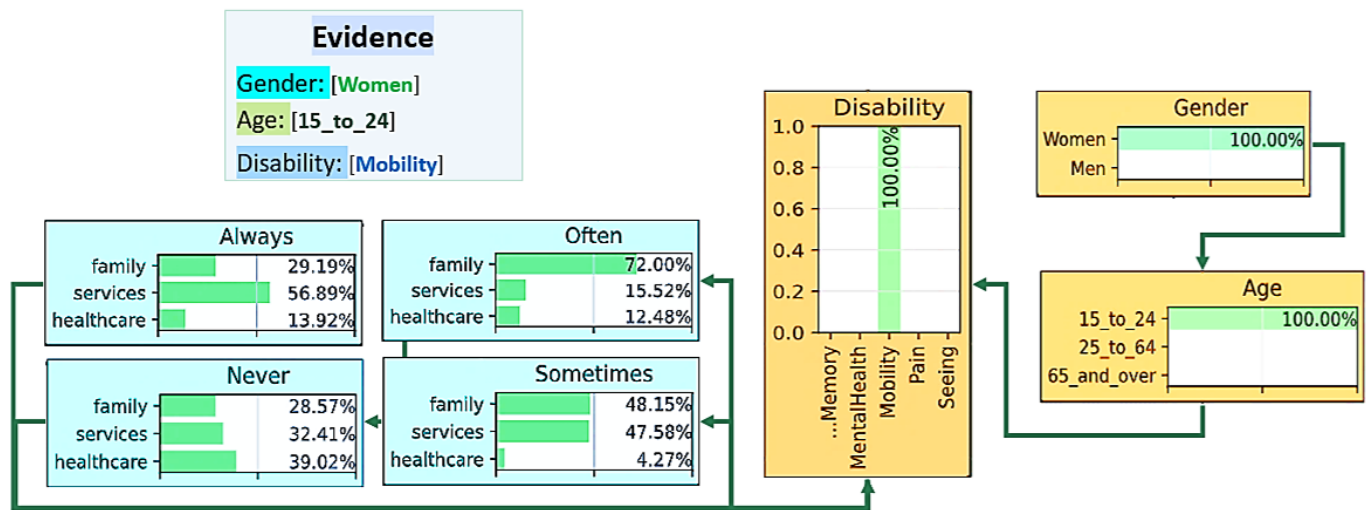


Figure 8. BN inference: probability of encountering barriers to communication via *interaction* with *family*, *healthcare* and other *services*. Only hard evidence (100% probability of the presence of an attribute) is randomly selected here (rectangles highlighted in orange). The inference of experiencing barriers by Women between the ages of 15 to 24 with *Mobility* disability is illustrated here.

Table 5. Causal inference: overall probability of encountering barriers for persons with *mobility* disabilities.

Frequency	Attitude (Family)	Attitude (Healthcare)	Attitude (Business)	Alarms	Lighting /Sound	Video Conference	Automated Message	Interaction (Family)	Interaction (Services)	Interaction (Healthcare)
Never	0.0310	0.0714	0.0432	0.0760	0.0326	0.0503	0.0601	0.0659	0.0668	0.0314
Sometimes	0.0334	0.0133	0.0354	0.0036	0.0169	0.0596	0.0182	0.0339	0.0777	0.0787
Often	0.0409	0.0595	0.0759	0.0105	0.0765	0.0113	0.0564	0.0188	0.0209	0.0630
Always	0.0950	0.0252	0.0534	0.0260	0.0533	0.0291	0.0088	0.0400	0.0490	0.0278

Frequency	Signs	Traffic Signals	Online (Booking)	Online (Service)	Door Ways	Side Walks	Wash Rooms	Queue	Phone (Healthcare)	Phone (Family)	Phone (Service)	Self-serve Technology	Social Media
Never	0.0660	0.0732	0.0462	0.0065	0.0623	0.0238	0.0771	0.0095	0.0358	0.0202	0.0442	0.0042	0.0023
Sometimes	0.0777	0.0308	0.0688	0.0733	0.0010	0.0299	0.0087	0.0788	0.0746	0.0188	0.0276	0.0690	0.0705
Often	0.0646	0.0768	0.0297	0.0120	0.0374	0.0073	0.0603	0.0166	0.0622	0.0376	0.0142	0.0770	0.0709
Always	0.0218	0.0316	0.0228	0.0445	0.0279	0.0561	0.0625	0.0794	0.0783	0.0650	0.0575	0.0321	0.0130

Results from this inference indicate that attitudinal barriers when dealing with family and friends (*attitude (family)*; Table 5) have the greatest likelihood of occurring among all other barriers for a person with mobility issues. *Queue* and accessing healthcare via phone (*phone (healthcare)*) also show significant probabilities for *always* encountering a barrier.

The overall analysis shows that individuals with conditions related to *mental health* and *pain* are more susceptible to barriers, among other conditions. Moreover, persons with disabilities between the ages of 15 and 24 years are most likely to have conditions related to *mental health*, while those aged between 25 and 64 years are most likely to report *pain* and *mobility* disabilities. For persons with disabilities aged over 65 years, BN analysis infers that *men* are 3% more prone to overall barriers compared to *women*.

7. Synthetic Data Generation from BN

The methodology completes with the description of approaches to generating the SD and evaluation of the performance of the BN recreated using the SD. This paper aims to lay the groundwork for future research by introducing a framework to generate SD that mimics the statistical properties of the PD [38]. These data can be not only used for more

precise inference and to overcome limited data but also for the protection of any sensitive information about persons with disabilities.

7.1. SD Generation: Purpose and Overview

Synthetic data (SD) generation is the process of creating artificial data that mimic the statistical patterns and properties of real-life data using algorithms, models and other techniques [59]. The goal is to have realistic data, similar to the original data, for a greater sample size. Such data are needed for a more comprehensive analysis and to overcome the scarcity or limitations of real-life data in areas that are usually dependent on survey-based collection.

Moreover, the SD generation method aids in effectively testing the performance of the BN created earlier in Section 6. The data generation process directly in the network allows for observation of the precision of the parameter learning capabilities of the BN.

Since the PD is collected and released occasionally [7], some limitation in analysis or prediction may occur. Nevertheless, merging SD with the PD can help to obtain more accurate CPTs in the BN. This also leads to better inference and a more robust network. Hence, this avenue was explored for future use in a more comprehensive data analysis and for time-series forecasting of accessibility trends. The purpose, framework for data generation and validation are illustrated in Figure 9 and described in the following sections.

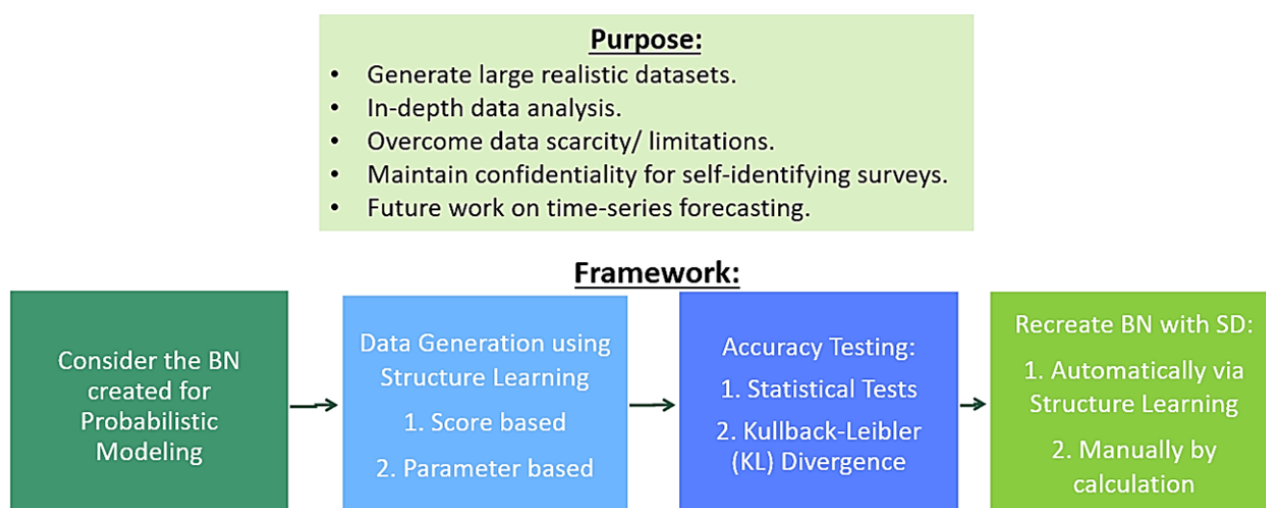


Figure 9. Synthetic data (SD) generation: purpose and methodological process for generative model application and validation.

7.2. Application of the SD Generation Method

The application of the SD generation technique, including the process for data generation and rebuilding of the BN model using the data, is detailed in this section.

7.2.1. Data Generation Process

To create the data, a subset of the PD is considered that only concerns barriers to *interaction* with *family*, *healthcare* and *services*. For the network in Figure 6, conditional probability were previously calculated and stored using the PD. Hence, these data were utilized to create the SD using the PyAgrum library in Python [54]. The BN was provided as a parameter for the learner in order to obtain the variables and the order of the labels for each variable.

With the network in place, the aim was to generate SD that maintain the statistical characteristics of the PD to achieve greater precision in the analyses [59,60]. The process is outlined in Figure 10. First, parameter and structure learning techniques are applied to

the BN to learn the variables and structure of the actual dataset. Parameter learning is the process of using data to learn the BN distributions. Structure learning picks up both the structure and parameters of the provided data [28].

Then, data are generated using the BN by specifying the desired data sample size and variable order. These SD are rearranged, with minor changes to the structure of the dataset to ensure that it completely aligns with the overall structural appearance of the PD. This also helps with the data validation process later.

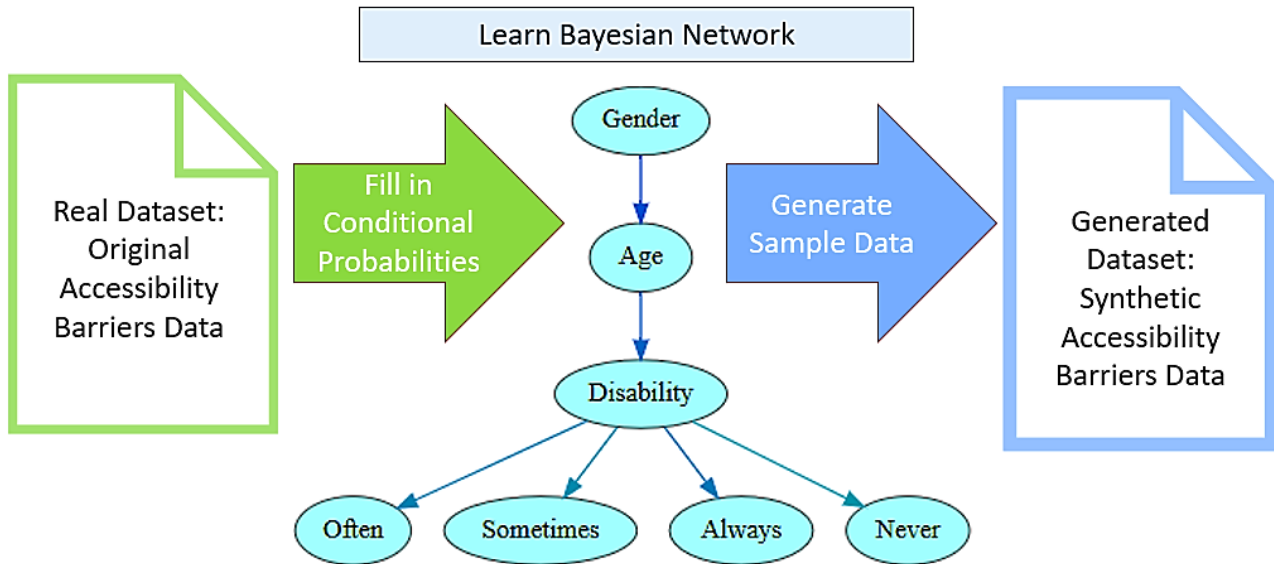


Figure 10. Synthetic data (SD) Generation: purpose and methodological process for the proposed framework.

7.2.2. Recreating the Network

The BN learning process gives access directly to the complete learning algorithm and its parameters, such as prior, scores and constraints [54]. It also facilitates the development of new learning algorithms through statistical tests and the creation of a database from the BN based on the conditional likelihood [54].

A. Automatic recreation via parameter learning:

To recreate the BN from SD, the type of learning algorithms (both parametric and score-based), priors (such as smoothing), constraints, order of the variables and conditional likelihood were specified. This way, the algorithm was able to learn the structure and CPTs of the BN from the provided data.

Structure learning refers to learning of the structure of the DAG from data. We employ this method to estimate the graph by determining how the variables depend on each other based on the SD. Parameter learning is applied when the graph structure is known in order to estimate the factors. A combination of the score-based and constraint-based approach was utilized for this.

- **Score-based learning:** To score and selecting a model, a score-based approach, namely the Bayesian Information Criterion (BIC), was employed. It is a statistical technique to estimate how well a given model fits a dataset and the model's complexity. It is derived from Bayesian probability and is appropriate for models under the MLE framework. The BIC statistic is expressed as follows [61]:

$$BIC = -2 * LL + \log(N) * k \quad (5)$$

Here $\log()$ has a base-e called the natural logarithm, LL is the log likelihood of the model, N is the number of examples in the training dataset and k is the number of parameters in the model. The score defined above is minimized, e.g., the model with the lowest BIC is selected [61].

- **Constraint-based learning:** For the constraint-based approach, the Multivariate Information-based Inductive Causation (MIIC) is used in recreating the BN. Starting from a complete graph, the method iteratively removes dispensable edges by uncovering significant information from indirect paths. This also assesses edge-specific confidences resulting from the randomization of provided data. Remaining edges are oriented based on the signature of causality in the data [54,62,63].

B. Manual recreation via direct probability calculation:

The direct way to recreate the BN from the SD is to calculate the probabilities for different variable conditions and fill in the CPTs with them for each BN node. The process is similar to how the probabilities are calculated for the CPTs in Figure 6. Once the SD are generated, it is essential to evaluate how well they preserves the PD's characteristics. Hence, the synthetic data generation process is evaluated by comparing specific attributes between these two datasets.

7.3. Accuracy Evaluation of SD Generated by the BN

The experiments and statistical tests performed to validate the accuracy of the SD created from the BN learned from the PD are described in this section.

7.3.1. Using Structure and Parameter Learning Algorithm

The SD generated using the BN in Figure 6 is compared with the PD in Figure 11. Here, adjustment techniques are applied on the automatically recreated BN using the SD introduced earlier in this chapter. The BN was recreated from the SD through structure learning techniques using conditional probabilities. Then, Laplace smoothing was added as a prior to prevent the multiplication of probabilities from resulting in zero, ensuring that no feature was entirely disregarded [54].

SD adjustment: Laplace adjustment or Laplacian smoothing refers to an interactive mesh smoothing process that adjusts parameters such as a the weighting factor and number of iterations to achieve real-time adjustment. It involves utilizing suitable data structures to allow for efficient implementation of the smoothing algorithm [60].

$$P_{i,empirical} = x_i / N \quad (6)$$

$$P_{(i,\alpha-smoothed)} = x_i + \alpha / (N + \alpha d) \quad (7)$$

A notional value (α) weighs into the posterior distribution. If the frequency of each item (i) is x_i out of N samples, the empirical probability of event i is Equation (6). The process of the posterior probability being additively Laplace-smoothed, as if to increase each count (x_i) by α [60], is expressed by Equation (7).

In Figure 11, the red line graph shows the PD, while the dark-green and light-green bar plots illustrate probabilities for all 10 *disability* types. The values corresponding to the BN learned from SD both before and after smoothing (Equations (6) and (7), respectively) seem to be closest to the values when the PD was used to create the BN. The prior to smoothing, the bar plots show considerable fluctuations for some disability types, such as *pain*. The value greatly diverges from conditional probabilities in the PD, with a substantial decline from 0.36 to 0.09. In particular, *dexterity* and *flexibility* show probabilities that are evidently higher (4.2% and 6.7%, respectively) than those for the PD.

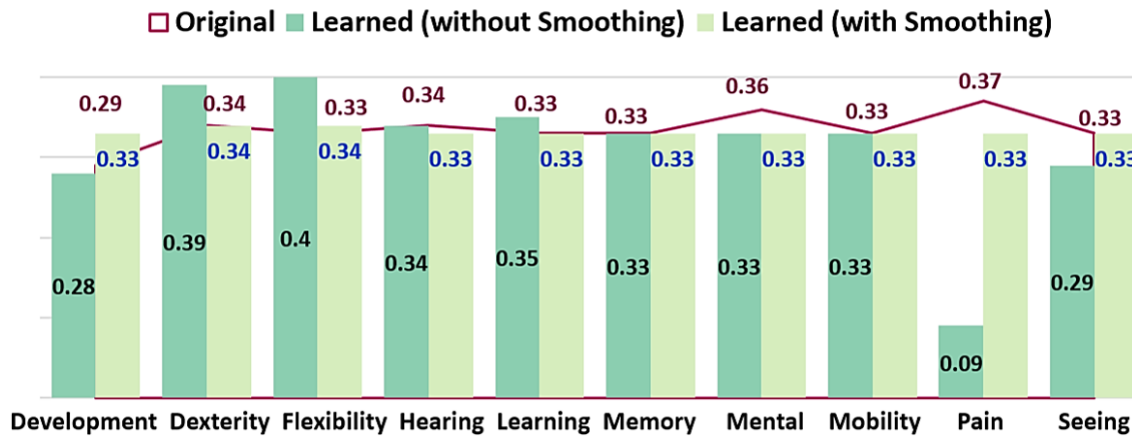


Figure 11. BN recreated from SD: comparison of conditional probabilities for *always: interaction (healthcare)* (Figure 6). Representation: red line plot—PD; dark- and light-green bar plots—SD prior to and post smoothing, respectively.

7.3.2. Hypothesis Testing

Statistical tests are performed to determine if the SD and PD are significantly different from each other. The assumptions that can be made [64] are whether:

- The two samples' data groups are independent;
- Respective data elements follow a normal distribution;
- Samples have equal variance (homogeneity assumption).

The original data include both categorical and numerical variables [40]. In reality, the SD include categorical variables with the same set of values or categories as the PD, but the numerical variable are different, since they are decimal values for probability, which, upon simulation, do not follow the same counts as the PD. Hence, the statistical tests differ based on the variable type. The numerical variables are tested using an independent T-test whereas the categorical data are tested with the Chi-square test. The results from these tests involve two hypotheses, H0 and H1 [64].

H0. *The variables have no relationship (independent).*

H1. *There is a relationship (dependent) between variables.*

All tests performed below use the *statistic* and *p*-value parameters to assess their results, similar to the SEM analysis earlier, where the concept of *statistic* is equivalent to RE in SEM. These tests how the SD conform to the hypothesis of no similarity relative to the original distribution in the PD.

1. Test for Numerical Data: To check the accuracy of the newly generated SD from the BN, we conduct a parametric test called a two-sample T-test in Python with the PD and the SD.

- **Two-sample independent T-test:**

The independent two-sample t-test (referred to as T-test here) allows for confirmation that the means of two datasets are equal [65] using the following formula:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}} \quad (8)$$

The T-test quantifies the difference between the arithmetic means of the two samples. Equation (8) includes the sample means (M_1 and M_2), variance or squared standard deviations (SD_1 and SD_2) and the sample sizes in the two data groups (n_1 and n_2) [66].

Before conducting the T-test, it needed to be checked if the given data groups had the same variance [67]. In Table 6, the quotient of variances is approximately 1, which qualifies for the homogeneity assumption [64]. The *statistic* and *p*-value parameters were ready to be calculated for numerical data with the condition met for the T-test.

Table 6. Accuracy evaluation for SD—numerical variable.

Variable	Variance (V)			Two-Sample T-Test	
	<i>o</i> = V (PD)	<i>s</i> = V (SD)	<i>o/s</i>	<i>Statistic</i>	<i>p</i> -Value
<i>Probability</i>	0.0013	0.0014	1.1228	0.5954	0.0468

The test results in a *p*-value of 0.0468 in Table 6, which is below the threshold of 0.05. This indicates that the respective distributions of the *Probability* numerical variable in the PD and SD are not equal (first hypothesis (H_0) in Section 7.3.2).

2. Test for Categorical Data: The distribution of a categorical variable in a sample often needs to be compared with the distribution of a categorical variable in another sample to examine their differences [65]. In this study, they are computed to contrast the PD and SD.

- **Pearson's chi-squared test:**

This is a hypothesis test that is used to determine whether there is a significant association between two categorical variables in the data [68]. The equation below expresses the test parameters, where χ^2 is chi-square or the *statistic* and O_i and E_i are observed and expected values, respectively.

$$\chi^2 = \sum (O_i - E_i)^2 / E_i \quad (9)$$

All categorical variables from both the PD and SD were accounted for in implementing the chi-square test to check whether there was a relation between the pair of variables with the same name from each dataset. This tests the null hypothesis (assumption) that the variables are independent.

The chi-square statistic is a non-parametric (distribution-free) tool that helps analyze differences within data distributions [69]. Like all non-parametric statistics, the chi-square is robust with respect to the distribution of the data and does not require equality of variance [70].

The results are included in Table 7; the *p*-values are insignificant for all variables (over the threshold of a *p*-value < 0.05, the chi-square test hypotheses do not hold, which implies that the values and distributions for the categorical variables are the same for both datasets).

Table 7. Accuracy evaluation for SD—categorical variables.

Parameter	Chi-Square Test Parameters				
	<i>Gender</i>	<i>Age</i>	<i>Disability</i>	<i>Frequency</i>	<i>Barriers</i>
<i>Statistic</i>	13.3469	18.7759	37.1769	19.6190	33.3070
<i>p</i> -value	0.9942	0.9269	0.1418	0.9044	0.2655

Hypothesis Testing Results: The results of the statistical tests, as included in Tables 6 and 7, indicate that there is significant difference between the two datasets for numerical data, whereas the categorical data are evidently similar. The *p*-values obtained with these statistical tests also confirm that. This aligns with the rationale that, evidently, the subcategories for each of these variables are exactly the same in both datasets. For

instance, the **age** variable only takes on any of the three subcategories from the list of *15_to_24*, *25_to_64* and *65_and_over* throughout the dataset in the SD. This is due to this data being generated from the BN that was initially learned from the PD.

7.3.3. Accuracy Estimation with the KL Divergence

To estimate the difference between the SD and PD, Kullback–Leibler (KL) divergence was utilized. It is used in AI as a loss function to compare predicted data with true values [71]. It can also be utilized to compare two BNs' probability distributions when one is a true distribution and the other one is an approximation of it [72].

The KL divergence can be expressed as the below equation to measure the dissimilarity of two probability distributions, P and Q [72].

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) \log\left(\frac{P(x)}{Q(x)}\right) \quad (10)$$

Here, X is a possible event in the probability space; $P(x)$ and $Q(x)$ are probabilities of x in P and Q , respectively; and $P(x)/Q(x)$ represents the relative likelihood of event x according to P in comparison with Q .

Exact and (Gibbs) approximated KL divergence

To compare the initial and previously recreated BN, the exact and Gibbs-approximated KL divergence are considered. Networks for both distributions have the same structure, but the values are different for every CPT.

- **Exact KL divergence:** Here, the brute-force KL divergence is calculated using Equation (10), which yields 0.0712, representing the exact distance between the true and randomly generated probability distribution [54].
- **Gibbs KL divergence:** For the approximated distance, the calculation uses pyAgrum's built-in function for KL divergence using Gibbs inequality [54], which states that the KL divergence is always positive ($KL(F|G) \geq 0$) unless $F(x) = G(x)$ [72]. Gibbs-approximated computation of the difference between the BN is depicted in Figure 12. In the computation of the distance for the KL divergence, the equations to convert the absolute value to log and vice versa are:

$$\begin{aligned} y &= \log_b x \\ x &= \log_b^{-1}(y) = b^y \end{aligned} \quad (11)$$

The divergence or distance between the original BN and the one recreated with the SD is plotted in Figure 12. The time period for each oscillation is 500 ms. Epsilon is a small value to address cases in the KL computation where division by zero occurs to avoid infinite distances.

In the plot, the divergence is calculated for each iteration and expressed on a logarithmic scale. As presented in Equation (11), y is the plotted divergence; b is the base of the logarithm, which is considered to be 10 here; and x is the actual divergence before converting to the log scale. Then, the anti-logarithm (inverse) can be calculated by raising the base (b) to the logarithm (y). In the figure, the longest distance can be seen at around -4.0 , which is equivalent to a value of 0.0010. As the distance approaches 0, the iteration comes to a halt in the approximation process in Figure 12 so that the logarithmic value does not become undefined.

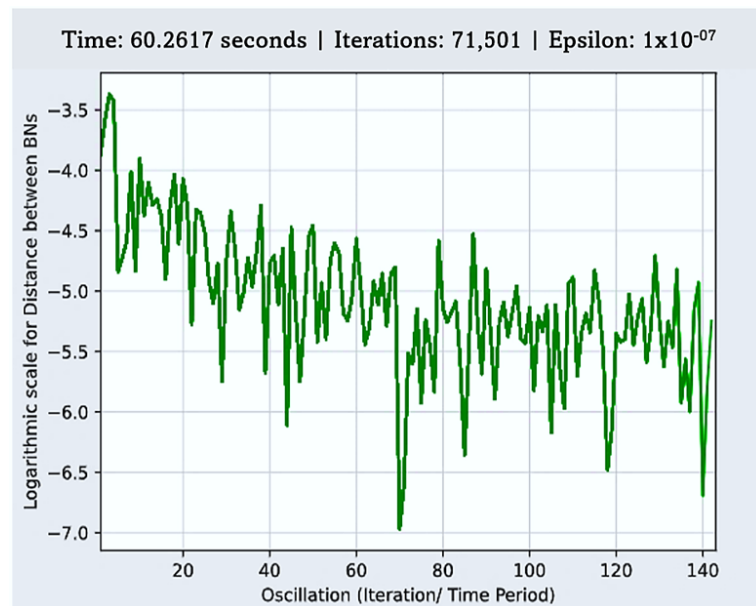


Figure 12. Gibbs -approximated KL divergence: the distance between respective BNs learned from the PD and SD is calculated and plotted here. The number of oscillations is represented on the horizontal axis with a time period of 500 millisecond. The logarithmic scale ($\log_{10}$) for the distance is represented by the vertical axis. The approximation stops when the actual distance approaches zero.

The KL divergence quantifies the ‘distance’ between probability distributions in the two BNs developed from the PD and SD. This is useful to see how much information is lost when one is used to approximate the other, as expressed by Equation 10. This also shows how the quality of parameter learning of the algorithm gets more precise and improved with every iteration.

7.4. Discussion on SD Generation

Applying parameter and structure learning techniques to create the SD helped in exploring the aspect of employing artificial data to explore different scenarios, validate models and test the hypotheses without using real-world data.

The accuracy with which these methods capture the underlying patterns of the real data greatly depends on the data generation method and adjustment techniques. Assessing the quality of the SD based on the selected generation method revealed the areas of improvement for continuous optimization of these BN learning algorithms.

This new dataset was then compared with the original PD through statistical testing to evaluate the accuracy of the former. The conclusion from the tests aligns with the real-life observation that categorical variables are the same in both datasets, while the numerical variables have different values. Moreover, KL divergence was approximated to visualize the distance between the two datasets, which eventually approaches zero after a few iterations, indicating equal distribution.

The synthetic data (SD) were then used to recreate the Bayesian network, and Laplace smoothing was applied to normalize the values. The conditional probabilities before and after the adjustment show that the probabilities learned with synthetic data are closer to the real data after adjustment.

Despite uncertainty around the necessary number of iterations for the algorithm to produce data with 100% similarity to the real data pattern, this method holds significant potential in the expansion and enhancement of insufficient and unbalanced data in any field of research. Overall, the methodology proposed in this paper covers the complete process of implementation and testing for data modeling and analysis.

8. Summary and Discussion

This investigation contributes to the comprehensive understanding of the nature and extent of the demographic–accessibility dynamic. This may be essential for legislators to plan and evaluate policies and programs to achieve social sustainability, a key objective of smart cities, by enabling full participation in society for persons with disabilities.

Each individual has their unique needs that can only be fulfilled by creating barrier-free communities, environments, services, technology and opportunities. Ensuring the removal of existing accessibility barriers and the prevention of new barriers will help realize smart cities and an overall society that are also inclusive. This way, all members of society can be empowered and equipped to reach their full potential in an environment that accommodates everyone, regardless of their diverse needs.

To facilitate the development of such inclusive smart cities, it is imperative that factors causing disadvantage be recognized. Identifying these crucial areas will accelerate the prioritization and allocation of resources to support those experiencing stress and pressure that come with these disadvantages. Moreover, to efficiently overcome discrimination and inequality, it is necessary that persons with disabilities be integral to the policy- and decision-making process. This study incorporates first-hand accounts from persons with disabilities themselves in the analysis, ensuring a realistic view of accessibility barriers. For decision making via data utilization and analytics, unique computational intelligence tools such as causal modeling are envisioned to be a part of smart cities. In particular, information on probable areas where issues are encountered by each group of individuals with diverse attributes is essential for the effective planning and development of specialized programs.

This study aims to determine the demographic and disability profiles of groups that are significantly impacted by barriers to further assist in risk mitigation. Based on the lived experiences of persons with disabilities, the proposed methodology can help design and model other aspects of the accessibility goals of cities and communities, such as employment disparities, lack of accommodation, social isolation and food security.

8.1. Difference Between Preliminary Investigation and Extended Study

This study utilizes the Canadian Disability Survey from 2022. The last survey cycle was in 2017, which was incorporated in the preliminary investigation for this research [15] on the nature of accessibility barriers linked to an individual's disability type.

While the majority of questions remained the same between the 2017 and 2022 Disability Survey, the content of the 2022 survey was expanded to add several new indicators. Some subject areas were revised and updated, while others were removed, with new survey modules added for 2022 in order to address data gaps and emerging data needs. The CSD 2022 includes information on gender and a total of 25 additional, more specified barrier scenarios. This is reflected in the accessibility barrier dataset in terms of frequency used in this study. The data used for the preliminary investigation based on the 2017 survey only included 11 scenarios. Similarly, the 6 types of disabilities represented in the 2017 survey were more narrowly defined in 2022 to include 10 disabilities.

The 2022 survey also classifies data for each barrier in terms of the frequency of encountering. Confidence intervals in data estimation in both whole numbers and percentage were introduced in the 2022 survey. Due to updating of the sorting criteria and method, this research was able to perform a more rigorous analysis of accessibility experiences compared to the preliminary study.

Unlike a complete framework with interconnected methods and the latest 2024 accessibility data used in this research, the preliminary study developed a computational model where each method is a standalone tool for confirmation of causality using an older dataset from 2021.

8.2. Summary of Experiments and Results

The application of EFA aimed to identify the factors that define the underlying relationships between observed variables. This dimension reduction method confirmed the correlation of data variables and resulted in more manageable data. The number of factors with eigenvalues greater than 1.0 (factors that explain more variance than a single variable) were selected for further analysis with SEM. Here, variance is a measure of how far a set of numbers is spread out its average value. The results were used to define the structure to categorize each observed variable under one of these factors.

Parameter analysis of the SEM tool confirmed that all barriers experienced at different frequencies have significant connections with the type of disability, except for the *online (business)* and *sidewalk (snow)* scenarios. The scenarios ‘always’ encountered in each barrier category that yielded the highest standard estimates are **communication**—*phone (family)*, **ICT**—*online (booking)*, **transportation**—*Washrooms* and **attitudinal**—*attitude (healthcare)*.

Causal Bayesian inference was performed on the model that was determined by the SEM process, excluding the abovementioned insignificant scenarios. Barriers with the greatest probability under each category turned out to be the same those determined for the SEM above. Overall diagnosis of the demographic attributes depending on reported barriers showed that individuals with *mental health* and *pain-related* disability are the most susceptible to experiencing barriers in general. The intersectionality of these attributes was also considered for inference.

To explore the possibility of using synthetic data for future investigations in cases of inconclusive or limited real-world data, the BN learned from real data served as a template for artificial data generation. By checking the similarities between the real and synthetic distribution samples through statistical tests, it was verified that the latter were different for numerical values but similar for categorical ones, attesting to the expected outcome based on observation of the two datasets. Furthermore, recreating the Bayesian model from synthetic data allowed for the application of adjustment techniques, which visibly improved the resulting conditional probabilities.

The experimental results confirm the initial hypothesis that probabilistic causal reasoning tools such as structural equation modeling and Bayesian networks are feasible methods for understanding varying accessibility experiences attributed to demographics and disability types. This study also explored synthetic data generation techniques with structures identical to real data for more comprehensive analysis, which helped in testing the precision of the probabilistic model.

9. Conclusions and Future Work

It is a human right for all individuals to have access to basic needs so that they can thrive and do more than just survive within a community. Learning about the aspects and areas associated with higher probabilities of facing obstacles can help to relieve anxiety around accessibility issues for persons with disabilities. Hence, the work presented here can significantly benefit individuals with accessibility anxiety by identifying significant scenarios through the modeling of real-world data to enhance preparedness and the predictability of barriers.

For persons with disabilities, it is crucial to be recognized as a vulnerable population, and disability should be regarded in research as another demographic attribute, such as gender or age. Incorporating first-hand accounts such as through surveys to exercise the inclusion of people with disabilities can enhance the scientific integrity of research findings to develop real-world policy applications.

As a stepping stone to address disparities through research, this work proposes a scientific approach for systematic evaluation of accessibility experiences of the millions

of people living with disabilities across Canada. The findings can provide a closer view of the degree of accommodation in federal services and transportation that can serve as a guide for prospective studies and detailed identification of the priority areas through causal reasoning. This work also provides insights into accessibility barriers, which are essential in designing inclusive smart cities.

For subsequent work in the future, synthetic data generation techniques can be adopted to expand real-world data to predict accessibility trends and to enable in-depth analysis in any social development research. This research lays the groundwork for projection of the level of accommodation available at any facility using advanced machine learning techniques to help with accessibility anxiety for people with disabilities. The synthetic data generation techniques for the expansion and replacement of real data used here can enable the production of large datasets for time-series forecasting.

This methodological approach incorporating causal modeling of data can be applied in any investigative research regarding vulnerable groups in the equity, diversity and inclusion paradigm. Inference derived from this framework paired with the relevant data can be used to develop and redesign improved assistive technology that would fit the population with special needs. Furthermore, future work could facilitate the development of means and technology for loved ones and caregivers to help and look after individuals with functional limitations.

Through this research, we hope that governments adopting smart city solutions for urban transformation incorporate smart tools and available applications to understand the needs and experiences of patrons in order to make federal sector organizations and services fully accessible. Moreover, frequent and efficient initiatives to share such information with the public would improve not only the quality of government services but also citizen well-being. This avenue of research could serve to streamline inter-organizational collaboration, supporting accessibility initiatives and humanitarian efforts in communities around the world to ensure equal participation through having meaningful options and the freedom to make their own choices for persons with disabilities.

Author Contributions: Conceptualization, M.Z., G.W. and S.Y.; methodology, S.Y.; software implementation, M.Z.; formal analysis, G.W., S.Y. and M.Z.; writing—original draft preparation, M.Z.; writing—review and editing, G.W., S.Y. and M.Z.; visualization, M.Z.; supervision and funding acquisition, S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Social Sciences and Humanities Research Council of Canada (SSHRC) through the project “Emergency Management Cycle-Centric R&D: From National Prototyping to Global Implementation” with Funding opportunity “New Frontiers in Research Fund” under Grant NFRF-2021-00277.

Institutional Review Board Statement: Not applicable. Data collection was performed as part of the Canadian Disability Survey by the Government of Canada. The authors declare no involvement in the process.

Informed Consent Statement: Not Applicable. The data used in this study were collected and published by Statistics Canada for the 2022 Canadian Survey on Disability. The authors of this paper confirm no involvement in the collection or publication process.

Data Availability Statement: The original dataset used for the case study is publicly available in the Statistics Canada Data hub at <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310089901> (accessed on 14 August 2024) and cited here [7].

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
BIC	Bayesian Information Criterion
BN	Bayesian Network
CDF	Cumulative Distribution Function
CPT	Conditional Probability Table
CRPD	Convention on the Rights of Persons with Disabilities
CSD	Canadian Survey on Disability
EFA	Exploratory Factor Analysis
ICT	Information and Communication Technology
KL	Kullback–Leibler
KMO	Kaiser–Meyer–Olkin
ML	Machine Learning
MIIC	Multivariate Information-based Inductive Causation
MLE	Maximum Likelihood Estimation
NA	Not Applicable
PD	Primary Dataset
RE	Regression Estimate
SD	Synthetic Data
SEM	Structural Equation Modeling
SRE	Standardized Regression Estimate

Appendix A

Figure A1 show the alternative structure that was considered for the SEM. The results from the EFA suggest four latent variables for the statistical modeling, which were not possible with this structure. Therefore, we proceeded with the structure shown in Figure 5 for all modeling and analysis purposes in this article.

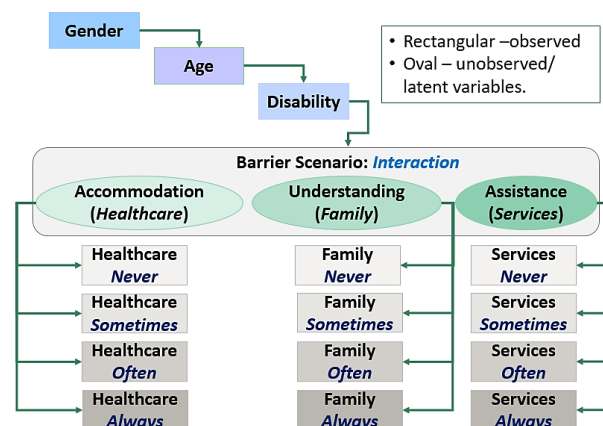


Figure A1. SEM structure 2: Each rectangle is an observed variable with multiple subcategories, whereas the oval shapes represent the latent variables. The latent variables of *accommodation (healthcare)*, *understanding (family)* and *assistance (services)* represent issues persons with disabilities may experience at frequencies of *never*, *sometimes*, *often* or *always*.

Appendix B

The survey questionnaire prepared and provided by the Government of Canada for the CSD 2022 is included Table A1 and a brief description of the different disability types are presented in Table A2 in alignment with the CSD data [73] released by Statistics Canada.

Table A1. CSD 2022 Questionnaire for Barriers encountered by Persons with Disabilities.

Introduction: “We now want you to think about accessibility barriers that might limit your access to different areas and activities in your daily life due to your condition. By barriers we mean something that could be removed, modified, or done differently” [74].		
i. In the past 12 months, have you experienced barriers with any of the following features inside or outside public spaces because of your condition(s)?	A. Entrances or exits to buildings	1 2 3 4 5
	B. Floorplans inside buildings	1 2 3 4 5
	C. Lighting or sound levels inside buildings	1 2 3 4 5
	D. Public washrooms	1 2 3 4 5
	E. Wait lines	1 2 3 4 5
	F. Self-serve technology	1 2 3 4 5
	G. Announcements or alarms	1 2 3 4 5
	H. Signs or directions	1 2 3 4 5
	I. Pedestrian signals at intersections or crosswalks	1 2 3 4 5
	J. Sidewalks when covered in ice or snow	1 2 3 4 5
	K. Sidewalks in general	1 2 3 4 5
ii. In the past 12 months, have you experienced any barriers in terms of behaviours, misconceptions or assumptions made about you from any of the following because of your condition(s)?	A. Family or close friends	1 2 3 4 5
	B. Staff of a business	1 2 3 4 5
	C. Medical or health care professionals	1 2 3 4 5
	D. Staff of government services or programs	1 2 3 4 5
iii. In the past 12 months, have you experienced barriers communicating in any of the following situations because of your condition(s)?	A. In person with family or close friends	1 2 3 4 5
	B. In person with medical or health care professionals	1 2 3 4 5
	C. In person with others such as the general public, customer service or government representatives	1 2 3 4 5
	D. By voice over the phone with family or close friends	1 2 3 4 5
	E. By voice over the phone with medical or health care professionals	1 2 3 4 5
	F. By voice over the phone with others such as the general public, customer service or government representatives	1 2 3 4 5
	G. With an automated phone message system	1 2 3 4 5
	H. Using video conferencing	1 2 3 4 5
	I. Using social media or online chat forums to interact with others	1 2 3 4 5
iv. In the past 12 months, have you experienced barriers using the Internet for any of the following activities because of your condition(s)?	A. Online banking or online shopping	1 2 3 4 5
	B. Online booking for appointments, services or reservations	1 2 3 4 5
	C. Online access to government information, services or supports	1 2 3 4 5

1 : not applicable; 2 : never; 3 : sometimes; 4 : often; 5 : always.

Table A2. Overview and Accessibility Recommendation for various types of Disabilities.

Disability	Description of Associated Difficulties	Recommended Accessibility Practices to be Incorporated
Developmental	Communicating or understanding written or spoken language, overlap with learning disability.	Reduce communication barriers by speaking at a steady pace, using clear, short sentences; avoid words or phrases that have multiple meanings.
Hearing	Hearing and how well a person hears or whether they hear at all.	Information in sign language; assistive listening devices; proper sound and lighting; closed captioning for videos and slide shows.
Mobility	Ability to move around, walk or use stairs.	Option to stand or sit at events and extra time to complete documents.
Flexibility	Ability to stand, pick things up, reach out/ upwards.	Accommodation at events; assistive Web features; speaking-based participation on videoconferencing platforms; Accessible parking
Dexterity	Affects way of using fingers, wrist, any extremities.	Assistive devices to perform tasks; Accessible sidewalks and restrooms.
Learning	Affects ability to learn, read, write, process information, organize or manage time.	Reduce auditory and visual distractions; providing information in simple, clear and concise language and in alternate formats.
Memory	Difficulty processing or retaining information.	Similar to learning and developmental disabilities
Mental health	Difficulty managing emotions, thoughts and sensations in different ways (usually not visible).	Avoid sudden noises, changes in lighting and other things that could trigger anxiety; Set aside a quiet, private room to sit or take medication.
Pain-related	Chronic pain that may fluctuate unpredictably over time, disrupting daily activities.	Adjust light, sound and seating arrangements; prohibit strong-smelling products; be aware of specific needs and conditions.
Seeing	Difficulty in seeing light, color, distance or texture or seeing at all.	Provide audio versions of Web pages and electronic documents; ensure color contrast between text and backgrounds.

References

- Gossett, A.; Mirza, M.; Barnds, A.K.; Feidt, D. Beyond access: A case study on the intersection between accessibility, sustainability, and universal design. *Disabil. Rehabil. Assist. Technol.* **2009**, *4*, 439–450. [CrossRef] [PubMed]
- Chen, T.; Ramon Gil-Garcia, J.; Gasco-Hernandez, M. Understanding Social Sustainability for Smart Cities: The Importance of Inclusion, Equity, and Citizen Participation as Both Inputs and Long-term Outcomes. *J. Smart Cities Soc.* **2022**, *1*, 135–148. [CrossRef]
- Wolbring, G.; Rybchinski, T. Social Sustainability and Its Indicators through a Disability Studies and an Ability Studies Lens. *Sustainability* **2013**, *5*, 4889. [CrossRef]
- Vallance, S.; Perkins, H.C.; Dixon, J.E. What is social sustainability? A clarification of concepts. *Geoforum* **2011**, *42*, 342–348. [CrossRef]
- Rebernik, N. Smart and Sustainable Cities Inclusive of People with Disabilities: An Exploratory Study Using Digital Tools and Participatory Practices. Ph.D. Thesis, University of Deusto, Bilbo, Spain, 2020.
- Makkonen, T.; Inkinen, T. Inclusive smart cities? Technology-driven urban development and disabilities. *Cities* **2024**, *154*, 105334. [CrossRef]
- Pianosi, R.; Presley, L.; Buchanan, J.; Lévesque, A.; Savard, S.A.; Lam, J. *Canadian Survey on Disability, 2022: Concepts and Methods Guide*; Technical report; Government of Canada: Ottawa, ON, Canada, 2023.
- Statistics Canada. *Canadian Survey on Disability, 2017 to 2022*; Technical report; Government of Canada: Ottawa, ON, Canada, 2023.
- Statistics Canada. *About an Accessible Canada*; Technical Report; Government of Canada: Ottawa, ON, Canada, 2023.
- Zhou, S.; Loiacono, E.T.; Kordzadeh, N. Smart Cities for People with Disabilities: A Systematic Literature Review and Future Research Directions. *Eur. J. Inf. Syst.* **2023**, *33*, 845–862. [CrossRef]
- Sugandha.; Freestone, R.; Favaro, P. The social sustainability of smart cities: A conceptual framework. *City Cult. Soc.* **2022**, *29*, 100460. [CrossRef]
- City of Toronto. 2019. Available online: <https://www.toronto.ca/city-government/accountability-operations-customer-service/long-term-vision-plans-and-strategies/connected-community/> (accessed on 24 November 2024).
- United Nations. *Article 1—Purpose*; Technical Report; Department of Economic and Social Affairs: New York, NY, USA, 2016.
- Grecoa, A. An introduction to causal inference. *EGE Rev. Expr. Graf. Edif.* **2020**, *6*, 18–27. [CrossRef]
- Zakir, M.; Wolbring, G.; Yanushkevich, S. A Causal Approach to Investigating Accessibility Experiences of Persons with Disabilities. In Proceedings of the 2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS), Marrakech, Morocco, 23–24 October 2024; pp. 1–8. [CrossRef]
- Statistics Canada. *Canadian Survey on Disability, 2017*; Technical report; Government of Canada: Ottawa, ON, Canada, 2018.
- Gkontzis, A.F.; Kotsiantis, S.; Feretzakis, G.; Verykios, V.S. Enhancing Urban Resilience: Smart City Data Analyses, Forecasts, and Digital Twin Techniques at the Neighborhood Level. *Future Internet* **2024**, *16*, 47. [CrossRef]
- Gkontzis, A.F.; Kalles, D.; Paxinou, E.; Tsoni, R.; Verykios, V.S., A Big Data Analytics Conceptual Framework for a Smart City: A Case Study. In *Building on Smart Cities Skills and Competences*; chapter Internet of Things; Springer: Cham, Switzerland, 2022; pp. 259–279. [CrossRef]
- Lim, C.; Kim, K.J.; Maglio, P.P. Smart cities with big data: Reference models, challenges, and considerations. *Cities* **2018**, *82*, 86–99. [CrossRef]
- Bringolf, J. The smart city and disability. In *Research Papers for Built Environment*; Academic Paper; Centre for Universal Design Australia: Moorebank, NSW, Australia, 2024. Available online: <https://universaldesignaustralia.net.au/the-smart-city-and-disability/> (Accessed on 2 December 2024).
- People with Disability Australia. Social Model of Disability. Available online: <https://pwd.org.au/resources/models-of-disability/> (accessed on 4 October 2024).
- Hosseinpour, A.R.; Bergen, N.; Kostanjsek, N.; Kowal, P.; Officer, A.; Chatterji, S. Socio-demographic patterns of disability among older adult populations of low-income and middle-income countries: Results from World Health Survey. *Int. J. Public Health* **2015**, *61*, 337–345. [CrossRef] [PubMed]
- Statistics Canada. *World Report on Disability*; Technical Report; World Health Organization: Geneva, Switzerland, 2011.
- Wolbring, G.; Lillywhite, A. Burnout through the Lenses of Equity/Equality, Diversity and Inclusion and Disabled People: A Scoping Review. *Societies* **2023**, *13*, 131. [CrossRef]
- Church, R.L.; Marston, J.R. Measuring Accessibility for People with a Disability. *Geogr. Anal.* **2010**, *35*, 83–96. [CrossRef]
- Americans with Disabilities Act (ADA). *ADA Standards for Accessible Design*; Technical Report; U.S. Department of Justice: New York, NY, USA, 2010.
- Wald, M. AI Data-Driven Personalisation and Disability Inclusion. *Front. Artif. Intell.* **2021**, *3*, 571955. [CrossRef] [PubMed]
- Kitson, N.K.; Constantinou, A.C. Learning Bayesian networks from demographic and health survey data. *J. Biomed. Inform.* **2021**, *113*, 103588. [CrossRef]

29. Tarka, P. An overview of structural equation modeling: Its beginnings, historical development, usefulness and controversies in the social sciences. *Qual. Quant.* **2018**, *52*, 313–354. [CrossRef]
30. Abraham, S.; Mir, B.A.; Suhara, H.; Mohamed, F.A.; Sato, M. Structural equation modeling and confirmatory factor analysis of social media use and education. *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 32. [CrossRef]
31. Abid, N.; Pond, L.; Yanushkevich, S. Causality Exploration in Modeling Engineering Student Satisfaction. In Proceedings of the 2024 IEEE Global Engineering Education Conference (EDUCON), Kos Island, Greece, 8–11 May 2024; pp. 1–10. [CrossRef]
32. Diez-Mesa, F.; de Oña, R.; de Oña, J. Bayesian Networks and Structural Equation Modelling to develop service quality models: Metro of Seville case study. *Transp. Res. Part Policy Pract.* **2018**, *118*, 1–13. [CrossRef]
33. Lai, K.; Yanushkevich, S. Causal Models Applied to the Patterns of Human Migration due to Climate Change. In Proceedings of the 2023 IEEE Symposium Series on Computational Intelligence, Mexico City, Mexico, 5–8 December 2023; pp. 1644–1649.
34. Gupta, S.; Kim, H.W. Linking structural equation modeling to Bayesian networks: Decision support for customer retention in virtual communities. *Eur. J. Oper. Res.* **2008**, *190*, 818–833. [CrossRef]
35. Oliveira, H.C.R.; Yanushkevich, S.; Almekhlafi, M. Sensitivity Analysis of Stroke Predictors Using Structural Equation Modeling and Bayesian Networks. In Proceedings of the 2022 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Ottawa, ON, Canada, 15–17 August 2022; pp. 1–8. [CrossRef]
36. Shaposhnyk, O.; Lai, K.; Wolbring, G.; Shmerko, V.; Yanushkevich, S. Next Generation Computing and Communication Hub for First Responders in Smart Cities. *Sensors* **2024**, *24*, 2366. [CrossRef]
37. Balsa-Barreiro, J.; Cebrián, M.; Menéndez, M.; Axhausen, K. Leveraging Generative AI Models in Urban Science. *Curr. Top. Behav. Neurosci.* **2024**, *23*, 143–149. [CrossRef]
38. Statistics Canada. *Barriers to Accessibility for Persons with Disabilities Aged 15 Years and Over, by Disability Type, Age Group and Gender*; Technical Report; Government of Canada: Ottawa, ON, Canada, 2024.
39. Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **2015**, *521*, 452–459. [CrossRef] [PubMed]
40. Statistics Canada. Statistics: Power from Data! Available online: <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-eng.htm> (accessed on 15 September 2024).
41. Babu, D. Machine Learning Exercises. Available online: <https://github.com/Dhamodaran-Babu/Machine-Learning-Exercises/tree/master/12.Factor%20Analysis> (accessed on 15 July 2024).
42. University of Toronto. Exploratory Factor Analysis. Available online: <https://www.utstat.toronto.edu/~brunner/oldclass/431s23/textbook/Ch2.pdf> (accessed on 18 August 2024).
43. Crossman, A. Principal Components and Factor Analysis. 2023. Available online: <https://www.thoughtco.com/principal-factor-analysis-3026699> (accessed on 5 November 2024).
44. Swami, V.; Maïano, C.; Morin, A.J. A guide to exploratory structural equation modeling (ESEM) and bifactor-ESEM in body image research. *Body Image* **2023**, *47*, 101641. [CrossRef] [PubMed]
45. Kaiser, H.F.; Rice, J. Little jiffy, mark IV. *Educ. Psychol. Meas.* **1974**, *34*, 111–117. [CrossRef]
46. Yoo, C.; Oh, S. Combining Structure Equation Model with Bayesian Networks for predicting with high accuracy of recommending surgery for better survival in Benign prostatic hyperplasia patients. In Proceedings of the 20th International Congress on Modelling and Simulation, MODSIM, Adelaide, Australia, 1–6 December 2013.
47. Violato, C.; Beran, T.N. Structural Equation Modeling in Medical Research: A primer. *BMC Res. Notes* **2010**, *3*, 267. [CrossRef]
48. Igolkina, A.A.; Meshcheryakov, G. Semopy: A Python Package for Structural Equation Modeling. *Struct. Equ. Model. Multidiscip. J.* **2020**, *27*, 952–963. [CrossRef]
49. Siegel, A.F.; Wagner, M.R. Chapter 12—Multiple Regression: Predicting One Variable From Several Others. In *Practical Business Statistics*, 8th ed.; Academic Press: Cambridge, MA, USA, 2022; pp. 371–431. [CrossRef]
50. Dahiru, T. P-value, a true test of statistical significance? A cautionary note. *Ann. Ib. Postgrad. Med.* **2008**, *6*, 21–26. [CrossRef]
51. Hebert, S.; Lee, V.; Morabito, M.; Polan, J. 13.5: Bayesian Network Theory. 2023. Available online: [https://eng.libretexts.org/Bookshelves/Industrial_and_Systems_Engineering/Chemical_Process_Dynamics_and_Controls_\(Woolf\)/13%3A_Statistics_and_Probability_Background/13.05%3A_Bayesian_network_theory](https://eng.libretexts.org/Bookshelves/Industrial_and_Systems_Engineering/Chemical_Process_Dynamics_and_Controls_(Woolf)/13%3A_Statistics_and_Probability_Background/13.05%3A_Bayesian_network_theory) (accessed on 28 August 2024).
52. Albright, E.A. Probability: Joint, Marginal and Conditional Probabilities. 2024. Available online: <https://sites.nicholas.duke.edu/statsreview/jmc/> (accessed on 15 October 2024).
53. Kendall's Advanced Theory of Statistics, D.T. *Stuart, Alan and Ord, Keith*; John Wiley & Sons: Hoboken, NJ, USA, 2010.
54. Ducamp, G.; Gonzales, C.; Wuillemin, P.H. aGrUM/pyAgrum: A Toolbox to build models and algorithms for Probabilistic Graphical Models in Python. In Proceedings of the 10th International Conference on Probabilistic Graphical Models, PMLR, Online, 23–25 September 2020; Jaeger, M., Nielsen, T.D., Eds.; Volume 138, pp. 609–612.
55. Hayes, A. Posterior Probability: Definition, Formula for Calculation. 2023. Available online: <https://www.investopedia.com/terms/p/posterior-probability.asp> (accessed on 28 September 2024).
56. Pearl, J. An introduction to causal inference. *Int. J. Biostat.* **2010**, *6*, 7. [CrossRef]

57. Fenton, N.; Neil, M. *Risk Assessment and Decision Analysis with Bayesian Networks*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2018. [CrossRef]
58. Bours, M.J. Bayes' rule in diagnosis. *J. Clin. Epidemiol.* **2021**, *131*, 158–160. [CrossRef]
59. Ali, M. Synthetic Data is the Future of Artificial Intelligence. 2023. Available online: <https://moez-62905.medium.com/synthetic-data-is-the-future-of-artificial-intelligence-6fcfd2ce1a14> (accessed on 2 September 2024).
60. Preim, B.; Botha, C. *Visual Computing for Medicine*, 2nd ed.; Morgan Kaufmann: Boston, MA, USA, 2014; p. iii. [CrossRef]
61. Brownlee, J. Probabilistic Model Selection with AIC, BIC, and MDL. 2020. Available online: <https://machinelearningmastery.com/probabilistic-model-selection-measures/> (accessed on 15 September 2024).
62. Tocci, T.; Cabeli, V.; Li, H.; Simon, F. GitHub—miicTeam/miic_R_package: Learning Causal or Non-Causal Graphical Models Using Information Theory. Available online: https://github.com/miicTeam/miic_R_package (accessed on 25 September 2024).
63. Verny, L.; Sella, N.; Affeldt, S.; Singh, P.; Isambert, H. Learning causal networks with latent variables from multivariate information in genomic data. *PLoS Comput. Biol.* **2017**, *13*, e1005662. [CrossRef]
64. Fein, E.C.; Gilmour, J.; Machin, T.; Hendry, L. *Statistics for Research Students*; University of Southern Queensland: Toowoomba, Australia, 2022.
65. Campbell, M.J. *Statistics at Square One*; BMJ Publication Group: London, UK, 1998.
66. Price, P.C.; Chiang, I.C.A.; Jhangiani, R.S. *Research Methods of Psychology—2nd Canadian Edition*; BCcampus: Victoria, BC, Canada, 2015.
67. Reddy, T. Statistical Functions (scipy.stats). 2024. Available online: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html (accessed on 15 July 2024).
68. Pearson, K.X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1900**, *50*, 157–175. [CrossRef]
69. Light, C. Tutorial: Pearson's Chi-Square Test for Independence. 2008. Available online: <https://www.ling.upenn.edu/~clight/chisquared.htm> (accessed on 15 September 2024).
70. McHugh, M.L. The chi-square test of independence. *Biochem. Med.* **2023**, *23*, 143–149. [CrossRef] [PubMed]
71. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; The MIT Press: Cambridge, MA, USA, 2012.
72. Moral, S.; Cano, A.; Gómez-Olmedo, M. Computation of Kullback-Leibler Divergence in Bayesian Networks. *Entropy* **2021**, *23*, 1122. [CrossRef] [PubMed]
73. Employment and Social Development Canada. *Annex: Understanding Disabilities*; Technical Report; Government of Canada: Ottawa, ON, Canada, 2022.
74. Statistics Canada. Canadian Survey on Disability—2022: Questionnaire. Available online: https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&lang=en&Item_Id=1400719#qb1403532 (accessed on 15 August 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.