**UNIVERSIDADE FEDERAL DE MINAS GERAIS**
**Instituto de Ciências Exatas**
**Programa de Pós-Graduação em Ciência da Computação**

Guilherme Fernandes Marchezini

# Counterfactual inference and its Application in Mental Health Care

Belo Horizonte
2021

Guilherme Fernandes Marchezini

# Counterfactual inference and its Application in Mental Health Care

**Final Version**

Thesis presented to the Graduate Program in Computer Science of the Federal University of Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Anisio Mendes Lacerda

Belo Horizonte
2021

2021, Guilherme Fernandes Marchezini.
Todos os direitos reservados

Marchezini, Guilherme Fernandes.

M317c Counterfactual inference and its application in mental health care [recurso eletrônico] / Guilherme Fernandes Marchezini – 2021.
1 recurso online (59 f. il., color.) : pdf.

Orientador: Anísio Mendes Lacerda.

Dissertação (Mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação.
Referências: f. 52-55.

1. Computação – Teses. 2. Estatística matemática – Teses. 3. Redes neurais (Computação) – Teses. 4. Saúde mental - Controle preditivo – Teses. 5. COVID-19 Pandemia, 2020- – Aspectos psicológicos – Teses. I. Lacerda, Anísio Mendes. I. Universidade Federal de Minas Gerais Exatas, Departamento de Ciência da Computação. III. Título.

CDU 519.6*85(043)

Ficha catalográfica elaborada pela bibliotecária Irénquer Vismeg Lucas Cruz
CRB 6/819 - Universidade Federal de Minas Gerais - ICEx


UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

# FOLHA DE APROVAÇÃO

## Counterfactual inference and its Application in Mental Health Care

**GUILHERME FERNANDES MARCHEZINI**

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. ANISIO MENDES LACERDA - Orientador
Departamento de Ciência da Computação - UFMG

PROFA. ANNA HELENA REALI COSTA
Departamento de Engenharia de Computação e Sistemas Digitais - USP

PROFA. GISELE LOBO PAPPA
Departamento de Ciência da Computação - UFMG

PROF. ADRIANO CÉSAR MACHADO PEREIRA
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 10 de Setembro de 2021.

# Acknowledgments

Gostaria de agradecer a todos que fizeram parte dessa jornada, todos que me ajudaram de forma direta ou indireta. Agradecer o apoio, as ideias e também as críticas. Agradeço, e muito, pela paciência, pelo tempo, pelo esforço e também pela diversão proporcionada nesse caminho.

Confesso que esse trabalho não foi fácil, mas não poderia deixar de falar da diversão que foi. O contraste do trabalho até de madrugada vendo repetidos erros e quebrando a cabeça pra entender, com o sorriso de um resultado esperado. O contraste também de ver um resultado inesperado que me deixava perdido com o resultado inesperado que trazia luz e explicava tudo.

Quero começar então agradecendo pela amizade de todos que me motivaram a continuar perseguindo minha vontade, que me apoiaram quando queria jogar tudo pro alto, e que me ouviram em minhas incontáveis teorias erradas. Agradeço especialmente a Jéssica, minha namorada e amiga, por ouvir praticamente todos os dias desses últimos anos sobre esse trabalho, por me apoiar com carinho em dias que eu ficava frustrado, por me incentivar quando eu queria parar, por me ajudar a parar quando precisava descansar e também por comemorar comigo os resultados positivos. Mesmo não sendo da área de computação, sempre se interessou por entender o que eu estava fazendo e como as coisas funcionavam. Quero agradecer também por me explicar várias coisas da área da saúde, em um momento da pesquisa que infelizmente não chegou ao trabalho final, mas que com certeza ajudou no desenvolvimento.

Queria agradecer ao Anísio, meu orientador, que por muitas vezes conversou e me direcionou, principalmente quanto eu estava perdido. Pelo incentivo a investigar mais, a pensar mais, por discutir comigo, trazer novas ideias e agregar. Agradeço principalmente por me falar que eu tinha ali uma ideia inovadora e que ele iria nesse caminho comigo. Agradeço também a Gisele, minha orientadora, pela ajuda no desenvolvimento e escrita desse trabalho. Também à UFMG e a todos os seus professores e profissionais, por ter tornado possível essa oportunidade, e à FAPEMIG por apoiar a realização desde trabalho por meio do projeto PPM-00408-18.

Agradeço aos meus amigos Victor, Daniel e Passos, por me ouvir tantas vezes sobre meu trabalho, sobre as dificuldades dele, sobre os resultados e também por todas as zoações, porque nem tudo é seriedade, né? Agradeço aqui ao Pedro e Rodrigo, que além de colegas de trabalho, se tornaram amigos e tiveram uma participação grande nesse trabalho, mesmo que talvez eles não saibam, nas várias conversas na copa, tomando café,

discutindo tantas e tantas ideias e que muitas acabaram entrando nesse trabalho.

Agradeço a minha família, pela motivação de perseguir esse meu sonho e pela paciência nos dias que estava estressado. Agradeço ao meu pai, Ronaldo, por me ouvir, principalmente quando eu estava teorizando usos práticos, que me fizeram tentar buscar mais e mais algo que fosse útil e ao mesmo tempo tangível o suficiente. À minha mãe Vanessa, pela preocupação comigo e minha saúde, me chamando quando estava muito tarde pra descansar e repor as energias. Ao meu irmão Rafael, pelos momentos de descontração aqui em casa.

Me dou ao direito de me agradecer. Por toda força de vontade, por todo esforço e dedicação a esse trabalho. Hoje olho esse trabalho, finalmente completo, não só como uma dissertação de mestrado, mas como uma obra do meu empenho. E, talvez, o maior agradecimento a mim e a todos não sejam essas palavras aqui, seja o orgulho de uma dissetação completa, uma pequena contribuição à ciência.

Finalmente agradeço a Deus, por estar do meu lado, por me ajudar nessa caminhada, por todos os 'acasos' que contribuíram pra eu tirar minhas conclusões. Obrigado por atender minhas preces e por realizar mais um sonho meu.

E com um sorriso no rosto, que esbanja todo meu orgulho, eu posso finalmente dizer: 'Meu trabalho está pronto'.

"*Look how far I've come, the wars that I have won, I think out loud: Victorious and Proud.*"

(Falling in Reverse)

# Resumo

Esse trabalho tem como objetivo modelar predições contrafactuais em cenários nos quais, além das variáveis observáveis (i.e. endógenas), existem variáveis latentes (i.e. exógenas) que afetam as predições e, consequentemente, as respostas das perguntas contrafactuais. Essa situação é comum em problemas da área de saúde, incluindo saúde mental. Para isso, propomos um arcabouço onde o problema supracitado é modelado como uma tarefa de regressão multivariada, e o modelo contrafactual utiliza as variáveis observáveis e também as variáveis latentes, que nesse trabalho são referenciadas como fator de individualidade do paciente ($\varphi$). No campo de saúde mental, o foco em abordagens individuais é fundamental, visto que experiências passadas podem mudar como uma pessoa vê ou lida com situações atuais, mesmo que essa individualidade não possa ser diretamente medida. Ao que tange o conhecimento dos autores, essa é a primeira abordagem contrafactual que considera *variáveis observáveis e latentes*, para responder de forma *determinística* as perguntas contrafactuais, do tipo: Se eu alterar o apoio social do paciente, o quanto eu posso alterar seu nível de ansiedade? Este *framework* combina conceitos de aprendizado profundo de representações e inferência causal para inferir o valor de $\varphi$ e capturar *efeitos não-lineares e multiplicativos* das variáveis causais. Experimentos foram feitos tanto em bases sintéticas quanto em uma base real, referente a saúde mental das pessoas durante a pandemia de COVID-19. Nessa última, foi predito como as mudanças das ações e percepções das pessoas poderiam levar a desfechos diferentes em relação a sintomas de saúde mental e a qualidade de vida. Os resultados mostraram que o modelo aprende o fator de individualidade, possibilitando taxas de erro inferiores a 0.05 em análises contrafactuais, enquanto suas predições estavam de acordo com a literatura médica. Esse modelo tem potencial de recomendar tratamentos personalizados e impactar diretamente a qualidade de vida de pacientes com problemas de saúde mental.

**Palavras-chave:** inferência contrafactual; redes neurais artificiais; saúde mental; causalidade.

# Abstract

This work deals with the problem of modeling counterfactual reasoning in scenarios where, apart from the observed endogenous variables, we have a latent variable that affects the outcomes and, consequently, the results of counterfactual queries. The existence of latent variables is a common setup in healthcare problems, including mental health. We propose a new framework where the aforementioned problem is modeled as a multivariate regression and the counterfactual model accounts for both observed and a latent variable, where the latter represents what we call the patient individuality factor ($\varphi$). In mental health, focusing on individuals is paramount, as past experiences can change how people see or deal with situations, but individuality cannot be directly measured. To the best of our knowledge, this is the first counterfactual approach that considers *both observational and latent variables* to provide *deterministic* answer to counterfactual queries, such as: what if I change the social support of a patient to what extent can I change his/her anxiety? The framework combines concepts from deep representation learning and causal inference to infer the value of $\varphi$ and capture both *non-linear and multiplicative effects* of causal variables. Experiments are performed on both synthetic and real-world datasets, where we predict how changes in people's actions may lead to different outcomes in terms of symptoms of mental illness and quality of life. Results show the model was able to learn the individually factor to predict counterfactual with errors below 0.05 and also answers counterfactual queries that are supported by the medical literature. The model has the potential to recommend small changes in people's lives that may completely change their relationship with mental illness.

**Keywords:** counterfactual inference; artificial neural networks; mental health; causality.

# List of Figures

# List of Tables

# Contents

# Chapter 1

# Introduction

The use of machine learning and artificial intelligence has gained more popularity in both the scientific community and the industry, and it is having a profound impact on human decision making. This is mainly due to the availability of a large amount of data, the proposal of more advanced algorithms for data exploration, and the improvement in computing power and storage. Because of that, there is an increasing demand for investigating and contributing to the theoretical advancement and practical success of the use of these techniques both by researchers and practitioners [12].

A common type of question raised in the decision-making process in almost all domains of knowledge is contemplative: *"What if?"*. For example, a physician treating a patient has to choose between treatment A or B. He recommends treatment A, but her condition worsens and she dies. A typical question is: "What would have happened if she had been treated with B instead of A?" Counterfactual inference of hypothetical scenarios is an approach that can be used to answer these types of questions [35].

Counterfactual inference is the third step of the model of causality presented by Pearl [35] – called the ladder of causal knowledge – that is based on three steps. The first step is *association*, where the model observes the facts and makes conclusions from them. Most of the machine learning models are at this level, as they are capable of making predictions based on what they have seen but without deeply understanding the causal relation. For example, a model of association has the capability of predicting if a patient has a disease or not given the patient vital signs and results of exams, but cannot say if the abnormal results are the consequence of the illness or the condition that made it predisposed to contract the disease.

The second step of the causality model is *action*, where the system must be able not only to understand the result of a given situation but also the causal structure behind it. This enables the system to understand what will happen if a certain action is taken. In the previous example, a model at the second step has the information (which can be learned or given as input) that the disease causes changes in the vital signs and impacts results of exams. Therefore, the model can predict that, giving a treatment to cure the disease will also stabilize vital signs, while giving treatment to stabilize the vital signs will not cure the disease.

Finally, the third step is the *counterfactual*, a contemplative step where the system will try to answer questions of an alternative and fictional scenario. This scenario occurs in the same circumstance as the real one, but different actions would have been taken. Let us return to the patient who received treatment $A$ and the outcome event of her death (d), where our evidence is $E = d$. Here the model will be able to give either a probability or deterministic value of the outcome of the fictional world – if the patient would have survived ($E = s$) or not ($E = d$) if she had received treatment $B$ instead of $A$.

Research on models capable of working in the last two steps of the causality ladder had a significant increase after Pearl [35] introduced a formal mathematical language to them, the *structural causal model* (SCM), which still is the standard framework for calculating the outcome of counterfactual queries. The SCM is often associated with a graphical causal model, and since the causal relationship is not cyclic it is often represented with a directed acyclic graph (DAG), which represents the causal relations of the system variables.

Counterfactual queries are not only important from a mathematical point of view, but also from a psychological side. Although most of the time 'what if' questions are associated with negative feelings and effects, researches have shown that beneficial effects may also emerge from counterfactual thinking, mainly because the act of thinking of how it might have been can suggest knowledge and paths of what may yet to be [38].

The function of counterfactuals is seen not only as a benefit to make individuals feel better, but also to facilitate individual improvement and allow people to better both themselves and their surroundings in globally meaningful ways [37]. As these findings show the importance of counterfactual thinking, we understand that using methods to calculate and bring this kind of approach to the maximum number of applicable situations possible is an important step to give insights and contribute to the artificial intelligence field.

The search for counterfactual queries requires one to have all information regarding a given scenario. This makes these queries difficult to be computed, as the system must replicate all factors that were present in the original scenario to compute what would have happened. This necessity of knowing all the information can be better illustrated, again, in a medical case. Suppose a doctor gives a medicine to treat a patient by analyzing a set of known variables, but the treatment does not have the expected effect. The difference from the expected result to the real one can come from the individuality of the patient, and this same individuality could also affect the result of other treatments. Therefore, this individuality has to be taken into account by a model in order to understand the real effects of an action, but it is not necessarily an observable variable.

## 1.1 Objectives

In this context, this work has as its main objective to answer counterfactual queries considering both observational and latent data. For that, the following specific objectives are tackled:

- We propose a new framework that models the counterfactual problem as a multivariate regression, where the counterfactual considers the existence of both observable variables and a latent variable $\phi$, which represents the individuality factor.
- The method is validated in a set of synthetic datasets, given that the nature of counterfactual examples do not make it possible to mathematically measure the correctness of the results without the previous knowledge of the individuality ($\phi$) value.
- We modeled the counterfactual reasoning in a real-world mental health study, focused on understanding how the conditions are affecting people in the pandemic context of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

## 1.2 Contributions

This work has contributions for the areas of computer science and psychology, as detailed below.

### 1.2.1 Computer Science

The recent advancements in the first steps of the causal ladder are relying upon working and using causality together with machine learning techniques [2]. This motivated us to present a new framework based on a deep neural network that can answer questions on the three levels of Pearl's causal hierarchy, i.e., association, action, and structural counterfactual, which is validated in a set of synthetic simulated dataset. This implies that we move from understanding and predicting behaviors and consequences of a given situation to being able of understanding the consequences of actions in the future.

We propose to model the counterfactual problem as a multivariate regression, where the counterfactual considers the existence of both observable variables and a latent variable $\phi$, which represents the individuality factor. All variables considered in the model are exogenous, except $\phi$, since it is not observable. To the best of our knowledge this is the first counterfactual approach that considers both *observational and latent variables* and can captures both *non-linear and multiplicative* effects of causal variables.

The framework combines concepts from deep representation learning and causal inference to infer the value of $\phi$ and provide *deterministic* answers to counterfactual queries. The model is based on a new neural network architecture able to estimate the value of the latent variable.

### 1.2.2 Medical and Psychological

Concerning the application of the model to a real health problem, while not being easy to verify the results, the practical use of counterfactuals has been explored in healthcare applications in the past [34], mainly in the estimation of individual treatment effect (ITE) also using synthetic or semi-synthetic data [49, 40, 4]. In this case, we have observational data, called endogenous variables, which contain past actions (medications), their outcomes (evidences), and records of the patients. The counterfactual inference works to remove bias or other cofounding from models, since the decision mechanism is not known (why each medication was given to each patient), i.e., it's common for patients with worse health to receive a different treatment than patients who are with better health. This bias can mislead the counterfactual analysis of how a given patient would have been affected had it had a different treatment.

Different from the ITE studies, our focus is not on treatment (therefore we do not have observations of actions), we focus on how changing one's behavior in a contextual situation (which is represented in our observed variables) may affect both quality of life - measured using World Health Organization Instrument to evaluate Quality of Life (WHOQOL-BREF) and symptoms related to somatization conditions, measured by the Brief Symptoms Inventory (BSI) instrument [11].

Both WHOQOL and BSI are produced based on questionnaires, composed of a set of questions related to their recent quality of life and mental health. These two instruments are commonly used for assessing mental health. When dealing with these two instruments we assume that there is a common latent variable that directly affects both of them, and name it the individuality factor of a patient. The analysis focusing on individuals is essential for mental health since the way people see and deal with different

situations can vary according to one's past experiences. Moreover, it is well-known that individual psychotherapy and specific drug treatment are successful in roughly half of the patients [47, 21]. This individual factor is given as the latent variable $\phi$, since it is inherent to each person and therefore cannot be directly measured.

The use of personalized mental health treatment, where individual factors are considered, has the potential to be more effective than traditional clinical and statistical approaches, in which the objective normally is to improve the community benefits and explain the overall variance, which is formally testing for "*group effects*" in the "*majority of a clinical group*" [6]. Two important factors need to be identified when dealing with adjusted treatments: the most appropriate set of interventions and the way it will be delivered to a particular individual. Finally, it is important to understand the causal effects of treatments for individuals, as it can lead to more effective just-in-time adaptive interventions, generating more engaging treatments [30].

Targeting the analysis of mental health considering individuality, we worked with data obtained from a mental health questionnaire answered by 153,514 subjects in the year 2020 to assess the somatization, psychological, physical, and environmental effects of pandemics in their lives. For this problem, we show we can compute accurate predictions in diagnostics and interventions for mental health. While it was possible to verify the accuracy of the prediction, the counterfactual evaluation is challenging in its own nature, therefore we had specialists analyzing the counterfactual queries, which in this dissertation appear supported by medical literature.

## 1.3 Text Organization

The rest of this text is organized as follows:

a) Chapter 2 introduces the necessary concepts to understand this dissertation and reviews and compares related work.

b) Chapter 3 introduces the proposed framework and the method ICoNet - Individuality factor discover Counterfactual Network

c) Chapter 4 shows the experimental evaluation of ICoNet, where we performed experiments with both a synthetic and the real-world dataset from the psychology. For this second problem, we evaluate both populational and individual counterfactuals.

d) Chapter 5 draws some final conclusions and points directions of future work.

# Chapter 2

# Background and Related Work

This chapter presents the background knowledge necessary for understanding this work. We will first explain the importance of causality and its structure, followed by the functionality of counterfactual inference. Finally, we will present the uses of causality in Health Care.

## 2.1 Causality

Two main question are often raised when studying causality: "why?" and "how?". These questions are important because, while the applicability of machine learning methods has been rising, their lack of causal reasoning can be considered their weak point. Causal reasoning goes beyond the association of data, using the model inputs to predict outputs but also trying to understand why a given input affects the output and how it affects it.

For example, we can use machine learning to predict the weather – if it is raining or not – by looking at the number of open umbrellas in the street. While this algorithm will have great accuracy in most cases, it can mislead researchers to think that closing the umbrellas will lead the rain to stop falling. This is an extreme case where causality is already obvious and known, and therefore researchers know this assumption is absurd, but when the relationship between two variables is unknown or subtle, there is a clear possibility of this approach to become error-prone.

Causal discovery leads us into another direction when compared to associations: the algorithms are not trying to exclusively optimize the prediction and neither use all variables to predict one output. These algorithms try to find which variables cause the others, making it possible to understand that the rain causes people to open an umbrella to not get wet, and that when it is raining and people go out without an umbrella, they get wet. While the cause and consequence relations are explicit, this model does not lose the ability of association, being able, for example, to determine the number of umbrellas

in the street given the intensity of the rain. Hence, causal discovery goes beyond simple associations, being able to predict consequences for actions.

The causal action prediction will be able to infer, for example, what would happen if it is raining and I go out with a closed umbrella. It is also able to conclude that closing the umbrella in the rain will not make the rain stop. We will discuss later how the causal action inference, called the *DO operator*, works.

Models able to perform causal action prediction can go beyond the causal answers of what would happen if we take a specific action. They can answer important and contemplative questions that are inherent to human nature, called counterfactual. The counterfactual tackles the questions of possible consequences of different past actions, "What would have happened if I had invested more time in that project?", "If the doctor had given the patient another treatment, would the patient be alive now?". Answers to these questions can be given using the causal structures and information of the factual outcome. The causal relationship between variables has many practical uses, but they were seen as impossible to be calculated until [35] introduced the *structural causal model* (SCM), detailed in the next section.

### 2.1.1 Causal structures

The *structural causal model* (SCM) [35] describes the features of the world and how they interact with each other. The introduction of SCM and the mathematical language enabled a new look at the area and a set of new studies.

Formally, an SCM is composed of two sets of variables $U$ and $V$ and a set of functions $f$ that assigns each variable in V a value according to the values of other variables in the model. The difference between $V$ and $U$ is their nature. $U$ are exogenous variables, which means they are external to the model and we chose (or cannot) explain how they are caused. They are root nodes, and cannot be descendants of any other variables. The variables in $V$ are endogenous, which means that they are a direct cause of at least one other variable in $U$ or $V$. The variables in $V$ cannot be descendant and cause of the same variable in either $U$ or $V$, meaning that there are no causal cycles. We can summarize a SCM using Definition 1.

**Definition 1** (Structural Causal Model (SCM))**.** *A structural causal model* $\mathfrak{M}$ *is a 4-tuple* $\langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}) \rangle$*, where (i)* $\mathbf{U}$ *is a set* $\{U_1, \ldots, U_d\}$ *of background variables (aka exogenous variables) that are determined by factors outside the model; (ii)* $\mathbf{V}$ *is a set* $\{V_1, \ldots, V_d\}$ *of endogenous variables, which are determined by other variables in the model* $\mathbf{U} \cup \mathbf{V}$*; (iii)* $\mathcal{F}$ *is a set of functions* $\{f_1, \ldots, f_d\}$ *(aka causal mechanisms) that represents structural*

*assignments* $v_i := f_i(pa_i, u_i)$*, where* $pa_i$ *is the set of parents of* $v_i$ *(its direct causes); and (iv)* $P(\mathbf{U})$ *is a probability function defined over the domain of* $\mathbf{U}$*.*

Every SCM can be associated with a graphical causal model. In this model, each node represents a variable and the edges between nodes represent the functions in $f$, making the graphical model $\mathfrak{G}$ represent the relation of all variables in $\mathfrak{M}$. It is important to note that if a variable $X$ contains the value of a variable $Y$ within its formula $f_x$, then there will be a direct edge connecting $Y$ to $X$.

Since we work with a model $\mathfrak{M}$ with no cycles, we can use a *direct acyclic graph* (DAG) $\mathfrak{G}$ to represent the model $\mathfrak{M}$. Using a DAG can make it easier to interpret the information, since the edges are directed. We can now define that a variable $X$ is a direct cause of $Y$ if there is an edge from $X$ to $Y$, and if $Y$ is a descendant of $X$, then $X$ is a cause of $Y$ (this does not apply in rare intransitive cases, but they are out of the scope of this work and will not be discussed. For more details on that the reader is referred to [36]).

We have directed edges from causes to effects, i.e., from each node in $pa_i$ to $v_i$ for $i \in \{1, \ldots, d\}$ . The acyclic factor also makes that every SCM $\mathfrak{M}$ implies a unique observation distribution $P_{\mathfrak{C}}(\mathbf{X})$ which factorizes over $\mathfrak{G}$. Note that this satisfies the causal Markov assumption: each variable is independent of its non-effects given its direct causes, i.e. $P_{\mathcal{C}}(\mathbf{V}) = \prod_{i=1}^{d} P_{\mathfrak{C}}(v_i|pa_i)$. While $\mathfrak{G}$ is similar to a Bayesian network, they differ because the conditional factors imply causal relationships and go beyond the dependence assumption. These relations enable the SCM framework to predict the effects of causal action inference, also called *interventions*.

Intervention refers to a situation where we are analyzing what would happen if we act on the system, e.g., will my headache get better if I use this medicine? Formally these interventions are referring to situations where the variables are being manipulated externally, which does not necessarily reflect a situation that we have registered in the data. This type of intervention that fixes a constant $a$ (medicine A) to $x_i$ (medicine taken) is denoted by $\texttt{do}(x_i := a)$ and called an atomic intervention.

### 2.1.2 Counterfactual inference

Beyond the prediction of what will happen with a certain intervention with the `do` operator, the causal language can also predict counterfactual outputs. Counterfactual refers to situations where part of them is untrue. For example, in "Should I have woken earlier to finish my work?", the factual information is that I woke at a certain hour and

was unable to finish my task, but I am contemplating: if I had hypothetically woken up earlier, would I have had enough time to finish it?

When using counterfactuals, we are comparing two outcomes (e.g., work finished or not) under the same conditions, where they differ in only one aspect: the antecedent, which in the example above corresponds to waking up earlier than I did. It is interesting to notice that this analysis is not taking a similar situation of another person to compare, we are comparing the same situation. It is important to distinguish these two approaches, because when we use another situation for comparison, we can have different aspects that were not in the model and can make the results differ, i.e., another person can be less productive and therefore could not be able to finish the work.

The information not included in the model is referred to as *latent variables*, and they are common in a wide range of situations. Some of these situations exist because of the difficulty of taking measures of complex information like stress, past traumas, etc. Another context is when there is information that was not recorded or could not be accessed. Following, we formally define counterfactuals.

**Definition 2** (Unit-level Counterfactuals [35]). *Let $\mathfrak{M}$ be a structural causal model and $\mathfrak{M}_{\mathbf{X};\mathtt{do}(a)}$ a modified version of $\mathfrak{M}$, with the equation(s) of $X$ replaced by $X = a$. Denote the solution for $Y$ in the equations of $\mathfrak{M}_{\mathbf{X};\mathtt{do}(a)}$ by symbol $Y_{\mathfrak{M}_{\mathbf{X};\mathtt{do}(a)}}(u)$. The counterfactual $Y_a(u)$ is given by $Y_x(u) \triangleq Y_{\mathfrak{M}_{\mathbf{X};\mathtt{do}(a)}}$.*

Given this definition, we can effectively explain the data, since it is possible to manipulate each variable and get the new output. We will now formalize the three-step procedure necessary to compute *deterministic* counterfactual queries:

**Theorem 1** (Counterfactual Computation [35]). *Given an SCM $\mathcal{C}$ and evidence $e$, the counterfactual can be evaluated using the three following steps:*

1. ***Abduction***– *Use evidence $E = e$ to determine the value of $U$.*
2. ***Action***– *Modify the SCM $\mathcal{C}$, by removing the structural equations for the variables in $X$ and replacing them with the appropriate functions $X = a$, to obtain the modified model $\mathcal{C}_{\mathbf{X};\mathtt{do}(a)}$.*
3. ***Prediction***– *Use the modified model $\mathcal{C}_{\mathbf{X};\mathtt{do}(a)}$, and the value of $U$ to compute the value of $Y$, the consequence of the counterfactual.*

In this dissertation, as previously formalized in Def. 2, we focused on deterministic counterfactuals, which means that the analysis refers to contemplative questions of a single unit of the population, e.g., a patient, and we consider this exact patient with it on individuality's independent of the chance of it having it or not, which differs from the probabilistic counterfactual analysis where it is analyzed considering probable individuality's.

## 2.2 Counterfactual Inference and Health Care

While causal inference concepts have evolved using statistical concepts, these techniques have been recently introduced to deep learning [39]. For instance, there are investigations in disentangled representation [28], causal reinforcement learning [49] and recourse [24].In principle, causality for deep learning models may provide means to answer interventional and counterfactual questions, although current methods on deep-based counterfactuals are limited by modeling only cause-effect relationships [42], or by not providing computation of deterministic counterfactuals [34].

Currently, the most common approach to causal inference on the unit level is the estimation of individual treatment effect (ITE) [50, 40]. ITE is a model with three main variables: (i) the individual's context, with information about the analyzed sample, i.e. patient information, (ii) outcomes, which is the variable that is being analyzed, and (iii) the binary interventions (aka treatments), which impact the observed outcome. The results of how the outcome would have changed with a different treatment are usually applied in contexts where the validation with real data is impractical. To solve this problem researchers focused on using synthetic or semi-synthetic data to validate the theoretical analysis of generalization error bounds.In contrast to these methods, we assume only the individual's context and outcomes are available and focus on learning the latent features that affect the outcomes. This is a significant difference because, while ITE focuses on the bias between treatment and results, they do not assume latent features in the model.

Notice that causal inference techniques have also been applied to sequential data. For example, Zhang and Bareinboim [49] analyzed the dynamic treatment regime, a model that takes a sequence of decision rules to determine the most appropriate treatment for patients, called dynamic treatment regime. They proposed to use reinforcement learning to find an optimal set of dynamic treatment regimes given the causal diagram of the underlying, unknown environment. A factor model able to predict the treatment effects over time was proposed by Bica et al. [4], where the factor model is built with a time series decofounder that uses a recurrent neural network with multitasking output.

As far as we know, the attempts to consider personalized interventions in the context of mental health are limited by the work presented in Paredes et al. [33]. This work used machine learning to learn recommendations that match interventions to individuals and their circumstances over time to reduce stress. Even though they were able to improve engagement and local efficacy by matching the right intervention to the context of the user, they were unable to consider causal interventions. We propose to model the causal variables relating to patients and outcomes, enabling interventions as counterfactual queries. We are proposing an agnostic framework rather than focusing on a specific

Table 2.1: Comparison of your work with other from the literature.

| | Three Levels of Causation | Deterministic Counterfactual | Tabular Data | Latent Vars. | Non-linear |
|---|---|---|---|---|---|
| [4] | | ✓ | | | |
| [16] | | ✓ | | | ✓ |
| [34] | ✓ | | | | ✓ |
| [49] | | | ✓ | | ✓ |
| [50] | | | ✓ | | ✓ |
| **Ours** | ✓ | ✓ | ✓ | ✓ | ✓ |

illness (e.g., stress), and we also enable personalized interventions in both mental health and other well-being dimensions.

The most similar work to ours we are aware of is Pawlowski et al. [34], which proposed a model for counterfactual inference in high-dimensional data. Two main differences from their work to ours are that they focused on probabilistic counterfactuals while we focus on deterministic counterfactuals, and they use high-dimensional data (i.e. images), while we analyze tabular data. Similar to the method proposed here, they also consider a counterfactual result at an individual level, but they work on brain images of patients generated by probabilistic abduction while we work on deterministic counterfactual interventions on mental health.

One important and useful aspect of counterfactual inference is the ability to consider questions at the unit level. Therefore, when modeling and predicting counterfactual queries it is necessary to consider all characteristics of each individual that affect the outcome, and these characteristics can be observed or latent. As far as we know, we are the first work that proposed to model latent variables that affect multiple outcomes.

Modeling the presence of these latent variables and the causal relations, we can answer questions on all three levels of the hierarchy of causation. The causal relations make it possible to answer intervention questions, while the ability to abduct the latent information makes it possible to distinguish individuals and answer deterministic counterfactuals. Table 2.1 summarizes how previously works compare to the method we propose here.

# Chapter 3

# A new Framework for Counterfactual Inference

This chapter introduces a new framework that combines concepts of deep learning representation and structural causal models. Here we will mathematically define the applicability and requirements it needs to be able to (i) infer associations between dependent and independent variables; (ii) abduct the values of the latent individuality factor; (iii) enable interventions at the population level; and (iv) compute unit level deterministic counterfactuals.

## 3.1 Problem Setting

We define our problem as computing deterministic structural counterfactuals in multivariate regression scenarios. We work with observation data, where we have the response vectors (the observed effects) $\mathbf{y}_i = (y_{i1}, \ldots, y_{ik}) \in \mathbb{R}^k$, and the predictor vectors (the causes) $\mathbf{x}_i = (x_{i1}, \ldots, x_{id}) \in \mathbb{R}^d$. A crucial property of our model is that it assumes the presence of an individuality factor that affects the response vectors and which is both latent and immutable. We model the individuality factor of each instance as latent vectors (non-observed causes) $\boldsymbol{\varphi}_i = (\varphi_{i1}, \ldots, \varphi_{il}) \in \mathbb{R}^l$. The DAG $\mathfrak{G}$ of the resulting SCM is presented in Fig. 3.2. From the graph, we can learn the causal mechanisms to estimate the impact of populational and counterfactual interventions [35]. Following the convention of Buesing et al. [5], calculated values are shown in black boxes, observed variables in black circles, and unobserved variables in white.

The counterfactual queries are defined at the unit level, where the causal mechanisms are changed, while the exogenous and latent variables are kept the same for the observed instance. Assuming a structural causal model $\mathcal{C}$ that factorizes as a directed acyclic graph (DAG), we specify the causal mechanisms $\mathcal{F}$ using a deep neural network. We also assume the full specification of $\mathcal{F}$ and $\mathcal{F}^{-1}$, such that $\mathcal{F}(\mathcal{F}^{-1}(\mathbf{x})) = \mathbf{x}$. Hence, one can de-

termine the distinct values of individuality factors that give rise to a particular realization of the endogenous variables, $\{X_i = \mathbf{x}_i^F\}_{i=1}^N$, as $\mathcal{F}^{-1}(\mathbf{x}^\mathcal{F})$. To compute the counterfactual queries, we assume a set of actions, $A$, in a world $\mathcal{C}$, which yields to an updated world model $\mathcal{C}_A$ with structural equations $F_A = \{F_i\}_{i \neq I} \cup \{X_i := a_i\}_{i \in I}$. As a consequence, we can compute *any* structural counterfactual queries $\mathbf{x}^{SCM}$, which automatically accounts for inter-variable causal dependencies, for an instance $\mathbf{x}_A^F$ as $\mathbf{x}^{SC\mathcal{F}} = \mathcal{F}(\mathcal{F}^{-1}(\mathbf{x}^\mathcal{F}))$, read as: "*given the model* $\mathcal{C}$ *and having observed* $\mathbf{x}^\mathcal{F}$*, what is the value of the endogenous variables if the set of actions* $A$ *is performed*?".

Summarizing, we make the following assumptions:

(i) There are no other causes of $Y$ except the latent individuality factor $\varphi$, i.e., we assume *determinism*: a set of functions $f_y$ exists such that $Y = f_y(X, \varphi)$.

(ii) $X$ and $U$ are independent: $X \perp\!\!\!\perp \varphi$.

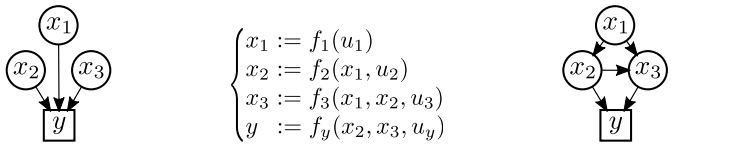
(iii) $\varphi$ variables are unchanged by hypothetical counterfactual interventions, i.e. they are immutable regardless of being latent.

## 3.2 Counterfactual Neural Network

We divided the computation of our counterfactual framework into three phases. In the first step, we have the objective to create a DAG, it is important to correctly define the causal relations, define the predictor variables that have a causal effect on the response variables. This is necessary because we will only use variables that have a causal effect, therefore we use *causal graph discovery algorithm* that generates a causal graph. Hence, given the causal graph in Fig. 3.1c, we would consider the variables $x_2$ and $x_3$ (direct causal mechanisms to $y$), and ignore variable $x_1$, which is available in the dataset but it is not a direct cause to $y$.

The second step is to run a *base model* $h$ with the objective of using the variables that have causal effects to predict the response variables. Even though it can have an agnostic base model, it must have the capacity of fitting the curve and be trained to reduce the error close to 0. A non-linear problem, for example, cannot have a linear regression as a base model, because the residual error will not reflect the information that is expected in the next step of the framework.

In the final step, we propose the individuality factor discovery counterfactual network, *ICoNet*. This network has a structure that takes the residual error from the base model together with the predictor and response variable to predict the counterfactuals.




(a) Regressor-centric view

(b) SCM $\mathcal{C}$

(c) Causal graph $\mathfrak{G}$ for $\mathcal{C}$

Figure 3.1: A view frequently adopted in regular regression models (a) assumes variables are independently manipulated inputs to a given fixed and deterministic regressor $h$. In the causal approach to counterfactual inference taken in this work, we rather view variables as causally related to each other by a structural causal model (SCM) $\mathcal{C}$ (b) with associated causal graph $\mathfrak{G}$ (c).

The main important part is that in its learning process it is forced to learn the latent individuality factors $\varphi$ shared by the output variables. Thus, with the learned value of $\varphi$ the ICoNet enables us to change the predictor's variables of interest and perform the counterfactual prediction.

In section 3.2.2 we delve into the generation process of the residuals of the interpolar, where the main rationale of the proposed method is that it corresponds to the effects of the latent variables that directly affect $Y^F$. In section 3.2.3 we present, ICoNet, a deep neural network that uses predictors and response variables together with the residuals to learn the latent variables to predict counterfactuals.

### 3.2.1 Causal Graph Discovery

One requirement to compute counterfactual queries is the causal relations among the variables, but these relations usually are not directly read from the data alone. As the correct relations must be used, it is necessary to be given an SCM from the professional knowledge of the real problem or to be inferred using algorithms. In most cases the practical applications causal graph is unknown even for the specialists and learning a directed acyclic graph from observational data is an NP-hard problem [8, 9].

The field of causal discovery is concerned with the problem of inference of causal relations [13, 51] and has been applied in a wide range of practical contexts. One area where it was successfully applied was with health-related problems, where the understanding of the causal impact of factors is important to better treat patients. The use of causal DAGS applied to health care was surveyed by Tennant et al. [44]. Another important study was presented by Shen et al. [41] which worked with data about Alzheimer's disease and focused on investigating if the causal discovery algorithms could use observational

clinical data and would generate the same relations that are already previously known. The results showed that a causal discovery framework should be used with longitudinal data providing full prior knowledge available.

To find the practical causal graphs in this work it was used the framework NO TEARS [51], which formulates the graph discovery learning problem as a continuous constrained optimization task, leveraging an algebraic characterization of DAGs. More formally, NO TEARS tries to solve the following continuous optimization problem:

$$\begin{aligned} \min_{\mathbf{G} \in \mathbb{R}^{d\times d}} \quad & \frac{1}{2N}||\mathbf{X} - \mathbf{X}\mathbf{G}^T||_2^2 + \lambda||\mathbf{G}||_1 \\ \text{s.t.} \quad & h(\mathbf{G}) = 0, \end{aligned} \tag{3.1}$$

where $\mathbf{X}$ is the input data, the function $h(\cdot) : \mathbb{R}^{d\times d} \to \mathbb{R}$ is continuous such that $h(\mathbf{G}) = 0$ iff the graph is acyclic, and $\mathbf{G} = (\mathbf{g}_1, \ldots, \mathbf{g}_d)'$ is the weighted adjacency matrix.

### 3.2.2 Base Model

One main component of our framework is that we assume a *base model*, or *interpolator*, where the objective is to minimize the error, trying to achieve almost zero training error. It is important to notice that we are considering two main situations where the counterfactual is possible, the first one the base model, which is capable to reduce the training and testing error to 0. This is indicative that all the individual's characteristics are already being measured, therefore, if it is using only the variables with direct causal effect, changing the variables to feasible values and predicting will output a mathematical valid counterfactual. Since it is straightforward to use this with current machine learning techniques, we will not do further investigations.

The second case, that is the focus of this work, it is impossible to have 0 error in all training and test examples because there are latent variables that directly affect the response variables, therefore when running a machine learning model that tries to minimize the error, we must have a near-zero error. This step of the framework is agnostic of the base model, the only requirement is that the model is capable of correctly finding the global minimum of the optimization criteria, this forces the model to only have the error introduced by the latent variables.

Even though we could use any interpolator we opted for a multilayer perceptron (MLP) network trained with the $L_1$ norm as the loss function and the ADAM optimizer. Even though deep neural network learning involves several parameters, it was shown by Zhang et al. [48] deep learning is frequently observed to generalize well in practice. The

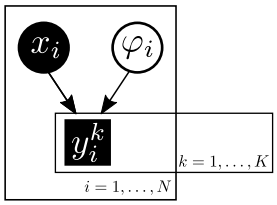
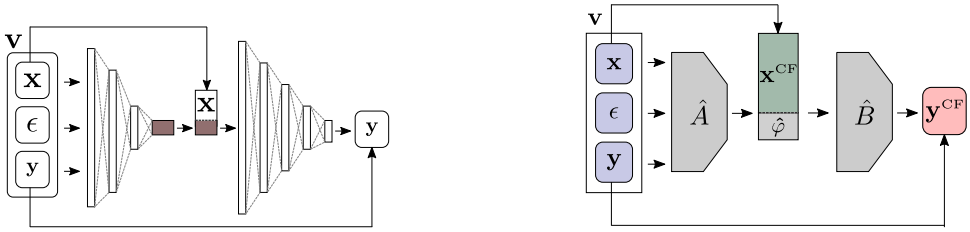



Figure 3.2: Causal DAG with multiple outcome $y^k$, features $x$, and latent individuality factor $\varphi$. Calculated values appear in black boxes, observed variables in black circles, and unobserved variables in white.




(a) Counterfactual network model training.

(b) Counterfactual model prediction.

Figure 3.3: An illustration of the proposed counterfactual model, highlighting the factual (i.e. $\mathbf{v} = \langle \mathbf{x}, \epsilon, \mathbf{y} \rangle$) and counterfactual inputs (i.e. $\mathbf{x}^{CF}$) to compute the factual (i.e. $\mathbf{y}$) and counterfactual outcome (i.e. $\mathbf{y}^{CF}$).

results showed by Belkin et al. [3] also contradict the conventional wisdom that interpolation (vanishing training error) leads to overfitting or poor generalization, its survey analyzed that the concepts of interpolation and overfitting are distinct, but interpolation does not contradict generalization.

In our framework we train a different base model for each response variable in our multiple regression setting, therefore we compute the training error $\epsilon_{\mathbf{k}} = \mathbf{y_k} - \hat{\mathbf{y}}_{\mathbf{k}}$, for each of our $k$ variables, where $\hat{\mathbf{y}}_{\mathbf{k}}$ stands for the prediction of the base model. Since we consider a good fitting model and can achieve almost zero error in training, we can attribute the error to the latent individuality of each unit. We exploit the interpolation property of neural networks to isolate the individuality factor $\boldsymbol{\varphi}$ shared by our outcome variables using the all calculated errors $\epsilon_{\mathbf{k}}$.

### 3.2.3 Individuality factor discover Counterfactual Network

We propose a new network architecture to compute structural deterministic counterfactual queries based on the individuality factor, named Individuality factor discover Counterfactual Network (ICoNet). The network input is composed of three main information, (i) the predictor variables $\mathbf{x}$, here is important that it is inputted only the causal

related variables and that they are independent,(ii) the response variables $\mathbf{y}$ which are dependent and causal related with $\mathbf{x}$ and the latent variable, (iii) the error $\epsilon = \mathbf{y} - \hat{\mathbf{y}}$, where $\hat{\mathbf{y}}$ refers to the base model prediction. This info is included in the input vector $\mathbf{v}$.

The ICoNet is trained with the main objective of predicting the response variables $\mathbf{y}$ with L1-distance as the loss function since we are searching for the lowest mean absolute error. To abduct the individuality factor created by the latent variables and enable counterfactual predictions, the network architecture was modeled in a way that we create two bottlenecks. Since the variables $X$ are independent, the first bottleneck forces an intermediate layer to learn the info from the latent variables that are necessary to correctly predict $Y$. This variable, in our case, is the individuality factor $\varphi$ – represented by a single neuron shown by a full brown square in Figure 3.3a). The $\varphi$ is abducted because we merge, with a merge layer, the end of the first bottleneck, which is a single neuron, to the observed variables $X$. This merged layer is followed by a sequence of feed-forward layers, which will form the second bottleneck.

It is important to note that one main requirement of the architecture is having a multiple regression problem, because having more than one output being mapped to a single neuron where the variables are independent will force the training to learn $\varphi$ at the end of the first bottleneck. Therefore, the same architecture will not work in cases where there is just one response variable $Y$ (simple multivariate regression cases). It will also not work if the architecture parameters are set with a higher number of neurons at the end of the first bottleneck than there are output variables $Y$ to be predicted.

When having the same amount or more neurons at the end of the first bottleneck than the number of output variables $Y$, the learned information is trivial, as the model can learn a representation of the values of $Y$. This is a well-known characteristic of autoencoders, an architecture that also relies on the bottleneck to learning representations [31].

The first bottleneck has three layers with Relu activation, with 40, 20, and 5 followed by one linear unit. They are followed by the merge layer with 1 unit, which concatenates the final unit with the $X$ input. The second bottleneck has 4 layers with Relu activation and 500, 500, 250, and 50 units, respectively. The output is a sigmoid layer with two neurons (one for each output), as the datasets have their outputs normalized in $[0, 1]$. The architecture was defined trying to maintain the bottlenecks and have more neurons in the first layer than the input layer. The results were given by the first tried, since it was not the scope of this work to deepen in the different configuration possible.

**Model training.** Formally, we define our counterfactual neural network as the solution of an optimization problem considering the following t-tuple $\langle \mathcal{A}, \mathcal{B}, \mathbf{V}, \Delta \rangle$ where:

1. $\mathcal{A}$ is a class of functions from $\mathbb{R}^{d+2k}$ to $\mathbb{R}^{d+1}$.
2. $\mathcal{B}$ is a class of functions from $\mathbb{R}^{d+1}$ to $\mathbb{R}^{d+2k}$.

3. $\mathbf{V}$ input vector comprises the following three vectors: (i) $\mathbf{X}$: a set of $n$ training vectors in $\mathbb{R}^d$; (ii) $\mathbf{Y}$: the corresponding set of $n$ target vectors in $\mathbb{R}^k$; and (iii) $\mathbf{E}$: the set of $n$ base model errors in $\mathbb{R}^k$.

4. $\Delta$ is a dissimilarity or distortion function defined over $\mathbb{R}^{d+2k}$.

For any $A \in \mathcal{A}$ and $B \in \mathcal{B}$, the counterfactual network transforms an input $\mathbf{v} \in \mathbb{R}^{d+2k}$ into an output vector $A \circ B(\mathbf{v})$. The model training refers to find $A \in \mathcal{A}$ and $B \in \mathcal{B}$ that minimize the overall distortion function:

$$\min_{\mathbf{A},\mathbf{B}} \sum_{i=1}^{n} \mathbb{E}(\mathbf{v}_i, x_i, \mathbf{y}_i) = \min_{\mathbf{A},\mathbf{B}} \sum_{i=1}^{n} \Delta(A \cdot [B(v_i), x_i], y_i) \tag{3.2}$$

**Counterfactual prediction.** After the model is trained, it can be used to perform counterfactual predictions. Figure 3.3b illustrates this process, where the counterfactual prediction depends on the factual instances and previously learned counterfactual network parameters. The network receives as input the same factual inputs $\mathbf{x}$, targets $\mathbf{y}$, and errors $\epsilon$. Additionally, the model is given the hypothetical or counterfactual values of instance features, i.e., $\mathbf{x}^{CF}$, to the merge layer. Given the previous inputs, the structural model (i.e., ICoNet) can learn the target counterfactuals $\mathbf{y}^{CF}$.

# Chapter 4

# Experimental Analysis

This chapter evaluates the proposed framework considering two different scenarios. In the first, composed of 53 synthetic datasets, the latent information is known, and hence we can better evaluate the proposed approach. In the second, we use real data from a questionnaire that deals with mental health and quality of life during the COVID pandemic. The results of this last approach were validated by experts and published studies in the area.

## 4.1 Experimental Setup

In both scenarios considered in this study, data was used using a train-test division, being 70% for the former and 30% for the latter. We did not perform any cross-validation since our main objective is not to make factual predictions. Experiments were run on one GPU (GEFORCE 1060 6GB), and the framework was implemented in Python 3.6.12 and Keras 2.4.3. The code and more implementation details are available online[1].

A base model with the same architecture was used for all datasets. The architecture used for the base network was an MLP with 5 tanh layers of 50 units, trained for 250 epochs with a 0.001 learning rate, the Adam optimizer, and mini-batches of size 128. The counterfactual network, ICoNet, was trained for 500 epochs, with a learning rate of 0.001 and mini-batches of size 128. As the objective of the optimization is to stop after the double descent we did not need to use early stopping. It will be shown further in this section. This architecture was defined using mainly the default values and a number of layers that could be a good predictor, without further experimentation.

[1] https://github.com/marchezinixd/ICoNet

## 4.2 Experiments with synthetic data

The objective of the experiment with synthetic data is to show that the model can abduct the individuality factor and answer counterfactual questions. We focused on showing how the model deals in different ways with linear or non-linear problems and additive or multiplicative. For the experiment was generated data for both $X$ and $Y$={ $y_1$,$y_2$} and the individuality factor $\varphi$, which is latent in our real-world problem. Since we are dealing with a controlled experiment, it is possible to use the value of $X$ and the known $\varphi$ applied to the function and verify the correctness of the counterfactual result.

We generated 53 simulated datasets, where the function was borrowed from the experimental setup followed by Stegle et al. [43]. The datasets were made using the model function $Y = (X + \beta X^3)e^{\alpha E} + (1 - \alpha)E$, where the random variable $X$ was sampled from a normal distribution with mean 0 and variance 1, and we consider $E$ as our individuality factor $\varphi$, as it is independent of $X$ and affects $Y$.

It is important to notice that this formula was used because its parameters can alter both the non-linearity and the additive or multiplicative nature of the function. The $\beta$ value will control the linearity of the function, having $\beta = 0$ as completely linear and $\beta = 1$ completely non-linear, the values in between control the strength of each. While the $\alpha$ value will control the nature of the relationship with the individuality factor, with $\alpha = 0$ we have an additive factor and $\alpha = -1$ or $\alpha = 1$ will be completely multiplicative, the value in between also shows the strength of each.

The 53 independent datasets were generated considering different values of $\alpha$ and $\beta$. Each dataset has $N$ = 10,000 samples, and were generated following three different scenarios. In the first, $\beta = 0$ and $\alpha$ ranges from 0 to 1 with a step of 0.1. In the second, $\alpha = 0$ and $\beta$ ranges from -1 to 1 in steps of 0.1. In the last, $\alpha = \beta$ as values range from -1 to 1, with steps of 0.1. In all cases, an equal number of instances were assigned values of $E$ equals 1, 2, and 3.

The model requires the prediction of two $Y$ , to maintain the consistency we generated both using the same equation. Since the equations shared the individuality factor $\varphi$, the differentiation comes from the values for $X$ that were generated separately, meaning that they can (and in most cases will) be different.

Another characteristic of the synthetic data is that we already know how the causality is done and we do not need to run a causal graph discovery algorithm, here is even simpler because we only have a single variable that is the direct cause of $y$. With this info, $X$ is inputted in the base models and predicted $\hat{\mathbf{y}}$. Note that although we know the value of $\varphi$, it is not used in any step of the prediction of the base model, neither on the ICoNet, it is only used to calculate the metrics.

As we observe in Figure 4.1, when running the proposed method in our synthetic

dataset (a similar behavior is found for the real-world mental health dataset), both training and test error decrease as a function of model complexity, measured in terms of epochs, and consequently, parameter convergence. This behavior follows a *double descent* curve, which is a mark of overparameterized interpolator models. And therefore does not need to use early stopping to search for the optimal point before diverging the first time.

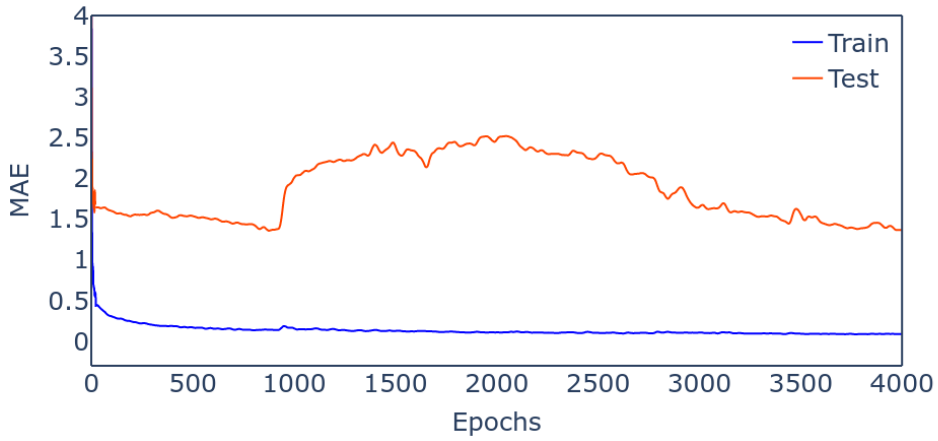
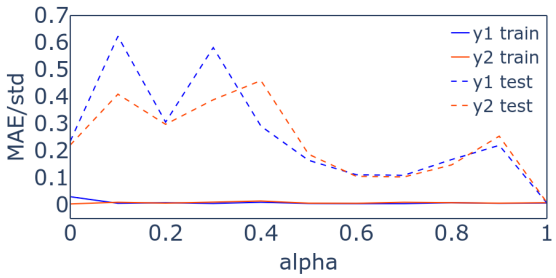



Figure 4.1: Double descent behavior of the base model in the test loss.

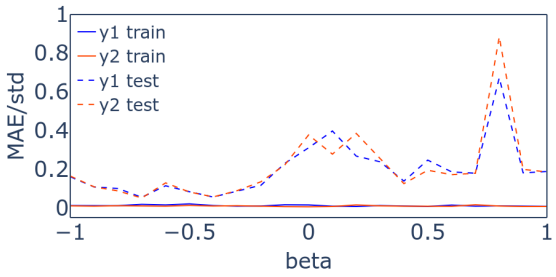
The first verification of the capability of the model is checking the train and test metric, we used the normalized mean absolute error (MAE) with the formula MAE/standard deviation for $y_1$ and $y_2$. It was necessary to normalize the MAE because as the formula varies in the different tests, it changes the range of response values. Fig. 4.2 shows the values of normalized mean absolute error for $y_1$ and $y_2$ considering the three scenarios used to generate the synthetic datasets. We can observe that all three scenarios had a low training error, while the test error differs significantly. In Fig. 4.2a the test error starts high when $\alpha$ is low and decreases as $\alpha$ increases and the individuality factor adds a multiplicative effect in $Y$. This shows that while the algorithm is having a good result in linear cases it can perform better with the multiplicative individuality factor.

Fig. 4.2b shows that as $\beta$ gets closer to 0 (the linear case), the error increases, meaning that the model with additive error is learning to generalize better for non-linear cases. It is interesting that while the highest peak was in 0.8, showing that the model could not generalize well, the situation where it consistently had higher values was in the linear case which, in theory, it is simpler to learn. Finally, in Fig. 4.2c we observe that the test error was kept below 0.4 in almost all series, the highest peak was a little above this threshold, but it was not consistent, showing that it could be a specific case where the model could not generalize well and needed more training. We can see consistent peaks when $\alpha$ and $\beta$ around 0, but the value is always below 0.3, which is considered a low value. In sum, we analyzed that the model can generalize in most cases, especially

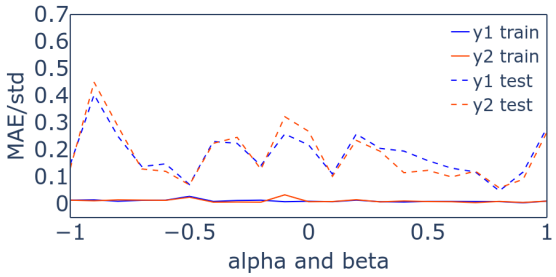



(a) $\beta = 0$.

3




(b) $\alpha = 0$.



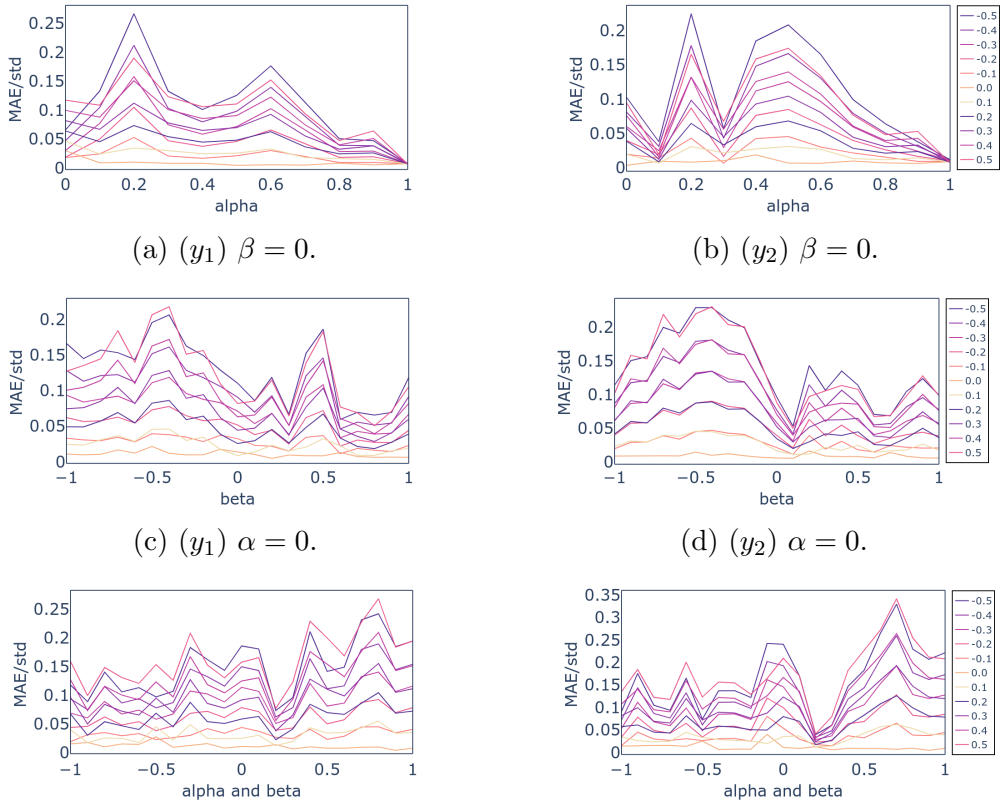

(c) $\alpha = \beta$.

Figure 4.2: Comparison of train and test *normalized association errors* for different combinations of $\alpha$ and $\beta$.

when dealing with non-linear and multiplicative cases that usually are the characteristics of real-world models.

Next, we verify the counterfactual capability of the ICoNet, which will evaluate the correctness of the counterfactual prediction. The test was made changing the value of $X$, following the formula $X^{cf} = X + q$, where $q$ values ranged from -0.5 to 0.5 in 0.1 steps. With this value inputted, we generate the response for both $Y_1$ and $Y_2$, and because the value of $\varphi$ maintains unchanged and known, we can calculate $Y^{cf}$, hence the MAE value.

Fig. 4.3 shows the values of MAE/std for the 10 counterfactual queries created using 10 different values for $X$. Note that in all cases we had values below 0.3, which is considered low. Looking at the results for both Y1 and Y2 in Figs 4.3a and 4.3b we can analyze the effects of moving from additive to multiplicative effect of the individuality factor on the counterfactual prediction. For both $Y_1$ and $Y_2$ we can see that as $\alpha$ goes




(a) $(y_1)$ $\beta = 0$.

(b) $(y_2)$ $\beta = 0$.

(c) $(y_1)$ $\alpha = 0$.

(d) $(y_2)$ $\alpha = 0$.

(e) $(y_1)$ $\alpha = \beta$.

(f) $(y_2)$ $\alpha = \beta$.

Figure 4.3: Comparison of *normalized counterfactual errors* for simulated data with different combinations of $\alpha$ and $\beta$: the left column shows the values for $y_1$ and the right column for $y_2$.

towards 1 the counterfactual error tends to go to $\approx 0.005$. Fig. 4.3a had a more stable results around 0.1 until $\alpha = 0.8$, Fig. 4.3b had a little higher results around 0.15, with lower peaks in 0.1 and 0.3. This shows that as the function has a higher multiplicative effect, in the case where the function is linear, the counterfactual tends to have better results, even in the extreme values of $X$ for the counterfactual queries.

In Figs 4.3c and 4.3d, we have $\alpha$ set to 0.0 always maintain the additive factor, while we have the behavior of the counterfactual changing from linear to non-linear as the valuers diverge from 0. Here we have two different patterns, a higher one when $\beta$ is negative and another when it is in positive value. In the former, we have higher error values and higher variance, with both $Y_1$ and $Y_2$ going a little above 0.2 with $\beta = -0.4$. The latter with $\beta$ with positive values we see it with a lower variance between the different counterfactual queries, and the two most extreme always around 0.1. Another interesting analysis is that while in Fig. 4.3d when $\beta = 0$ we have the error $\approx 0.05$ and in Fig. 4.3c $\approx$ 0.1, while they are both low values, it can indicate that in some cases the ICoNet learns better the relation of the variables and individuality with one $Y$. We can conclude that it was harder for the algorithm to learn the non-linear function when $\beta < 0$ compared with $\beta > 0$, but it still could learn well both linear and non-linear counterfactual.

Finally, Figures 4.3e and 4.3f show the cases where the counterfactual queries had the worse result, with the largest error overall, getting near to 0.35 standard deviations. In this graph, the lowest values of MAE/std happened when $\alpha = \beta = 0.2$ for Y1 and Y2, i.e., when we had a situation close to linear function with an additive individuality factor. This graph is the one that brings the most amount of info about the behavior of the algorithm, going from the simplest, that is, linear with additive individuality factor, to non-linear with multiplicative, which is more common in real-world problems. It also brings situations where it has both linear and non-linear factor and multiplicative and additive factors which are complex cases. We can see that for both $Y_1$ and $Y_2$ we had good performance when $\alpha$ and $\beta$ were equal to -1 and 1, the non-linear multiplicative cases, with errors below 0.25 even in the most extreme counterfactual query. It also had good results in the linear additive case, values in between 0 and 1 it is interesting to see that in both functions we had a growing pattern from until 0.7 or 0.8, after reaching the peak they go down to better results in the extreme. While the values between -1 and 1 we had a linear constant behavior around 0.15 and a smaller variance of the counterfactual queries.

The results in Figures 4.2 and 4.3 demonstrate that there is not a necessary correlation between good counterfactual prediction and the ability to generalize factual predictions. There are situations where we had good counterfactual predictions and the test error of the model was high. For example, when $\alpha = 0.3$ and $\beta = 0.0$ we had a high test error for the factual prediction, with $\approx 0.58$ standard deviations, while even the most extreme change in the counterfactual prediction was still an error below 0.15. Also the counterfactual result of $\alpha = \beta = 0.7$ was one of the highest counterfactual errors, getting to $\approx 0.25$ for $Y_1$ and 0.3 for $Y_2$ while the test error was below 0.1. One possibility that can explain the disconnection between test cases and counterfactuals is that the model probably needs to overfit the case to learn how to abduct individuality.

Overall, the results also showed that as the counterfactuals diverge from the original values, the counterfactual error tends to grow. This behavior may be related to the standard problems in machine learning, such as model effectiveness or lack of representative training data. Since the data was generated with a normal function it is expected that most of the values are in a certain range, and applying changes that sum or subtract can lead a subtle amount of samples to fall in the range that did not have enough examples to correctly learn the function. Another possibility is related to the bottleneck, this approach has some similarities to auto-encoders, and it is well-known that it can create non-continuous representations, which could lead to small changes in the factual values to cause very different impacts in the predictions. While in auto-encoders some strategies were developed, i.e., variational auto-encoders, the adaptation of them to our problem must be further studied, since the addition of noise or uncertainty can result in a probabilistic or non-deterministic approach, that is a different objective from ours.

## 4.3 Mental Health Case Study

This section applied the proposed framework to a mental health dataset, the objective was to calculate all three levels of Pearl's hierarchy of causal knowledge. The first step, *association*, showing that given the observed data we can predict a value close to the real one. In this step, we can quantitatively measure the correctness of the results, as the real expected result is factually observable. The second step, *intervention*, here is studied the impact of intervening in the whole population and seeing the statistic impact. Finally, in the third step, we will analyze the *individual counterfactual*, here we took samples and analyzed what would have happened if they had had different preconditions. It is verified if ICoNet can correctly predict different results for people that had the same factual collected information and the different impact counterfactual have on them. For the last two steps it is not possible to quantitatively measure the results, they can only be evaluated qualitatively[35], as they are based on counterfactual outputs, that refer to a contemplative study. Therefore, these results were validated by psychologists and psychiatrists.

The chosen problem is related to two well-known instruments for measuring symptoms of mental health diseases and quality of life: the Brief Symptom Inventory (BSI) and WHOQOL-BREF (World Health Organization Instrument to evaluate Quality of Life: Brief Version), respectively. The info came from questionnaires, where WHOQOL-BREF is composed of 26 questions and the BSI by 53. The scores are generated based on a linear combination of answers provided by people in the questionnaires.

BSI is an instrument designed to identify psychological symptoms in nine dimensions Somatization, Obsession-compulsion, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism. Each of these BSI dimensions has its score that is calculated using a combination of the 53 questions, but no question answer is used in more than one dimension. In our work, we focus on BSI-somatization, where somatization is the development and persistence of symptoms unexplained by medical causes, and it is well known that somatization is the result of both depression and anxiety [19]. This domain knowledge of the causal relationship between depression and anxiety being the causes of somatization was the reason we limited our study to this domain. The other variables still do not have well-established causal relations between them.

The second score, WHOQOL-BREF, measures people's quality of life considering four domains: physical health, psychological, social relationships, and environment. Similar to the BSI score the four domains scores are calculated using a linear combination of the values of specific questions in the questionnaire. Each question is also only used by one domain. We work with all domains except social relationships, as it includes only

three questions. The small number of questions would affect the analysis that was made in these problems, as one question can already have a large amount of the score's info.

One important characteristic of this data is that it requires the abduction of the individuality factor to enable the counterfactual calculation. This is necessary because there are known latent variables that were not measured. In this case, they are unviable or really hard to obtain since the effects come from genetics and life experiences [15], that affect both BSI and WHOQOL, which make these scores a perfect fit for the causal model proposed here.

Studies have shown that exists an association between somatization and behavior symptoms [46, 18]. Although there is plenty of evidence, they cannot explain how it works, this lack of explicability represents a necessity of causal models. These models could give a better understanding of the alignment between the underlying biology and behavior, and how changes in the scenario would impact the outcome. As it seems important to understand behavior and psychopathology, the results generated by the counterfactual may help in understanding the effects of these changes. In the population, analysis is possible the see the extension of the impact given different scenarios, while our individual analysis can show how people with similar conditions would have different impacts, even when checking equal counterfactual queries.

### 4.3.1 Dataset

The dataset was built from the Psychological Covid impact study, which was divided into sections that include psychological and socioeconomic questions. It was answered online from May-2020 to July-2020. We extracted only 53 BSI and 26 WHOQOL answers, having a completely anonymized dataset. Filtering the samples that did not completely answer all questions in the two cited sections, we have 153,514 valid individuals. The BSI questions use the 5-points Likert scale from 0 (not at all) to 4 (extremely), while the WHOQOL uses a 1 to 5 Likert scale.

As previously stated, the scores we want to predict and generate counterfactuals are calculated with a linear combination of the values of specific questions, most of the current regressors can easily predict the perfect score. This scenario is not appropriate to test the proposed framework, as this prediction would already have all the necessary information to predict and there are no latent variables (individuality factor), therefore we performed a data preparation step.

To prepare the data, we first used the mapping from questions to the domain for both instruments, then separated the question of each domain and calculated the

correlation between the values of the calculated scores and its respective questions. Using the correlation score, we selected the two most correlated to the score. It was necessary to have highly correlated questions as we need to have a model where the base model uses the inputs to give a close, yet not perfect, prediction. Finally, we disregard the rest of the correlated questions and merged only these two with a set of other questions extracted from the causal graph, since we heavily rely on the correctness of the causality between the answers and scores. It is important to notice that this pre-processing was done to each of the domains, and some of them had little changes that will be further specified.

The first step of our framework is the inference of the causal graph, we chose to use NOTEARS as it is a well-known algorithm that was proposed by Zheng et al. [51]. We separately build a causal graph for each domain, since it enables a better fine-tuning of the algorithm. For the BSI-Somatization score, we restricted our search space to questions classified in anxiety, depression, and somatization categories. This choice was based on the causal graph produced by the authors in [17]. Since the Somatization questions were already analyzed in the preprocessing step, taking only the two most correlated, we took all questions that directly affected them with a weight greater than 0.05. The weight corresponds to the strength of a causal relation. The 6 variables identified were merged with the two most correlated to the index, and the base model has as input 8 variables. It is important to notice that the questions that are direct causes of the two most correlated were not used, as they would be highly dependent.

For WHOQOL, we did not have any references to filter out variables, and hence we have used the answers to all questions to build the causal graph. For this reason, it was also not possible to filter any of the Three domains, therefore we tested the combination of BSI with each one of them. We applied the same process for both environmental WHOQOL (*Env*) and the Physical WHOQOL (*Phys*), with the only difference being the cut-off value that was set to 0.1 to *Phys*, as we observed that the weight values were much higher for this graph. For WHOQOL Psychological (*Psych*), the causal graph showed that one of the questions that are used in the formula is caused by almost all the others, using them all would not be good, as some of them are strongly dependent. Instead of adding all cause variables to the graph, we opt to include a third variable that was directly associated with the score. The number of input variables used by each model is shown in Table 2.

Table 4.1: Number of input and output variables and MAE measured for association relations with BSI and WHOQOL.

| **BSI** | | **WHOQOL** | | Train | | Test | |
|---|---|---|---|---|---|---|---|
| $\#X_1$ | $y_1$ | $\#X_2$ | $y_2$ | $y_1$ | $y_2$ | $y_1$ | $y_2$ |
| 8 | *Somat.* | 9 | *Env* | 0.0179 | 0.0079 | 0.0439 | 0.0081 |
| 8 | | 12 | *Phys* | 0.0007 | 0.0007 | 0.0165 | 0.0217 |
| 8 | | 10 | *Psych* | 0.0004 | 0.0005 | 0.0070 | 0.0094 |

### 4.3.2 Association

The first level of Pearl's causality hierarchy is association, the ability to correctly predict the factual value given the available info. As it was shown before we could not attest the direct correlation of a better generalization for association in unseen predictions with better counterfactuals, but it is a valid analysis to see the capacity of the framework to correctly predict the first step.

It is important to notice that although we are performing a factual regression where it is possible to calculate metrics, we did not compare the results with other methods as we use the desired output as the input of the method. This means that even with perfect results it does not mean that it could be used to predict unseen situations, as this is not the objective.

Table 4.1 shows the MAE for each of the three scenarios combining BSI (y1) with one of the WHOQOL domains (y2). It brings the results for both training and test for the two outputs variables of each case. BSI with Environment had good results for training in both Y, but the BSI generalization in the test was considerably higher, but it is still a low error, considering the values range from 0 to 1. From the results of this scenario, we can conclude that it had a great generalization for Environment and a little worst for BSI.

The second scenario shows the prediction of BSI Somatization and WHOQOL physical, where both training values were equally low, and although the test error is higher it is still in a considerably low range. Finally, it is shown Somatization combined with WHOQOL Psychological, wherein both training and testing showed the best results for BSI and really low values for WHOQOL. These results show that the base model can learn associations and correctly fit the function for the factual prediction task.

### 4.3.3 Population-based Interventions

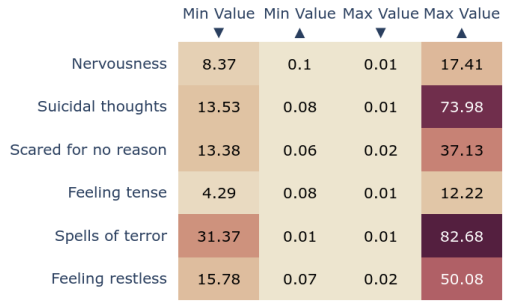
The population-based intervention has the objective to understand the impact of actions when intervening in the whole population. The experiment was done by changing one single variable and checking the percentage of people that had their scores increased, lowered, or maintained. Although we cannot guarantee the independence of the variables, and that this does not affect the result, we considered it while doing the test. Therefore when changing only a single variable and keeping the original value in the other it will not generate an unrealistic case where the input conflicts with itself.

Another premise of our model is that it works with causally related variables and there is no evidence of a strong causal relation between BSI features and WHOQOL scores nor between WHOQOL features and BSI scores. The only variable that directly affects both BSI and WHOQOL is the individuality factor, and since it is latent, it is kept the same for all tests. This premise is in accordance with the proposed causal model in Fig 6a and was taken into consideration in the first step of the framework, the outcomes do not have a causal edge between them, and the only expected information that they share is individuality. For this reason, we restricted our tests to analyzing the impact from BSI features to BSI score and from WHOQOL features to WHOQOL score. Remind that as BSI increases, we relate to worse cases of somatization, while a higher WHOQOL score is understood as better quality of life in the analyzed dimension.

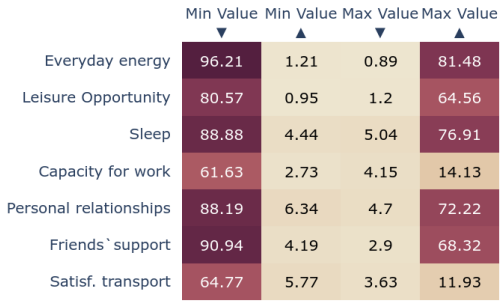
To analyze the impact of interventions in the whole population we tested extreme situations, excluding people where the factual value was already the desired extreme counterfactual query. This was done to create a direct analysis of the results with changes, where if we use the higher value we are always increasing, and if is the lowest all the counterfactual are decreasing. Although we still have different situations, the understanding and validation are simplified, since we do not have percentages where some people are in worse conditions and some in better, which is a more complex case. Another reason to exclude people whose original responses in the questionnaire were equal to the extreme value analyzed is that they would not have changes in their scores.

We consider scores change if their values increase or decrease by at least 0.01 (1%). Other variations were also tested but were both discarded as they were considered harder for psychological analysis by the experts. Considering 5% or higher we had almost all individuals of the population without changes in most cases. While using lower values like 0.1% was hard to justify the changes since they could simply be capturing the model's prediction variation.

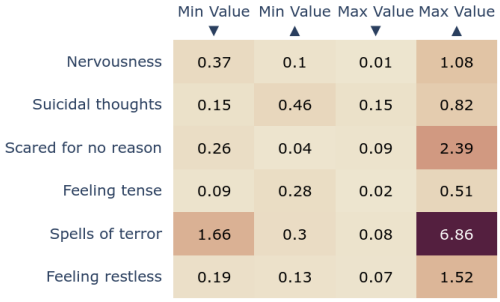
The figure shows 4.4 heatmaps, being $y_1$ and $y_2$ for each of the three cases, it is built with each line corresponding to a question in the questionnaire. The details of the questions are also described in Appendix A, while in the heatmap a simplified description



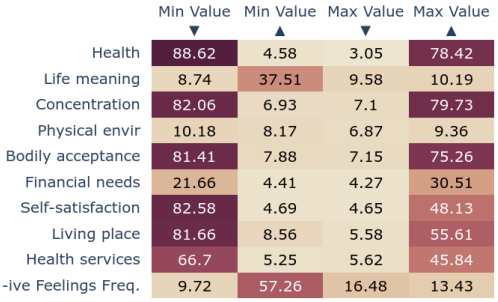

(a) Changes in BSI with Psych WHOQOL.



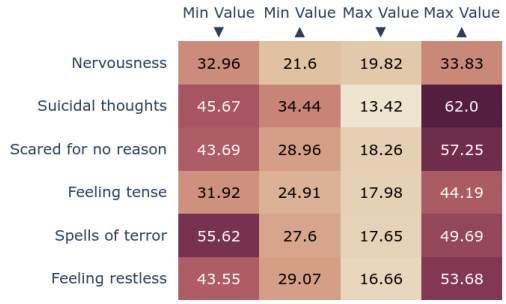

(b) Changes in Psych WHOQOL with BSI.



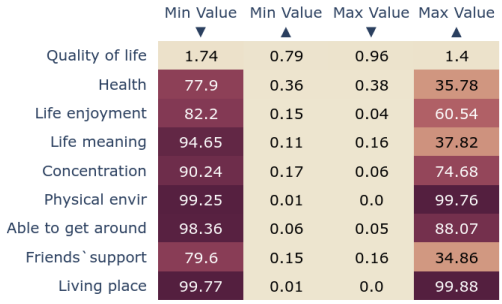

(c) Changes in BSI with Phys WHOQOL.




(d) Changes in Phys WHOQOL with BSI.




(e) Changes in BSI with Env WHOQOL.




(f) Changes in Env WHOQOL with BSI.

Figure 4.4: Heatmaps showing populational changes when counterfactualy changing values of variables. First and third columns show the percentage of the population that has the value of the score (indicated in the caption) decreased when the variable in the line receives minimum and maximum values, respectively. Second and fourth columns show the same percentage of the population that have their score values increased with these changes.

is given in favor of the analysis. As previously stated, we analyzed the situations where the original values were counterfactually changed to the minimum and maximum values, where in BSI they are, respectively, 0 and 4 and in WHOQOL they are respectively 1 and 5. The first and third columns of each map show the percentage of the population that have their score value (indicated in the caption) *decreased* when the variable in

the line received the minimum and maximum values, respectively. Following the same pattern, the second and fourth columns show the percentage of the population that have their score values *increased* with the changes: when the variable in the line received the minimum and maximum values. It is important to notice that the heatmaps do not show the questions that were originally used to calculate the scores, as their correlation is high and almost all populations had the expected impact.

First, we analyze the BSI results to compare the impact of the same feature in the three different scenarios, considering the different WHOQOL categories (All heatmaps in the first column in Figure 5 are BSI results). Note that the impacts follow the same pattern where the extremes (first and fourth columns) are always with a high value, but the percentages vary significantly from one model to the other. These variations are expected because, although the BSI score comes from the same function, they are combined with different WHOQOL, this forces the model to find different individual factors. As the individuality factor is latent, we cannot explain how it causally combines with the variables to result in the score, but it is expected that the impact will be different.

There is plenty of evidence showing the association between somatization and behavior symptoms [46, 18]. Observe that in Fig 4.4a 78% of the population had their BSI Somatization increased when it received the higher possible value in "suicidal thoughts" (BSI 9), this impact can be easily validated by a specialist. Wiborg et al. [46] showed that when analyzing primary care in patients with somatoform disorders it was found suicidality as a substantial problem. They also argued that it is vital for a better understanding and management of active suicidal ideation to have a better perception of these dysfunctional illnesses. It was also shown by [18] that somatization was observed to be an additional factor that brings the highest risk for those who bereave with suicide ideation together with symptomatology, post-traumatic stress disorder (PTSD), lower perceived social support, and secular-religious orientation.

Another result shown in Fig 4.4a, is that 80% of the population had their somatization score increased in the output of the counterfactual query that set "spells of terror" (BSI 45) to the maximum value. Note that BSI 45 is a physical symptom and can be seen as an interpretation of the work of [7], which showed that physical symptoms are a common manifestation of potentially treatable psychiatric disorders. It is interesting to notice that in neither our results nor theirs it was found a manifestation in all cases, but they were common, being found in most of the cases. Another analysis made in accordance with ours, gathered results from other works and showed that up to 50% of the patients with somatic symptoms have sufficient criteria for a diagnosis of anxiety and mood disorder [10, 20]. We also found percentages similar to this for questions that are symptoms of anxiety and mood disorder, where 50% and 37% of the population were affected with the increase of somatization, respectively by "feeling restless" (BSI 49) and "Scared for no reason" (BSI 12) when setting the counterfactual queries to the maximum

values.

The results in Fig 4.4c, show another situation, where having the BSI Somatization combined with the physical WHOQOL, all counterfactual changes had little populational impact. We can see that some patterns remain with "Scared for no reason" (BSI 12), "feeling restless" (BSI 49), and "spells of terror" (BSI 45) were still the highest values, but affecting a small population. Suicidal thoughts, in contrast, had a lower impact. These results can be a reflex of the variables used to predict BSI since we cannot guarantee the independence of these features from BSI and the questions added in the Physical WHOQOL. Without guaranteeing the independence of the variables used we can have a situation where we have better predictors, but they are not good causes, as the relation between good predictors and good causes is not necessarily true.

Conversely, in Figure 4.4e the analysis of the BSI, when combined with the WHOQOL environment, showed that all variables significantly impacted somatization, we can see that even symptoms of anxiety that were not so strong in 4.4a like "Feeling tense" (BSI) had approximately 44% in the counterfactual max value and 31% in the counterfactual min value. As it was previously stated, this relation around 50% are in accordance with [10, 20]. In this analysis, "suicidal thoughts" had again the largest impact in the population somatization.

While it was the BSI analysis with the highest populational expected impact it was also the one with more counter-intuitively changes. We can see that we have a significant increase/decrease in BSI values when changing variables for both their min/max values. Note that while analyzing "suicidal thoughts" with the minimum value we had a significant impact in the expected direction with approximately 45% of the population decreasing the result, but also 35% increasing, which is not expected. Similarly, we can see this happens with other anxiety-related variables like "nervousness", "scared for no reason", "feeling tense" and "spells of terror" where decreasing their value resulted in roughly 20% of the population impacted by an increase in the BSI Somatization. While counter-intuitive, they cannot be considered wrong, there are some studies linking the changes in somatization to different expressions and suppression of anger feelings for men and women, respectively [27]. This study also showed that insecure attachment mediated the link between childhood trauma and adult somatic symptoms. However, while knowing the existence of the mediation, the mechanism which links this insecure attachment with somatizationit is poorly understood. [26] and [45] also showed that proneness to experiencing negative emotions and suppression of negative emotions are associated with somatoform disorders. It can be understood that people that have these negative emotions, but counterfactually we are querying a world where they would suppress it, could correctly lead to an increase in somatization.

Turning now to the results of WHOQOL, we will first analyze WHOQOL psychological, which are shown in figure Figure 4.4b. Note that, when analyzing the popula-

tional impact of the reduction of the variables to the extreme, therefore seeing the worst response for each variable, we can see that all of them resulted in a score reduction in the majority of the population. These results are consistent with the ones reported in [25], where it was found that during the pandemics the average resilience was lower than the published norms, but it was greater among those who tended to get outside more often, exercise more, perceive more social support from family, friends and significant others, sleep better, and pray more often. We can relate, perceive more social support from family, friends, and significant others with both "Personal relationships" (WHOQOL 20) and "Friends'support" (WHOQOL 22), while going outside more and exercising more can relate to "Everyday energy" (WHOQOL 10) and "Leisure Opportunity" (WHOQOL 14). Better sleep status (WHOQOL 16) was also found by [23], related to lower social stress, better management of stress, and good social support. In the literature, it was also found that even the result with the lowest impact leading to the maximum, "Satisfaction with transport" (WHOQOL 25), is also associated with the increased quality of life [22].

The results shown in Figure 4.4d for physical WHOQOL are all intuitive and reported in the literature, except from "Life meaning" (WHOQOL 6) and "Negative feeling frequency" (WHOQOL 26), where reducing the sense of life meaning improves the quality of life for 37% and reducing the presence of negative feelings improved 52% of the population. These findings go into the affirmations of [14], who said that although it is expected the effect of sense of meaning in life in the future health, it currently lacks strong empirical support and confirmation. It exists, however, considerable evidence that persistent striving for meaningful accomplishment is indeed a key pathway to health and longevity. He concludes that it is a promising direction, but it needs further study. As cited before, [26] and [45] showed the relation of the suppression of negative feelings with higher somatoform disorders, therefore worse quality of life.

Finally, for the WHOQOL environment, we can observe that in both extremes all variables followed the intuitive impact, with the minimum extreme reducing the score and the maximum extreme raising it, except for the quality of life (WHOQOL 1), one probable reason of this low result is the vagueness of the question and it probably also is a consequence value from the other variables. The WHOQOL variables "Health satisfaction" (WHOQOL 2), "Friend's support" (WHOQOL 22), and "Having a meaningful life" (WHOQOL 6) are known features related to environmental well-being, and it is seen in the results that the majority of the population was impacted in the minimum value scenario (see Figure 4.4f)). [14] support this by showing results that maintaining positive relationships with others improve well-being and general health, they also shown the link of longevity with close relationships, having mastery over one's environment, experiencing spirituality and engaging in life.

Table 4.2: Individual-level counterfactual analysis.

| ID | Score | Feature | F Value | F Score | CF Value | CF Score |
|---|---|---|---|---|---|---|
| 1 | Physical | WHOQOL_2 | 3 | 0.7399 | 5 | 0.7903 |
| 2 | | | 3 | 0.6716 | 5 | 0.6996 |
| 3 | BSI_Psy | BSI_45 | 0 | 0.0712 | 4 | 0.0713 |
| 4 | | | 0 | 0.0355 | 4 | 0.0712 |
| 3 | Psychological | WHOQOL_19 | 3 | 0.6291 | 5 | 0.7355 |
| 4 | | | 3 | 0.7294 | 5 | 0.7838 |
| 5 | Environment | WHOQOL_5 | 4 | 0.7418 | 2 | 0.7218 |
| 6 | | | 4 | 0.8246 | 2 | 0.8044 |
| 7 | | | 4 | 0.7971 | 2 | 0.7785 |
| 8 | | | 4 | 0.7137 | 2 | 0.6921 |

### 4.3.4 Individual-level Counterfactuals

The search for understanding the path to creating a better lifestyle for an individual is constant in psychiatry. They pursue to understand how and what actions can and could have been done to improve one's life. The objective of this section is to shed some light on some of these questions. In this experiment, we show that while the known information of some individuals (input variables) was the same, the output was different, as different people will not react in the same manner to similar situations. This is why it is important to have a deterministic individuality factor for each of the data samples, capable of generating a more personalized counterfactual query. Running the factual or counterfactual query will always return the same result, where the objective is to be as close as possible to the desired, a different approach than the probabilistic counterfactual which depends on a probability function that and generate different results.

Our model also differs from factual predictions from classical machine learning, where the methods learn a consistent output prediction for the same input. It is important to notice that our model consumes both input and output variables to be able to learn this pattern, this is the reason it has the capability of this difference in the factual level. This is not a problem since our main objective is not to predict unseen situations but to attest to the capability of generating these results on a factual level, as it is an important factor for counterfactuals. An appropriate counterfactual should be capable of both predicting different outputs (considering the individuality as the latent variable) for the same input and having a different impact when counterfactually changing the values of input variables.

Table 4.2 shows a few examples of predictions for factual (F) and counterfactuals (CF) of users who have given the same answers to the questionnaire (i.e., have the same input variables) but to which the counterfactual network has predicted different values for BSI and WHOQOL scores. In the table, the IDs refer to individuals in the database,

the score indicates the value that will be predicted by the counterfactual network, the feature is the answer from the questionnaire which has its value being analyzed, "F" stands for the factual value and "CF" to the counterfactual values. Let us start with the comparison between individuals with IDs 1 and 2. They both had the same input values for all variables, only Y was different when predicting for the situation of BSI combined with WHOQOL Physical. The method predicted a different score for WHOQOL Physical, where ID 1 had 0.7399 and ID 2 0.6716. The data showed that the counterfactual prediction was also different between them, in the scenario that both ID 1 and 2 would have their WHOQOL 2 (Health) value been 5, with ID 1 would have had a new Phys. WHOQOL of 0.7903 while individual 2 would have it as 0.6996. This result shows how that counterfactual can generate impacts that would even grow the difference between them.

Another comparison in the table is between ID 3 and 4, when predicting the situation of BSI Somatization with WHOQOL Psychological, in this case, we are displaying the results of the two Y, as both were different between the individuals. The results were ID 3 with BSI being 0.0712 and WHOQOL Psychological 0.6291 and For ID 4 BSI being 0.0355 and WHOQOL Psychological 0.7294. These values show that individual 4 has a better psychological quality of life and, while they both have a low somatization, the score for individual 3 is a little higher. In the analysis, we asked two counterfactual questions, one for BSI, changing BSI 45 ("spells of terror") to 4, and checking what the BSI score would be. The other, when changing the WHOQOL 19 ("'Self-satisfaction") to 5 and see what the WHOQOL score would be. The first query showed that while the BSI of individual 4 increased to 0.0712, the BSI of individual 3 had almost no change, rising to 0.0713. The second analysis checks the impact of asking what would have happened if (s)he was in a situation where hi(er)s "self-satisfaction" (WHOQOL 19) would have been 5 instead of 3. We found that the impact on individual 3 would be a score increase of 0.1 while individual 4 would have had an increase of 0.06, a lot smaller. This shows that not only the algorithm is capable of predicting different values from the same input variables, as it is also capable of finding complex individuality impact in the counterfactual, where both BSI and WHOQOL change in different ways.

Finally, the table shows the comparison between ID 5, 6, 7, and 8 when predicting BSI Somatization and WHOQOL Environment. Note that it is possible to compare the four because they had the same input variables, while having different outputs, and the model captured this difference predicting different factual scores. We queried what would have been their score had they had a worse "life enjoyment" (WHOQOL 5). Note that in this case we are analyzing a reduction of one factor and we are not taking it to the extreme value. We can see that they all varied around their factual predicted value, but the variance was approximately the same – 0.02 for all of them. This shows that the model is able to predict the counterfactuals differently also with changes that are not in

the extremes, and the existence of situations where, even though the counterfactual had different values, the impact can be similar in different individuals.

The results showed that the deterministic factual prediction can predict different outputs from the same factual inputs and provides different counterfactual impacts when applying similar changes. These results are expected in the psychology area as shown by Friedman et al. [15], as they exemplify showing that the serotonin levels in the nervous system are affected by genetics and also alterable by life circumstances. Serotonin levels influence personality, i.e., neuroticism and conscientiousness and they also help regulate core bodily functions necessary for good health, i.e. appetite and sleep. Another work that shows that genetic individuality can make them respond differently to similar situations was presented by McCaffery et al. [29], where they showed that the covariation of depressive symptoms and coronary artery disease may be due to a common genetic vulnerability.

The individual counterfactual analysis brings insights to professionals of what can be done to help a specific patient. The ability to understand what could be changed for these individuals can develop strategies to improve treatment and anticipate actions. If the specialist knows an individual will have its BSI lowered the most by increasing a specific indicator, it could work on that indicator, e.g, sleep or social support. Although the method is not capable of acting and self-feeding with the responses, it is still a source of mathematical findings, with possible actions and insights for specialists to do further investigations.

# Chapter 5

# Conclusions and Future Work

This work introduced a new framework for counterfactual inference, focused on modeling multivariate regression problems that consist of both observed variables and a latent variable $\varphi$, which we defined as the individuality factor. The individuality factor is inherent to a unit in our data and affects the desired output, therefore it needs to be inferred to correctly calculate counterfactual queries.

The framework has three main components, Causal discovery, base networks, and our main contribution is in the third step with ICoNet, a network trained to learn the individuality factor and that can be used later to answer counterfactual queries. The framework is capable of supporting the three levels of the SCM hierarchy of causal knowledge, being able to make associations, interventions, and counterfactuals. It is important to notice that while it is possible to act in all three levels of causation, it has to be done in different components of the framework, since ICoNet focuses on counterfactuals and takes the factual result as an input, and does not make future predictions.

Results in 53 synthetic datasets show the network can learn to minimize both training and test error in different scenarios, going from linear relation and additive individuality to non-linear and multiplicative individuality. The second is expected to better represent real cases. Since we are using synthetic data, it is possible to correct calculate the counterfactual, and verify the correctness of the model. The Results showed that the model was able to learns the individuality factor and effectively uses the learned individuality to make counterfactuals very small errors.

When using the mental health dataset, the counterfactuals found by the model were used in two experiments. The first searched population impact with counterfactual queries. They were all validated by specialists, which analyzed both intuitive as well as counter-intuitive findings. This analysis can help decide the actions to be used in the whole population to achieve a specific mental health improvement.

The second experiment computed individual counterfactual queries, where it was possible to see that the model can use the learned individuality to predict different impacts for individuals with similar observable variables. This corroborated the usefulness of the approach for clinical interventions on unit-level (patients).

## 5.1 Future Work

While we have worked so far with simple counterfactual queries, which focused on changing a single variable, in the future the use of *recourse* [24] could give a more complex and objective analysis. Recourse uses optimization methods to find the minimum change necessary to achieve a different output. Here, combining recourse with our framework can allow us to verify combinations of variables with a specific objective, i.e., what are the minimum changes necessary that could have been made to avoid a patient to develop depression.

Another possible improvement is to create a continuous bottleneck for individuality extraction, as we cannot currently guarantee that small individuality changes will not have a big impact on the model. This approach resembles an autoencoder [1]. For example, a variational autoencoder could make the space continuous but it also introduces noise, making it a probabilistic approach. It is important to create within the framework a continuous space but maintain the deterministic nature of the predictions.

Finally, it would be interesting to test different structures on ICoNet. While it has already presented good results, a bigger and deeper structure could capture more complex relationships in the datasets, but it would also demand a higher amount of data. The context of the model could also be changed to understand its ability to deal with different types of latent variables, and the complexity of these latent variables can be combined with the observed ones. It would also be interesting to see the method with different practical uses, as counterfactual is not limited to the medical context.

# references

[1] Dana H Ballard. Modular learning in neural networks. In *AAAI*, volume 647, pages 279–284, 1987.

[2] Elias Bareinboim, JD Correa, Duligur Ibeling, and Thomas Icard. On pearl's hierarchy and the foundations of causal inference. *ACM Special Volume in Honor of Judea Pearl (provisional title)*, 2020.

[3] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *PNAS*, 116(32): 15849–15854, 2019.

[4] Ioana Bica, Ahmed Alaa, and Mihaela Van Der Schaar. Time series deconfounder: Estimating treatment effects over time in the presence of hidden confounders. In *International Conference on Machine Learning*, pages 884–895. PMLR, 2020.

[5] Lars Buesing, Theophane Weber, Yori Zwols, Nicolas Heess, Sebastien Racaniere, Arthur Guez, and Jean-Baptiste Lespiau. Woulda, coulda, shoulda: Counterfactually-guided policy search. In *ICLR*, 2018.

[6] Danilo Bzdok and Andreas Meyer-Lindenberg. Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3):223–230, 2018.

[7] Claudia Carmassi, Valerio Dell′Oste, Annalisa Cordone, Virginia Pedrinelli, Andrea Cappelli, Diana Ceresoli, Gabriele Massimetti, Cristiana Nisita, and Liliana Dell′Osso. Relationships between somatic symptoms and panic-agoraphobic spectrum among frequent attenders of the general practice in italy. *The Journal of nervous and mental disease*, 208(7):540–548, 2020.

[8] David Maxwell Chickering. Learning bayesian networks is np-complete. In *Learning from data*, pages 121–130. Springer, 1996.

[9] Max Chickering, David Heckerman, and Chris Meek. Large-sample learning of bayesian networks is np-hard. *Journal of Machine Learning Research*, 5, 2004.

[10] Liliana Dell'Osso, L Bazzichi, G Consoli, Claudia Carmassi, M Carlini, Enrico Massimetti, C Giacomelli, Stefano Bombardieri, and Antonio Ciapparelli. Manic spectrum symptoms are correlated to the severity of pain and the health-related quality

of life in patients with fibromyalgia. *Clinical & Experimental Rheumatology*, 27(5): S57, 2009.

[11] Leonard R Derogatis. Bsi brief symptom inventory. *Administration, scoring, and procedures manual*, 1993.

[12] Yanqing Duan, John S Edwards, and Yogesh K Dwivedi. Artificial intelligence for decision making in the era of big data–evolution, challenges and research agenda. *International Journal of Information Management*, 48:63–71, 2019.

[13] Frederick Eberhardt. Introduction to the foundations of causal discovery. *International Journal of Data Science and Analytics*, 3(2):81–91, 2017.

[14] Howard S Friedman and Margaret L Kern. Personality, well-being, and health. *Annual review of psychology*, 65, 2014.

[15] Howard S Friedman, Margaret L Kern, Sarah E Hampson, and Angela Lee Duckworth. A new life-span approach to conscientiousness and health: Combining the pieces of the causal puzzle. *Developmental psychology*, 50(5):1377, 2014.

[16] L. Graham, C.M.Lee, and Y. Perov. Copy, paste, infer: A robust analysis of twin networks for counterfactual inference. In *NeurIPS'19 CausalML Workshop*, 2019.

[17] Eva Grill, Florian Schäffler, Doreen Huppert, Martin Müller, Hans-Peter Kapfhammer, and Thomas Brandt. Self-efficacy beliefs are associated with visual height intolerance: a cross-sectional survey. *PloS one*, 9(12):e116220, 2014.

[18] Sami Hamdan, Natali Berkman, Nili Lavi, Sigal Levy, and David Brent. The effect of sudden death bereavement on the risk for suicide: The role of suicide bereavement. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 2019.

[19] Kaitlin A Harding, Karly M Murphy, and Amy Mezulis. Cognitive mechanisms reciprocally transmit vulnerability between depressive and somatic symptoms. *Depression research and treatment*, 2015, 2015.

[20] Peter Henningsen, Thorsten Jakobsen, Marcus Schiltenwolf, and Mitchell G Weiss. Somatization revisited: diagnosis and perceived causes of common mental disorders. *The Journal of nervous and mental disease*, 193(2):85–92, 2005.

[21] Stefan G Hofmann, Anu Asnaani, Imke JJ Vonk, Alice T Sawyer, and Angela Fang. The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive therapy and research*, 36(5):427–440, 2012.

[22] Housing, U.S. Department of Urban Development, and U.S. Dept. of Transportation Federal Transit Administration. Better coordination of transportation and housing

programs to promote affordable housing near transit. *SSRN*, 2010. URL https://ssrn.com/abstract=1583534.

[23] Yulian Jin, Zheyuan Ding, Ying Fei, Wen Jin, Hui Liu, Zexin Chen, Shuangshuang Zheng, Lijuan Wang, Zhaopin Wang, Shanchun Zhang, et al. Social relationships play a role in sleep status in chinese undergraduate students. *Psychiatry research*, 220(1-2):631–638, 2014.

[24] Amir-Hossein Karimi, Bernhard Schölkopf, and Isabel Valera. Algorithmic recourse: from counterfactual explanations to interventions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 353–362, 2021.

[25] William DS Killgore, Emily C Taylor, Sara A Cloonan, and Natalie S Dailey. Psychological resilience during the covid-19 lockdown. *Psychiatry research*, 291:113216, 2020.

[26] Kyung Bong Koh, Dong Kee Kim, Shin Young Kim, and Joong Kyu Park. The relation between anger expression, depression, and somatic symptoms in depressive disorders and somatoform disorders. *The Journal of clinical psychiatry*, 66(4):485–491, 2005.

[27] Liang Liu, Shiri Cohen, Marc S Schulz, and Robert J Waldinger. Sources of somatization: Exploring the roles of insecurity in relationships and styles of anger experience and expression. *Social Science & Medicine*, 73(9):1436–1443, 2011.

[28] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *international conference on machine learning*, pages 4114–4124. PMLR, 2019.

[29] Jeanne M McCaffery, Nancy Frasure-Smith, Marie-Pierre Dubé, Pierre Théroux, Guy A Rouleau, QingLing Duan, and Francois Lespérance. Common genetic vulnerability to depressive symptoms and coronary artery disease: a review and development of candidate genes related to inflammation and serotonin. *Psychosomatic medicine*, 68(2):187–200, 2006.

[30] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support. *Annals of B. Medicine*, 52(6):446–462, 2018.

[31] Andrew Ng et al. Sparse autoencoder. *CS294A Lecture notes*, 72(2011):1–19, 2011.

[32] World Health Organization et al. Whoqol-bref: introduction, administration, scoring and generic version of the assessment: field trial version, december 1996. Technical report, World Health Organization, 1996.

[33] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. Poptherapy:coping with stress through pop-culture. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pages 109–117, 2014.

[34] Nick Pawlowski, Daniel Coelho de Castro, and Ben Glocker. Deep structural causal models for tractable counterfactual inference. *Advances in Neural Information Processing Systems*, 33, 2020.

[35] Judea Pearl. *Causality*. Cambridge university press, 2009.

[36] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

[37] Neal J Roese. The functional basis of counterfactual thinking. *Journal of personality and Social Psychology*, 66(5):805, 1994.

[38] Neal J Roese. Counterfactual thinking. *Psychological bulletin*, 121(1):133, 1997.

[39] Bernhard Schölkopf. Causality for machine learning. *arXiv preprint arXiv:1911.10500*, 2019.

[40] Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *ICML*, pages 3076–3085, 2017.

[41] Xinpeng Shen, Sisi Ma, Prashanthi Vemuri, and Gyorgy Simon. Challenges and opportunities with causal discovery algorithms: application to alzheimer's pathophysiology. *Scientific reports*, 10(1):1–12, 2020.

[42] Sumedha Singla, Brian Pollack, Junxiang Chen, and Kayhan Batmanghelich. Explanation by progressive exaggeration. In *ICLR*, 2020.

[43] Oliver Stegle, Dominik Janzing, Kun Zhang, Joris M Mooij, and Bernhard Schölkopf. Probabilistic latent variable models for distinguishing between cause and effect. *NeuriPS*, 23:1687–1695, 2010.

[44] Peter WG Tennant, Wendy J Harrison, Eleanor J Murray, Kellyn F Arnold, Laurie Berrie, Matthew P Fox, Sarah C Gadd, Claire Keeble, Lynsie R Ranker, Johannes Textor, et al. Use of directed acyclic graphs (dags) in applied health research: review and recommendations. *medRxiv*, 2019.

[45] David Watson. Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology*, 57(1):120, 1989.

[46] Jan F Wiborg, Dorothee Gieseler, Alexandra B Fabisch, Katharina Voigt, Anne Lautenbach, and Bernd Löwe. Suicidality in primary care patients with somatoform disorders. *Psychosomatic medicine*, 75(9):800–806, 2013.

[47] Erik HF Wong, Frank Yocca, Mark A Smith, and Chi-Ming Lee. Challenges and opportunities for drug discovery in psychiatric disorders: the drug hunters' perspective. *Int. Journal of Neuropsychopharmacology*, 13(9):1269–1284, 2010.

[48] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*, 2016.

[49] Junzhe Zhang and Elias Bareinboim. Designing optimal dynamic treatment regimes: A causal reinforcement learning approach. In *ICML*, pages 11012–11022, 2020.

[50] Yao Zhang, Alexis Bellot, and Mihaela Schaar. Learning overlapping representations for the estimation of individualized treatment effects. In *International Conference on Artificial Intelligence and Statistics*, pages 1005–1014. PMLR, 2020.

[51] Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. Dags with no tears: Continuous optimization for structure learning. *NeuriPS*, 31:9472–9483, 2018.

# Appendix A

# Appendix

## A.1 WHOQOL quality of life assessment questions [32]

In Table A.1, we present the questions considered to describe the individuals in our experiments.

## A.2 BSI distress assessment topics [11]

In Table A.2, we detail all BSI factors considered to describe the individuals in our experiments.

Table A.1: Questions considered in this paper. All questions are answered following a range from 1 to 5, and scores in positive direction, i.e., higher scores denote higher quality of life.

| ID | WHOQOL Question |
|---|---|
| 1 | How would you rate your quality of life? |
| 2 | How satisfied are you with your health? |
| 5 | How much do you enjoy life? |
| 6 | To what extent do you feel your life to be meaningful? |
| 7 | How well are you able to concentrate? |
| 9 | How healthy is your physical environment? |
| 10 | Do you have enough energy for everyday life? |
| 11 | Are you able to accept your bodily appearance? |
| 12 | Have you enough money to meet your needs? |
| 14 | To what extent do you have the opportunity for leisure activities? |
| 15 | How well are you able to get around? |
| 16 | How satisfied are you with your sleep? |
| 17 | How satisfied are you with your ability to perform your daily living activities? |
| 18 | How satisfied are you with your capacity for work? |
| 19 | How satisfied are you with yourself? |
| 20 | How satisfied are you with your personal relationships? |
| 22 | How satisfied are you with the support you get from your friends? |
| 23 | How satisfied are you with the conditions of your living place? |
| 24 | How satisfied are you with your access to health services? |
| 25 | How satisfied are you with your transport? |
| 26 | How often do you have negative feelings such as blue mood, despair, anxiety, depression? |

Table A.2: Topics considered in this paper. All questions are answered following a 5-point scale of distress, ranging from "not-at-all" to "extremely".

| ID | BSI Factor |
|---|---|
| 1 | Nervousness or shakiness |
| 9 | Thoughts of ending your life |
| 12 | Suddenly scared for no reason |
| 23 | Nausea or upset stomach |
| 37 | Feeling weak in parts of your body |
| 38 | Feeling tense or keyed up |
| 45 | Spells of terror or panic |
| 49 | Feeling so restless you could not sit still |

Table A.3: Percentage of Variation in the BSI and WHOQOL scores as Counterfactual interventions are made in the variables indicated in the Id column. Ids proceed by an (*) indicate the variables directly used to generate the score.

| | Physical Intervention | | | | | | | Psychological Intervention | | | | | | | Environmental Intervention | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Min | | | Max | | | | Min | | | Max | | | | Min | | | Max | | | |
| Id | > | < | = | > | < | = | Id | > | < | = | > | < | = | Id | > | < | = | > | < | = |
| BSI | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0.1 | 0.37 | 99.53 | 1.08 | 0.01 | 98.91 | 1 | 0.1 | 8.37 | 91.53 | 17.41 | 0.01 | 82.58 | 1 | 21.6 | 32.96 | 45.44 | 33.83 | 19.82 | 46.35 |
| 9 | 0.46 | 0.15 | 99.39 | 0.82 | 0.15 | 99.03 | 9 | 0.08 | 13.53 | 86.39 | 73.98 | 0.01 | 26.01 | 9 | 34.44 | 45.67 | 19.9 | 62 | 13.42 | 24.57 |
| 12 | 0.04 | 0.26 | 99.7 | 2.39 | 0.09 | 97.51 | 12 | 0.06 | 13.38 | 86.56 | 37.13 | 0.02 | 62.85 | 12 | 28.96 | 43.69 | 27.35 | 57.25 | 18.26 | 24.49 |
| *23 | 0 | 99.98 | 0.01 | 100 | 0 | 0 | 23 | 0.02 | 76.38 | 23.61 | 99.54 | 0.01 | 0.45 | 23 | 6.69 | 88.88 | 4.43 | 94.91 | 3.43 | 1.66 |
| *37 | 0 | 99.99 | 0.01 | 100 | 0 | 0 | 37 | 0.04 | 80.44 | 19.51 | 99.86 | 0.01 | 0.13 | 37 | 6.11 | 89.9 | 3.99 | 96.5 | 2.35 | 1.15 |
| 38 | 0.28 | 0.09 | 99.63 | 0.51 | 0.02 | 99.47 | 38 | 0.08 | 4.29 | 95.63 | 12.22 | 0.01 | 87.78 | 38 | 24.91 | 31.92 | 43.16 | 44.19 | 17.98 | 37.82 |
| 45 | 0.3 | 1.66 | 98.04 | 6.86 | 0.08 | 93.06 | 45 | 0.01 | 31.37 | 68.62 | 82.68 | 0.01 | 17.31 | 45 | 27.6 | 55.62 | 16.78 | 49.69 | 17.65 | 32.67 |
| 49 | 0.13 | 0.19 | 99.68 | 1.52 | 0.07 | 98.41 | 49 | 0.07 | 15.78 | 84.15 | 50.08 | 0.02 | 49.9 | 49 | 29.07 | 43.55 | 27.39 | 53.68 | 16.66 | 29.66 |
| WHOQOL | | | | | | | | | | | | | | | | | | | | | |
| 2 | 4.58 | 88.62 | 6.8 | 78.42 | 3.05 | 18.52 | *6 | 0.31 | 99.2 | 0.5 | 93.14 | 0.1 | 6.76 | 1 | 0.79 | 1.74 | 97.47 | 1.4 | 0.96 | 97.64 |
| 6 | 37.51 | 8.74 | 53.75 | 10.19 | 9.58 | 80.23 | 10 | 1.21 | 96.21 | 2.58 | 81.48 | 0.89 | 17.63 | 2 | 0.36 | 77.9 | 21.74 | 35.78 | 0.38 | 63.84 |
| 7 | 6.93 | 82.06 | 11.01 | 79.73 | 7.1 | 13.16 | 14 | 0.95 | 80.57 | 18.47 | 64.56 | 1.2 | 34.24 | 5 | 0.15 | 82.2 | 17.65 | 60.54 | 0.04 | 39.42 |
| 9 | 8.17 | 10.18 | 81.65 | 9.36 | 6.87 | 83.78 | 16 | 4.44 | 88.88 | 6.68 | 76.91 | 5.04 | 18.06 | 6 | 0.11 | 94.65 | 5.24 | 37.82 | 0.16 | 62.02 |
| *10 | 1.4 | 96.64 | 1.96 | 98.98 | 0.4 | 0.62 | 18 | 2.73 | 61.63 | 35.64 | 14.13 | 4.15 | 81.72 | 7 | 0.17 | 90.24 | 9.59 | 74.68 | 0.06 | 25.26 |
| 11 | 7.88 | 81.41 | 10.71 | 75.26 | 7.15 | 17.58 | *19 | 0.13 | 99.54 | 0.33 | 99.66 | 0.08 | 0.26 | *9 | 0.01 | 99.25 | 0.74 | 99.76 | 0 | 0.23 |
| 12 | 4.41 | 21.66 | 73.93 | 30.51 | 4.27 | 65.22 | 20 | 6.34 | 88.19 | 5.46 | 72.22 | 4.7 | 23.08 | 15 | 0.06 | 98.36 | 1.58 | 88.07 | 0.05 | 11.88 |
| *17 | 0.65 | 98.43 | 0.92 | 99.63 | 0.15 | 0.21 | 22 | 4.19 | 90.94 | 4.87 | 68.32 | 2.9 | 28.79 | 22 | 0.15 | 79.6 | 20.25 | 34.86 | 0.16 | 64.98 |
| 19 | 4.69 | 82.58 | 12.73 | 48.13 | 4.65 | 47.22 | 25 | 5.77 | 64.77 | 29.46 | 11.93 | 3.63 | 84.43 | *23 | 0.01 | 99.77 | 0.23 | 99.88 | 0 | 0.12 |
| 23 | 8.56 | 81.66 | 9.78 | 55.61 | 5.58 | 38.81 | *26 | 0.22 | 99.28 | 0.5 | 99.59 | 0.13 | 0.27 | | | | | | | |
| 24 | 5.25 | 66.7 | 28.05 | 45.84 | 5.62 | 48.54 | | | | | | | | | | | | | | |
| 26 | 57.26 | 9.72 | 33.02 | 13.43 | 16.48 | 70.09 | | | | | | | | | | | | | | |