

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback Please support the ScholarWorks@UMBC repository by emailing [scholarworks-group@umbc.edu](mailto:scholarworks-group@umbc.edu) and telling us what having access to this work means to you and why it's important to you. Thank you.

# Structural Causal Model with Expert Augmented Knowledge to Estimate the Effect of Oxygen Therapy on Mortality in the ICU

**Md Osman Gani**

*Information Systems*

*University of Maryland, Baltimore County*

*Baltimore, Maryland, USA*

MOGANI@UMBC.EDU

**Shravan Kethireddy**

*Critical Care, Northeast Georgia Health System*

*Georgia, USA*

SHRAVAN.KETHIREDDY@NGHS.COM

**Marvi Bikak**

*Critical Care, Palos Health*

*Illinois, USA*

MARVI.BIKAK@GMAIL.COM

**Paul Griffin**

*Department of Industrial and Manufacturing Engineering*

*Penn State University*

*University Park, Pennsylvania, USA*

PMG14@PSU.EDU

**Mohammad Adibuzzaman**

*Regenstrief Center for Healthcare Engineering*

*Purdue University*

*West Lafayette, Indiana, USA*

MADIBUZZ@PURDUE.EDU

## Abstract

Recent advances in causal inference techniques, more specifically, in the theory of structural causal models, provide the framework for identification of causal effects from observational data in the cases where the causal graph is identifiable, i.e., the data generating mechanism can be recovered from the joint distribution. However, no such studies have been done to demonstrate this concept with a clinical example. We present a complete framework to estimate the causal effect from observational data by augmenting expert knowledge in the model development phase and with a practical clinical application. Our clinical application entails a timely and important research question, i.e., the effect of oxygen therapy intervention in the intensive care unit (ICU); the result of this project is useful in a variety of disease conditions, including severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) patients in the ICU. We used data from the MIMIC III database, a standard database in the machine learning community that contains 58,976 admissions from an ICU in Boston, MA, for estimating the oxygen therapy effect on morality. We also identified the covariate-specific effect to oxygen therapy from the model for more personalized intervention.

**Keywords:** Structure Causal Model, Causal inference, Oxygen therapy, Expert Augmented Knowledge, Critical Care

## 1. Introduction

Since the 1960s, randomized controlled trials (RCTs) are considered the gold standard to identify causation by regulatory bodies such as the US Food and Drug Administration (US FDA) and by the clinical communities (Greene and Podolsky, 2012). The key idea behind RCT is that, by random assignment of treatment or interventions, the confounding bias, i.e., the bias due to the assignment of treatment or presence of other variables, including the unobserved confounders, can be accounted for in the estimand. Despite the strength of RCTs to identify causation, RCTs are increasingly considered time-consuming, costly, and often infeasible for safety and efficacy reasons (Frieden, 2017). Furthermore, RCTs can be plagued with selection bias due to the strict eligibility criteria, or transportability bias, due to the study population being different from the target population. At the other end of the spectrum, advances in the technology and the adoption of computerized systems in routine healthcare has enabled the collection and curation of large volumes of data, such as electronic health records (EHR), during routine care, albeit with the presence of confounding biases. Therefore, researchers are increasingly trying to address the challenge of identifying causal relationships from large but biased observational datasets.

Recent advances in the theory of causal inference, more specifically, structural causal models (SCM) provide the framework for adjusting for different kinds of biases such as confounding, transportability, or selection bias under the general theory of data fusion with identifiability formulas, i.e., identification of causal effects, such as backdoor and front-door adjustment, in many cases even if the confounders are unobserved (Bareinboim and Pearl, 2016; Pearl et al., 2016; Pearl, 2009). However, this approach requires developing a graphical representation of the causal relationship between variables of interest with meticulous scrutiny, using structure learning algorithms as well as with expert knowledge from clinicians or existing literature (Raghu et al., 2018; Lederer et al., 2019). While there are other clinical use cases with SCM (Schlüter et al., 2019; Adegunsoye et al., 2019; Vitolo et al., 2018; Arendt et al., 2016; Hernández-Díaz et al., 2008; Schisterman et al., 2005), most of the literature developed the causal graph with domain expertise only. On the other hand, there are many algorithms that can learn the graphical structure from data with different assumptions. Consequently, the resultant causal graph can vary significantly based on the underlying assumptions (Raghu et al., 2018). There have been some recent works that tried to bridge between these two approaches: expert knowledge and automated data-driven graph development (Nordon et al., 2019; Rotmensch et al., 2017; Goodwin and Harabagiu, 2013). However, none of these existing approaches provides a complete framework for expert augmented causal graph generation with specific clinical application. We provide an approach for the development of the SCM with structure learning algorithms and expert augmentation of clinical knowledge in a principled way. We also present a clinical application to demonstrate the feasibility of conducting virtual trials – from observational data – under the proposed framework.

Our approach starts with processing observational data collected during routine health care such as EHR data from hospitals. These EHR systems or data repositories are queried based on the clinical question of interest, with explicit inclusion-exclusion criteria, to identify a cohort of eligible patients representative of the clinical setting. All the variables within the scope of the clinical question are extracted from the data repository. A graphical representa-

tion of cause and effect relationship between the variables can be learned from the extracted data, with structure learning algorithms augmented with domain knowledge through expert reviews. The resultant graph can be validated from the data through testable implications (e.g., conditional independence, Verma constraints). The key advantage of using SCM over statistical methods such as g-methods under the potential outcome framework (Hernán et al., 2008; Hernán and Robins, 2020) is that all the assumptions such as the assumptions on observed and unobserved confounders are explicitly encoded in the SCM that can be analyzed in a principled way with *do*-calculus (Pearl, 2009; Pearl and Mackenzie, 2018).

For the clinical application, we study the effect of liberal versus conservative oxygen therapy in the intensive care unit with respect to ICU mortality as the primary outcome. Although oxygenation has been used to treat critically ill patients for hundreds of years (Heffner, 2013), there has been a renewed interest (Panwar et al., 2016; Girardis et al., 2016; Investigators et al., 2020) in this topic as new evidence suggests not all patients may benefit from the same oxygenation strategy. To define the study question with a causal query that encodes eligibility or inclusion/exclusion criteria, we aimed at implementing the study protocol as described by Panwar et al. (Panwar et al., 2016) in a multi-center pilot RCT. We will refer to this study as the oxygen therapy RCT (OT-RCT) in this paper. We used MIMIC-III database to emulate the OT-RCT protocol using observational data and SCM Johnson et al. (2016). For the study, we used a conservative oxygenation strategy, target  $SpO_2$  is between 88–95%, compared to a liberal oxygenation strategy, target  $SpO_2$  is greater than or equal to 96%, for adult ICU patients receiving invasive mechanical ventilation (IMV). In the observational data, we implemented a clear separation in the mean  $SpO_2$  94.25(3.14) vs. 97.88(3.42),  $SaO_2$  91.39(9.20) vs. 91.78(11.84),  $PaO_2$  89.80(69 – 98) vs. 121.32(86 – 140), and  $FiO_2$  0.59(0.18) vs. 0.47(0.14) values between the conservative and the liberal groups respectively. Our analysis was performed in two different ways using the MIMIC database: i) correlation-based observational study (CB-OBS) and ii) structural causal model based observational study as a virtual RCT (SCM-VRCT).

The point estimate for 90-day mortality with SCM-VRCT was lower with the liberal oxygenation strategy compared to the conservative strategy. This is not consistent with the findings of the OT-RCT. The expectation of 90-day mortality for patients in the conservative oxygenation arm in the SCM-VRCT protocol (54%) was equal to the CB-OBS approach (54%). Also, given the severity of illness based on the physiological scores and age of the patients, the liberal oxygenation strategy in SCM VRCT has lower mortality 29.8%(95% CI 28.1 – 31.1%) compared to the conservative approach in SCM-VRCT 54.1%. Furthermore, within the liberal arm, the mortality in SCM-VRCT is lower compared to that of in the CB-OBS 34.1%(95% CI 33 – 35.3%). Another important observation is that the expectation of mortality is higher for the patients with a sequential organ failure assessment (SOFA) score greater than ten compared to those with scores less than or equal to ten. Conditioned on the severity of illness based on the SOFA score, there was no statistically significant difference in mortality for different oxygenation targets. Therefore, conditioned on the severity of illness, as measured by the SOFA score (SOFA score is greater than or equal to 10), SCM-VRCT supports the feasibility of a larger study to investigate the effect of conservative oxygenation to treat patients requiring invasive MV in ICUs.

### 1.1. Technical Significance

We provide a framework with a step by step approach from data curation, model development, model validation, and causal estimation and demonstrate the approach using a clinical application. Previous research has shown model development using SCMs with only a few variables that was mostly hypothesis-driven (Schlüter et al., 2019; Adegunsoye et al., 2019; Vitolo et al., 2018; Arendt et al., 2016; Hernández-Díaz et al., 2008; Schisterman et al., 2005). Our approach provides a framework to encode expert knowledge using human-in-the-loop model development with the SCM in a principled way. Furthermore, we provide the derivation for the identifiability equation using backdoor adjustment and do-calculus as well as the estimation process with inverse probability weighting (IPW).

### 1.2. Clinical Relevance

There has been a number of trials in recent years to investigate the effect of conservative versus liberal (conventional) oxygenation strategies on mortality and number of ventilator-free days among patients in ICUs receiving invasive mechanical ventilation (Panwar et al., 2016; Girardis et al., 2016; Investigators et al., 2020). Results have been mixed and the most recent study (Investigators et al., 2020) showed no significant effect on the number of ventilator-free days between the treatment and the control group. Consequently, the study has a significant impact on clinical practice. The approach can be generalized for other clinical questions and can be adopted by the broader clinical community if relevant data and domain expertise exists.

## 2. State of the Art

Ronald Fischer in his seminal work, “Design of Experiments”, first introduced randomized controlled trials (RCTs) as an experimental approach to identify causation by accounting for confounding biases (Fisher et al., 1960). The enactment of the 1962 Amendments to the Food, Drug and Cosmetics Act required that new treatments need to be proven efficacious in “adequate and well-controlled investigations” (Greene and Podolsky, 2012). In the 1970s, the FDA translated this into the requirement that an RCT is needed to validate the causal link between a new treatment (causal intervention) and a putative clinical outcome. The rationale for adopting RCTs as the means of gathering scientific evidence is that spurious associations due to factors extraneous to the relationship between the treatment and the outcome can be controlled for by randomizing the treatment assignment. Despite a host of benefits, it is widely acknowledged that RCTs are far from perfect. In practice, RCTs are usually time-consuming, overly expensive, not always entirely ethical, contain biases even with randomization, and are often only applicable to a narrow stratum of the population.

In the 1970s, the potential outcome framework for experimental studies was extended to non-experimental or observational studies through the introduction of the ignorability assumption and propensity score matching for adjustment of confounding biases (Rubin, 1974; Rosenbaum and Rubin, 1983). In the 1980s, sequential backdoor and g-methods were introduced with the potential outcome framework for time-varying exposures (Robins, 1986; Robins and Hernán, 2009). In the 1990s the potential outcome framework was first conceptualized through graphical models, namely, structural causal models, on the foundations

of Bayesian networks (Pearl, 1995). Subsequently, *do-calculus* was introduced with SCMs to mathematically transform observational and experimental studies. The strength of the SCM is that all the assumptions are explicitly specified in the model to provide a unified framework for an objective reproducible estimation approach. In recent years, a unified framework, called *data fusion*, for adjusting other types of bias such as selection bias and transportability bias have been proposed (Pearl and Bareinboim, 2014; Bareinboim and Pearl, 2016; Pearl, 2019).

Most of the applications of SCM based on clinical literature has been purely hypothesis-driven. In other words, the graphical model that describes the assumptions was generated with expert knowledge with a small set of variables (Schlüter et al., 2019; Adegunsoye et al., 2019; Vitolo et al., 2018; Arendt et al., 2016; Hernández-Díaz et al., 2008; Schisterman et al., 2005). Several recent works attempted to generate the causal graph from medical literature automatically (Nordon et al., 2019; Rotmensch et al., 2017; Goodwin and Harabagiu, 2013). While both of these approaches are useful, to the best of our knowledge, there are no hybrid approaches that develop the causal graph iteratively from both the clinical data and expert knowledge. A recent study provided a framework for expert augmented machine learning with boosting trees to extract expert knowledge (Gennatas et al., 2020), but not with SCMs.

We propose an approach for generating evidence from clinical data captured during routine healthcare such as electronic health records with expert augmented causal graphs. We used the Medical Information Mart for the Intensive Care (MIMIC) III (Johnson et al., 2016; Adibuzzaman et al., 2016, 2017) to answer the clinical question, whether liberal or conservative oxygen therapy help to improve survival rate in the intensive care unit. Numerous machine learning and clinical application studies have been published using this standard database (Ghassemi et al., 2015; Zhu et al., 2018) and the database is widely considered as a benchmark due to the high-quality granular data.

The clinical question of interest, i.e., the effect of liberal versus conservative oxygen therapy in the ICU has drawn widespread attention in recent years with a number of randomized trials. These trials investigated the effect of oxygenation strategies on mortality and the number of ventilator-free days among patients in ICUs receiving intermittent mandatory ventilation (IMV). In 2016, a pilot RCT (Panwar et al., 2016) to investigate the effect of different oxygenation targets during mechanical ventilation on mortality supported the feasibility of a conservative oxygenation strategy in patients but concluded with a need for a larger RCT to be performed to evaluate its efficacy. In another study (Girardis et al., 2016), a conservative oxygenation strategy was associated with lower ICU mortality compared to the conventional strategy. One important note about this study is that the findings were reported after an unplanned early termination of the trial due to the potential adverse effect in one of the study arms. They also emphasized the need for a multi-center trial that evaluates this intervention. The most recent study (Investigators et al., 2020) with a 1000 patients’ cohort found no significant effect on the number of ventilator-free days from the use of conservative oxygen therapy as compared to the conventional liberal oxygen therapy. Table 5 summarizes the study demographics and Table 6 summarizes the results of these studies, among others. The lack of a consensus from these studies shows the importance of the continued study of clinical practice and resulting outcomes on mortality (Hirase et al., 2019).

### 3. Background

The causal analysis goes beyond association to infer probabilities under experimental conditions to estimate the effect of interventions. One example of these changes can be external interventions that can be used to estimate treatment effects. The development of mathematical machinery with SCM in the last couple of decades can be used to adjust for biases such as confounding bias, selection bias, and estimate treatment effect and counterfactual (Bareinboim and Pearl, 2016). In this section, we provide the fundamentals of the mathematical machinery with SCM.

#### 3.1. Structural Causal Model (SCM)

An SCM  $M$ , represent the causal relationship between variables.  $M$  consists of two sets of variables  $U$  and  $V$ , called exogenous and endogenous variables respectively. It also includes a set of functions  $f$  that assign each endogenous variable in  $V$  a value based on the values of other variables in the SCM. A variable  $Y$  is directly caused by  $X$  if  $X$  is in the function  $f$  of  $Y$ . A variable  $X$  is a cause of another variable  $Y$  if it is a direct or any cause of  $Y$  (Pearl, 2009).

**Definition 1** *Structural Causal Model; A structural causal model is a 4-tuple  $M = \langle U, V, F, P(u) \rangle$ , where*

1.  $U$  is a set of background variables (also called exogenous) that are determined by factors outside the model.
2.  $V$  is a set  $\{V_1, V_2, \dots, V_n\}$  of endogenous variables that are determined by variables in the model, viz. variables in  $U \cup V$ .
3.  $F$  is a set of functions  $\{f_1, f_2, \dots, f_n\}$  such that each  $f_i$  is a mapping from the respective domains of  $U_i \cup PA_i$  to  $V_i$  and the entire set  $F$  forms a mapping from  $U$  to  $V$ . In other words, each  $f_i$  assigns a value to the corresponding  $V_i \in V$ ,  $v_i \leftarrow f_i(pa_i, u_i)$ , for  $i = 1, 2, \dots, n$ .
4.  $P(u)$  is a probability function defined over the domain of  $U$ .

Each SCM  $M$  is associated with a graph, known as a *graphical causal model*,  $G$ . Each  $G$  is a directed acyclic graph (DAG). This graphical causal model  $G$  consists of a set of nodes or vertices that represent the variables  $U$  and  $V$ , and a set of edges that represents the functions in  $f$ .  $G$  contains a node for each variable in  $M$  and a directed edge from  $X$  to  $Y$  if  $X$  is in the function  $f$  of  $Y$  ( $X$  is a direct cause of  $Y$ ) in  $M$ . Causal models and graphs, therefore, represent causal relationships and encode causal assumptions (Pearl, 2009).

#### 3.2. Causal Structure Learning

Causal structure learning algorithms are a family of algorithms that estimate the causal graph or certain aspects of it from a given data set (Heinze-Deml et al., 2018). Given a distribution, these algorithms find possible causal diagrams implied by the graphical structure that could generate it. We assume the underlying graph is a directed acyclic



graph (DAG). The underlying causal DAG is generally not identifiable but the Markov equivalence class of DAGs, i.e. the set of DAGs that encode the same set of d-separation relationships, can be identified (Pearl, 2009). The Markov equivalence class of DAGs can be described by completed partially DAGs (CPDAG) (Andersson et al., 1997).

There are many methods that have been developed to estimate the underlying DAG. These are broadly classified as 1) constraint-based methods, 2) score-based methods and 3) hybrid methods (Scutari, 2010). The constraint-based methods, PC (Peter Spirtes, 2000), rankPC (Harris and Drton, 2013), and FCI (Peter Spirtes, 2000), perform statistical tests of marginal and conditional independence to check dependency. On the other hand, the score-based methods, GES (Chickering, 2002), rankGES (Nandy et al., 2018), GIES (Alain Hauser, 2012), and rankGIES, optimize the search according to a score function. Hybrid methods such as Max-Min Hill-Climbing (MMHC) (Ioannis Tsamardinos, 2006) combine conditional tests with a score-based approach. Besides these three approaches, there are other methods such as structural equation models with additional restrictions, Linear Non-Gaussian Acyclic Models (LINGAM), (Shimizu et al., 2006) that exploit invariance properties (Rothenhäusler et al., 2015).

### 3.3. *do*-Calculus – Graphical Identification Criterion

We consider a query  $Q$ , the task of estimating the distribution of the outcome  $Y$  after intervening on the treatment variable  $X$ , mathematically written as  $P(Y = y | do(X = x))$ . The question is if we can define the conditions under which we can infer the results of the query when we have data,  $X, Y, Z$  from an observational study where  $X, Y$ , and  $Z$  are randomly sampled. This is a standard query of policy evaluation that needs to be estimated given  $P(y, x, z)$  by controlling for confounding bias (Bareinboim and Pearl, 2016).

The question at the center of causal inference is to estimate the effects of interventions such as medical treatments or policy actions (Hünermund and Bareinboim, 2019) (Bareinboim and Pearl, 2016). Interventions in an SCM  $M$  are denoted by a mathematical operator, namely the  $do(\cdot)$  operator. These interventions in SCMs, in the form of  $do(X = x)$ , are carried out by removing individual functions,  $f_i$ , from the model i.e. in the corresponding graph, severing all incoming arrows that enter the manipulated or intervening variable  $X$ . The distribution of  $Y$  after the intervention can be defined using the counterfactual notation as,

$$P(y | do(x)) \triangleq P(Y_x = y), \quad (1)$$

where  $Y_x = y$  means “ $Y$  would be equal to  $y$  if  $X$  had been  $x$ ” (Pearl, 2009). The backdoor and front-door criteria can be used to identify sets of covariates that should be adjusted for using simple graphical identification rules. However, to uncover all causal effects that can be identified from a given DAG requires symbolic machinery (Pearl, 1995). The problem of identification that asks whether the interventional distribution,  $P(Y = y | do(X = x))$ , can be estimated from the graph (along with the assumptions) and the available observational data. For queries in the form of  $Q$  or *do*-expression, identifiability can be decided procedurally using an algebraic method known as the *do*-calculus. It allows us to manipulate probabilistic distributions, both interventional and observational, through three inference rules whenever certain separation conditions hold in the causal graph  $G$  in the SCM  $M$ . A sufficient version of *do*-calculus is the adjustment criterion.



Do-calculus is a causal inference engine that takes three inputs ([Hünernmund and Bareinboim, 2019](#)):

1. A causal quantity  $Q$ , which is the query we want to answer;
2. A causal graphical model  $G$  that encodes the qualitative understanding about the structural dependencies between the variables under study;
3. A collection of datasets  $P(v|\cdot)$  that are available, such as observational, experimental, from selection-biased samples, or from different populations.

Based on these three inputs, using *do*-calculus and three inference rules, we can transform interventional probabilistic sentences into equivalent expressions.

Let  $X$ ,  $Y$ ,  $Z$ , and  $W$  be the arbitrary disjoint sets of nodes in a causal graph  $G$ . Also, let  $G_{\overline{X}}$  be the graph obtained by removing all arrows pointing to nodes in  $X$ . Likewise,  $G_{\underline{X}}$  is a graph obtained by deleting all arrows that are emitted by  $X$  in  $G$ . The notation  $G_{\overline{X}\underline{Z}}$  can be used to represent the previous two configurations together, the removal of both arrows incoming in  $X$  and arrows outgoing from  $Z$ . Given this notation, *do*-calculus has the following three rules that are valid for every interventional distribution compatible with the causal graph  $G$  ([Pearl, 2009](#); [Bareinboim and Pearl, 2016](#)).

**Do-Calculus Rule 1** : *Insertion/deletion of observations*

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}. \quad (2)$$

**Do-Calculus Rule 2** : *Action/observation exchange*

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}. \quad (3)$$

**Do-Calculus Rule 3** : *Insertion/deletion of actions*

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z(W)}}}, \quad (4)$$

where  $Z(W)$  is the set of  $Z$ -nodes that are not ancestors of any  $W$ -node in  $G_{\overline{X}}$ .

The above rules of *do*-calculus make it complete for general queries of the form  $Q = P(y|do(x), z)$ . It is guaranteed to return a solution for the identification problem, whenever such a solution exists ([Pearl, 1995](#); [Bareinboim and Pearl, 2012](#)). To establish identifiability of a causal query  $Q$ , we need to repeatedly apply the rules of *do*-calculus to  $Q$  until an expression is obtained that is *do*-expression free. This *do*-free expression makes  $Q$  estimable from non-experimental data.

## 4. Methods

We present an expert-augmented causal estimation framework to estimate the effect of an intervention from observational data as depicted in [Figure 1](#). We discuss the framework, in the following subsections, with respect to a clinical question using EHR data curated during routine care in the ICU.

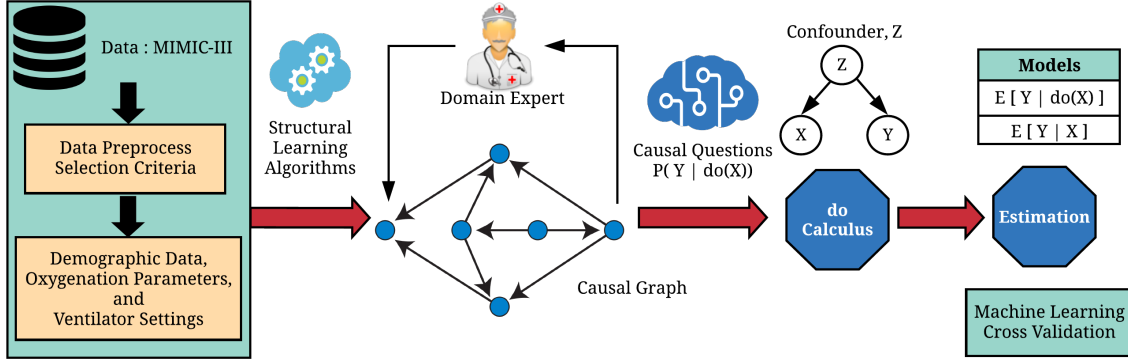


Figure 1: Schematic diagram for the causal estimation framework. The framework allows domain expertise to be encoded in the SCM.

#### 4.1. Cohort selection

We first consider the observational data collected during routine care such as from an EHR. EHRs contain data repositories from multiple sources or healthcare facilities. In order to identify a set of eligible patients for the experiment, inclusion-exclusion criteria are defined based on the clinical question of interest.

#### 4.2. Structure Learning Algorithms

Given a data set, eligible patients for the study, the framework estimates a possible causal graph that could be generated given the distribution using structure learning algorithms (SLAs) (Raghu et al., 2018). These SLAs learn the underlying causal graph at different levels of granularity under different assumptions. Based on these assumptions, multiple causal graphs were estimated using SLAs. We have considered the following seven SLAs for this framework: 1) PC Algorithm, 2) RFCI, 3) GES, 4) GDS, 5) TABU Search (Glover and Laguna, 1998), 6) MMHC, and 7) LINGHAM.

#### 4.3. Majority Voting

Seven different causal graphs were estimated using the SLAs discussed in subsection 3.2 and subsection 4.2. We then used a majority voting scheme on the causal relationships estimated using the SLAs. If a simple majority,  $m/2 + 1$  of algorithms support a causal relationship between  $X$  and  $Y$ , we consider the relationship for our model. Here  $m$  is the total number of SLAs under consideration i.e. 7. If it is supported by less than  $m/2 + 1$  algorithms, we discard that functional relation from the model. If we consider a causal DAG  $G$  has  $n$  nodes, then these nodes are enumerated using integers from 1 to  $n$ . Therefore, it can have  $n \times n$  or  $n^2$  number of directional edges each representing a causal relationship. Let's consider a matrix  $V$  of  $n \times n$  for a causal graph  $G$  that stores the number of votes received for each of the causal relationships between nodes  $i$  and  $j$ , two arbitrary nodes where  $1 \leq i, j \leq n$ .

For each element  $(i, j)$  in  $V$  we consider the voting procedure in [Equation 5](#).

$$V(i, j) = \begin{cases} 0 & \text{initially} \\ V(i, j) + 1 & \text{if a SLA supports a causal edge from } i \text{ to } j \\ V(i, j) + 0 & \text{if a SLA does not support a causal edge from } i \text{ to } j \end{cases} \quad (5)$$

Consider the pair  $(u, v)$  representing a directed edge from node  $u$  to node  $v$  in a causal graph. After counting votes for all the causal relationships in  $V$ , we draw a causal graph  $G$  using [Equation 6](#).

$$G = \begin{cases} G \cup (u, v) & \text{if } V(u, v) \geq \frac{m}{2} + 1 \\ G \cup \emptyset & \text{if } V(u, v) < \frac{m}{2} + 1 \end{cases} \quad (6)$$

where,  $1 \leq u, v < n$ .

#### 4.4. Expert Augmented Knowledge

To finalize the causal graph for estimation, we verify the model with expert consultation and review of the current clinical literature. This provides a systematic approach for encoding existing clinical knowledge into the graph. In this step, we review the initial estimation, through majority voting, of the causal graph  $G$  by domain experts. It allows us to review causal relationships estimated by majority voting and makes the necessary adjustment based on existing evidence. We perform either of the following five operations based on the expert knowledge described in [algorithm 1](#). The resultant causal graph is then reviewed again to make sure that it does not contain any cycles. Therefore, the final causal graph is always a DAG.

#### 4.5. Causal Questions

With the graph encoded with the causal relationships, we can now ask causal questions to understand how variables influence each other. This allows us to run virtual experiments to estimate treatment effects. The probability that  $Y = y$  when we intervene to make  $X = x$  is denoted by,  $P(Y = y | do(X = x))$ . That is to say if everyone in a population had their  $X$  value fixed at  $x$  then what is the population distribution of  $Y$ . Using do-expressions and a causal graph, we can ask causal questions and untangle the causal relationships. For our causal questions,  $X$  is the treatment variable and  $Y$  is the outcome.

#### 4.6. Causal Identification

We used causal identification formula such as *backdoor adjustment* derived using the rules of *do-calculus* to find a set of confounding variables,  $Z$  that must be adjusted depending on the causal query in the form of  $P(Y | do(X))$  and the graphical model ([Pearl, 2009](#)). The model automatically finds the confounding variables, can distinguish between confounders and mediators, and provides the causal identification formula. The causal identification formula provides a mathematical transformation between the observational reality and the corresponding experimental reality. It is possible that certain graphical model will not be able to provide the causal identification formula, and in those cases, the method will not be able to answer the query of interest. The resultant formula can be used to evaluate the causal effect.

---

**Algorithm 1** Encoding Expert Knowledge in to the Causal Graph,  $G$ 

---

**Input:**  $G$ ,  $ExpertKnowledge$ **Output:**  $G$ 

```

1: for Each edge  $(u, v)$  in  $G$  do
2:   if An evidence of a causal relationship from  $u$  to  $v$  then
3:     Keep the edge  $(u, v)$  in  $G$ 
4:   else if An evidence of a causal relationship from  $v$  to  $u$  then
5:     Change the orientation of the edge to  $(v, u)$ 
6:   else if An evidence of no causal relationship between  $u$  and  $v$  then
7:     Remove the edge  $(u, v)$  from  $G$ 
8:   else if No evidence of a causal relationship between  $u$  and  $v$  then
9:     Keep the edge  $(u, v)$  in  $G$ 
10:  end if
11: end for
12: for Each node  $u$  in  $G$  do
13:   for Each node  $v$  in  $G - \{u\}$  do
14:    if the edge  $(u, v)$  not in  $G$  then
15:      if An evidence of a causal relationship from  $u$  to  $v$  then
16:        Add the edge  $(v, u)$  to  $G$ 
17:      end if
18:    end if
19:  end for
20: end for

```

---

#### 4.7. Estimation

Once we have these three sets of variables, treatment  $X$ , confounders  $Z$ , and outcome  $Y$ , we can compute the conditional probabilities in the adjustment formula using the probability distribution to compute the interventional effect. Computing each of the conditional probabilities in the adjustment formula is computationally expensive and may not lead to feasible results with small sample sizes. Instead, we use machine learning algorithms to estimate the probability distribution from the data, that is we use treatment  $X$  and confounders  $Z$  as input to an ML algorithm to estimate the outcome  $Y$ . For this step, we split our data into training and testing sets, each containing  $X$ ,  $Y$ , and  $Z$ . We train two ML models, one based on observational data and another based on experimental data, using the training dataset. We use the testing set to predict  $Y$  from  $X$  and  $Z$  using the trained ML models. After that, we use bootstrapping to randomly subsample  $X$  and predicted  $Y$  for  $k$  iterations. This gives us an estimation of the effect of  $X$  on  $Y$  and used to estimate the mean and confidence interval.

## 5. Results

There are approximately 2 to 3 million patients in the US that receive invasive MV in an ICU (Wunsch et al., 2010; Adhikari et al., 2010) with an estimated annual cost of \$15-27 billion. MV is highly associated with morbidity (Kahn et al., 2010) and mortality (Metnitz

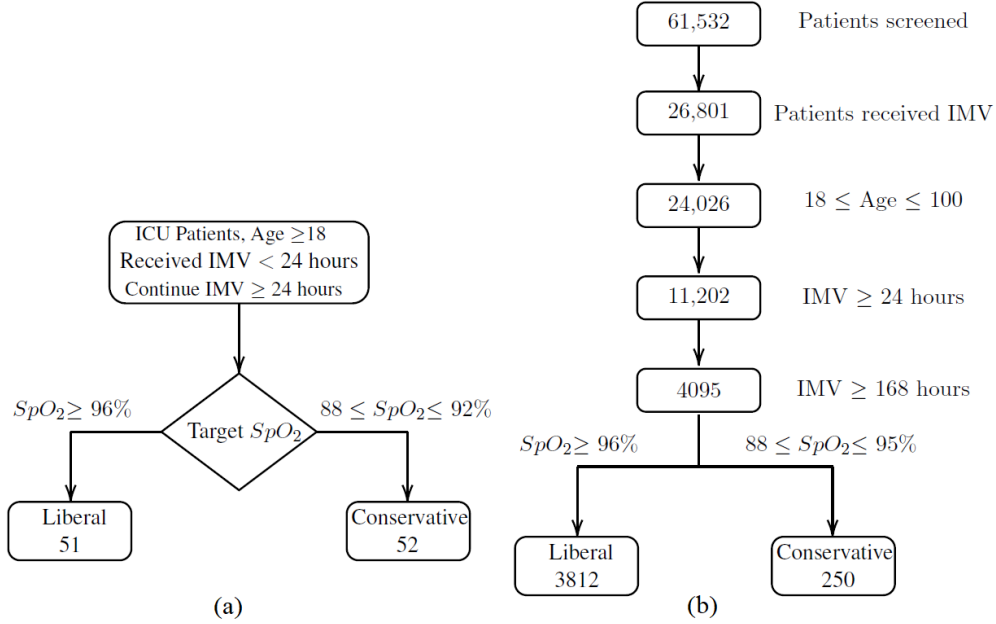


Figure 2: Inclusion-exclusion criteria for (a) pilot RCT and (b) for the observational study with the MIMIC database.

et al., 2009). It is important to maintain safe levels of tissue oxygenation estimated through peripheral oxygen saturation ( $SpO_2$ ) (Vincent and De Backer, 2013) for MV patients in the ICU, most of whom also receive supplemental oxygen therapy (OT) (Panwar et al., 2016).

Current practice and recommendations related to oxygenation targets for ICU patients are not based on strong evidence. Rather, it is primarily based on normal physiologic values in healthy adults at sea level,  $SpO_2$ ,  $SaO_2$ , and  $PaO_2$ , which are approximately 90%, 95 – 97%, and 88 – 100mmHg, respectively (Crapo et al., 1999). For acutely ill patients  $SpO_2$  target recommendations vary from 90% to 94 – 98%. Also, for patients with acute respiratory distress syndrome (ARDS),  $SpO_2$  targets of 88 – 95% are considered acceptable levels (Network, 2000; Meade et al., 2008). This liberal approach to oxygen therapy is now being challenged due to the increasing recognition of the potential harm of excessive  $FiO_2$  in the form of hyperoxemia and tissue hyperoxia (Panwar et al., 2013; Suzuki et al., 2013).

Recent trials have been inconclusive in determining the effect difference between the liberal and conservative oxygen therapy on mortality (Panwar et al., 2016; Girardis et al., 2016) with a need for large trials. Another recent study found no significant effect on the number of ventilator-free days for the use of conservative oxygen therapy compared to conventional oxygen therapy (Investigators et al., 2020). In the clinical practice guideline for oxygen therapy for acutely ill patients, it is strongly recommended not to administer oxygen therapy at higher than 96% saturation with a few exceptions (Siemieniuk et al., 2018). It is also strongly recommended not to provide oxygen therapy at or above 93% saturation

for patients with acute stroke or myocardial infarction. We have presented oxygenation strategies for multiple studies based on target  $SpO_2$  in Table 5.

Our goal is to determine the feasibility of a conservative oxygenation strategy as an alternative to the conventional liberal oxygenation approach. We use the computational model, discussed in section 4, developed using causal inference methods to estimate the effect of different oxygenation strategies for adult patients receiving MV in the ICU. We used SCMs to identify and represent causal relationships between variables. The overall framework is presented in Figure 1. We describe the experimental details in the following subsections.

### 5.1. Selected Cohort

Table 1: Summary statistics of the patients. The first two columns have summary statistics for the cohort selected from MIMIC database and the last two columns have summary statistics for the OT-RCT.

Variable	SCM - VRCT		OT-RCT	
	Conservative	Liberal	Conservative	Liberal
Age	60.82(15.6)	62.94(16.22)	62.4(14.9)	62.4(17.4)
Male Sex	150 (60%)	2197(57.09%)	32(62%)	33(65%)
BMI	28.1 (10.5)	27.8 (10.7)	27.6(10.3)	27.6(10.1)
Diagnosis type				
Trauma	25(10%)	466(12.22%)	3(6%)	2(4%)
Medical	15(6%)	152(4%)	39(75%)	41(80%)
Surgery	8(3.2%)	93(2.43%)	10(19%)	8(16%)
APSOII	50.44(36-61.75)	54.05(38-67)	79.5(61-92.5)	70(50-84)
SOFA	6.96(3.43)	6.27(3.73)	7.9(2.9)	7.4(3.1)
Smoker	36(14.4%)	345(9.05%)	10(19%)	14(27%)
COPD	21(8.4%)	154(4.04%)	11(21%)	5(10%)
Ischemic HD	11(4.4%)	151(3.96%)	6(12%)	5(10%)
ARDS	13(5.2%)	130(3.4%)	17(33%)	10(20%)
SpO2	94.25(3.14)	97.88(3.42)	95(3)	96(3)
FiO2	0.59(0.18)	0.47(0.14)	0.44(0.2)	0.44(0.18)
SaO2	91.39(9.20)	91.78(11.84)	95.5(3)	96(2.7)
PaO2	89.80(69-98)	121.32(86-140)	81(68-109)	82(75-104)
PaCO2	49.19(13.03)	41.78(9.67)	38(7)	39(6)
pH	7.37(0.09)	7.39(0.08)	7.36(0.07)	7.37(0.07)
Lactate	2.23(1.2-2.5)	2.46(1.2-2.7)	1.99(1.4-2.9)	1.65(1.2-2.6)
Hemoglobin	10.07(1.92)	9.86(1.68)	110(23)	115(23)
PEEP	10.8(4.82)	7.02(3.31)	8.2(3)	7.3(3)
VT	505.48(133.3)	525.56(2241.31)	8(1.8)	8(1.9)
Peak Air Prs.	29.83(9.39)	24.98(8.76)	22(6)	21(5)
No. of Patients	250	3812	52	51

We closely followed the study guideline and selection criteria used in OT-RCT. This trial was conducted at four multidisciplinary ICUs in Australia, New Zealand, and France. The goal was to investigate whether a conservative oxygenation strategy is a feasible alternative to a liberal oxygenation strategy among ICU patients requiring invasive MV. A total of 103 adult patients, age greater than or equal to 18 years, were eligible for the study as they were given invasive MV for less than 24 hours and their attending clinician expected the invasive MV to continue for at least the next 24 hours. These patients were randomly

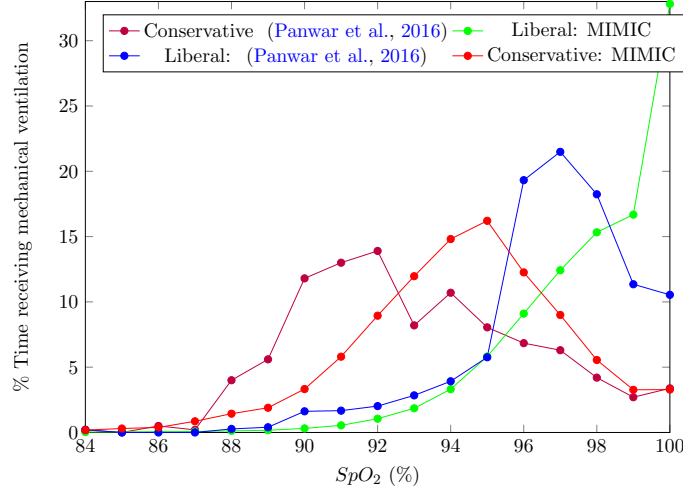


Figure 3: Pooled frequency histogram of the percentage of time spent at each  $SpO_2$  level in both oxygenation groups for OT-RCT (Panwar et al., 2016) and our study, SCM-VRCT.  $SpO_2$  is oxygen saturation as measured by pulse oximetry.

treated with either a conservative oxygenation strategy with target  $SpO_2$  of 88–92% and a liberal oxygenation strategy with target  $SpO_2$  of greater than or equal to 96%. There were 52 and 51 patients in the conservative and liberal arms respectively. The adjusted hazard ratio for 90-day mortality in the conservative arm was 0.77(95%CI, 0.40 – 1.50,  $P = 0.44$ ) overall. There were no significant between-group differences and no harm was associated with the use of the conservative oxygenation strategy.

For our study, we considered a large ICU database containing data routinely collected from patients in the US. Adult patients of age at least 18 years old were considered. We used the MIMIC-III dataset for model development and evaluation (Johnson et al., 2016). The database comprises a total of 61,532 ICU stays at Beth Israel Deaconess Medical Center in Boston, MA in which 53,432 stays are for adult patients and 8,100 stays are for neonatal patients. The database includes data logged using the CareVue and Metavision electronic health record (EHR) systems spanning from June 2001 to October 2012, a little over 11 years.

We included adult patients with eligibility criteria described in Figure 2 (b). We closely followed the inclusion-exclusion criteria in Figure 2 (a) employed in OT-RCT (Panwar et al., 2016). After the exclusion of ineligible cases, we included 4,062 ICU patients from the database. Out of 4,062 patients, 3,812 patients received liberal oxygenation and 250 patients received conservative oxygenation. Patient demographics and clinical characteristics are shown in Table 1. We also present the demographics and clinical characteristics of the patients from the OT-RCT in Table 1 for comparison. Patients spent the majority of the time within the intended target range in both liberal and conservative groups. The pooled frequency histogram of the percentage of time spent at each  $SpO_2$  level for both groups is shown in Figure 3.



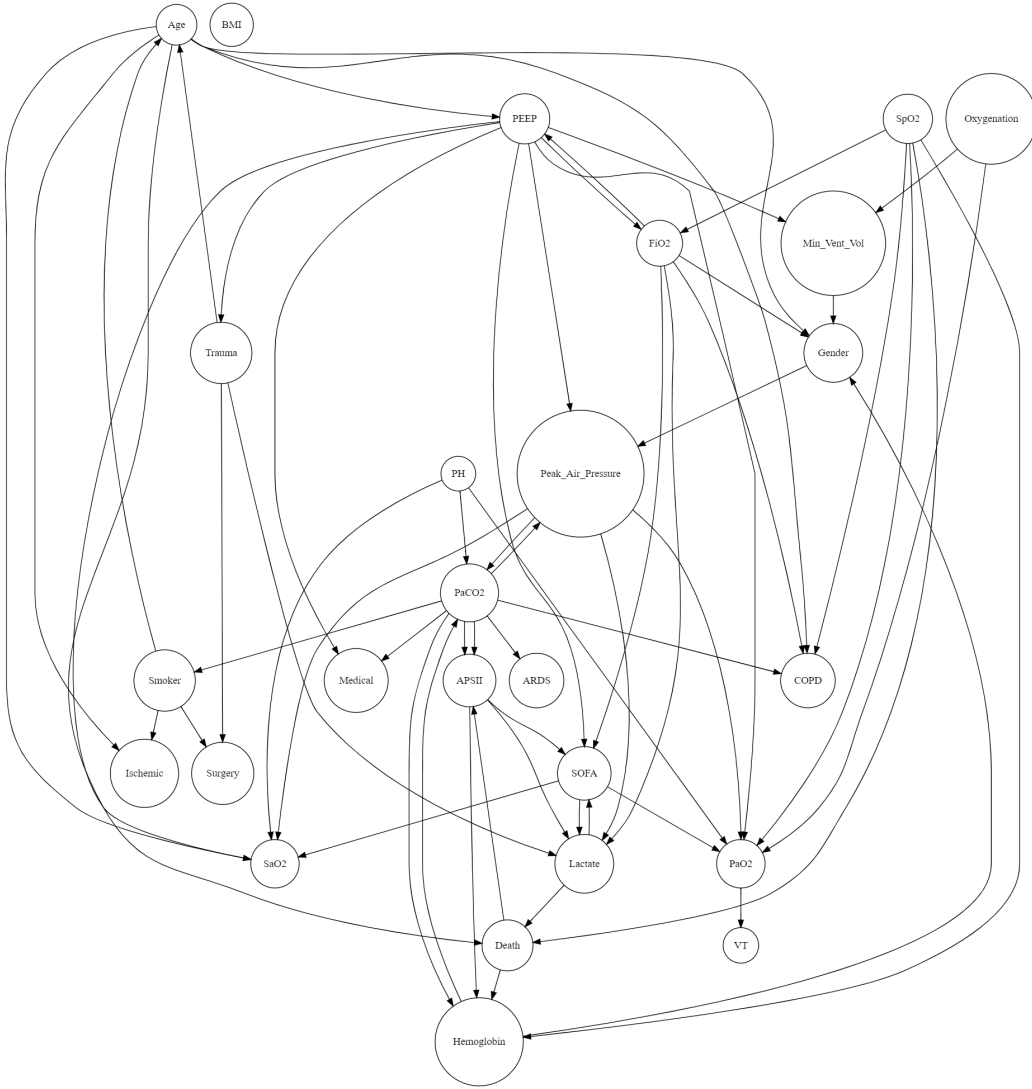


Figure 4: Causal Graph  $G$  based on the majority voting of SLAs. *oxygenation* is the treatment variable, and *death* is the outcome variable.

Our inclusion-exclusion criteria ensure that our study, SCM-VRCT has a similar distribution of  $SpO_2$  in the liberal and conservative arms for observational data compared to the OT-RCT. We extracted a set of 26 variables, including demographics, ventilator settings, and oxygenation parameters, from the dataset. Data on oxygenation parameters and ventilator settings were extracted every 4 hours and for each patient, we included at least 168 hours of measurements.

## 5.2. Structure Learning Algorithm

Our study has focused on estimating the effect of different oxygenation strategies by establishing the causal relationship between measured oxygenation parameters, ventilator settings, disease state, and demographic information. Demographic data and computed mean were then used as features to estimate causal graphs using SLAs. For each algorithm, the data for the variables in [Table 1](#) were used to learn the causal relationship among them and then estimated a causal graph. We estimated 7 causal graphs using SLAs we discussed in [subsection 3.2](#).

## 5.3. Majority Voting

Based on the estimation by SLAs, we computed the matrix  $V$  for voting, using [Equation 5](#). The resultant matrix is shown in [Table 2](#), which has 26 rows and 26 columns. Each row has a variable for which we have the vote count in the columns for all 26 variables.

For example, the first row has the number of votes for variable *age* and the second row has the number of votes for *gender*. The value 6 at row 5 (variable *trauma*) and column 4 (variable *surgery*) means that the causal relationship from *trauma* to *surgery* has received 4 out of the total number of votes, 7. As this vote count is greater than or equal to  $7/2 + 1$  or 4, the majority voting, we include this causal relationship in our final causal graph. The estimated graph using the majority voting ([Equation 5](#)) based on the vote count in [Table 2](#) is shown in [Figure 4](#).

## 5.4. Encoding Expert Knowledge

We developed the causal graph from the result of the majority voting of estimated graphs by structural learning algorithms. We then incorporated clinical domain knowledge by consulting the domain expert, a critical care physician. We reviewed each causal relationship estimated by the algorithms. This review phase of each of the edges from variable  $X$  to variable  $Y$  involves three operations: 1) keep the edge if it is supported by the domain knowledge, 2) change the direction or orientation of the edge if that is supported by the domain knowledge, and 3) remove the edge if it is not supported by the domain knowledge. [Table 3](#) shows the process of incorporating the domain expert’s input.  $Node_{from} \rightarrow Node_{to}$  means  $Node_{from}$  causes  $Node_{to}$ . For example, in row 8, *age* causes *COPD*. The column, majority voting (MJV) is “Yes” which means that it was estimated by the majority voting of the structural learning algorithms, whereas a “No” indicates it was not. The *Expert Knowledge* column represents the addition of a new causal relationship, removal of an edge, or change of the orientation of an existing causal edge followed by the reference of the domain knowledge in the next column. In total, we added 16 new edges, removed 10 edges, and changed directions of 14 edges as a result of the process. We then review the graph to ensure that the graph is acyclic with clinician consultation. This process allows us to make sure that the causal graph encodes the clinical knowledge along with the data generating process.

Table 2: Vote Counts For All Causal Edges

	Age	Gender	BMI	Surgery	Trauma	Medical	APSiii	SOFA	Smoker	COPD	Ischemic HD	ARDS	Death	Oxygenation	$SpO_2$	$FiO_2$	$SaO_2$	$PaO_2$	$PaCO_2$	pH	Lactate	Hemoglobin	Peep	VT	Peak Air Pressure	Minute Ventilation
Age	0	4	0	0	3	0	0	0	1	6	5	0	6	0	1	0	5	2	0	1	2	0	4	0	2	0
Gender	3	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	0	3	0	0	4	2
BMI	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Surgery	0	0	0	0	2	0	0	0	3	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0	2	0
Trauma	4	1	0	6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	5	0	3	0	0	0
Medical	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3	2	0	0	0
APSiii	1	1	2	0	0	0	0	6	0	0	0	0	3	1	2	2	0	3	3	0	5	4	0	0	0	0
SOFA	0	2	0	2	2	0	3	0	0	2	0	0	1	1	1	3	5	4	0	3	4	2	3	0	1	0
Smoker	4	0	0	5	0	0	0	0	0	2	4	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
COPD	2	0	0	0	0	0	0	0	2	0	0	0	0	0	1	2	0	1	1	0	3	0	0	0	0	0
Ischemic HD	3	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
ARDS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Death	2	0	0	0	0	0	5	0	0	0	0	0	0	0	1	1	0	0	0	0	3	4	0	0	2	0
Oxygenation	1	0	0	0	2	0	1	1	0	2	0	0	2	0	3	1	0	5	0	2	2	2	1	0	0	4
$SpO_2$	2	0	0	0	1	0	1	2	1	4	0	0	6	2	0	4	2	5	2	2	0	5	3	0	2	2
$FiO_2$	0	5	0	0	0	0	1	6	2	5	0	0	2	1	3	0	0	1	2	2	5	1	4	2	3	0
$SaO_2$	3	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0
$PaO_2$	1	0	0	0	0	0	1	3	0	1	0	0	0	2	3	1	0	0	0	2	0	3	1	5	3	0
$PaCO_2$	0	0	0	3	0	5	4	0	5	5	5	6	0	0	0	1	2	0	0	3	0	4	0	0	4	0
pH	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	5	6	4	0	2	0	1	0	0	0
Lactate	0	0	0	0	3	0	3	4	0	3	0	0	5	0	0	2	0	0	0	3	0	0	0	0	1	1
Hemoglobin	0	5	0	0	0	2	3	1	0	1	0	2	3	0	2	1	0	2	4	2	2	0	0	0	0	0
Peep	3	2	0	0	5	6	1	5	0	0	0	3	0	1	3	5	4	6	2	1	1	0	0	0	5	4
VT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1
Peak AP	1	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	7	4	4	1	7	0	3	0	0	0
Min. Vent.	0	5	0	0	0	0	0	0	0	3	0	0	0	1	1	1	0	0	0	0	1	0	3	0	0	0

Table 3: Expert Knowledge encoding in the causal graph

Node <sub>from</sub>	Node <sub>to</sub>	MJV <sup>1</sup>	Expert Knowledge	Literature (Evidence)
Age	Gender	Yes	Removed	Domain Expert
Age	Trauma	Yes	Changed Orientation	Domain Expert; (Ottochian et al., 2009)
Age	Smoker	Yes	Changed Orientation	Domain Expert
APSi	Death	Yes	Changed Orientation	Domain Expert; (Le Gall et al., 1984)
APSi	Peak AP	No	Added	Domain Expert
ARDS	PaCO <sub>2</sub>	Yes	Changed Orientation	Domain Expert; (Pham et al., 2017)
ARDS	Medical	No	Added	Domain Expert
BMI	Oxygenation	No	Added	Domain Expert; (Pham et al., 2017)
BMI	Peak AP	No	Added	Domain Expert; (Pham et al., 2017)
BMI	PaCO <sub>2</sub>	No	Added	Domain Expert; (Pham et al., 2017)
COPD	FiO <sub>2</sub>	Yes	Changed Orientation	Domain Expert; (Pham et al., 2017)
COPD	PaCO <sub>2</sub>	Yes	Changed Orientation	Domain Expert; (Pham et al., 2017)
COPD	Medical	No	Added	Domain Expert;
COPD	Peak AP	No	Added	Domain Expert; (Pham et al., 2017)
FiO <sub>2</sub>	Gender	Yes	Removed	Domain Expert
FiO <sub>2</sub>	Oxygenation	No	Added	Domain Expert; (Pham et al., 2017)
Gender	Min. Vent.	Yes	Changed Orientation	Domain Expert; (Pham et al., 2017)
Gender	Hemoglobin	Yes	Changed Orientation	Domain Expert
Hemoglobin	Death	Yes	Changed Orientation	Domain Expert
Hemoglobin	APSi	Yes	Changed Orientation	Domain Expert
Hemoglobin	Oxygenation	No	Added	Domain Expert
SpO <sub>2</sub>	Medical	No	Added	Domain Expert
Lactate	Medical	No	Added	Domain Expert;
Min Vent.	Oxygenation	Yes	Changed Orientation	Domain Expert; (Pham et al., 2017)
Oxygenation	Death	No	Added	Domain Expert
PaCO <sub>2</sub>	Ischemic HD	Yes	Removed	Domain Expert
PaCO <sub>2</sub>	Hemoglobin	Yes <sup>2</sup>	Removed	Domain Expert
PaCO <sub>2</sub>	Peak AP	Yes <sup>2</sup>	Removed	Domain Expert
PaO <sub>2</sub>	Oxygenation	Yes	Changed Orientation	Domain Expert; (Pham et al., 2017)
Peak AP	Oxygenation	No	Added	Domain Expert; (Pham et al., 2017)
PEEP	SOFA	Yes	Removed	Domain Expert
PEEP	Medical	Yes	Removed	Domain Expert
PEEP	Peak AP	Yes	Removed	Domain Expert
PEEP	Min. Vent.	Yes	Removed	Domain Expert;
PEEP	FiO <sub>2</sub>	Yes <sup>2</sup>	Changed Orientation	Domain Expert; (Mora Carpio AL, 2020)
Smoker	Surgery	Yes	Removed	Domain Expert
Smoker	PaCO <sub>2</sub>	Yes	Changed Orientation	Domain Expert; (Munro, 2006)
Smoker	Medical	No	Added	Domain Expert
SOFA	Lactate	Yes <sup>2</sup>	Changed Orientation	Domain Expert; (Jansen et al., 2009)
SpO <sub>2</sub>	Oxygenation	No	Added	Domain Expert; (Pham et al., 2017)
VT	Oxygenation	No	Added	Domain Expert; (Pham et al., 2017)

The final graph after incorporating domain knowledge with expert reviews is shown in Figure 5. The graph represents the causal relationship between the 26 variables under consideration after majority voting using SLAs and the encoding of the expert knowledge.

<sup>1</sup>Majority Voting

<sup>2</sup>Bidirectional edge

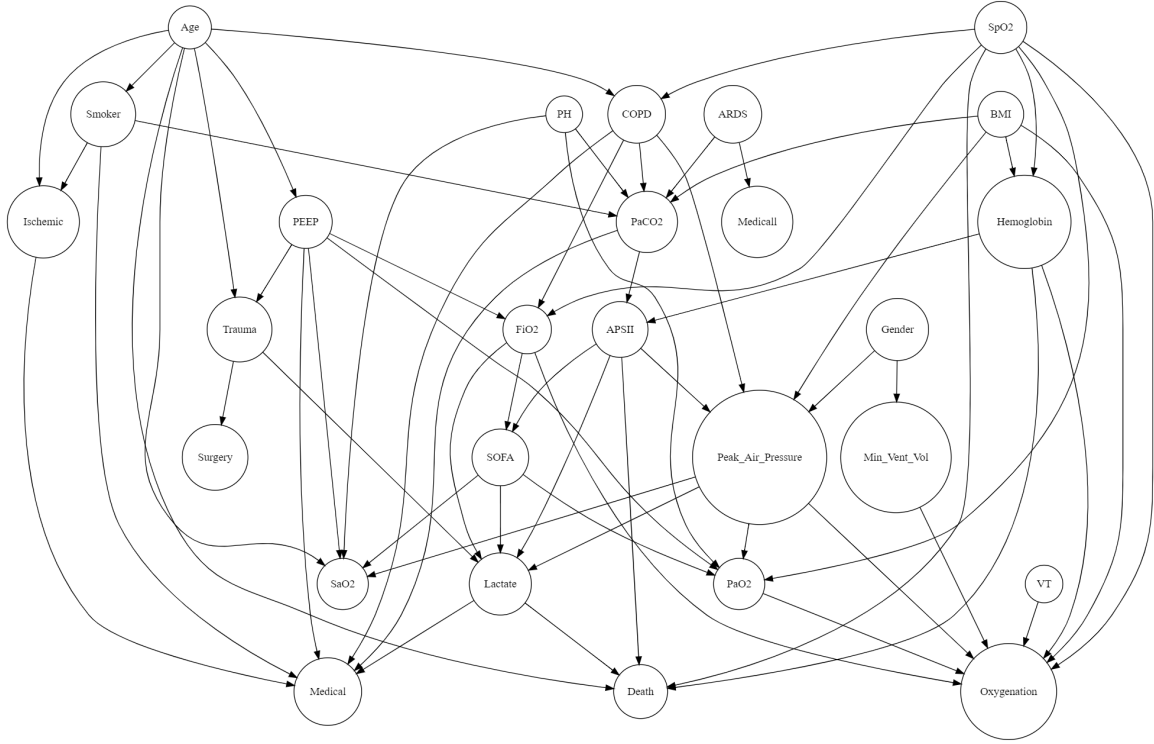


Figure 5: Causal Graph  $G$  after incorporating domain knowledge through expert reviews. *oxygenation* is the treatment variable, and *death* is the outcome variable.

### 5.5. Conservative versus liberal oxygenation strategies

The percentage time spent on any mandatory mode of MV during the study time period in the conservative and liberal group is 51%, hence patients spent the majority of the time within the intended target range for both of the groups. Overall, patients in the conservative group spent more time off-target than the patients in the liberal group. Mean  $SpO_2$ ,  $PaO_2$ , and  $FiO_2$  were well separated for all patients between these groups. Patients in the conservative group spent more time at a  $SpO_2$  of 90 – 95% than those in the liberal group (Figure 3).

We performed a series of experiments based on the causal graph and the selected cohort from the observational data. For each of the following oxygenation strategies, discussed in the subsequent subsections, we have obtained a formula with an admissible set by adjustment using the rules of *do-calculus*. To illustrate the full picture, the derivation of the query in subsection 5.5.3 presented in the appendix, Appendix A and Appendix B. We also present the summary of the experimental results for CB-OBS and SCM-VRCT in Table 4

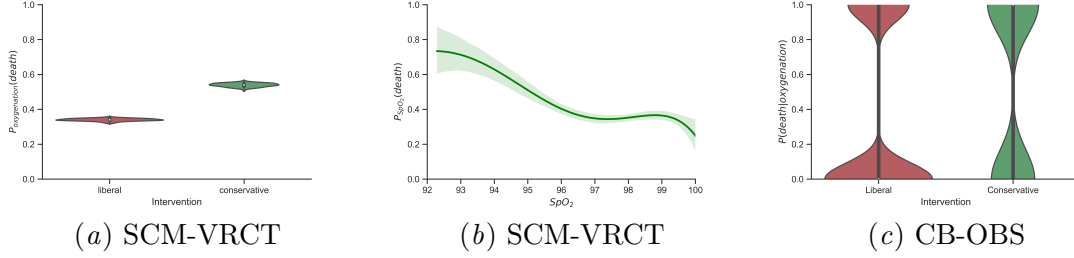


Figure 6: Estimation of queries on the causal graph for both observational and experimental dataset. (a)  $P_{\text{oxygenation}}(\text{death})$ , (b)  $P_{\text{SpO}_2}(\text{death})$  and (c)  $P(\text{death}|\text{oxygenation})$

#### 5.5.1. EFFECT ON MORTALITY FOR CONSERVATIVE AND LIBERAL OXYGENATION STRATEGIES

In our first experiment, we estimated the effect of conservative and liberal oxygenation on mortality,  $P_{\text{oxygenation}}(\text{death})$ . The expectations of mortality for conservative and liberal oxygenation strategies are shown in Figure 6 for both the CB-OBS and SCM-VRCT. For the CB-OBS, we found that the expectation of mortality is higher, 0.54, for the conservative oxygenation than the expectation of the mortality, 0.362 (95% CI [0.347, 0.377]) for the liberal oxygenation, as shown in 6(c)subfigure. However, for the SCM-VRCT, the expectation for mortality for the conservative oxygenation strategy is also 0.54 (95% CI [0.521, 0.558]) which is the same as the expectation we estimated in the CB-OBS. The expected mortality for the liberal approach in SCM-VRCT is 0.34 (95% CI [0.33, 0.353]), as shown in 6(a)subfigure. The variance for the effect of oxygenation on mortality in SCM-VRCT is significantly lower, 0.01 compared to the results in CB-OBS, 0.49. We also present the effect of  $SpO_2$  on mortality for SCM-VRCT in 6(b)subfigure

#### 5.5.2. AGE-SPECIFIC EFFECT ON MORTALITY FOR OXYGENATION STRATEGIES

In the second covariate-specific experiment, we estimated the effect of oxygenation on mortality conditioned on age,  $P_{\text{oxygenation}}(\text{death}|\text{age})$ . The expectations of the mortality for conservative and liberal oxygenation for both the CB-OBS and SCM-VRCT are shown in Figure 7. For CB-OBS, we found that the age of the patient does not have any effect on the expected mortality. The expectation of mortality in SCM-VRCT for the liberal oxygenation strategy conditioned on age is 0.298 (95% CI [0.285, 0.313]), see 7(a)subfigure, which is lower compare to non-age specific CB-OBS, 0.362 (95% CI [0.347, 0.377]). On the other hand, the expectations of the mortality for conservative oxygenation for SCM-VRCT and CB-OBS are 0.529 (95% CI [0.505, 0.55]) and 0.54 (95% CI [0.478, 0.60]), respectively.

#### 5.5.3. DISEASE SEVERITY-SPECIFIC EFFECT ON MORTALITY FOR OXYGENATION STRATEGIES

We also estimated the effect of oxygen therapy on mortality conditioned on APSIII and SOFA scores,  $P_{\text{oxygenation}}(\text{death}|\text{APSIII}, \text{SOFA})$ . The expectation of mortality for conservative and liberal oxygenation for CB-OBS and SCM-VRCT are shown in Figure 7. In

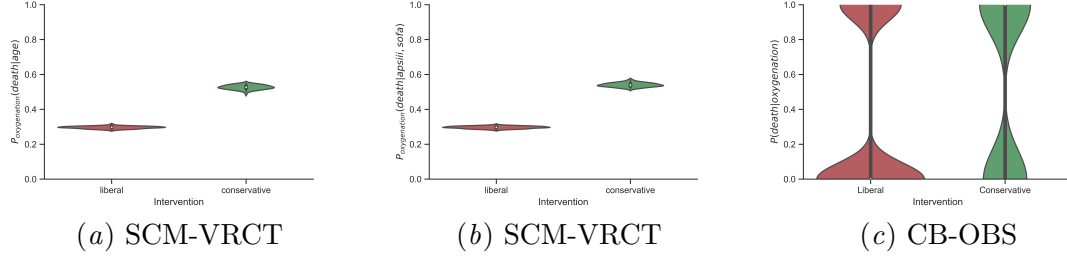


Figure 7: Estimation of queries on the causal graph for both observational and experimental dataset. (a)  $P_{oxygenation}(death|age)$ , (b)  $P_{oxygenation}(death|apsiii, sofa)$  and (c)  $P(death|oxygenation)$

CB-OBS, as we discussed in the previous section, the expectation of mortality is higher, 0.54 (95% CI [0.478, 0.60]), for the conservative oxygenation than the expectation of mortality, 0.362 (95% CI [0.347, 0.377]) for the liberal oxygenation. The conservative strategy in SCM-VRCT results in 0.539 (95% CI [0.518, 0.562]), see 7(c)subfigure. However, in the SCM-VRCT, the expectation of the mortality in the liberal strategy is 0.298 (95% CI [0.281, 0.311]) which is significantly lower than the expectation we estimated in the CB-OBS. Therefore, given the severity of illness based on the physiological scores, the liberal approach lowers the expectation of mortality. This finding is consistent with the results of several other studies reported in Table 6.

#### 5.5.4. DISEASE SEVERITY-SPECIFIC EFFECT ON MORTALITY CONDITIONED ON SPECIFIC SOFA SCORE

Finally, we estimated the effect of oxygenation on mortality conditioned on two different ranges of SOFA score,  $P_{oxygenation}(death|SOFA \leq 10)$  and  $P_{oxygenation}(death|SOFA > 10)$ . Although the expectation of mortality decreases as we increase the  $SpO_2$ , as shown in 7(c)subfigure, but the expectations of mortality for conservative and liberal oxygenation are different for two different ranges of SOFA scores. In the lower range of SOFA score ( $SOFA \leq 10$ ), we found that the expectation of mortality is higher, 0.42 for the conservative oxygenation than the expectation of mortality, 0.35 for the liberal oxygenation. However, in the upper range of SOFA score ( $SOFA > 10$ ), we found that the expectation of mortality is higher, 0.58 for the conservative oxygenation than the expectation of mortality, 0.48 for the liberal oxygenation. One important observation here is that the expected outcome is higher for patients with a SOFA score greater than 10 compared to patients with

Table 4: Summary Results of CB-OBS and SCM-VRCT

Virtual Experiments	Study	Liberal			Conservative		
		Mean	Std	Upper 95% CI	Mean	Std	Upper 95% CI
$P(death Oxygenation)$	CB-OBS	0.362	0.481	0.377	0.54	0.499	0.60
$P_{Oxygenation}(death)$	SCM-VRCT	0.34	0.01	0.355	0.54	0.01	0.56
$P_{Oxygenation}(death age)$	SCM-VRCT	0.298	0.01	0.313	0.529	0.01	0.55
$P_{Oxygenation}(death apsiii, sofa)$	SCM-VRCT	0.298	0.01	0.311	0.54	0.01	0.563



Table 5: Comparative Analysis of Study Demographics

Study	Investigators et al. (2020)	Panwar et al. (2016)	et al. Girardis (2016)	et al. Eastwood (2016)	et al. Suzuki (2013)	MIMIC	
Type of study	RCT	RCT	RCT	RCT	RCT	CB-OBS	SCM-VRCT
Included Patients	Age $\geq$ 18 admitted to ICU on MV with MV time $< 2$ hours	Age $\geq$ 18 admitted to ICU on MV with expected MV time $\geq 24$ hours	Age $\geq$ 18 admitted to ICU with expected ICU LOS $\geq 72$ hours	Age $\geq$ 18 admitted to ICU on MV post CA	Age $\geq$ 18 admitted to ICU on MV with expected MV time $\geq 48$ hours	Age $\geq$ 18 admitted to ICU on MV with expected MV time $\geq 24$ hours	Age $\geq$ 18 admitted to ICU on MV with expected MV time $\geq 24$ hours
SpO <sub>2</sub> Minimum Target							
Conservative	90%	88%	94%	88%	90%	88%	88%
Conventional (Liberal)	97%	96%	97%	97%	97%	96%	96%
No of patients							
Conservative	484	52	216	50	54	250	2028
Conventional	481	51	218	50	51	3812	2034
Age, SD(IQR)							
Conservative	58.1(16.2)	62.4 (14.9)	63 (51-74)	65(59-77)	56 (16)	60.82 (15.6)	60.78 (0.55)
Conventional	57.5 (16.1)	62.4 (17.4)	65 (52-76)	67(50-71)	59 (17)	62.94 (16.22)	62.95(0.34)
Female Gender (%)							
Conservative	178 (36.8%)	20 (38.5%)	95 (44.0%)	21 (42.0%)	22 (40.7%)	100 (40.0%)	801(10%)
Conventional	179 (37.2%)	18 (35.3%)	93 (42.7%)	16 (32.0%)	13 (25.5%)	1615 (42.4%)	847 (42%)
Mechanical Ventilation N, (%)							
Conservative	484(100%)	52 (100%)	143 (66.2%)	50 (100%)	54 (100%)	250 (100%)	2000(100%)
Conventional	481(100%)	51 (100%)	148 (67.9%)	50 (100%)	51 (100%)	3812 (100%)	2000(100%)
APACHE III score (IQR)							
Conservative	NR (APACHE II)	79.5 (61-92.5)	NR	121 (105-142)	62 (49-92)	NR (APS III)	NR (APS III)
Conventional	NR (APACHE II)	70 (50-84)	NR <sup>3</sup>	125 (107-141)	68 (42-94)	NR (APS III)	NR (APS III)

a SOFA score smaller than or equal to 10. Therefore, given the severity of illness based on the SOFA score, there is no significant difference in the mortality for different oxygenation targets. This finding also supports the feasibility of a larger study to investigate the effect of conservative oxygenation to treat patients requiring invasive MV in ICUs.

## 6. Discussion

The study was conducted and validated on the MIMIC III ICU database, which contains data routinely collected from adult patients in the US. We included adult patients fulfilling the requirement for the study based on the OT-RCT study. After the exclusion of ineligible cases, we included 4,062 patients for the study. The time-series data of oxygenation parameters and ventilator settings were then used to compute the mean mortality over the entire period. Patient demographics and clinical characteristics are shown in Table 1. We extracted a set of 26 variables, 24 variables following the OT-RCT, including demographics, oxygenation parameters and ventilator settings, and 2 variables, one for oxygenation (treatment) and another for mortality (outcome).

We assessed the feasibility of a conservative oxygenation strategy (target  $SpO_2$  88-95%) compared to a liberal oxygenation strategy (target  $SpO_2 \geq 96\%$ ) during MV for adult ICU

<sup>3</sup>Not reported

Table 6: Comparative Analysis of Study Outcome Measures

Study	Investigators et al. (2020)	Panwar et al. (2016)	Girardis et al. (2016)	Eastwood et al. (2016)	Suzuki et al. (2013)	CB-OBS	MIMIC SCM-VRCT
ICU Mortality, N (%)							
Conservative	NR	13 (25%)	25 (11.6%)	15 (30%)	NR	NR	NR
Conventional	NR	12 (24%)	44 (20.2%)	16 (32%)	NR	NR	NR
28-day Mortality, N (%)							
Conservative	NR	NR	NR	27 (54%)	9 (16.7%)	122 (48.8%)	NR
Conventional	NR	NR	NR	26 (52%)	16 (31.4%)	949 (24.89%)	NR
90-day Mortality, N (%)							
Conservative	166 (34.7%)	21 (40%)	NR	NR	NR	135 (54%)	1078(53.9%)
Conventional	156 (32.5%)	19 (37%)	NR	NR	NR	1380 (36.2%)	682(34.1%)
180-day Mortality, N (%)							
Conservative	170 (35.7%)	NR	NR	NR	NR	NR	NR
Conventional	164 (34.5%)	NR	NR	NR	NR	NR	NR
Hospital Mortality, N (%)							
Conservative	NR	NR	52 (24.2%)	28 (56%)	NR	NR	NR
Conventional	NR	NR	74 (33.9%)	27 (54%)	NR	NR	NR
ICU LOS, Days (IQR)							
Conservative	NR	9 (5-13)	6 (4-10)	4 (2-7)	NR	NR	NR
Conventional	NR	7 (4-12)	6 (4-11)	5 (4-9)	NR	NR	NR
Hospital LOS, Days (IQR)							
Conservative	NR	20 (10-25)	21 (13-38)	9 (3-17)	NR	NR	NR
Conventional	NR	16 (7-30)	21 (12-34)	9 (4-24)	NR	NR	NR
New infections, N (%) - Search database for Bacteremia (Girardis looked at this), Nosocomial							
Conservative	NR	NR	39 (18.1)	NR	31 (57.4)	NR	NR
Conventional	NR	NR	50 (22.9)	NR	28 (54.9)	NR	NR
New non-respiratory organ failure, N (%) - 3 organs, if possible: Creatine baseline vs. after 2 weeks or 30 days; Troponin baseline vs. after 1 week; Lactate baseline vs. 1 week							
Conservative	NR	NR	38 (17.6)	NR	16 (29.6)	NR	NR
Conventional	NR	NR	58 (26.6)	NR	22 (43.1)	NR	NR

patients. The study protocol followed the protocol in the OT-RCT. To achieve treatment feasibility, this study implemented a clear separation in the mean  $SpO_2$ ,  $SaO_2$ ,  $PaO_2$ , and  $FiO_2$  values between the groups. The point estimate for 90-day mortality was lower with the liberal oxygenation strategy. This is not consistent with the findings of the OT-RCT. However, in the SCM-VRCT, the expectation of the mortality in the liberal strategy is significantly lower than the expectation we estimated in the CB-OBS. We conclude that given the severity of illness based on the physiological scores, the liberal approach lowers the expectation of mortality. This finding is consistent with the results of several other studies reported in Table 6.. The result of the estimation are presented in Figure 6 (a). Another important observation is that the expected outcome is higher for patients with a SOFA score greater than 10 compared to patients with a SOFA score smaller than or equal to 10. That is, given the severity of illness based on the SOFA score, there is no significant difference in the mortality for different oxygenation targets. This finding also supports the feasibility of a larger study to investigate the effect of conservative oxygenation to treat patients requiring invasive MV in ICUs.

Several methodological limitations that we have in the current model are currently being addressed by us for future work with the theoretical foundations. These include time-varying causal model with SCM, survival analysis with SCM, and incorporating unobserved confounders in the model, among others.

It is important to note that although we demonstrated this method on oxygenation therapy for MV patients in the ICU, the technique can be applied generally for a broad set of clinical questions. We are currently applying this method for a range of questions including

estimating the impact of antipsychotic drugs on delirium in the ICU and estimating the impact of timing of antibiotics for patients with chronic pulmonary obstructive disease.

## 7. Acknowledgement

This study was partially supported by the Regenstrief Center for Healthcare Engineering at Purdue University. We would like to express our sincere gratitude to Professor Elias Bareinboim for his insights on the methodological framework. Part of the analysis (Appendix [Appendix B](#)) was achieved with the [causalfusion.net](#) software developed by Dr. Bareinboim.

## References

- Ayodeji Adegunsoye, Justin M Oldham, Shashi K Bellam, Steven Montner, Matthew M Churpek, Imre Noth, Rekha Vij, Mary E Strek, and Jonathan H Chung. Computed tomography honeycombing identifies a progressive fibrotic phenotype with increased mortality across diverse interstitial lung diseases. *Annals of the American Thoracic Society*, 16(5):580–588, 2019.
- Neill KJ Adhikari, Robert A Fowler, Satish Bhagwanjee, and Gordon D Rubenfeld. Critical care and the global burden of critical illness in adults. *The Lancet*, 376:1339–1346, 2010. doi: 10.1016/S0140-6736(10)60446-1. PMID: 24171518.
- Mohammad Adibuzzaman, Ken Musselman, Alistair Johnson, Paul Brown, Zachary Pitluk, and Ananth Grama. Closing the data loop: An integrated open access analysis platform for the MIMIC database. In *Computing in Cardiology Conference (CinC), 2016*, pages 137–140. IEEE, 2016.
- Mohammad Adibuzzaman, Poching DeLaurentis, Jennifer Hill, and Brian D Benneyworth. Big data in healthcare—the promises, challenges and opportunities from a research perspective: a case study with a model database. In *AMIA Annual Symposium Proceedings*, volume 2017, page 384. American Medical Informatics Association, 2017.
- Peter Buhlmann Alain Hauser. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. *Journal of Machine Learning Research*, 13:2409–2464, 2012. URL <http://www.jmlr.org/papers/volume13/hauser12a/hauser12a.pdf>.
- Steen A. Andersson, David Madigan, and Michael D. Perlman. A characterization of markov equivalence classes for acyclic digraphs. *The Annals of Statistics*, 25(2):505–541, 1997. ISSN 00905364. URL <http://www.jstor.org/stable/2242556>.
- Linn Håkonsen Arendt, Cecilia Høst Ramlau-Hansen, Allen J Wilcox, Tine Brink Henriksen, Jørn Olsen, and Morten Søndergaard Lindhard. Placental weight and male genital anomalies: a nationwide danish cohort study. *American journal of epidemiology*, 183(12):1122–1128, 2016.
- Elias Bareinboim and Judea Pearl. Transportability of causal effects: completeness results. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pages 698–704, 2012.
- Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.
- David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, 2002. URL <http://www.jmlr.org/papers/volume3/chickering02b/chickering02b.pdf>.
- Robert O Crapo, Robert L Jensen, Mathew Hegewald, and Donald P Tashkin. Arterial blood gas reference values for sea level and an altitude of 1,400 meters. *American Journal of Respiratory and Critical Care Medicine*, 160(5):1525–1531, 1999.

- Glenn M Eastwood, Aiko Tanaka, Emilo Daniel Valenzuela Espinoza, Leah Peck, Helen Young, Johan Mårtensson, Ling Zhang, Neil J Glassford, Yu-Feng Frank Hsiao, Satoshi Suzuki, et al. Conservative oxygen therapy in mechanically ventilated patients following cardiac arrest: a retrospective nested cohort study. *Resuscitation*, 101:108–114, 2016.
- Ronald Aylmer Fisher et al. The design of experiments. *The design of experiments.*, (7th Ed), 1960.
- Thomas R Frieden. Evidence for health decision making—beyond randomized, controlled trials. *New England Journal of Medicine*, 377(5):465–475, 2017.
- Efstathios D Gennatas, Jerome H Friedman, Lyle H Ungar, Romain Pirracchio, Eric Eaton, Lara G Reichmann, Yannet Interian, José Marcio Luna, Charles B Simone, Andrew Auerbach, et al. Expert-augmented machine learning. *Proceedings of the National Academy of Sciences*, 117(9):4571–4577, 2020.
- Marzyeh Ghassemi, Leo Anthony Celi, and David J Stone. State of the art review: the data revolution in critical care. *Critical Care*, 19(1):118, 2015.
- Massimo Girardis, Stefano Busani, Elisa Damiani, Abele Donati, Laura Rinaldi, Andrea Marudi, Andrea Morelli, Massimo Antonelli, and Mervyn Singer. Effect of Conservative vs Conventional Oxygen Therapy on Mortality Among Patients in an Intensive Care Unit: The Oxygen-ICU Randomized Clinical Trial. *JAMA*, 316(15):1583–1589, 10 2016. ISSN 0098-7484. doi: 10.1001/jama.2016.11993. URL <https://doi.org/10.1001/jama.2016.11993>.
- Fred Glover and Manuel Laguna. *Tabu Search*, pages 2093–2229. Springer US, Boston, MA, 1998. ISBN 978-1-4613-0303-9. doi: 10.1007/978-1-4613-0303-9\_33. URL [https://doi.org/10.1007/978-1-4613-0303-9\\_33](https://doi.org/10.1007/978-1-4613-0303-9_33).
- Travis Goodwin and Sanda M Harabagiu. Automatic generation of a qualified medical knowledge graph and its usage for retrieving patient cohorts from electronic medical records. In *2013 IEEE Seventh International Conference on Semantic Computing*, pages 363–370. IEEE, 2013.
- Jeremy A Greene and Scott H Podolsky. Reform, regulation, and pharmaceuticals—the kefauber–harris amendments at 50. *New England Journal of Medicine*, 367(16):1481–1483, 2012.
- Naftali Harris and Mathias Drton. Pc algorithm for nonparanormal graphical models. *Journal of Machine Learning Research*, 14(69):3365–3383, 2013. URL <http://jmlr.org/papers/v14/harris13a.html>.
- John E Heffner. The story of oxygen. *Respiratory Care*, 58(1):18–31, 2013. ISSN 0020-1324. doi: 10.4187/respcare.01831. URL <http://rc.rcjournal.com/content/58/1/18>.
- Christina Heinze-Deml, Marloes H Maathuis, and Nicolai Meinshausen. Causal structure learning. *Annual Review of Statistics and Its Application*, 5:371–391, 2018.

- MA Hernán and JM Robins. Causal inference: What if. *Boca Raton: Chapman & Hill/CRC*, 2020.
- Miguel A Hernán, Alvaro Alonso, Roger Logan, Francine Grodstein, Karin B Michels, Meir J Stampfer, Walter C Willett, JoAnn E Manson, and James M Robins. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology (Cambridge, Mass.)*, 19(6): 766, 2008.
- Sonia Hernández-Díaz, Allen J Wilcox, Enrique F Schisterman, and Miguel A Hernán. From causal diagrams to birth weight-specific curves of infant mortality. *European journal of epidemiology*, 23(3):163–166, 2008.
- Takashi Hirase, Eric S Ruff, Iqbal Ratnani, and Salim R Surani. Impact of conservative versus conventional oxygenation on outcomes of patients in intensive care units: A systematic review and meta-analysis. *Cureus*, 11(9), 2019.
- Paul Hünermund and Elias Bareinboim. Causal inference and data-fusion in econometrics. *arXiv preprint arXiv:1912.09104*, 2019.
- The ICU-ROX Investigators, the Australian, and New Zealand Intensive Care Society Clinical Trials Group. Conservative oxygen therapy during mechanical ventilation in the icu. *New England Journal of Medicine*, 382(11):989–998, 2020. doi: 10.1056/NEJMoa1903297. URL <https://doi.org/10.1056/NEJMoa1903297>.
- Constantin F Aliferis Ioannis Tsamardinos, Laura E Brown. The max-min hill-climbing bayesian network structure learning algorithm. *Machine Learning*, 65:31–78, 2006. doi: 10.1007/s10994-006-6889-7.
- Tim C. Jansen, Jasper van Bommel, Roger Woodward, Paul G. H. Mulder, and Jan Bakker. Association between blood lactate levels, sequential organ failure assessment subscores, and 28-day mortality during early and late intensive care unit stay: A retrospective observational study\*. *Read Online: Critical Care Medicine — Society of Critical Care Medicine*, 37(8), 2009. ISSN 0090-3493. URL [https://journals.lww.com/ccmjournal/Fulltext/2009/08000/Association\\_between\\_blood\\_lactate\\_levels,.7.aspx](https://journals.lww.com/ccmjournal/Fulltext/2009/08000/Association_between_blood_lactate_levels,.7.aspx).
- Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3:160035, 2016.
- Jeremy M. Kahn, Nicole M. Benson, Dina Appleby, Shannon S. Carson, and Theodore J. Iwashyna. Long-term Acute Care Hospital Utilization After Critical Illness. *JAMA*, 303(22):2253–2259, 06 2010. ISSN 0098-7484. doi: 10.1001/jama.2010.761. URL <https://doi.org/10.1001/jama.2010.761>.
- Jean-Roger Le Gall, Philippe Loirat, Annick Alperovitch, Paul Glaser, Claude Granthil, Daniel Mathieu, Philippe Mercier, Remi Thomas, and Daniel Villers. A simplified acute physiology score for icu patients. *Critical care medicine*, 12(11):975–977, 1984.

- David J Lederer, Scott C Bell, Richard D Branson, James D Chalmers, Rachel Marshall, David M Maslove, David E Ost, Naresh M Punjabi, Michael Schatz, Alan R Smyth, et al. Control of confounding and reporting of results in causal inference studies. guidance for authors from editors of respiratory, sleep, and critical care journals. *Annals of the American Thoracic Society*, 16(1):22–28, 2019.
- Maureen O. Meade, Deborah J. Cook, Gordon H. Guyatt, Arthur S. Slutsky, Yaseen M. Arabi, D. James Cooper, Andrew R. Davies, Lori E. Hand, Qi Zhou, Lehana Thabane, Peggy Austin, Stephen Lapinsky, Alan Baxter, James Russell, Yoanna Skrobik, Juan J. Ronco, Thomas E. Stewart, and for the Lung Open Ventilation Study Investigators. Ventilation Strategy Using Low Tidal Volumes, Recruitment Maneuvers, and High Positive End-Expiratory Pressure for Acute Lung Injury and Acute Respiratory Distress Syndrome: A Randomized Controlled Trial. *JAMA*, 299(6):637–645, 02 2008. ISSN 0098-7484. doi: 10.1001/jama.299.6.637. URL <https://doi.org/10.1001/jama.299.6.637>.
- Philipp G. H. Metnitz, Barbara Metnitz, Rui P. Moreno, Peter Bauer, Lorenzo Del Sorbo, Christoph Hoermann, Susana Afonso de Carvalho, V. Marco Ranieri, and on behalf of the SAPS 3. Investigators. Epidemiology of mechanical ventilation: Analysis of the saps 3 database. *Intensive Care Medicine*, 35(5):816–825, 2009. ISSN 1432-1238. doi: 10.1007/s00134-009-1449-9. URL <https://doi.org/10.1007/s00134-009-1449-9>.
- Mora JI Mora Carpio AL. Ventilator management. *StatPearls [Internet]. Treasure Island (FL)*, January 2020. URL <https://www.ncbi.nlm.nih.gov/books/NBK448186/>.
- Nancy Munro. Weaning smokers from mechanical ventilation. *Critical Care Nursing Clinics of North America*, 18(1):21 – 28, 2006. ISSN 0899-5885. doi: <https://doi.org/10.1016/j.ccell.2005.10.001>. URL <http://www.sciencedirect.com/science/article/pii/S0899588505000882>. Tobacco Use and Smoking Cessation in Acute and Critical Care.
- Preetam Nandy, Alain Hauser, Marloes H Maathuis, et al. High-dimensional consistency in score-based and hybrid structure learning. *The Annals of Statistics*, 46(6A):3151–3183, 2018.
- The Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *New England Journal of Medicine*, 342(18):1301–1308, 2000. doi: 10.1056/NEJM200005043421801. URL <https://doi.org/10.1056/NEJM200005043421801>. PMID: 10793162.
- Galia Nordon, Gideon Koren, Varda Shalev, Benny Kimelfeld, Uri Shalit, and Kira Radinsky. Building causal graphs from medical literature and electronic medical records. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1102–1109, 2019.
- Marcus Ottochian, Ali Salim, Joseph DuBose, Pedro G. R. Teixeira, Linda S. Chan, and Daniel R. Margulies. Does age matter?: The relationship between age and mortality in penetrating trauma. *Injury*, 40(4):354–357, Apr 2009. ISSN 0020-1383. doi: 10.1016/j.injury.2008.10.015. URL <https://doi.org/10.1016/j.injury.2008.10.015>.



- Dr R. Panwar, G. Capellier, N. Schmutz, A. Davies, D. J. Cooper, M. Bailey, D. Baguley, D. V. Pilcher, and R. Bellomo. Current oxygenation practice in ventilated patients—an observational cohort study. *Anaesthesia and Intensive Care*, 41(4):505–514, 2013. doi: 10.1177/0310057X1304100412. URL <https://doi.org/10.1177/0310057X1304100412>. PMID: 23808511.
- Rakshit Panwar, Miranda Hardie, Rinaldo Bellomo, Loïc Barrot, Glenn M Eastwood, Paul J Young, Gilles Capellier, Peter WJ Harrigan, and Michael Bailey. Conservative versus liberal oxygenation targets for mechanically ventilated patients. a pilot multicenter randomized controlled trial. *American journal of respiratory and critical care medicine*, 193(1):43–51, 2016.
- J Pearl and D Mackenzie. The book of why: The new science of cause and effect. *Hachette Book Group*, 2018.
- Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.
- Judea Pearl. *Causality*. Cambridge university press, 2009.
- Judea Pearl. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3):54–60, 2019.
- Judea Pearl and Elias Bareinboim. External validity: From do-calculus to transportability across populations. *Statistical Science*, pages 579–595, 2014.
- Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.
- Richard Scheines Peter Spirtes, Clark Glymour. *Causation, Prediction, and Search. Adaptive Computation and Machine Learning Series*. MIT Press, Cambridge, MA, 2nd edition, 2000.
- Tài Pham, Laurent J. Brochard, and Arthur S. Slutsky. Mechanical ventilation: State of the art. *Mayo Clinic Proceedings*, 92(9):1382–1400, Sep 2017. ISSN 0025-6196. doi: 10.1016/j.mayocp.2017.05.004. URL <https://doi.org/10.1016/j.mayocp.2017.05.004>.
- Vineet K Raghu, Allen Poon, and Panayiotis V Benos. Evaluation of causal structure learning methods on mixed data types. *Proceedings of machine learning research*, 92:48, 2018.
- James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512, 1986.
- James M Robins and Miguel A Hernán. Estimation of the causal effects of time-varying exposures. *Longitudinal data analysis*, 553:599, 2009.
- Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

- Dominik Rothenhäusler, Christina Heinze, Jonas Peters, and Nicolai Meinshausen. Back-shift: Learning causal cyclic graphs from unknown shift interventions. In *NIPS*, 2015.
- Maya Rotmensch, Yoni Halpern, Abdulhakim Tlimat, Steven Horng, and David Sontag. Learning a health knowledge graph from electronic medical records. *Scientific reports*, 7(1):1–11, 2017.
- Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.
- Enrique F Schisterman, Brian W Whitcomb, Germaine M Buck Louis, and Thomas A Louis. Lipid adjustment in the analysis of environmental contaminants and human health risks. *Environmental health perspectives*, 113(7):853–857, 2005.
- Daniela K Schlüter, Rowena Griffiths, Abdulfatah Adam, Ashley Akbari, Martin L Heaven, Shantini Paranjothy, Anne-Marie Nybo Andersen, Siobhán B Carr, Tania Pressler, Peter J Diggle, et al. Impact of cystic fibrosis on birthweight: a population based study of children in denmark and wales. *Thorax*, 74(5):447–454, 2019.
- Marco Scutari. Learning bayesian networks with the bnlearn r package. *Journal of Statistical Software, Articles*, 35(3):1–22, 2010. ISSN 1548-7660. doi: 10.18637/jss.v035.i03. URL <https://www.jstatsoft.org/v035/i03>.
- Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(Oct):2003–2030, 2006.
- Reed A C Siemieniuk, Derek K Chu, Lisa Ha-Yeon Kim, Maria-Rosa Güell-Rous, Waleed Alhazzani, Paola M Soccia, Paul J Karanicolas, Pauline D Farhoumand, Jillian L K Siemieniuk, Imran Satia, Elvis M Irusen, Marwan M Refaat, J Stephen Mikita, Maureen Smith, Dian N Cohen, Per O Vandvik, Thomas Agoritsas, Lyubov Lytvyn, and Gordon H Guyatt. Oxygen therapy for acutely ill medical patients: a clinical practice guideline. *BMJ*, 363, 2018. ISSN 0959-8138. doi: 10.1136/bmj.k4169. URL <https://www.bmj.com/content/363/bmj.k4169>.
- Satoshi Suzuki, Glenn M. Eastwood, Leah Peck, Neil J. Glassford, and Rinaldo Bellomo. Current oxygen management in mechanically ventilated patients: A prospective observational cohort study. *Journal of Critical Care*, 28(5):647 – 654, 2013. ISSN 0883-9441. doi: <https://doi.org/10.1016/j.jcrc.2013.03.010>. URL <http://www.sciencedirect.com/science/article/pii/S0883944113000713>.
- Jean-Louis Vincent and Daniel De Backer. Circulatory shock. *New England Journal of Medicine*, 369(18):1726–1734, 2013. doi: 10.1056/NEJMra1208943. URL <https://doi.org/10.1056/NEJMra1208943>. PMID: 24171518.
- Claudia Vitolo, Marco Scutari, Mohamed Ghalaieny, Allan Tucker, and Andrew Russell. Modeling air pollution, climate, and health data using bayesian networks: A case study of the english regions. *Earth and Space Science*, 5(4):76–88, 2018.

- Hannah Wunsch, Walter T. Linde-Zwirble, Derek C. Angus, Mary E. Hartman, Eric B. Milbrandt, and Jeremy M. Kahn. The epidemiology of mechanical ventilation use in the united states. *Critical Care Medicine — Society of Critical Care Medicine*, 38(10): 1947–1953, 2010. doi: 10.1097/CCM.0b013e3181ef4460. PMID: 24171518.
- Tingting Zhu, Alistair EW Johnson, Yang Yang, Gari D Clifford, and David A Clifton. Bayesian fusion of physiological measurements using a signal quality extension. *Physiological measurement*, 39(6):065008, 2018.

## Appendices: Derivation of Identifiability of the causal effect of *oxygenation* on *death* given $\{apsiii, sofa\}$

### Appendix A. Formulation

To keep the derivation steps brief and concise, we define the following sets of variables in Table 7 that are combinations of the 24 variables we have in our model. The large number of variables, 24 variables, we have in the model makes the subscripts in the equations long and clumsy. We use these sets in the derivation section.

Set	Variables
$U =$	$\{age, apsiiii, ards, bmi, copd, death, fio2, gender, hemoglobin, ischemicHd, lactate, medical, minVentVol, oxygenation, paco2, pao2, peakAirPressure, peep, ph, sao2, smoker, sofa, spo2, surgery, trauma, vt\}$
$A =$	$\{oxygenation, apsiiii, sofa, smoker, copd, spo2, fio2, hemoglobin, peep, peakAirPressure\}$
$B =$	$\{smoker, copd, spo2, fio2, hemoglobin, peep, peakAirPressure\}$
$C =$	$\{age, gender, bmi, trauma, smoker, copd, ards, spo2, fio2, paco2, ph, hemoglobin, peep, peakAirPressure, lactate\}$
$D =$	$\{age, gender, bmi, trauma, apsiiii, sofa, smoker, copd, ards, death, spo2, fio2, paco2, ph, hemoglobin, peep, peakAirPressure, lactate\}$
$E =$	$\{bmi, ards, age, smoker, ph, spo2, copd, paco2, gender, hemoglobin, apsiiii, peakAirPressure, peep, trauma, fio2, lactate\}$
$F =$	$\{bmi, ards, age, smoker, ph, spo2, copd, paco2, gender, hemoglobin, apsiiii, peep, peakAirPressure, trauma, fio2, lactate, minVentVol, sofa, pao2, vt, oxygenation\}$
$G =$	$\{lactate, fio2, trauma, peep, peakAirPressure, apsiiii, hemoglobin, gender, paco2, copd, spo2, ph, smoker, age, ards, bmi\}$
$H =$	$\{oxygenation, vt, pao2, sofa, minVentVol, lactate, fio2, trauma, peep, age, ards, peakAirPressure, apsiiii, hemoglobin, gender, paco2, copd, spo2, ph, smoker, bmi\}$
$I =$	$\{fio2, trauma, peep, peakAirPressure, apsiiii, hemoglobin, gender, paco2, copd, spo2, ph, smoker, age, ards, bmi\}$
$J =$	$\{gender, trauma, surgery, smoker, ischemicHd, hemoglobin, ph, minVentVol, ards, bmi, medical, paco2, apsiiii, peakAirPressure, lactate, sofa, sao2, pao2, vt, oxygenation, death\}$
$K =$	$\{age, copd, peep, fio2, trauma, surgery, smoker, ischemicHd, ph, minVentVol, ards, bmi, paco2, apsiiii, peakAirPressure, lactate, medical, sofa, sao2, pao2, vt, oxygenation, death\}$
$L =$	$\{ph, spo2, gender, hemoglobin, copd, fio2, surgery, smoker, ischemicHd, minVentVol, ards, bmi, paco2, apsiiii, peakAirPressure, lactate, medical, sofa, sao2, pao2, vt, oxygenation, death\}$
$M =$	$\{peep, trauma, gender, hemoglobin, fio2, surgery, ischemicHd, minVentVol, apsiiii, peakAirPressure, lactate, medical, sofa, sao2, pao2, vt, oxygenation, death\}$
$N =$	$\{gender, peep, trauma, surgery, smoker, ischemicHd, fio2, hemoglobin, ph, minVentVol, ards, bmi, paco2, apsiiii, peakAirPressure, lactate, medical, sofa, sao2, pao2, vt, oxygenation, death\}$
$O =$	$\{bmi, ards, age, smoker, ph, spo2, copd, paco2, gender, hemoglobin, apsiiii, peakAirPressure, peep, trauma, fio2\}$
$P =$	$\{peep, trauma, fio2, surgery, ischemicHd, minVentVol, peakAirPressure, lactate, medical, sofa, sao2, pao2, vt, oxygenation, death\}$
$Q =$	$\{bmi, ards, age, smoker, ph, spo2, copd, paco2, gender, hemoglobin, apsiiii\}$
$R =$	$\{peep, trauma, fio2, surgery, ischemicHd, minVentVol, lactate, medical, sofa, sao2, pao2, vt, oxygenation, death\}$
$S =$	$\{apsiii, hemoglobin, gender, paco2, copd, spo2, ph, smoker, age, ards, bmi\}$

The qualitative knowledge of causal relationships in the domain is represented by a causal model shown in The treatment variable is *oxygenation* and the outcome variable is *death*. We show that the causal effect  $do(oxygenation =$

oxygenation) on death given  $\{apsiii, sofa\}$ , written as  $P_{oxygenation}(death|apsiii, sofa)$ , is identifiable from a distribution over the observed variables  $P(U)$ .

## Appendix B. Derivation

**Theorem 2** *The causal effect of oxygenation on death given  $\{apsiii, sofa\}$  is identifiable from  $P(U)$  and is given by the formula*

$$P_{oxygenation}(death|apsiii, sofa) = \sum_B P(death|A) P(B)$$

**Proof**

$$P_{oxygenation}(death|apsiii, sofa) \tag{7a}$$

$$= \frac{P_{oxygenation}(death, apsiii, sofa)}{P_{oxygenation}(apsiii, sofa)} \tag{7b}$$

$$= P_{oxygenation, pao2, vt, minVentVol}(death, apsiii, sofa) \tag{7c}$$

$$= \sum_C P_{surgery, ischemicHd, minVentVol, medical, sao2, pao2, vt, oxygenation}(D) \tag{7d}$$

$$\tag{7e}$$

Eq. (7b) follows from the definition of conditional probability and Eq. (7c) from the third rule of do-calculus with the independence

$(pao2, vt, minVentVol \perp death, apsiii, sofa | oxygenation)_{G_{oxygenation, pao2, vt, minVentVol}}$  (refer to Fig. 8). Eq. (7d) follows from summing over  $\{C\}$  and Eq. (??) from C-component factorization.

**Task 1: Compute  $P_{U-\{age\}}(age)$**

$$P_{U-\{age\}}(age) \tag{8}$$

$$= P(age) \tag{9}$$

Eq. (9) follows from the third rule of do-calculus with the independence

$(U - \{age\} \perp age)_{G_{U-\{age\}}}$  (refer to Fig. 9).

**Task 2: Compute  $P_{U-\{peep\}}(peep)$**

$$P_{U-\{peep\}}(peep) \tag{10}$$

$$= P_{age}(peep) \tag{11}$$

$$= P(peep|age) \tag{12}$$

Eq. (11) follows from the third rule of do-calculus with the independence

$(U - \{age, peep\} \perp peep | age)_{G_{U-\{peep\}}}$  (refer to Fig. 10). Eq. (12) follows from the second rule of do-calculus with the independence  $(age \perp peep)_{G_{age}}$  (refer to Fig.

11).

**Task 3: Compute  $P_{U-\{gender\}}(gender)$**

$$P_{U-\{gender\}}(gender) \quad (13)$$

$$= P(gender) \quad (14)$$

Eq. (14) follows from the third rule of do-calculus with the independence  $(U - \{gender\} \perp gender)_{G_{U-\{gender\}}}$  (refer to Fig. 9).

**Task 4: Compute  $P_{U-\{SpO_2\}}(spo2)$**

$$P_{U-\{SpO_2\}}(spo2) \quad (15)$$

$$= P(spo2) \quad (16)$$

Eq. (16) follows from the third rule of do-calculus with the independence  $(U - \{SpO_2\} \perp spo2)_{G_{U-\{SpO_2\}}}$  (refer to Fig. 9).

**Task 5: Compute  $P_{U-\{copd\}}(copd)$**

$$P_{U-\{copd\}}(copd) \quad (17)$$

$$= P_{spo2,age}(copd) \quad (18)$$

$$= P(copd|age, spo2) \quad (19)$$

Eq. (18) follows from the third rule of do-calculus with the independence  $(N \perp copd|spo2, age)_{G_{U-\{copd\}}}$  (refer to Fig. 12). Eq. (19) follows from the second rule of do-calculus with the independence  $(age, spo2 \perp copd)_{G_{age,spo2}}$  (refer to Fig. 13).

**Task 6: Compute  $P_{U-\{fio2\}}(fio2)$**

$$P_{U-\{fio2\}}(fio2) \quad (20)$$

$$= P_{spo2,age,copd,peep}(fio2) \quad (21)$$

$$= P(fio2|peep, copd, age, spo2) \quad (22)$$

Eq. (21) follows from the third rule of do-calculus with the independence  $(J \perp fio2|spo2, age, copd, peep)_{G_{U-\{fio2\}}}$  (refer to Fig. 14). Eq. (22) follows from the second rule of do-calculus with the independence  $(peep, copd, age, spo2 \perp fio2)_{G_{peep,copd,age,spo2}}$  (refer to Fig. 15).

**Task 7: Compute  $P_{U-\{hemoglobin\}}(hemoglobin)$**

$$P_{U-\{hemoglobin\}}(hemoglobin) \quad (23)$$

$$= P_{spo2,gender}(hemoglobin) \quad (24)$$

$$= P(hemoglobin|gender, spo2) \quad (25)$$

Eq. (24) follows from the third rule of do-calculus with the independence  $(K \perp \text{hemoglobin} | \text{spo2}, \text{gender})_{G_{\overline{U - \{\text{hemoglobin}\}}}}$  (refer to Fig. 16). Eq. (25) follows from the second rule of do-calculus with the independence  $(\text{gender}, \text{spo2} \perp \text{hemoglobin})_{G_{\text{gender}, \text{spo2}}}$  (refer to Fig. 17).

**Task 8: Compute  $P_{U - \{\text{ph}\}}(\text{ph})$**

$$P_{U - \{\text{ph}\}}(\text{ph}) \tag{26}$$

$$= P(\text{ph}) \tag{27}$$

Eq. (27) follows from the third rule of do-calculus with the independence  $(U - \{\text{ph}\} \perp \text{ph})_{G_{\overline{U - \{\text{ph}\}}}}$  (refer to Fig. 9).

**Task 9: Compute  $P_{U - \{\text{trauma}\}}(\text{trauma})$**

$$P_{U - \{\text{trauma}\}}(\text{trauma}) \tag{28}$$

$$= P_{\text{age}, \text{peep}}(\text{trauma}) \tag{29}$$

$$= P(\text{trauma} | \text{peep}, \text{age}) \tag{30}$$

Eq. (29) follows from the third rule of do-calculus with the independence  $(L \perp \text{trauma} | \text{age}, \text{peep})_{G_{\overline{U - \{\text{trauma}\}}}}$  (refer to Fig. 18). Eq. (30) follows from the second rule of do-calculus with the independence  $(\text{peep}, \text{age} \perp \text{trauma})_{G_{\text{peep}, \text{age}}}$  (refer to Fig. 19).

**Task 10: Compute  $P_{U - \{\text{smoker}\}}(\text{smoker})$**

$$P_{U - \{\text{smoker}\}}(\text{smoker}) \tag{31}$$

$$= P_{\text{age}}(\text{smoker}) \tag{32}$$

$$= P(\text{smoker} | \text{age}) \tag{33}$$

Eq. (32) follows from the third rule of do-calculus with the independence  $(U - \{\text{age}, \text{smoker}\} \perp \text{smoker} | \text{age})_{G_{\overline{U - \{\text{smoker}\}}}}$  (refer to Fig. 20). Eq. (33) follows from the second rule of do-calculus with the independence  $(\text{age} \perp \text{smoker})_{G_{\text{age}}}$  (refer to Fig. 11).

**Task 11: Compute  $P_{U - \{\text{ards}\}}(\text{ards})$**

$$P_{U - \{\text{ards}\}}(\text{ards}) \tag{34}$$

$$= P(\text{ards}) \tag{35}$$

Eq. (35) follows from the third rule of do-calculus with the independence  $(U - \{\text{ards}\} \perp \text{ards})_{G_{\overline{U - \{\text{ards}\}}}}$  (refer to Fig. 9).



**Task 12: Compute**  $P_{U-\{bmi\}}(bmi)$

$$P_{U-\{bmi\}}(bmi) \quad (36)$$

$$= P(bmi) \quad (37)$$

Eq. (37) follows from the third rule of do-calculus with the independence  $(U - \{bmi\} \perp bmi)_{G_{U-\{bmi\}}}$  (refer to Fig. 9).

**Task 13: Compute**  $P_{U-\{paco2\}}(paco2)$

$$P_{U-\{paco2\}}(paco2) \quad (38)$$

$$= P_{bmi, ards, age, smoker, ph, spo2, copd}(paco2) \quad (39)$$

$$= P(paco2 | copd, spo2, ph, smoker, age, ards, bmi) \quad (40)$$

Eq. (39) follows from the third rule of do-calculus with the independence  $(M \perp paco2 | bmi, ards, age, smoker, ph, spo2, copd)_{G_{U-\{paco2\}}}$  (refer to Fig. 21). Eq. (40) follows from the second rule of do-calculus with the independence  $(copd, spo2, ph, smoker, age, ards, bmi \perp paco2)_{G_{\underline{copd, spo2, ph, smoker, age, ards, bmi}}}$  (refer to Fig. 22).

**Task 14: Compute**  $P_{U-\{apsiii\}}(apsiii)$

$$P_{U-\{apsiii\}}(apsiii) \quad (41)$$

$$= P_{Q-\{apsiii\}}(apsiii) \quad (42)$$

$$= P(apsiii | S - \{apsiii\}) \quad (43)$$

Eq. (42) follows from the third rule of do-calculus with the independence  $(P \perp apsiii | Q - \{apsiii\})_{G_{U-\{apsiii\}}}$  (refer to Fig. 23). Eq. (43) follows from the second rule of do-calculus with the independence  $(S - \{apsiii\} \perp apsiii)_{G_{\underline{S-\{apsiii\}}}}$  (refer to Fig. 24).

**Task 15: Compute**  $P_{U-\{peakAirPressure\}}(peakAirPressure)$

$$P_{U-\{peakAirPressure\}}(peakAirPressure) \quad (44)$$

$$= P_Q(peakAirPressure) \quad (45)$$

$$= P(peakAirPressure | S) \quad (46)$$

Eq. (45) follows from the third rule of do-calculus with the independence  $(R \perp peakAirPressure | Q)_{G_{U-\{peakAirPressure\}}}$  (refer to Fig. 25). Eq. (46) follows from the second rule of do-calculus with the independence  $(S \perp peakAirPressure)_{G_{\underline{S}}}$  (refer to Fig. 26).

**Task 16: Compute  $P_{U-\{lactate\}}(lactate)$**

$$P_{U-\{lactate\}}(lactate) \quad (47)$$

$$= P_O(lactate) \quad (48)$$

$$= P(lactate|I) \quad (49)$$

Eq. (48) follows from the third rule of do-calculus with the independence  $\left( \begin{smallmatrix} surgery, ischemicHd, minVentVol, medical, sofa, \\ sao2, pao2, vt, oxygenation, death \perp lactate | O \end{smallmatrix} \right)_{G_{U-\{lactate\}}}$  (refer to Fig. 27). Eq. (49) follows from the second rule of do-calculus with the independence  $(I \perp lactate)_{G_I}$  (refer to Fig. 28).

**Task 17: Compute  $P_{U-\{death\}}(death)$**

$$P_{U-\{death\}}(death) \quad (50)$$

$$= P_F(death) \quad (51)$$

$$= P(death|H) \quad (52)$$

Eq. (51) follows from the third rule of do-calculus with the independence  $(surgery, ischemicHd, medical, sao2 \perp death | F)_{G_{U-\{death\}}}$  (refer to Fig. 29). Eq. (52) follows from the second rule of do-calculus with the independence  $(H \perp death)_{G_H}$  (refer to Fig. 30).

**Task 18: Compute  $P_{U-\{sofa\}}(sofa)$**

$$P_{U-\{sofa\}}(sofa) \quad (53)$$

$$= P_E(sofa) \quad (54)$$

$$= P(sofa|G) \quad (55)$$

Eq. (54) follows from the third rule of do-calculus with the independence  $\left( \begin{smallmatrix} minVentVol, pao2, vt, oxygenation, death, surgery, \\ ischemicHd, medical, sao2 \perp sofa | E \end{smallmatrix} \right)_{G_{U-\{sofa\}}}$  (refer to Fig. 31). Eq. (55) follows from the second rule of do-calculus with the independence  $(G \perp sofa)_{G_G}$  (refer to Fig. 32).

Substituting Eq. (9), Eq. (12), Eq. (14), Eq. (16), Eq. (19), Eq. (22), Eq. (25), Eq. (27), Eq. (30), Eq. (33), Eq. (35), Eq. (37), Eq. (40), Eq. (43), Eq. (46), Eq. (49), Eq. (52), and Eq. (55) back into Eq. (??), we get

$$P_{oxygenation}(death|apsiii, sofa) = \sum_B P(death|A) P(B) \quad (56)$$

■

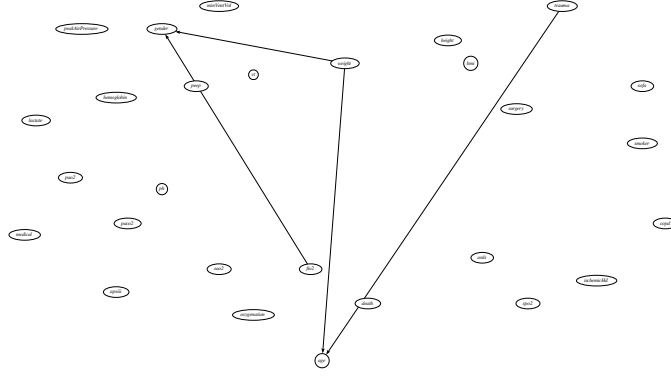


Figure 8: Causal Graph  $G_{\text{oxygenation, pao2, vt, minVentVol}}$ .

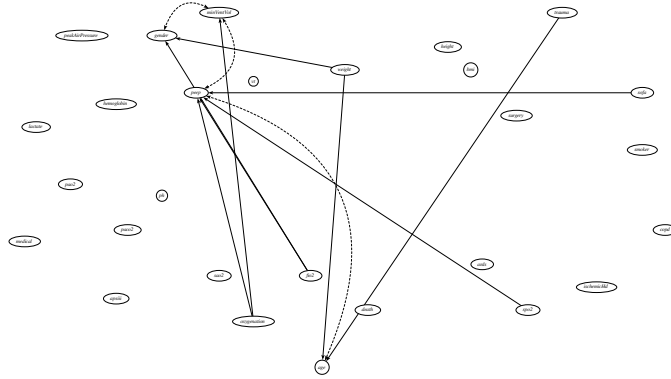


Figure 9: Causal Graph  $G_{U-\{age\}}, G_{U-\{gender\}}, G_{U-\{spo2\}}, G_{U-\{ph\}}, G_{U-\{ards\}}, G_{U-\{bmi\}}$ .

## Appendix C. Figures

The subgraphs used in the derivation of the causal effect of *oxygenation* on *death* given  $\{apsiii, sofa\}$  are as follows:

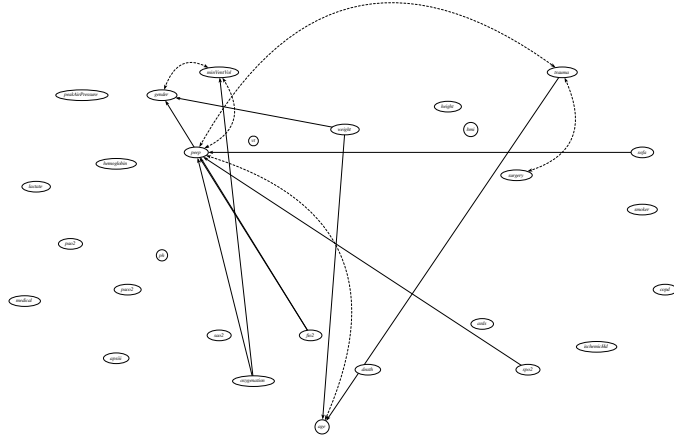


Figure 10: Causal Graph  $G_{U-\{peep\}}$ .

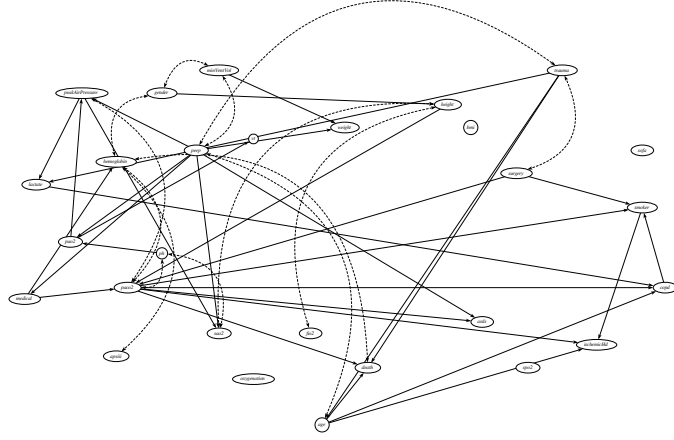


Figure 11: Causal Graph  $G_{\underline{age}}$ .

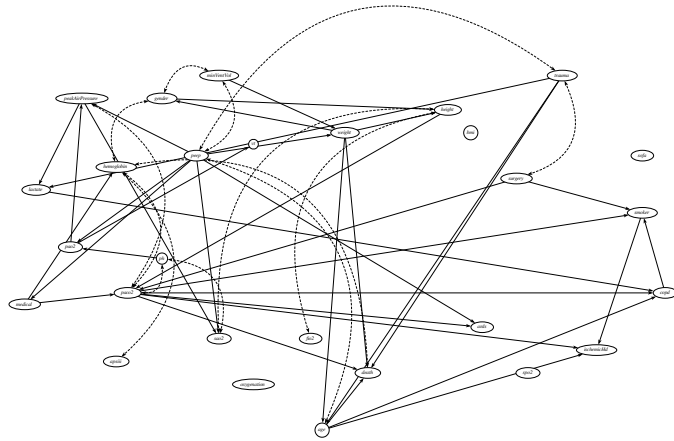


Figure 12: Causal Graph  $G_{U-\{copd\}}$ .

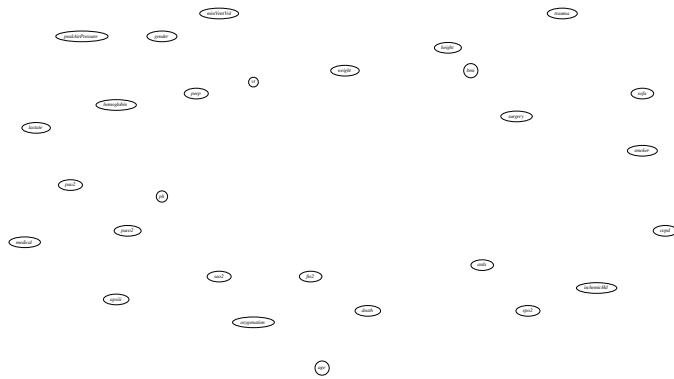


Figure 13: Causal Graph  $G_{age,spo2}$ .

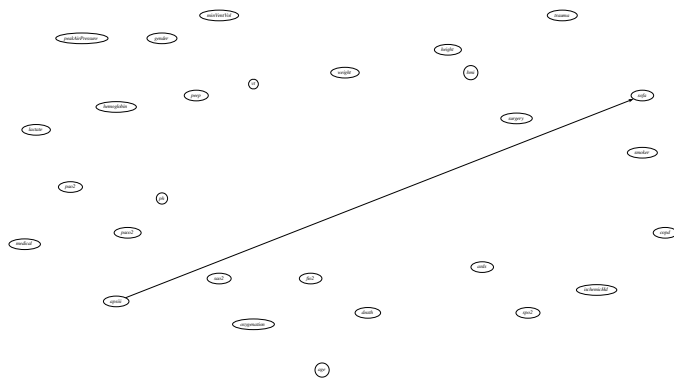


Figure 14: Causal Graph  $G_{U-\{fio2\}}$ .

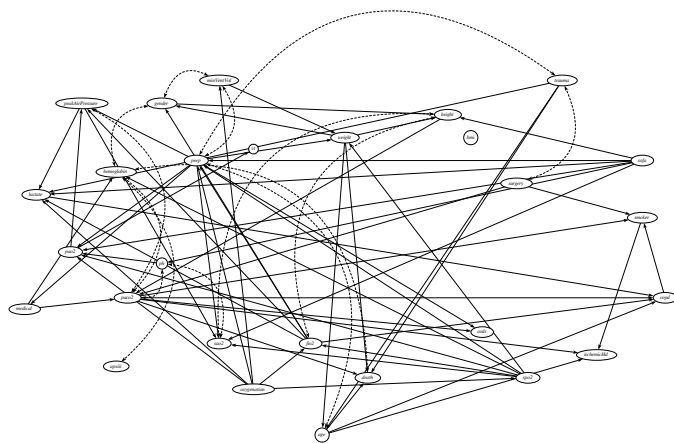


Figure 15: Causal Graph  $G_{\underline{peep, copd, age, spo2}}$ .

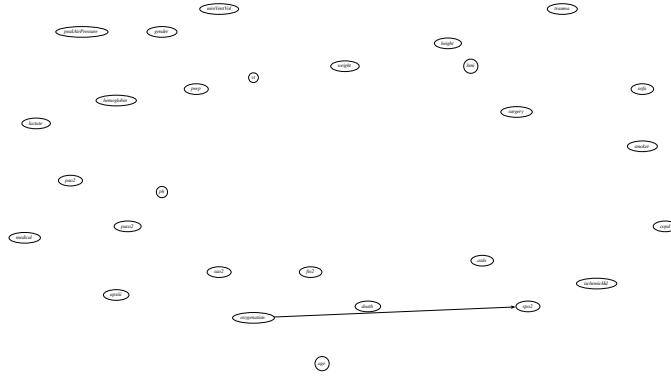


Figure 16: Causal Graph  $G_{U-\{hemoglobin\}}$ .

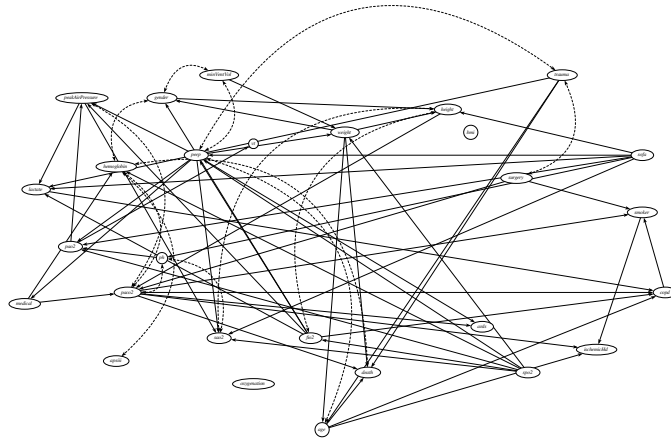


Figure 17: Causal Graph  $G_{gender,spo2}$ .

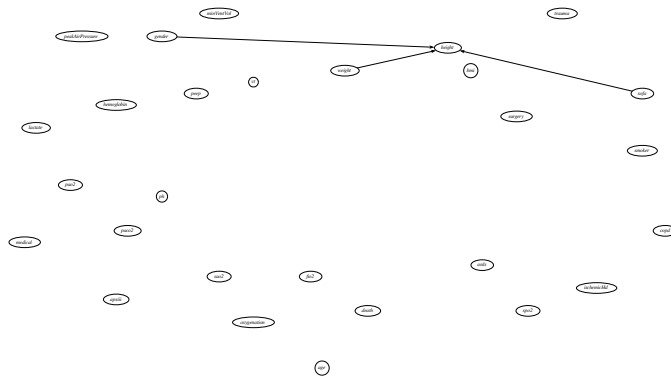


Figure 18: Causal Graph  $G_{U-\{trauma\}}$ .

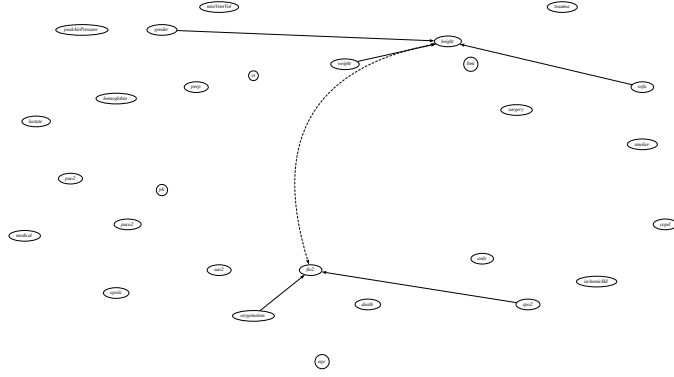


Figure 19: Causal Graph  $G_{\underline{peep}, \underline{age}}$ .

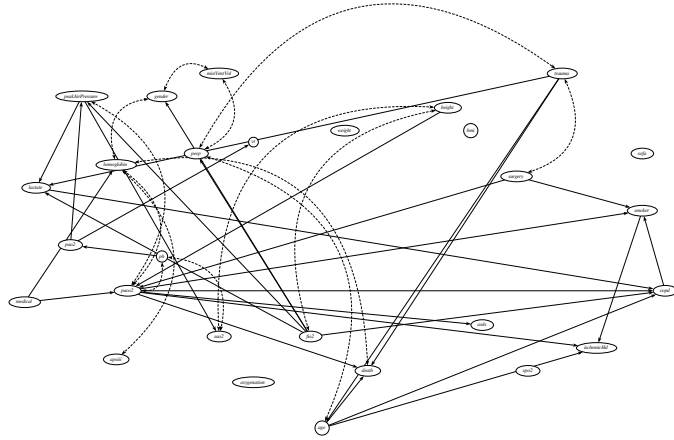


Figure 20: Causal Graph  $G_{\overline{U - \{smoker\}}}$ .

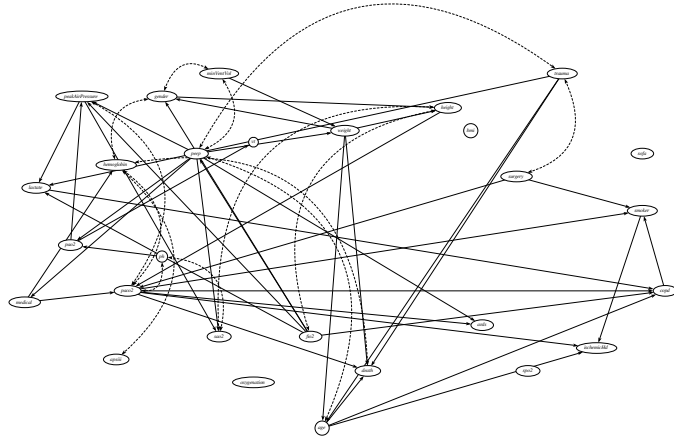


Figure 21: Causal Graph  $G_{\overline{U - \{paco2\}}}$ .

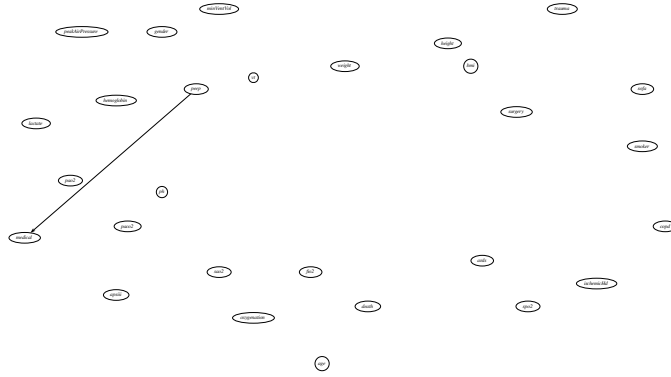


Figure 22: Causal Graph  $G_{\underline{copd,spo2,ph,smoker,age,ards,bmi}}$ .

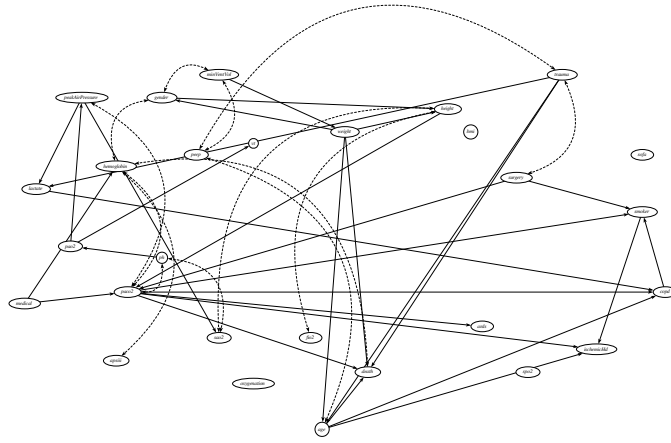


Figure 23: Causal Graph  $G_{U-\{apsiii\}}$ .

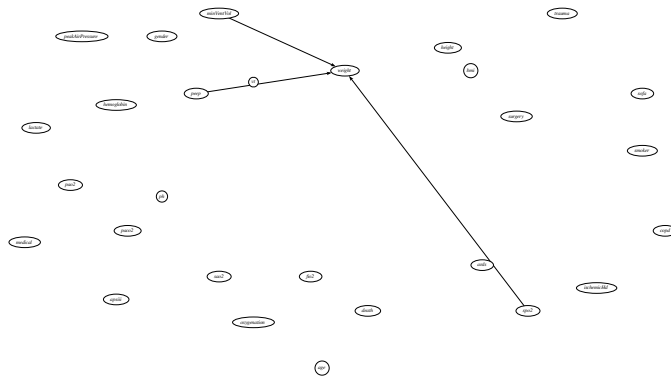


Figure 24: Causal Graph  $G_{\underline{S-\{apsiii\}}}$ .



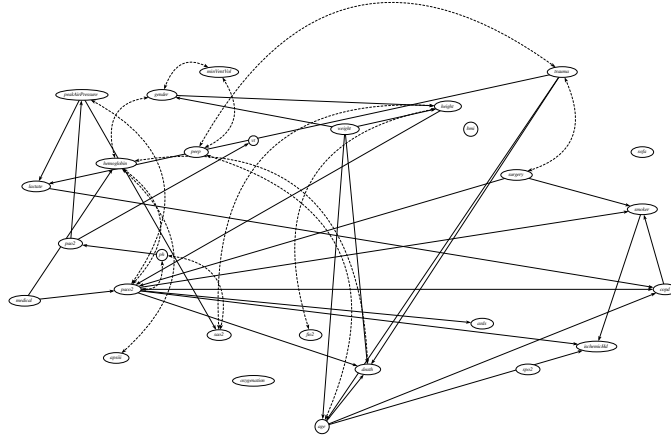


Figure 25: Causal Graph  $G_{\overline{U - \{peakAirPressure\}}}$ .

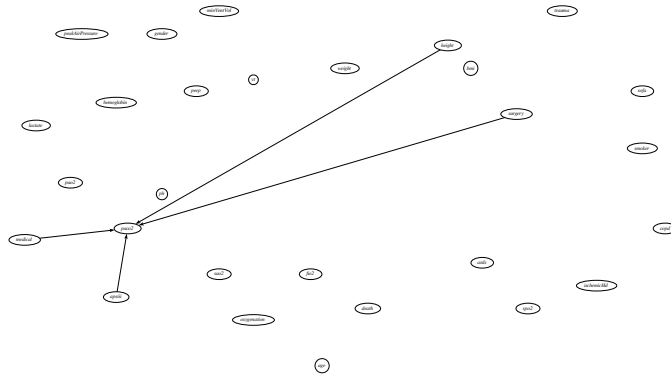


Figure 26: Causal Graph  $G_{\underline{S}}$ .

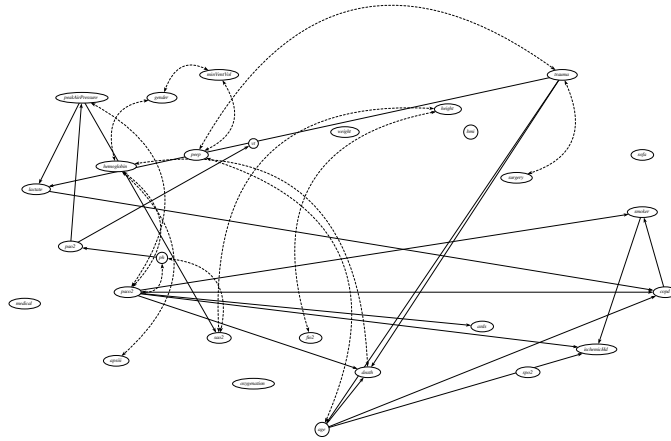


Figure 27: Causal Graph  $G_{\overline{U - \{lactate\}}}$ .

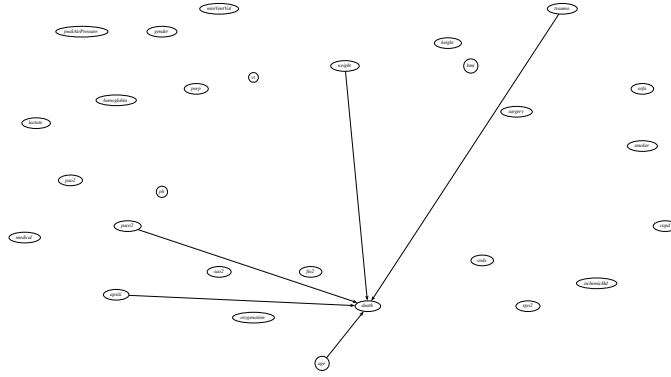


Figure 28: Causal Graph  $G_I$ .

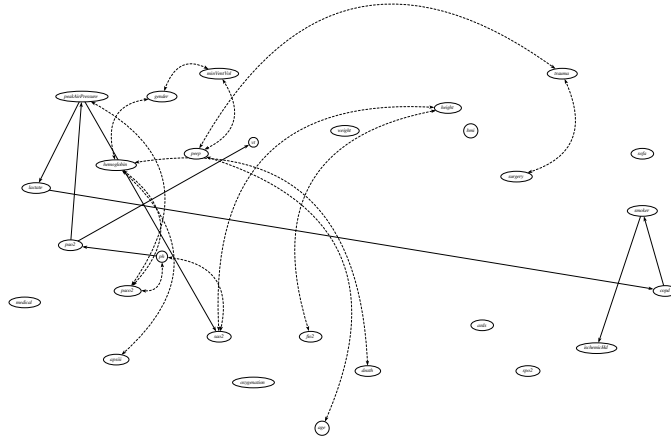


Figure 29: Causal Graph  $G_{U - \{death\}}$ .

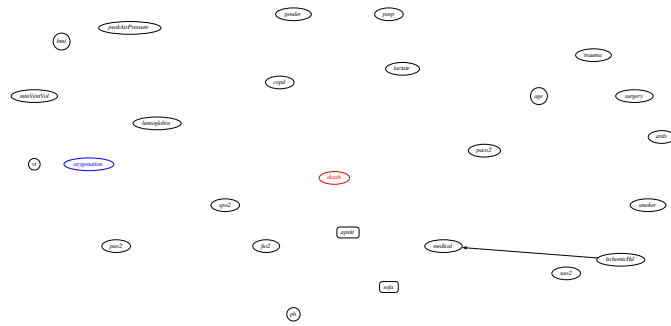


Figure 30: Causal Graph  $G_H$ .

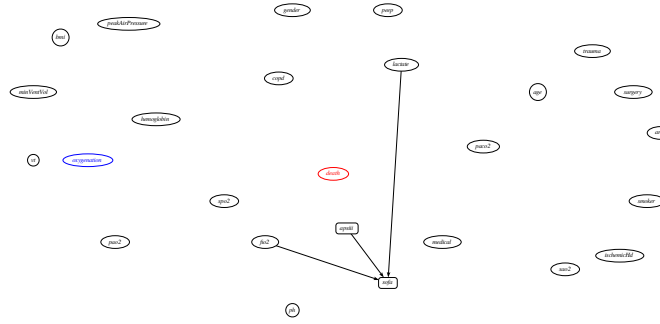


Figure 31: Causal Graph  $G_{U-\{sofa\}}$ .

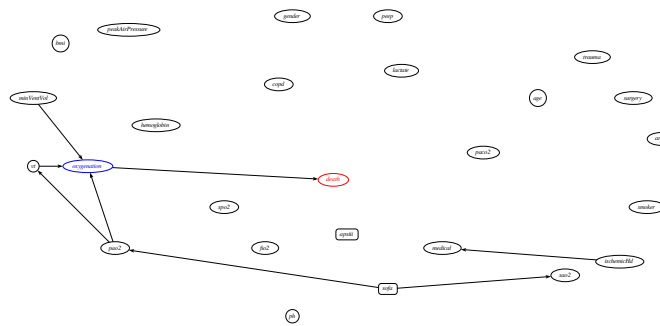


Figure 32: Causal Graph  $G_{\underline{G}}$ .