Design and Validation of Structural Causal Model: A focus on SENSE-EGRA Datasets

Gabriel Terna Ayem[a, 1, *], Augustine Shey Nsang [a, 2], Bernard Igoche Igoche [b, 3], Garba Naankang [c, 4]

[a] Computer Science Department, School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria.
[b] Computer Science Department, School of Computing, University of Portsmouth, United Kingdom.
[c] True-life Engineering Department, School of Aerospace, Transport, and Manufacturing, Cranfield University, United Kingdom.
[1] gabriel.ayem@aun.edu.ng, [2] augustines.nsang@aun.edu.ng, [3] bern.igoche@port.ac.uk, [4] naankang.garba@cranfield.ac.uk
* corresponding author

ARTICLE INFO

Article history
Received August 25, 2023
Revised September 15, 2023
Accepted October 15, 2023

Keywords
Causality
structural causal validation
model assumption
observational datasets
model testing.

ABSTRACT

Designing and validation of causal model correctness from a dataset whose background knowledge is gotten from a research process is not a common phenomenon. In fact, studies have shown that in many critical areas such as healthcare and education, researchers develop models from direct acyclic graphs without testing them. This phenomenon is worrisome and is bound to cast a dark shadow on the inference estimates that many arise from such models. In this study, we have design a novel application-based SCM for the first time using the background knowledge gotten from the American university of Nigeria (AUN), Yola, on the letter identification subtask of Early Grade reading Assessment (EGRA) program on Strengthen Education in Northeast Nigeria (SENSE-EGRA) project dataset, which was sponsored by the USAID. We employed the conditional independence test (CIT) criteria for the model's correctness validation testing, and the results shows a near perfect SCM.

1. Introduction

From time immemorial till date, human actions, processes and indeed scientific explorations have been predicated on the premise of cause and effect. In the primordial era, the savaged and primitive man sought ways to articulate and uncover this phenomenon of cause and effects; and not having equipment, enough facts or the sine-quo-non to ascertain this phenomenon of knowing what actions (causes) that produces the effects especially in incidences that were agonizing to him such as certain ebullitions of some sickness's concomitant with mysterious deaths. Thus, the ability to know the right action to influence his environment or predict his future made man an idiosyncratic specie from the rest of the animals. Thus, driving the savaged man from his initial state of higgledy-piggledy to embrace the practice of magic, astrology, and certain fetish ways to achieve the causation phenomenon in order to overcome his bewildered state. Gradually, as societies evolved and advanced, and mankind himself advanced from his primitive and savaged state to his current state of scientific and technological advancement. Thus, establishing his hegemony on earth over and above every other specie; the same motives of trying to influence his environment and predict his future still stands. Nonetheless the methods of achieving it have evolved; as magic arts wanes to scientific logic, and astrology metamorphosed to astronomy and other technological innovations such as computer predictions, simulations etc., became the modern genies that are aberrations from the fetish ways of predicting the future. Albeit, in this current era, the science of trying to ascertain causality or causation in human processes and actions is still a daunting and a nontrivial task; as the traditional scientific way of ascertaining this act is resident with the randomized controlled experiment or randomized controlled trial (RCT) method. This RCT method and idea is credited to Fisher [1]. Thus, this standard framework for causal discovery known as RCT always involves setting some (usually half) of the sampled population of study and given them a treatment (an intervention) under the same conditions, while the second half of the study population is left untreated (not intervened on) or controlled under the same or similar conditions, in order to slay any possible confounding or lurking variable, which is often the factor that jeopardizes a proper juxtaposition of these two sampled population in the RCT experiments. As fascinating as this method of RCT is, there are events and circumstances that makes this kind of experiments too expensive, infeasible or even unethical to perform. A good instance is to perform a RCT on a hypothesize query that seeks to uncover the health benefits, or otherwise of smoking on a certain population. This is an unethical experiment to conduct under RCT, because it would involve setting half of the population under review to smoke (treated) and the other not to smoke (controlled). Hence, with this obstacles posed by RCT, many researchers have resorted to the discovery and inferring of causal structures from purely observational dataset, or from a combination of both data and RCT [2, 3].

However, in spite of the successes recorded by causal models using observational dataset, many causal models designs are not tested or validated for correctness as far as the extant literature would reveal. In fact, a recent study by Tennant, Murray [4], that investigated model testing in healthcare sector, revealed that among the 200 articles reviewed, not a single one of them was tested or validated for correctness. Thus, if these models are to be further used in the estimation or evaluation of causal inference of such projects, the estimation results may leave a lot of room for dispute and doubt. Thus, in this study, we have designed an application-based novel SCM from the background knowledge gotten from the American university of Nigeria (AUN), Yola’ s project on the letter identification subtask of Early Grade reading Assessment (EGRA) program on Strengthen Education in Northeast Nigeria (SENSE), which was sponsored by the United State Agency for International Development (USAID), which occurred between 2021 to 2202. We employed the conditional independence test (CIT) criteria for the testing of our SENSE-EGRA SCM correctness and the results shows a near perfect model. See Table 1.

1.1 Study Contributions

The main contributions of this work is as follows:

(i) Theoretical insight into structural causal model (SCM) framework,

(ii) Development of an application-based novel SCM for the SENSE-EGRA dataset

(iii) Model correctness validation using the conditional independent test (CIT) criteria

(iv) Experiment Reproducibility. See appendix links to data and CIT codes for reproduction of the experiment.

1.2 Study Structure

In section 2, the basic theoretical concept of causal model is discussed. Section 3 discusses direct acyclic graphs and their relations to causality and the Bayesian network factorization. Section 4 presents some of the main assumptions driving SCMs. Section 5 presents our experiment setup as its relates the design of our SENSE-EGRA SCM. Section 6 presents our model correctness validation testing results using the CIT criteria. And finally, section 7 wraps up the study and give direction on future work.

2. Basic Concept of Causal Models

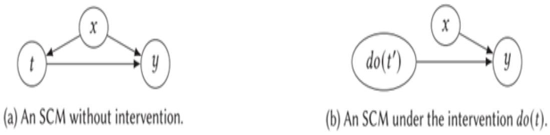
In this section, the various forms of causality are defined, followed by the two major framework used for causality, which are the structural causal model (SCM) framework and the potential outcome or Rubin causal model (RCM) framework; with a juxtaposition of both frameworks. The section concludes with how causal interventions are executed in dataset with the SCM framework.

2.1 Causal Model: It is an abstraction of mathematics that describes quantitatively the relations of causality that exist among variables in an observable dataset [5]. These mathematical models are derived from the domain and background knowledge embodied in the DAG, and they evince the causal relations within the observable dataset [6-8].

2.2 Types of Causal Models: Two types of casual models exists for causality, which are (i) Structural causal model (SCM) proposed by Pearl [7] and (ii) Potential outcome framework also called Rubin causal model (RCM) [9, 10]. However, the study scope is limited to the SCM and not the RCM.

2.2.1 An SCM: The framework for causality based on SCM gives a holistic understanding of the theory of cause and effect. It is composed of two parts: the causal diagram (or graph) that encodes background domain knowledge and assumptions of the distribution (the dataset), and the Bayesian network factorization (BNF) or structural equations part, which models or algorithmised (mathematically) the relations among the study variables based on the causal assumptions from the graph [5, 11-13]. This works focuses more on the SCM with a more detail explication of the connections of the graphs and the dataset in subsequent sections.

2.3 Causal Relations with SCM: Determining the causal relations that exists among variables in an observational study in a purely probabilistic distribution is an ambiguous and daunting task. If a conditional probability distribution such as $P(Y|X)$ for instance, represent the conditional probability distribution of obesity (Y) given a particular level of sugar intake (X). This distribution relation is ambiguous in terms of an experimental setting (RCT) where sugar intake was ascertained by randomization or by merely through an observational process. In his book on causality, Pearl [7] in order to differentiate the mere conditional observational probability distribution (I,e., statistical association/correlation) and interventional conditional probability distribution (which is a causal association), introduced the do-operator of the do-calculus to differentiate interventional distribution from observational' . Hence, the expression $P(Y|X)$ can now be regarded as mere conditional observational association which depict how the probability of Y (obesity) will change, if someone were to observe the sugar intake (X). While $P(Y|do(X = x))$ is regarded as the interventional conditional probability distribution (which is a causal association), depicting the probability of obesity (Y) given that a measure unit of sugar (x) were taken (purposefully and not observed). Hence, making the observation and intervention distinct: $P(Y|X = x) \neq P\big(Y\big|do(X = x)\big)$. The practical difference between the two may be the existence of a variable(s) Z (individual gene tar for instance) that may be confounding the relations, which exists in some back-door path: See figure 1 DAG for confounding relations. In the intervention distribution, the causal effects is determined given difference values of the treatment/control X (i.e., when sugar is taken and when sugar is not taken) and this can be measured and compared in the interventional distribution, written as: $P\big(Y\big|do(x = 1)\big)$, and $P\big(Y\big|do(x = 0)\big)$ where 1 and 0 stands for treatment and no treatment (control) respectively for an individual instance, which is called the *individual treatment effect* (ITE). Thus, when this process involves all sampled or all instances of the population, the causal intervention is defined in terms of the average treatment effects (ATE) for the instances of the population. Written in terms of the expectation as: $\tau(1,0) = E[Y|do(x = 1)] - E[Y|do(x = 0)]$. Also, conditional average treatment effects (CATE) can be taken for subpopulation group in a similar manner as well. Thus, it can be seen that this kinds of intervention model' s the RCT experiment that determines causality in observational dataset [14, 15]. In spite of the clear distinction describing and differentiating these two processes by Pearl [7], not every dataset can be neatly categorized into this distinction of observational and interventional dataset, as some experiments may not clearly or wholly show the value of the variable that is intervened on in the dataset. Thus, due to this two distinctions, which are obfuscated in the distributions, it has become imperative to represent causal models explicitly in terms of directed acyclic graph (DAG) or simply causal graph as proposed by Pearl [14]. Causal graph in SCM are very essential component which make it easier to identify the causality from dataset; hence, we discuss them in the next section.



(a) An SCM without intervention.

(b) An SCM under the intervention $do(t)$.

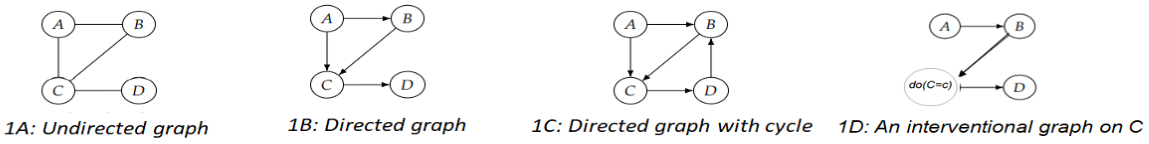
Figure 1. Shows a SCM with (b) intervention and without (a) an intervention

3. Causal Graph

In circuit design, the economic value of the components used must be taken into consideration. Before creating the circuit and system, a block diagram is first designed. This is to achieve the goal of having a circuit that leads to the desired outcome, as illustrated in the block diagram shown in the figure 4.

This section presents causal graphs as is applicable in SCM. Fundamental concepts in graph such as the popular backdoor adjustment criteria and the Bayesian network factorization (BNF) are elicited and explicated.

3.1 Causal Graph Composition: A causal graph (denoted as $G = (V, E)$, consists of two or more nodes (also called vertices) representing a random variable sets $(V), where\ V = X_1,\ X_2, X_3, \dots X_n$ and a number of connecting lines among the nodes called edges (E). These random variables may include the observed and unobserved (if the exists) variable alongside the treatment and outcome variables. In figure 2: 1A is an undirected graph due to the lack of directional arrows on them. While 1B the graph is directed because of the arrow direction. And 1C shows a directed graph with a cycle [16] and finally 1D shows and intervention graph on variable C. A directed edge from A to B (written as: A → B) is interpreted as, B is caused by A or (A is the potential cause of B) [5]. Hence, with a causal graph an hypothesized causal query can clearly be modelled through the causal pathways in the graph, and all dependent/independent relations as it relates all variables associated with the query are known. And this graph model can be factorized using the Bayesian network factorization or the structural equations; based on some assumptions to obtain a causal estimand of the conditional probability distribution from which it can be used with the observed dataset to ascertain the causal estimate of the hypothesized query [14, 17].



1A: Undirected graph

1B: Directed graph

1C: Directed graph with cycle

1D: An interventional graph on C

Figure 2. shows an undirected, directed, directed with cycle, and intervention graph

A ***path*** in the graph is an oriented order of adjacent edges irrespective of the direction of the adjoining nodes. For instance, $A - C - B$ is considered as a path in figure 2 1A and $A \longrightarrow C \longleftarrow B$ is also a path in figure 2 1B. A directed path is that in which all edges are directed or pointing in the same direction. E.g., the path, $A \longrightarrow C \rightarrow B$ in figure 2 1B is regarded as directed. Most causal algorithms work best with the directed acyclic graphs (DAGs) condition as shown in figure 2 1B and a few causal algorithms work with the cyclic graph condition as shown in figure 2 1C [5, 11-13].

3.2 Three Cardinal Relations in Graphs: A descendant of a node A is a node $C \in V$, such that there is direct edge from A to C (written as: $A \longrightarrow C$) in the DAG G. This corresponds to A being an ancestor (parent of) C. The progenies (A and B) of a node C, are the nodes in V with a directed edges connecting C, (designated as: $A \longrightarrow C \longleftarrow B$). This child and two parents relationship designated as $A \longrightarrow C \longleftarrow B$, is also called a ***collider*** [18, 19] or ***immorality*** [8, 16] is the first basic relation that can exists among variables represented in DAG. A second relation exists called a ***mediator or chain***, where a parent node A (usually exogenous) that produces a child node C, where C in turn produces another child B (which is a grand descendant of A) [8, 14, 17]. finally, a third relationship exists where a node C, which is a single parent having two descendants A and B (written as: $A \longleftarrow C \longrightarrow B$) is called a ***fork*** or common cause confounder. Thus, these three relations (collider, chain/mediator and fork) are the three common relations that exist in an observational dataset and can be mirrored or expressed in a DAG, forming the building block or structure in causal graph for determining relationship (causal or associational) in an observational settings [8, 11, 14, 17, 20, 21].

3.3 Causal Connection & the Backdoor Adjustment Criteria in a Graph: *D-separation* and *d-connection* are the processes that define a sets of variable V's connectivity in a causal graph G [21]. The D in the d-separation and d-connection stands for *dependency* and it is a process of establishing independency or dependency from two or more variables that are independent or otherwise on a third variable C in in a DAG which is a reflection in the dataset. For instance, in the case of a fork ($A \longleftarrow C \longrightarrow B$), or a chain/mediator ($A \longrightarrow C \rightarrow B$), the variable C is a link between both A and B. Hence, once you condition on the linking variable C, you will block or close the dependency relationship that exist between paths A and B. That is to say, paths A and B will become independent conditioned on C, written as: $A \coprod B|C$. Albeit the reverse is the case, when it comes to the collider or immorality structure ($A \longrightarrow C \longleftarrow B$), as the paths A and B are already independent or blocked in their current state (i.e., $A \coprod B \nmid C$: A is independent of B not conditioned on C), without the need for conditioning on any variable including C. Hence, once you condition on C, a relationship between A and B is induced (i.e., A and B becomes dependent conditioned on C. written as: $A \coprod B|C$). This process of blocking the flow of unwanted association on non-causal pathways in order to determine causality only through a causal pathway is called the ***backdoor adjustment criteria*** [22, 23]. Pearl [21], defined the process of d-separation and d-connection for backdoor adjustment criteria in a DAG G formally as follows: A path connecting two variables A and B is said to be d-separated or blocked if and only if: (i) the path contains a fork such as : ($A \longleftarrow C \longrightarrow B$) or chain/mediator such as: ($A \longrightarrow C \rightarrow B$) that has been conditioned on C. Written as: (A$\coprod_G$ B|C), and (ii) the path between A and B contain a collider on C, such as ($A \longrightarrow C \longleftarrow B$) that has not been conditioned on, alongside any descendant of collider C, that is not conditioned on as well. Written as: ($A \coprod_G B \nmid C$) or just $A \coprod_G B$. This same process of d-separation and the backdoor adjustment criteria from the graph G can be utilized to determine dependencies/independencies of variables in the distribution (or dataset), which is a factorization of the d-separation in the graph using the Bayesian Network Factorization (BNF). The d-separation in the distribution is written as: A$\coprod_p$ B|C, or A$\coprod_p$ B|C for independency and dependency conditions respectively, similar to the d-separation in the graph with the subscript $_{\mathbf{P}}$ to distinguish it from the graph's d-separation criteria, which is represented by the subscript $_{\mathbf{G}}$. This can further be used to determine causal relations in the distribution as whole.

On the other hand, a path from A and B through C, is said to be ***d-connected***, unblocked or open when it is not d-separated [17, 21].

3.4 The Bayesian Network Factorization (BNF) in Graphs: The DAGs are interpreted in two part. i.e., the probabilistic and the causal interpretations. The probabilistic inference sees the directional arrows on the DAG G as showing a probabilistic dependences or associations among the variables of study, while the lack of arrows corresponds to the conditional independence asserted by the study variables [24]. Based on some assumptions, the simplest being the *Markovian condition*, which states that each study variable is considered independent of all its non-descendants in the graph with the exception of its direct parent. Usually written as $A\coprod B|C$. Hence, based on the assumption, the joint probability distribution function $P(v) = P(v_i, \ldots, v_n)$ factorizes based on the BNF as:

$$P(v) = \prod_i^n P(v_i|pa_i) \quad (1)$$

Where $v_i = 1, \ldots, n, and\ pa_i$ denotes the parent of the variable $v_i$in the graph

Thus, based on the BNF of equation (1), the graph in figure 2:1B for instance, the probability distribution of it (i.e.,1B) can be factorized and summarized, based on the Markov assumption as follows:

$$P(A,B,C) = P(A)P(B|A)P(C|B,A)P(D|C) \quad (2)$$

This contrasts the normal Bayesian probability distribution network which uses the chain rule without the graph and the Markov assumption, written as:

$$P(A,B,C) = P(A)P(B|A)P(C|B,A)P(D|C,B,A) \quad (3)$$

The difference in equation (2) and (3) is in the last product conditional probability of D, where equation (2) reduces the conditioning probability to only its immediate parent node C, based on the position of equation (1) and as captured in the graph of figure 2:1B. While equation (3) assumes no graph and factorizes the distribution using the chain rule. Hence, the probability of D, given (or conditioned on:) C, B and A are used as elicited in equation (3).

3.5 Causal Identifiability with BNF Intervention Graphs: The second interpretation of the graph is called a causal interpretation. In this scenario, the arrows direction in the DAG G represents the influence of causality among the variables. Here the BNF of equation (1) above is still essential, but the arrows are assumed to evince a separate process in the data generated. Hence, after eliciting causal path from the DAG G, the conditional probability of the distribution $P(v_i|pa_i)$ which is generated based on the graph G, and which is a statistical estimand, can be estimated from the data. The relations of conditional dependency expressed by the BNF formula of equation (1) does not necessarily leads to causal inference (due to the mixtures of confounding variables sometimes). However equation (1) can be extended to cater for interventions (which are causal in their implementation) as presented by Pearl in [7]. Using the do-operator of the do-calculus as an intervention on the desired variable (or node) the difference between mere conditional distribution (correction), written as: $P(Y|X = x)$, and the causal intervention of the conditional distribution, written as: $P(Y|do(X = x))$, in the graph and subsequently the data can be clearly distinguished. For instance, if the graph in figure 2, were derived from the query hypothesis of determining the effects of shoe size X on the reading ability Y of children. The age variable Z, confounds the relationship between reading ability Y and shoe size X, making them to have statistical correlation as shown in figure 1(a). But when you carry out an intervention on the shoe size X such as $P(Y|do(X = x))$, the age variable Z that confounds the relations is severed, and the conditional probability of the BNF produces an estimand which is given as $P\big(Y\big|do(X = x)\big) = P(Z)P(X|Z)P(Y|Z,X)$. Which is summarized by getting rid of the factor for probability of X in the BNF to get: $P\big(Y\big|do(X = x)\big) = \sum_z P(Y|Z,X)\, P(Z)$. With this causal intervention estimand, using the d-separation and the backdoor criteria, the shoe size X will be set to a treatment unit of 1 and no treatment (control) unit of 0, while conditioning on a certain age Z say 8years. Thus, the difference between the treatment and no treatment of shoe size $(X{:}\, 0,1)$ generated from conditioning on a certain age $(Z = 8)$ for the set of $\boldsymbol{Z}$ variable in the dataset can be calculated as the ATE, given mathematically in terms of their expectation as: $\tau(1,0) = E[Y|do(x = 1)] - E[Y|do(x = 0)]$, which translate to the causal estimate or causal inference estimation on the effect of shoe size X on reading ability Y in children. This estimate would likely be zero (no effect), thus killing the lurking variable (age) and exposing the spurious association (correlation) that exists between shoe size X and reading ability Y. Note however that if the confounding variable Z is unobserved or not part of the distribution (the dataset), the causal identification of X on Y cannot be feasible to obtain in the data, even though it is revealed in the graph. This do-operator which translate to intervention and causality in data differentiates mere association (correlation) that is used in machine learning algorithms.

With SCM, counterfactual hypothesized queries which are carried out on an individual level of the sampled dataset can also be estimated, using some techniques proposed by Pearl [25, 26] which transcend the do-operator of the do-calculus, which only work with i.i.d condition [27]. Although counterfactual causal effects would not be covered in this work.

4. Assumptions in SCM

The highest carbon monoxide (CO) levels were found in Arabica Special coffee, with an ADC level of 662 carbon monoxide gas. The lowest carbon monoxide (CO) levels were detected in Liberica coffee, with an ADC level of 105 carbon monoxide gas. The Artificial Neural Network method can identify Liberica coffee with a success rate of 98%, Arabica coffee with 100%, and Robusta coffee with 98%.

This section covers the three major assumptions often used for causality, especially with i.i.d datasets, thus driving the process of causality in observational data setting with the SCM framework. These assumptions are: (i) The Markov assumption, (ii) The Acyclicity assumption (iii) The causal sufficiency assumption. These assumptions are summarized as follows:

4.1 The Markov Assumption: This assumption states that, a parent node in a DAG G representing a variable is considered independent of all its non-descendant in the graph with the exception of its direct parent. This assumption ensures that causal estimand for the identification of the causal relations is generated from the graph to the data, using the BNF or the structural equation of functional causal model (FCM). This estimand which is modeled using the Markov condition when it is sufficient (i.e., all confounding variable identified), becomes the basis for which the probability distribution, which is a statistical estimand can be estimated from the dataset. Equation (1) is a representation of the Markov condition. The Markov assumption when combined with the causal edge assumption that states that: in a DAG G, all adjacent nodes are dependent; can generally be referred to as the *minimality assumption* [10, 16, 28].

4.2 The Acyclicity Assumption: It is the phenomenon that ensures that the set of adjoining variables nodes $\boldsymbol{V}$ in the causal graph does not form a cycle, a feedback loop or go back in time as shown in figure 2:1C, but are rather directed and acyclic as shown in figure 2:1B [29, 30].

4.3 The Causal Sufficient Assumption: This condition states that in a given causal graph G, there are no variables confounding relationships that is unobserved among the study variables. That is to say, the causal sufficiency assumption ensures that all variables that may be confounding or having a hidden effect on the hypothesized query variable of treatment and outcome (t, y) are identified and explicitly shown on the graph, whether or not they are observed in the distribution of the dataset [31-33]. Hence, these are the assumption that are employed in the development of our SENSE-EGRA SCM.

5. Experiment Design

The EGRA (SENSE) dataset on the subtask of letter identification for grade 2 students of two northeast states of Nigeria under study is made of 1114 records, collected from a population of over 200 primary schools in the said states. 19 columns are of interest for our design of the SCM and analysis. These columns are further grouped into 5 distinctive groups which are: A set of input features or covariates (X) where X stands for $State, LGA, Gender, Age$ etc.; the output feature

LI_3 (Y), the treatment variable T $(Treament)$ and two other assessment or evaluation features (LI_1, and LI_2) respectively. See the appendix for more details on the dataset-encoded meanings.

Thus, based on the above-discussed methodology in section 2, we designed the EGRA- SENSE SCM of figure 3, and validated for correctness with the dataset as shown in equation 4 below, and the result is presented in Table 1:

$$LI\_1 \perp X|LI\_2, T$$

$$LI\_2 \perp T|X$$

$$LI\_3(Y) \perp T|LI\_1, LI_2,$$

$$LI\_3(Y) \perp X|LI\_2, T$$

$$LI\_3(Y) \perp X|LI\_1, LI\_2, \quad (4)$$

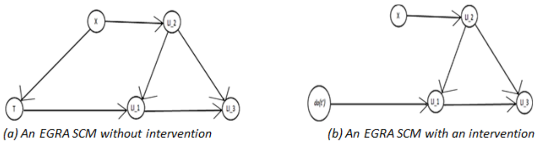
Thus, the estimand and the back-door adjustment criteria, which identified the admissible set of covariates required for adjustment in our EGRA- SENSE SCM, as shown in figure 3(a) is given as:

$$\mathrm{P(T, X, LI\_2, LI\_3) = P(LI\_3|X, LI\_2, T)} \quad (5)$$

And the corresponding NPSEM generated from mutilated DAG as shown in figure 3(b) for our SENS-EGRA SCM designating an intervention distribution is given as:

$$x = \mathrm{f}_x(U_x),\ t = t',\ li_2 = \mathrm{f}_{li_2}(x, U_{li_2}), li_1 = \mathrm{f}_{li_2}(t, U_{li_1}), li\_3 = \mathrm{f}_{li_3}(li_1, li\_2, U_{li\_3}) \quad (6)$$

Notice that LI_1 is not conditioned on, since from the DAG, it is considered a post-treatment or mediator variable. Pearl et.al [7, 24, 25, 34], advised against conditioning on such post-treatment or mediator variables. Section 6, presents the result of the conditional independence test (CIT) implemented in an *R* package called Daggity [35]



(a) An EGRA SCM without intervention

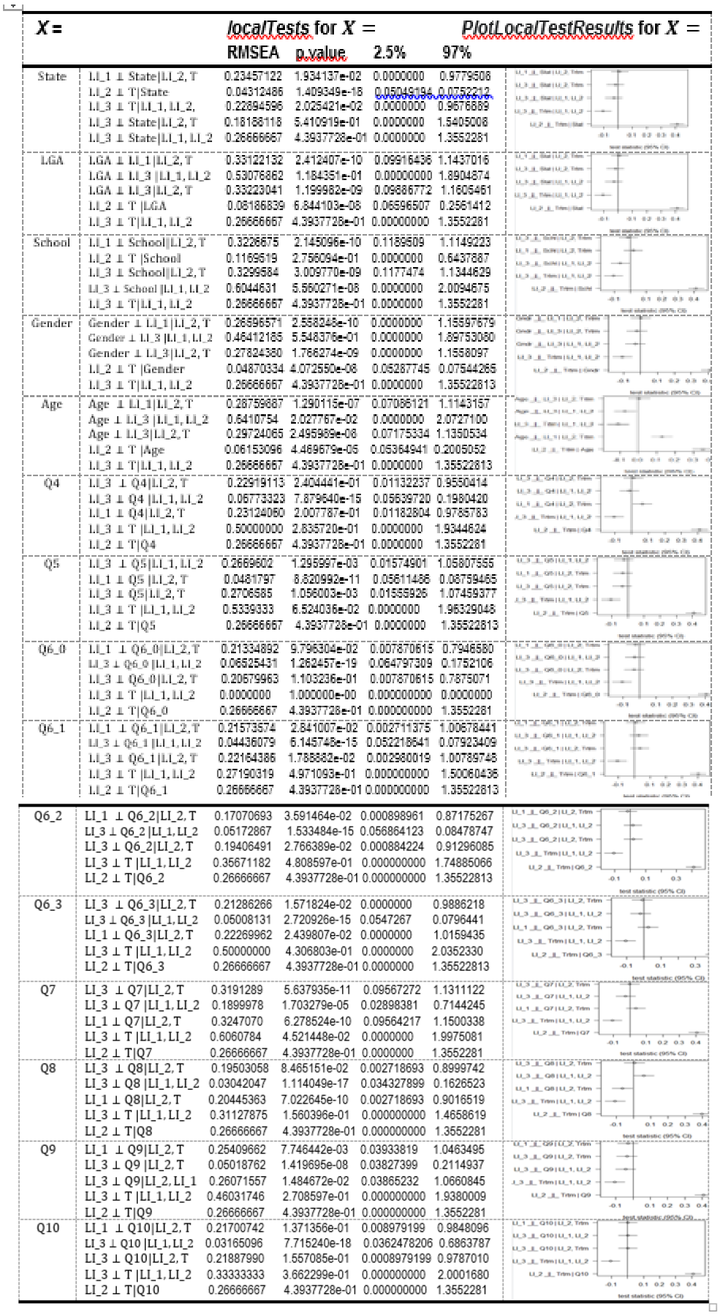
(b) An EGRA SCM with an intervention

Figure 3. Shows our EGRA- SENSE- SCM with (b) and without (a) intervention

6. Results Presentation and Discussion

6.1 CIT Results for SENSE-EGRA SCM

SCM is a qualitative process that is subjective based on background knowledge. Hence, experts advise validation and testing of the model with the dataset to ensure its correctness [4, 24, 25, 36-39]. One of the most pervasive validation tests for SCM is the use of conditional independence testing (CIT) criteria [24, 25, 36-39]. Thus, once the validation process is over, the adjustment criteria can be applied to the SCM. Pearl et al.[24, 25, 37], proposed two adjustment criteria (the backdoor and front-door) depending on the structure of the SCMs in a concept called the d- separation (dependency separation). This concept when properly applied to the SCM is sufficient to identify the estimand (mathematics formula) for adjusting covariates and estimating the causal impact of the intervention. For our experiment, we implemented the CIT using the identified conditional independencies set of equation 4 and applied the back-door adjustment criteria for eliminating confounding bias as shown in equations 5 and 6 respectively. Table 1 below shows the result of the CIT performed on the dataset in order to verify and validate the correctness of our EGRA- SENSE SCM, implemented in the R package tool of [35].



X =		localTests for X =				PlotLocalTestResults for X =
		RMSEA	p.value	2.5%	97%	
State	LI_1 ⊥ State|LI_2, T	0.23457122	1.934137e-02	0.0000000	0.9779508	
	LI_2 ⊥ T|State	0.04312486	1.409349e-18	0.05049194	0.0752212	
	LI_3 ⊥ T|LI_1, LI_2,	0.22894596	2.025421e-02	0.0000000	0.9676889	
	LI_3 ⊥ State|LI_2, T	0.18188118	5.410919e-01	0.0000000	1.5405008	
	LI_3 ⊥ State|LI_1, LI_2	0.26666667	4.3937728e-01	0.0000000	1.3552281	
LGA	LGA ⊥ LI_1 |LI_2, T	0.33122132	2.412407e-10	0.09916436	1.1437016	
	LGA ⊥ LI_3 |LI_1, LI_2	0.53076862	1.184351e-01	0.00000000	1.8904874	
	LGA ⊥ LI_3|LI_2, T	0.33223041	1.199982e-09	0.09886772	1.1605461	
	LI_2 ⊥ T |LGA	0.08186839	6.844103e-08	0.06596507	0.2561412	
	LI_3 ⊥ T|LI_1, LI_2	0.26666667	4.3937728e-01	0.00000000	1.3552281	
School	LI_1 ⊥ School|LI_2, T	0.3226675	2.145096e-10	0.1189609	1.1149223	
	LI_2 ⊥ T |School	0.1169519	2.756094e-01	0.0000000	0.6437887	
	LI_3 ⊥ School|LI_2, T	0.3299584	3.009770e-09	0.1177474	1.1344629	
	LI_3 ⊥ School |LI_1, LI_2	0.6044631	5.560271e-08	0.0000000	2.0094675	
	LI_3 ⊥ T|LI_1, LI_2	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Gender	Gender ⊥ LI_1|LI_2, T	0.26596571	2.558246e-10	0.0000000	1.15597679	
	Gender ⊥ LI_3 |LI_1, LI_2	0.46412165	5.548376e-01	0.0000000	1.89753080	
	Gender ⊥ LI_3|LI_2, T	0.27824380	1.766274e-09	0.0000000	1.1558097	
	LI_2 ⊥ T |Gender	0.04870334	4.072550e-08	0.05287745	0.07544265	
	LI_3 ⊥ T|LI_1, LI_2	0.26666667	4.3937728e-01	0.0000000	1.35522813	
Age	Age ⊥ LI_1|LI_2, T	0.28759687	1.290115e-07	0.07086121	1.1143157	
	Age ⊥ LI_3 |LI_1, LI_2	0.6410754	2.027767e-02	0.0000000	2.0727100	
	Age ⊥ LI_3|LI_2, T	0.29724065	2.495969e-08	0.07175334	1.1350534	
	LI_2 ⊥ T |Age	0.06153096	4.469679e-05	0.05364941	0.2005052	
	LI_3 ⊥ T|LI_1, LI_2	0.26666667	4.3937728e-01	0.0000000	1.35522813	
Q4	LI_3 ⊥ Q4|LI_2, T	0.22919113	2.404441e-01	0.01132237	0.9550414	
	LI_3 ⊥ Q4 |LI_1, LI_2	0.06773323	7.879640e-15	0.05639720	0.1980420	
	LI_1 ⊥ Q4|LI_2, T	0.23124060	2.007787e-01	0.01182804	0.9785783	
	LI_3 ⊥ T |LI_1, LI_2	0.50000000	2.835720e-01	0.0000000	1.9344624	
	LI_2 ⊥ T|Q4	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Q5	LI_3 ⊥ Q5|LI_1, LI_2	0.2669602	1.295997e-03	0.01574901	1.05807555	
	LI_1 ⊥ Q5 |LI_2, T	0.0481797	8.820992e-11	0.05611486	0.06759465	
	LI_3 ⊥ Q5|LI_2, T	0.2706585	1.056003e-03	0.01555926	1.07459377	
	LI_3 ⊥ T |LI_1, LI_2	0.5339333	6.524036e-02	0.0000000	1.96329046	
	LI_2 ⊥ T|Q5	0.26666667	4.3937728e-01	0.0000000	1.35522813	
Q6_0	LI_1 ⊥ Q6_0|LI_2, T	0.21334892	9.796304e-02	0.007870615	0.7946580	
	LI_3 ⊥ Q6_0 |LI_1, LI_2	0.06525431	1.262457e-19	0.064797309	0.1752106	
	LI_3 ⊥ Q6_0|LI_2, T	0.20679963	1.103236e-01	0.007870615	0.7875071	
	LI_3 ⊥ T |LI_1, LI_2	0.0000000	1.000000e-00	0.000000000	0.0000000	
	LI_2 ⊥ T|Q6_0	0.26666667	4.3937728e-01	0.000000000	1.3552281	
Q6_1	LI_1 ⊥ Q6_1|LI_2, T	0.21573574	2.841007e-02	0.002711375	1.00678441	
	LI_3 ⊥ Q6_1 |LI_1, LI_2	0.04436079	6.145748e-15	0.052218641	0.07923409	
	LI_3 ⊥ Q6_1|LI_2, T	0.22164386	1.788882e-02	0.002980019	1.00789746	
	LI_3 ⊥ T |LI_1, LI_2	0.27190319	4.971093e-01	0.000000000	1.50060436	
	LI_2 ⊥ T|Q6_1	0.26666667	4.3937728e-01	0.000000000	1.35522813	
Q6_2	LI_1 ⊥ Q6_2|LI_2, T	0.17070693	3.591464e-02	0.000898961	0.87175267	
	LI_3 ⊥ Q6_2 |LI_1, LI_2	0.05172867	1.533484e-15	0.056864123	0.08478747	
	LI_3 ⊥ Q6_2|LI_2, T	0.19406491	2.766389e-02	0.000884224	0.91296085	
	LI_3 ⊥ T |LI_1, LI_2	0.35671182	4.808597e-01	0.000000000	1.74885066	
	LI_2 ⊥ T|Q6_2	0.26666667	4.3937728e-01	0.000000000	1.35522813	
Q6_3	LI_3 ⊥ Q6_3|LI_2, T	0.21286266	1.571824e-02	0.0000000	0.9886218	
	LI_3 ⊥ Q6_3 |LI_1, LI_2	0.05008131	2.720926e-15	0.0547267	0.0796441	
	LI_1 ⊥ Q6_3|LI_2, T	0.22269962	2.439807e-02	0.0000000	1.0159435	
	LI_3 ⊥ T |LI_1, LI_2	0.50000000	4.306803e-01	0.0000000	2.0352330	
	LI_2 ⊥ T|Q6_3	0.26666667	4.3937728e-01	0.0000000	1.35522813	
Q7	LI_3 ⊥ Q7|LI_2, T	0.3191289	5.637935e-11	0.09567272	1.1311122	
	LI_3 ⊥ Q7 |LI_1, LI_2	0.1899978	1.703279e-05	0.02898381	0.7144245	
	LI_1 ⊥ Q7|LI_2, T	0.3247070	6.278524e-10	0.09564217	1.1500338	
	LI_3 ⊥ T |LI_1, LI_2	0.6060784	4.521448e-02	0.0000000	1.9975081	
	LI_2 ⊥ T|Q7	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Q8	LI_3 ⊥ Q8|LI_2, T	0.19503058	8.465151e-02	0.002718693	0.8999742	
	LI_3 ⊥ Q8 |LI_1, LI_2	0.03042047	1.114049e-17	0.034327899	0.1626523	
	LI_1 ⊥ Q8|LI_2, T	0.20445363	7.022645e-10	0.002718693	0.9016519	
	LI_3 ⊥ T |LI_1, LI_2	0.31127875	1.560396e-01	0.000000000	1.4658619	
	LI_2 ⊥ T|Q8	0.26666667	4.3937728e-01	0.000000000	1.3552281	
Q9	LI_1 ⊥ Q9|LI_2, T	0.25409662	7.746442e-03	0.03933819	1.0463495	
	LI_3 ⊥ Q9 |LI_2, T	0.05018762	1.419695e-08	0.03827399	0.2114937	
	LI_3 ⊥ Q9|LI_2, LI_1	0.26071557	1.484672e-02	0.03865232	1.0660845	
	LI_3 ⊥ T |LI_1, LI_2	0.46031746	2.708597e-01	0.000000000	1.9380009	
	LI_2 ⊥ T|Q9	0.26666667	4.3937728e-01	0.000000000	1.3552281	
Q10	LI_1 ⊥ Q10|LI_2, T	0.21700742	1.371356e-01	0.008979199	0.9848096	
	LI_3 ⊥ Q10 |LI_1, LI_2	0.03165096	7.715240e-18	0.0362478206	0.6863787	
	LI_3 ⊥ Q10|LI_2, T	0.21887990	1.557085e-01	0.0008979199	0.9787010	
	LI_3 ⊥ T |LI_1, LI_2	0.33333333	3.662299e-01	0.000000000	2.0001680	
	LI_2 ⊥ T|Q10	0.26666667	4.3937728e-01	0.000000000	1.3552281	

Figure 4. Shows the result of the CIT identified in Equation 10 for each of the variables X

6.2 CIT Results Discussion

When testing for conditional independence, between two or more variables, it is required that their conditional dependency be zero [35]. Hence with the use of the R tool of Ankur [35], as used in this work, the root mean square error of approximation (RMSEA) and the p-value results that are close to zero (our p-value threshold is set at 0.05) validate the assumptions evinced by the SCM. While the values of the RMSEA and p-value that deviates significantly from zero or that are statically significant, reveal the model’ s inaccuracy or lack of conditional dependency among them.

Thus the R tool produced by Ankur [35], package functions *LocalTests()* and the *PlotLocalTestResults()* are used for the analysis of the CIT. The *LocalTests()* tests the CIT for each of the feature variables X under the five conditional independence conditions identified in our EGRA- SENSE SCM of equation 4 for variable $X = State, LGA, Gender, Age$ etc., at a confidence interval of 95% for all test cases as shown in column 2 of table 1. While the *PlotLocalTestResults()* function *plots the results of the localTests() function* as shown in column 3. All the results indicate negative p-values and zero-scale RMSEA values. Thus, validating the correctness of our EGRA-SENSE SCM, as no conditional dependency exceeds 0.4 in all test cases as shown in *PlotLocalTestResults()* graphical output in column 3 of the table; meaning their dependence is nearly zero.

7. Conclusion and Future Work

7.1 Conclusion

In this study, we have designed a novel application-based SCM from the background knowledge gotten from the American university of Nigeria (AUN), Yola's project on the letter identification subtask of Early Grade reading Assessment program on Strengthen Education in Northeast Nigeria (SENSE-EGRA), which was sponsored by the United State Agency for International Development (USAID), which occurred between 2021 to 2202. We employed the conditional independence test (CIT) criteria for the testing and validation of the models 'correctness, and the results shows a near perfect model. The main contribution of this work is in the explication of the theoretical insight into the structural causal model (SCM) framework, the development and correctness validation testing of an application-based novel SCM for the SENSE-EGRA dataset.

7.2 Future Work

For future works, we shall use the developed SENSE-EGRA SCM alongside some adjustment and matching estimation techniques, such as ordinary least square regression adjustment, propensity score by (weighting, stratification and matching) to deal with confounding and selection biases in order to estimate the causal inference of SENSE-EGRA intervention program of the American University of Nigeria, Yola, Adamawa State, Nigeria under the sponsorship of USAID.

Conflict of Interest Declaration

All authors have no financial or proprietary interests in any material discussed in this work.

Appendix

SENSE-EGRA Dataset/Composition/CIT Codes

The entire materials needed for the reproduction of this study can be accessed on our GitHub page at: https://github.com/Sadaju-Codes/SENSE-EGRA_Project.git

References

[1] Box, J.F., *RA Fisher, the Life of a Scientist.* Revue Philosophique de la France Et de l, 1980. **170**(4).

[2] Benson, K. and A.J. Hartz, *A comparison of observational studies and randomized, controlled trials.* New England Journal of Medicine, 2000. **342**(25): p. 1878-1886.

[3] Silverman, S.L., *From randomized controlled trials to observational studies.* The American journal of medicine, 2009. **122**(2): p. 114-120.

[4] Tennant, P.W., et al., *Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations.* International journal of epidemiology, 2021. **50**(2): p. 620-632.

[5] Guo, R., et al., *A survey of learning causality with data: Problems and methods.* ACM Computing Surveys (CSUR), 2020. **53**(4): p. 1-37.

[6] Hitchcock, C. and M. Rédei, *Reichenbach's common cause principle.* 2020.

[7] Pearl, J., *Causality*. 2009: Cambridge university press.

[8] Peters, J., D. Janzing, and B. Schölkopf, *Elements of causal inference: foundations and learning algorithms*. 2017: The MIT Press.

[9] Neyman, J., *Sur les applications de la theorie des probabilites aux experiences agricoles: essai des principes (Masters Thesis); Justification of applications of the calculus of probabilities to the solutions of certain questions in agricultural experimentation. Excerpts English translation (Reprinted).* Stat Sci, 1923. **5**: p. 463-472.

[10] Rubin, D.B., *Estimating causal effects of treatments in randomized and nonrandomized studies.* Journal of educational Psychology, 1974. **66**(5): p. 688.

[11] Spirtes, P., C. Glymour, and R. Scheines, *Discovery algorithms for causally sufficient structures*, in *Causation, prediction, and search*. 1993, Springer. p. 103-162.

[12] Greenland, S., J. Pearl, and J.M. Robins, *Causal diagrams for epidemiologic research.* Epidemiology, 1999: p. 37-48.

[13] Lauritzen, S.L., *Causal Inference from.* Complex stochastic systems, 2000: p. 63.

[14] Eberhardt, F., *Introduction to the foundations of causal discovery.* International Journal of Data Science and Analytics, 2017. **3**(2): p. 81-91.

[15] Halpern, J.Y., *The Book of Why, Judea Pearl, Basic Books (2018)*. 2019, Elsevier.

[16] Neal, B., *Introduction to causal inference from a machine learning perspective.* Course Lecture Notes (draft), 2020.

[17] Elwert, F., *Graphical causal models*, in *Handbook of causal analysis for social research*. 2013, Springer. p. 245-273.

[18] Nogueira, A.R., et al., *Methods and tools for causal discovery and causal inference.* Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2022: p. e1449.

[19] Yao, L., et al., *A survey on causal inference.* ACM Transactions on Knowledge Discovery from Data (TKDD), 2021. **15**(5): p. 1-46.

[20] Glymour, C., K. Zhang, and P. Spirtes, *Review of causal discovery methods based on graphical models.* Frontiers in genetics, 2019. **10**: p. 524.

[21] Pearl, J., *Probabilistic reasoning in intelligent systems: networks of plausible inference*. 1988: Morgan kaufmann.

[22] Gultchin, L., et al. *Differentiable causal backdoor discovery*. at the International *Conference on Artificial Intelligence and Statistics*. 2020. PMLR.

[23] Correa, J. and E. Bareinboim. *Causal effect identification by adjustment under confounding and selection biases*. in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2017.

[24] Tian, J. and J. Pearl, *A general identification condition for causal effects*. 2002: eScholarship, University of California.

[25] Pearl, J. and D. Mackenzie, *The book of why: the new science of cause and effect*. 2018: Basic books.

[26] Pearl, J., *Theoretical impediments to machine learning with seven sparks from the causal revolution.* arXiv preprint arXiv:1801.04016, 2018.

[27] Richardson, T.S. and J.M. Robins. *Single world intervention graphs: a primer*. in *Second UAI workshop on causal structure learning, Bellevue, Washington*. 2013. Citeseer.

[28] Zhang, J. and P. Spirtes, *Intervention, determinism, and the causal minimality condition.* Synthese, 2011. **182**(3): p. 335-347.

[29] Hauser, A. and P. Bühlmann, *Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs.* The Journal of Machine Learning Research, 2012. **13**(1): p. 2409-2464.

[30] Hauser, A. and P. Bühlmann, *Jointly interventional and observational data: estimation of interventional Markov equivalence classes of directed acyclic graphs.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2015. **77**(1): p. 291-318.

[31] Mayrhofer, R. and M.R. Waldmann, *Sufficiency and necessity assumptions in causal structure induction.* Cognitive science, 2016. **40**(8): p. 2137-2150.

[32] Zhang, J. and W. Mayer, *Weakening faithfulness: some heuristic causal discovery algorithms.* International journal of data science and analytics, 2017. **3**(2): p. 93-104.

[33] Zhang, J. and P.L. Spirtes, *Strong faithfulness and uniform consistency in causal inference.* arXiv preprint arXiv:1212.2506, 2012.

[34] Shpitser, I., T. VanderWeele, and J.M. Robins, *On the validity of covariate adjustment for estimating causal effects.* arXiv preprint arXiv:1203.3515, 2012.

[35] Ankan, A., I.M. Wortel, and J. Textor, *Testing graphical causal models using the R package "dagitty".* Current Protocols, 2021. **1**(2): p. e45.

[36] Thoemmes, F., Y. Rosseel, and J. Textor, *Local fit evaluation of structural equation models using graphical criteria.* Psychological methods, 2018. **23**(1): p. 27.

[37] Pearl, J. and T.S. Verma, *A theory of inferred causation*, in *Studies in Logic and the Foundations of Mathematics*. 1995, Elsevier. p. 789-811.

[38] Pearl, J. and E. Bareinboim. *Transportability of causal and statistical relations: A formal approach*. in *Twenty-fifth AAAI conference on artificial intelligence*. 2011.

[39] Pearl, J., *Causal inference in statistics: An overview.* Statistics surveys, 2009. **3**: p. 96-146.