Bayesian Analysis (2007) **2**, Number 3, pp. 557–590

Cluster Allocation Design Networks

Ana Maria Madrigal*

Abstract. When planning and designing a policy intervention and evaluation, it is important to differentiate between (future) policy interventions we want to evaluate, F_T, affecting "the world", and experimental allocations, A_T, affecting "our picture of the world". The policy maker usually has to define a strategy that involves policy assignment and recording mechanisms that will affect the (conditional independence) structure of the data available. Causal inference is sensitive to the specification of these mechanisms. Influence diagrams have been used for causal reasoning within a Bayesian decision-theoretic framework that introduces interventions as decision nodes (Dawid 2002). Design Networks expand this framework by including experimental design decision nodes (Madrigal and Smith 2004). They provide semantics to discuss how a design decision strategy (such as a cluster randomised study) might assist the identification of intervention causal effects. The Design Network framework is extended to Cluster Allocation. It is used to assess identifiability when the experimental unit's level is different from the analysis unit's level, and to discuss the evaluation of cluster- and individual-level future policies. Cases of 'pure' cluster (all individuals in a cluster receiving the same intervention) and 'non-pure' cluster (only a subset receiving the policy) are discussed in terms of causal effects. The representation and analysis of a simplified version of a Mexican social policy programme to alleviate poverty (Progresa) is performed as an illustration of the use of Bayesian hierarchical models to make causal inferences relating to household and community level interventions.

Keywords: Cluster allocation, Influence diagrams, Causal inference, Identification of policy effects, DAGs

1 Introduction

Different data sets provide different types of information. Different queries might require different information to obtain answers. When using data for learning, it is important to consider the conditions and circumstances under which the data were collected. The distributions that can be learnt (or not) might vary among apparently similar data sets. This is an important consideration to the analyst before learning model parameters. Consider the case in which we have two data sets that contain records of whether or not children in a population take food nutrition supplements (FS) and whether or not they have gained weight. The first data set comes from a census sample, and the second comes from an experiment where half the children were given food supplements and half of them were not. Suppose we are interested in learning the prevalence of children taking supplements in the population, $p(FS)$. It is clear that learning from the second data set that $p(FS) = 0.5$ only reflects an experimental choice and not a population prevalence, as would be the case if we were to use the first data set (e.g. showing how

*University of Warwick, UK mailto:am.madrigal@warwickgrad.net

parents 'naturally' choose to give FS to children). The collecting strategy defines the structure of the data set. A perfect design of experiments will give a more organised layout of the data. An observational study might be more 'disorganised', as data is allowed to arise naturally. In this paper, we focus our attention on a particular type of query related to policy intervention evaluations and discuss which data set structures do or do not let us answer causal queries and extract appropriate causal effects (e.g. the causal effect of FS on children's weight). Discussions about the identifiability of causal effects have been usually phrased as 'Is the causal effect of T on Y identifiable?' (see Pearl 2000; Lauritzen 2001; Dawid 2002). In this paper, the role of data structures is made explicit by phrasing the identifiability question as 'Is the causal effect of T on Y identifiable *from the data available*?'.

Intervention has to do with 'perturbing' the dynamics of a system. If we say that a system consists of components which influence each other and that its dynamics describe the way these components interact with each other in an equilibrium state, some examples of systems might be consumption-expenditure patterns, road traffic in a town or the human body. The system at present has some *pre-intervention* dynamics attached to it. When we intervene a system, by introducing a promotion-advertisement campaign, by adding a red light at a corner, or by giving medicine, we are introducing a new component into a system that will imply new *post-intervention* dynamics. The intervention might have both qualitative effects, modifying the structure of the system (maybe by 'blocking' the interaction between two of its components), and quantitative effects, modifying the value of the components. One of the main interests consists in describing if and how the intervention affects the system. Evaluation of the intervention effects is required and it is usually measured in terms of a response variable, such as sales-awareness, number of accidents, or health condition.

Discussions of causal reasoning have been made usually assuming that the graph representing the system implicitly includes the underlying (experimental) mechanism that is generating the data (see Pearl 2000). Then, in this fixed 'natural' or 'idle' system, whether the future policy intervention F_T effect is identifiable and can be obtained is evaluated. Randomised allocation of treatments to units is a well known practice within medical clinical trials but, because of ethical, social and financial issues, complete randomisation within an experiment designed to evaluate a social policy will usually be unfeasible. Knowing the details of the policy assignment mechanism and a well-planned recording of the data become very relevant issues in order to obtain all the information needed to measure the right 'causal' effects (see Rubin 1978). Influence Diagrams (IDs) are used to represent the system dynamics and interventions graphically; a review of the main features of the framework used is made in Section 2. Our interpretation of causal effects for interventions is Bayesian decision-theoretic, where an intervention on a system is regarded as a decision. Dawid (2002)'s extended influence diagrams are augmented by including 'experimental design' decisions nodes within the set of intervention strategies to create what we call a Design Network (DN), to provide semantics to discuss how a 'design' strategy (such as clustering) might assist the systematic identification of intervention causal effects, to give a taxonomy for design decisions, and to show how these decisions might alter the graphical (conditional independence) struc-

ture used to evaluate the causal effect of policy F_T. It is maintained that experimental design decisions are intrinsic to any causal analysis of policy intervention strategies. Design Networks were introduced in Madrigal and Smith (2004) for random allocation, and their main characteristics are presented in Section 3. Design Networks for cluster allocation are discussed in Section 4; the propositions can be derived from the discussion in Appendix A.

This research was motivated by a Mexican Social Policy Programme (Progresa) whose objective is to alleviate poverty. It consists of a three-stage mechanism to target its eligible population, based on community and household characteristics. The policy involves a collection of interventions at different levels (community, household and individual). All households are recipients of the community-level interventions (e.g. health infrastructure and services). Actions at household and individual level (e.g. extra monetary support and nutritional supplements) affect only 'poor' (eligible) households and vary according to household/individual demographics, so not all units in the community (cluster) are intervened equally. This motivates the discussion about the data structure arising from a cluster allocation, the distinction of 'overall' and 'total' effects, the differences in the inference of cluster- and individual-level interventions, and the nested structures in design and analysis. The design of the study included a *randomised* cluster allocation for treatment and control communities. To illustrate some features of causal analysis in a cluster allocation setting, this paper presents in Section 5, a hierarchical model analysis based on Spiegelhalter (2001). Formulation is performed for the evaluation of cluster- and individual-level interventions based on Progresa data.

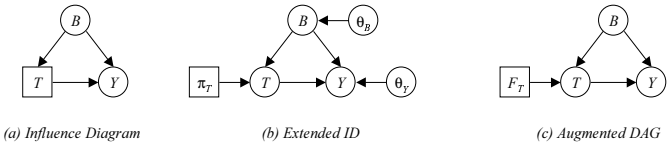
2 Intervention Graphical Framework and Causal Inference

Influence diagrams (IDs) have been used for over 20 years to form the framework for both describing (see Howard and Matheson 1981; Oliver and Smith 1990) and also devising efficient algorithms to calculate the effects of decisions (see Jensen 2001) in complex systems which implicitly embody strong conditional independence assertions. However, it is only recently that they have been used to explain causal relationships (Dawid 2000, 2002), and been shown to be much more versatile than Causal Bayesian Networks (Pearl 1993, 1995).

The simplest form of external intervention is when a single variable X is forced to take on some fixed value x'. This is known as an 'atomic intervention' and, following Pearl (2000), it is denoted by $do(X = x')$. The atomic intervention replaces the original mechanism: $p(x \mid pa(x))$ by $p(x \mid pa(x); do(X = x')) = 1$ **if** $X = x'$ where $pa(x)$ denotes the parent nodes of X. This conditioning by intervention formula has appeared in various forms (see Pearl 1993; Spirtes et al. 2000; Robins 1986). It cannot be asserted in general that the effect of *setting* the value of X to x' is the same as the effect of *observing* $X = x'$. Only in limited circumstances (as when the node for X has no parents in the graph) will conditioning by intervention and conditioning by observation coincide. Graphically, interventions are represented by deleting the arrows that enter

the intervened node in the original graph, making explicit the fact that when the value is set externally, the parents' values are not relevant post-intervention. Pearl's $do(\cdot)$ corresponds to an external intervention. By recognising interventions as decisions, the Bayesian decision-theoretical framework embeds Pearl's *doing* operation and provides a stronger framework for causal inference. The strong links between decision theory and Pearl's causal model have been discussed by Heckerman and Shachter (2003). Those who are familiar with Bayesian decision theory will find comfort, as I have, in these connections.

Dawid (2002) points out that, traditionally, in IDs conditional distributions are given for random nodes, but no description is supplied of the functions or distributions involved at the decision nodes, which are left arbitrarily at the choice of the decision maker. If we choose to provide some descriptions of the decision rules, then any given specification of the functions or distributions at decision nodes constitutes a decision strategy, π. Decisions determine what we may term the partial distribution, p, of random nodes given decision nodes which is not in general the same as the associated conditional distributions (see Cowell et al. 1999, section 2.3). If E and D denote the set of random events and the set of decisions, respectively, then the full joint specification p_π, consisting of decision strategy π and partial distribution p for all $e \in E$ and $d \in D$ is given by $p_\pi(e, d)$. The graphical representation of p_π can be made by using *extended* IDs that incorporate non-random parameter nodes ($\theta_e = p(e \mid pa^0(e))$ and strategy nodes $\left(\pi_d = \pi(d \mid pa^0(d))\right)$ representing the mechanisms that generate random and decision nodes respectively. Here, $pa^0(.)$ denotes the set of domain parents of X (i.e. parents in the original non-extended version of the ID). In what he calls *augmented* DAGs, Dawid incorporates intervention nodes F where $F_X = x$ corresponds to 'setting' the value of node X to x (in Pearl's language: $F_X = do(X = x)$), and he introduces a new value $\emptyset$ such that when $F_X = \emptyset$, X is left to have its 'natural' distribution, termed by Pearl the 'idle' system. Figure 1 shows, for a simple case, the usual representation of IDs as well as its extended and augmented versions, for the set (T, B, Y) where $T = (T_1, T_2, .., T_s)$ represents a set of policy variables (treatments), $B = (B_1, B_2, .., B_r)$ is a set of background variables (potential confounders) and Y is a response variable.



(a) Influence Diagram *(b) Extended ID* *(c) Augmented DAG*

Figure 1: Extended influence diagrams and augmented DAGs

Causal reasoning is related to prediction in the face of intervention. It relates to the idea that a variable is a 'cause' if setting this variable to a specific value (by intervention) changes the distribution of the response. Causal enquiries about the 'effect of T on Y' are seen as relating to (comparisons between) the distributions of Y given $F_T = do(T = t')$ for various settings of t'. The intervention node F of the augmented DAG is used as an 'auxiliary' variable to discuss the identifiability of these effects under certain DAG

structures. In particular, there is interest in establishing if the causal effect of F_T on Y can be identified and estimated correctly from the available data. The structure of the data available is defined by the set of conditional independencies that are derived from the graph.

Definition (Conditional independence) *If X, Y, Z are random variables with a joint distribution $P(\cdot)$, we say that X is conditionally independent of Y given Z under P, if for any possible pair of values (y, z) for (Y, Z) such that $p(x, y) > 0$, $P(x \mid y, z) = P(x \mid z)$. This can be written following Dawid (1979)'s notation as $(X \perp\!\!\!\perp Y \mid Z)_P$.*

The discussion is conducted in terms of the relevance of learning the strategy that gave the value t' to T, namely whether it arose from the original experimental setting $\pi(t \mid b)$ ($F_T = \emptyset$) or whether it was set externally ($F_T = do(T = t')$). Conditional independencies of the form $(Y \perp\!\!\!\perp F_T \mid T, \cdot)_{d_E}$ are used for this. Different examples of identifiable and unidentifiable situations are discussed by Pearl (2000), Lauritzen (2001) and Dawid (2002), each with their particular framework and notation. Imagine the set (T, B, Y) is available to us in the data set Δ. Figures 2*(a)* and 2*(b)* show the cases where B is said to be *irrelevant* for Y and where B is said to be *white noise* of Y (with respect to T) respectively. The case where B is an *intermediate* variable between T and Y (i.e. T affects B and B affects Y) is shown in Figure 2*(c)*. In these three structures the definition of absolute non-confounding given by $Y \perp\!\!\!\perp F_T \mid T$ holds (see Dawid 2002, §7). This asserts that the distribution of Y given T will be the same, whether T arose 'naturally' or T is set by intervention. Thus the causal effect can be estimated directly from the data available, Δ, using $p(y \mid t', F_T = do(T = t')) = p(y \mid t', F_T = \emptyset) = p(y \mid t')$. The definition of non-confounding $(Y \perp\!\!\!\perp F_T \mid T)$ does not hold for the structure shown in Figure 2*(d)*. In this latter system, B is said to act as a *confounder*, as it affects both treatment T and response Y. So, in order to obtain the causal effect, we are required to know (or observe) the marginal distribution $p(B)$. If this is the case, then the causal effect $p(y \mid F_T = t')$ can be obtained using the 'back-door formula' (Pearl 1993) which 'adjusts' for B such that $p(y \mid F_T = t') = \sum_b p(y \mid t', b) p(b)$.

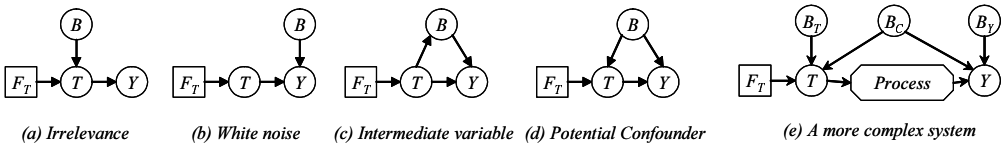


Figure 2: Possible basic structures

Social policies will usually be more complex systems including all irrelevant (B_T) and white-noise (B_Y) background variables, possible confounders (B_C) and an intermediate process, as shown in Figure 2*(e)*. Most of the examples in social policy interventions F_T involve (a collection of) atomic or contingent interventions. Therefore, the intermediate process might involve both intermediate variables affected by T and possible actions G

that will be triggered when the policy T is done. The 'overall' causal effect will include all direct and indirect effects of $do(T = t')$ on Y.

3 Introducing experimental nodes

3.1 Policy versus Experimental Decisions

When planning and designing a policy intervention and evaluation, the policy maker will have to define a strategy that involves 'policy intervention' actions ($D_T = \{d'\}$) and 'experimental design' actions ($D_E = \{d^*\}$). The former includes decisions related to how the policy is implemented and what (which doses and to whom) will be provided. The latter is related to the evaluation of the policy and includes experimental design decisions that define the (chosen or controlled) conditions under which the study is carried out and the data (Δ) recorded. If $\mathcal{D} = \{d'_1, .., d'_{\mathcal{D}_\mathcal{T}}, d^*_1, .., d^*_{\mathcal{D}_\mathcal{E}}\}$ are the components of a particular decision strategy $\pi_\mathcal{D}$, the interest lies in describing $\pi_D(\mathcal{D} \mid E)$. In this sense, we say that policy intervention actions (D_T) are concerned with intervening 'the world', while experimental design actions (D_E) relate to intervening the statistician's 'view of the world'.

It is important to differentiate between 'choosing a policy' and 'choosing a design', as the goals of these interventions are different. The 'success' of a policy intervention D_T is measured in terms of its efficacy to provoke 'better' values on the response variable Y through its overall effects reflected by $p(Y \mid F_T = do(T = t'); D_E)$. The efficacy of an experimental intervention, D_E, is measured in terms of its ability to isolate the policy effect as much as possible. Making an explicit representation of both types of interventions will assist decisions of the *experimenter* and considerations of the *analyst*, when the aim is to evaluate the causal effect of policy F_T.

When we, as data-collectors, approach the world, the data we collect depend on our way of approaching it. The data we observe in the database (available data, Δ_{d_E}) will reflect the experimental design decisions $D_E = d_E$ made (or deliberately 'not made') at the time of its collection through $p(data \mid D_E)$. Two extreme cases of designed studies might be, on the one hand, the 'perfect' experiment where all factors are controlled, balanced and randomised and, on the other hand, the complete observational study with all the relations that happen in 'natural' conditions (approximated by a census of all population). The available literature discusses broadly the cases for completely experimental data (see, for example, Chaloner and Verdinelli 1995; Wu and Hamada 2000) or completely observational data (e.g. Rosenbaum 2002). Although in the social sciences access to perfect experimental data is usually not feasible, the data is not always completely observational. In some cases, controls are taken at the time of the design/collection of data, which gives rise to partially experimental data. In this work, we consider D_E to include any experimental conditions that might involve a decision by the experimenter (data collector). The choice of 'no control at all' leads to observational data ($\Delta_\emptyset$) which is assumed to be a (degenerate) special type of experimental data.

Experimental design interventions, $D_E = \{M, R(B)\}$, contain the mechanisms $M =$

$\{M_E, M_S, M_T\}$ through which units are selected and assigned to eligible, sample and treatment groups, and the recording mechanism $R(B)$ that determines whether the background variables are observed and available to us in the data Δ_{d_E}. In addition, implementation details which refer to the logistics and how the study will be carried out are important, as they can introduce some biases. A complete description of these mechanisms is presented in Madrigal (2004). In this paper we focus on the treatment assignment mechanisms M_T.

As an example, imagine a policy will be implemented to increase the nutritional state (Y) of 'poor' children in a certain geographical area. Suppose there are two different brands of food supplements (FS) in green or red packages. A decision to sign a contract with the food supplement provider(s) for as long as the policy takes place has to be made. Imagine the policy maker is faced with four possible *policy interventions*: Policy 0 (t'_0): 'Do not give any FS'; Policy 1 (t'_1): 'Give green FS'; Policy 2 (t'_2): 'Give red FS'; and Policy 3 (t'_3):'Give green FS to young babies; and give red FS to older children'. Once a policy is chosen, all children in the target population will be under the same policy. In this case, *policy intervention strategies* (D_T) are defined for the same target population (namely, children in poverty), and the future policy interventions (F_T) are given by t'_0, t'_1, t'_2 and t'_3. When evaluating the policy intervention effect we obtain the 'overall effect' of each of the policies. Although the policy interventions act on children through the actual FS given, it is important to bear in mind that questions 'Is policy t'_i giving better results than policy t'_j?' are different from the question 'Is the green FS working better than the red FS?'. In this case, they will coincide when we are comparing policies t'_1 and t'_2, but to draw conclusions about the effects of green and red FS from a comparison between, say, t'_0 and t'_3 could be dangerous, as in t'_3, the effect of FS is confounded with age. The policy maker, as an experimenter, has to choose the *experimental design strategy* (D_E) used to collect data Δ_{d_E}. This data is used to evaluate policy intervention strategies (D_T) and compare the effects of policies $F_T = do(Policy = t'_s)$. Imagine that policy makers in principle have in mind the implementation of contingent policy t'_3 (against the option of not providing any food supplement at all t'_0). First, the experimental levels $\{t^*\}$ have to be set. These are allocated through action $A_T = do(Policy = t^*)$. Choosing some experimental levels $\{t^*\}$ to be equal to future policy levels $\{t'\}$, such that $t^*_1 = t'_0$ and $t^*_2 = t'_3$, ensures the positivity condition (see Appendix A), and then $\{t^*\} = \{t^*_1, t^*_2\} = \{t'_0, t'_3\}$. Imagine the allocation of policies is done randomly with probability of one half. This random intervention could be expressed as $A_{\theta_T} = do(\theta_T = \theta^*_T)$ such that it fixes $\theta^*_T = p(A_T = do(Policy = t^*_m)) = \frac{1}{2}$ for m=1,2. Policy allocation is randomised and it is defined by the *experimental design strategy*, D_E.

Dawid (2002)'s framework, although open to different strategies for setting the value of a treatment $T = t'$, including randomised or atomic definitions, does not allow us to represent in the same graph and formulae both the atomic (future) policy intervention $F_T \in D_T$ (allocating treatment $T = t'$ with probability one) and the (contingent or randomised) experimental allocation strategy followed when collecting data $A_T \in D_E$ (allocating treatment $T = t^*$ according to θ). Neither does it allow us to represent the impact on the (graphical) data structure of the experimental actions. Therefore, an

extension is needed.

3.2 Design Networks: Basics

In its simplified version, let $D_E = \{A, R(B)\}$ where A contains all the policy assignment mechanisms and $R(B)$ contains the recording mechanism, such that $R(B^q) = 1$, for $q = 1, 2...Q$, if variable B^q is recorded and $R(B^q) = 0$ if B^q is either unobservable or not recorded. *Assignment nodes* A and *recording nodes* R can be included in the DAG as decision nodes to create a *design network (DN)*. The design network shows the 'natural' (experimental) mechanisms that generate the data available Δ_{d_E}. In general, no matter whether the data has been collected already or we are planning the design to generate the data, D_E represents decisions to be made at the data collection time.

Consider the set (T, B, Y). For simplicity, suppose that T and Y are univariate, that B does not contain intermediate variables between T and Y (i.e. B consists of pre-intervention variables not affected by T), and that the future policy is an atomic intervention $F_T = do(T = t')$. Figure 3*(a)* shows the usual influence diagram representation of this case, and Figure 3*(b)* gives the corresponding design network. Note that A blocks all the paths going from B to the policy node T. This follows from the assumption that A captures *all* the allocation mechanisms for T that might be influenced by the background variables B, so that A is the only parent of the policy node T in the design network. Recording nodes, $R(B)$, are added for each background variable B^q, introducing the decision to record B^q versus not to record it. A double circle containing a dashed and solid line is given to each background node B^q to show its potential observability. It is assumed that policy T and response variable Y will be recorded. Figure 3*(c)* shows an *augmented design network* in which the future atomic intervention node F_T is added to the design network.

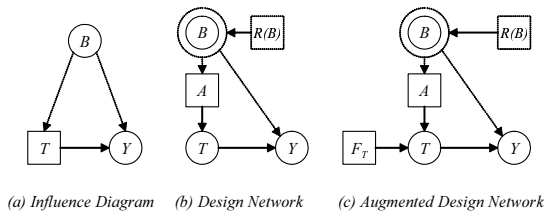


Figure 3: Design Networks

By representing simultaneously the design nodes $D_E = \{A, R(B)\}$ and the future intervention node F_T, the augmented design network is useful to make conclusions about two different tasks involved in policy evaluation and design: (1) the identifiability of the causal effect of T on Y, given a design $d_E \subset D_E$; and (2) the choice of a design strategy d_E to collect data when the interest is to evaluate the effect of intervention F_T. As mentioned before, the identifiability of intervention F_T depends on the data available (determined by mechanisms d_E); and the efficacy of experimental design (d_E) is always

to be determined with respect to the effects it tries to isolate (here F_T). Thus, D_E and F_T are and should always be read in the light of each other, and the augmented design network allows us to do that.

For causal reasoning in task (1), to discuss the identifiability of the causal effect of T on Y, we are interested in comparing the relevance of the choice of F_T given particular experimental conditions d_E^*, namely comparing $p(y \mid t', F_T = do(T = t'); D_E = d_E^*)$ and $p(y \mid t', F_T = \emptyset; D_E = d_E^*)$. This provides us with expressions and guidelines for *control via analysis* of possible confounders. On the other hand, in task (2), when planning the data collection by choosing an experimental design, we are interested in the relevance (or irrelevance) of the choice of experimental conditions d_E, (with respect to the identifiability of F_T) for different experimental choices $d_E \subset D_E$. So, we are interested in making comparisons between different settings of d_E^* and then choosing the optimal design from all experimental designs available in D_E. This provides us with guidelines for *control via design*.

Augmented DAGs and the set of conditional independencies derived from them have been used for causal reasoning in task (1) (see Dawid 2002). If two augmented DAGs derived from experimental conditions d_{E1} and d_{E2} share the same conditional independence statements, then they are equivalent for causal reasoning. Assignment actions A might affect the original collection of conditional independence statements. Recording decisions will have an effect on the set of variables that will be available to us through the available (sic) experimental data. Thus $R(B)$ will not introduce any new (in)dependencies in the structure, but will be relevant when discussing the potential identifiability of effects given assignment actions $A = a^*$. Some additional general remarks about the Decision Networks framework can be found in Appendix A.

4 Causal graphical analysis for cluster allocation

Most of the literature in Cluster Randomised Trials (CRTs) has emphasised the fact that Fisher's principle is violated, as the experimental unit does not coincide with the analysis unit, and the difference in levels where the experimental allocation generating data available, Δ, is at cluster level and the analysis is undertaken for a response at individual level. When introducing the need for the evaluation of a future intervention F_T using data generated from a (past) experiment d_E, it is important to acknowledge the fact that the future intervention level might differ from the experimental level. In general, the future intervention ($F_T = do(T = t')$), the experimental allocation ($A_T = do(T = t^*)$) and the response variable (Y) could each be at cluster/individual level and would not necessarily coincide. The experimental level will define the data structure and the conditional independence statements reflected in the 'experimental' causal graph through d_E. The future intervention F_T level will define the 'future' causal graph structure.

The interest could lie in the causal effect at cluster level or at individual level. Responses at cluster level will summarise what is observed at a community level, while responses at individual level are usually more of interest to describe what is the effect in,

say, a household within a community. If we are interested in the intervention effect on an individual level response, depending on how the future intervention will be implemented, there are two possible (causal) intervention effects we might be interested to identify: namely, the distribution of the individual outcome given a clustered intervention, $P(Y_{jk} \mid F_{Tj} = do(T_j = t'))$, and the distribution of the individual outcome given an individual intervention, $P(Y_{jk} \mid F_{Tjk} = do(T_{jk} = t'))$. The former would try to estimate the effect of a cluster-level intervention (usually the interest in social policy) and the latter would try to estimate the effect of an individual-level intervention (as could be the goal of many medical trials).

4.1 Cluster design networks

In terms of design decision strategies, a cluster-randomised study implicitly involves two design decisions: (1) the decision of clustering (i.e. to allocate the treatments to clusters of individuals) and (2) the decision of randomising (i.e. to perform the allocation using a random procedure).

When an intervention at cluster level occurs we distinguish between two cases. The first is related to the case in which the intervention affects all individuals in the cluster: for example, when the improvement of health services is undertaken at community level and all families within a community are subject to the same infrastructure. In this case, individuals within the cluster cannot choose not to be affected by the policy intervention. In this paper this case will be referred to as a 'pure-cluster' intervention (and denoted by $d_C = 1$). The second case refers to the situation in which, although an intervention is allocated at cluster level, not all individuals, but only a subset of them within a cluster, will be subject to the intervention. Actions, in this second case, are 'done' at individual-level to individuals within a cluster, and thus individual characteristics might have an influence in the individual's allocation of treatment

An example of the latter is when the intervention affects only eligible individuals. A cluster policy of this type could be seen as: 'all eligible individuals k in cluster j will receive policy t' ' via $A_{Tj} = do(T_j = t')$. So, allocation of policy is done at cluster level and two eligible individuals in the same cluster cannot be allocated different policies (contrary to what would happen if the policy allocation was done at individual level). In the case of Progresa, this will correspond to the case where only poor households are receiving extra money for nutrition and educational grants. These actions are aimed at household-level; however, not all households in a community are poor and therefore not all households within a community receive the same treatment, only the eligible ones. In a more general setting, the individual choice of treatment might depend on some possibly unobserved background variables and not necessarily only on some previously defined (and observed) eligibility criteria. For example, imagine that some health centres are allocated a certain restricted quantity of food supplements to be distributed among families visiting them, but the amount of food supplements is not enough to cover all families. Then, the fact that a family is receiving the food supplement or not could depend on the (unobserved) nurses' choice or on a first-come-first-served basis. In any case, when different units within a cluster do not necessarily receive the same treatment

and this is dependent on certain individual-level background variables, the experiment will be referred as a 'non-pure' cluster allocation (and denoted by $d_C = 0$).

4.2 Effects of cluster allocation design decisions

A design network for the general cluster setting for a cluster-level future intervention F_{Tj} is presented in Figure 4. As we are discussing clusters of individuals, naturally variables are not all at the same level and a simple DAG cannot be used without further notation The levels are represented in the graph by squares, following Spiegelhalter's notation (see WinBUGS), meaning that the same graphical structure applies for each of the observations at the same level. Design decisions can then be taken at both individual and cluster level. The structure has been kept similar to that used above, but now we have the situation replicated for the two levels involved. Let cluster j (for $j = 1, 2, ...J$) have K_j units and let T_j and T_{jk} be variables for intervention status (Treatment / Control) at cluster and unit level respectively. Similarly, let B_j and B_{jk} represent the background variables at cluster and unit level and Z_j some recorded cluster-level covariates that might be affected by the policy. Nodes A_j and A_{jk} will correspond to the assignment mechanisms to allocate policy at cluster and individual levels respectively. Action $A_{Tj} = do(T_j = t^*)$ defined in A_j will imply $T_j = 1$ if the value t^* corresponds to the policy taking place in cluster j, and will imply $T_j = 0$ if the value t^* corresponds to 'control'. This will work similarly for the individual-level case. The recording mechanisms could be defined over the set of cluster and individual background variables, $R(B_j)$ and $R(B_{jk})$, respectively. The response Y_{jk} will correspond to that observed for individual k in cluster j.

The decision of running a 'pure' cluster allocation experiment ($d_C = 1$) will imply that the intervention is done equally to all members in the cluster. So, once the treatment for cluster j T_j is fixed by action $A_{Tj} = do(T_j = t^*)$, this fully implies actions $A_{Tjk} = do(T_{jk} = t^*)$, and so the values of the treatments T_{jk} for all K_j individuals in cluster j. Then $t_j = t_{jk} = t_{jk'}$ for all individuals $k, k' = 1, 2, ...K_j$ in cluster j. So, the effect of pure-clustering prohibits individual covariates from influencing the choice of treatment, breaking any links that could be present from B_{jk} (any background individual-level covariates) to T_{jk} in the graph. When 'non-pure' cluster allocation takes place ($d_C = 0$), although the experimental allocation is made at cluster level, individual k within cluster j might be receiving treatment or not depending on some individual-level covariates B_{jk}, and thus t_{jk} might differ from $t_{jk'}$ for $k \neq k'$.

The assignment mechanism nodes A_j and A_{jk} could be expanded. This is not done in Figure 4, to keep the (already complex) graph as simple as possible. The individual assignment mechanism A_{jk} is considered to be dependent on the actual policy that was allocated to cluster j, T_j, and (possibly) on some individual background variables. Thus, the action assigning policy t^* to individual k $A_{Tjk} = do(T_{jk} = t^*)$ is considered to be dependent on T_j and B_{jk} such that $\theta_{Tjk} = p\left(A_{Tjk} = do(T_{jk} = t^*) \mid T_j, B_{jk}\right) = q(T_j, B_{jk})$. The 'pure-cluster' case will imply that the individual assignment mechanism does not depend on individual background variables B_{jk} and
$\theta_{Tjk} = p\left(A_{Tjk} = do(T_{jk} = t^*) \mid A_{Tj} = do(T_j = t^*), d_C = 1\right) = 1$ and the following propo-

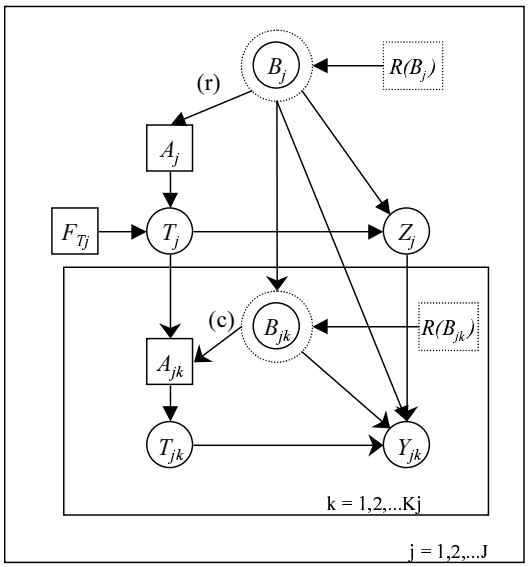


Figure 4: Design Network for cluster allocation and cluster-level future intervention F_{Tj}

sition is thus established.

Proposition (Pure-Clustering) *If the experimental design strategy d_E includes the action of performing a 'pure cluster' experiment such that $\{d_C = 1\} \in d_E$, then the 'structural' effect of $d_C = 1$ on the 'original' set of conditional independencies, is to introduce the set of conditional independencies $(T_{jk} \perp\!\!\!\perp B_{jk})_{d_C=1}$ that will hold on the data Δ_{d_E} generated by d_E.*

Figure 5 includes a close-up of the individual-level plateau in Figure 4, in which the design network has been extended for node A_{jk} and variables Z_{jk} introduced. Now let us refer to the situation when a 'pure' cluster experimental intervention is not feasible, but when we have a non-pure cluster experiment such that, within each cluster j, individual policy allocation follows a deterministic rule based on individual-level observed covariates Z_{jk}. The final policy allocated to an individual through A_{jk} will be a function of Z_{jk}, and any other possible influences on T_{jk} from background variables B_{jk} (other than Z_{jk}) are eliminated. The prevalences of T_{jk} in the experimental data available Δ will depend on the policy allocated to the cluster T_j and Z_{jk} but not on B_{jk}.(e.g. $\theta_{Tjk} = p\left(A_{Tjk} = do(T_{jk} = t^*) \mid T_j, Z_{jk}\right) = q(T_j, Z_{jk})$) The structure obtained is similar to the stratified allocation presented in Madrigal (2005), and arrow (c) will be deleted when this 'deterministic' allocation takes place.

Proposition *If a 'non-pure cluster' experiment includes a design strategy in which policies at individual level are allocated following a 'deterministic' rule defined by the exper-*

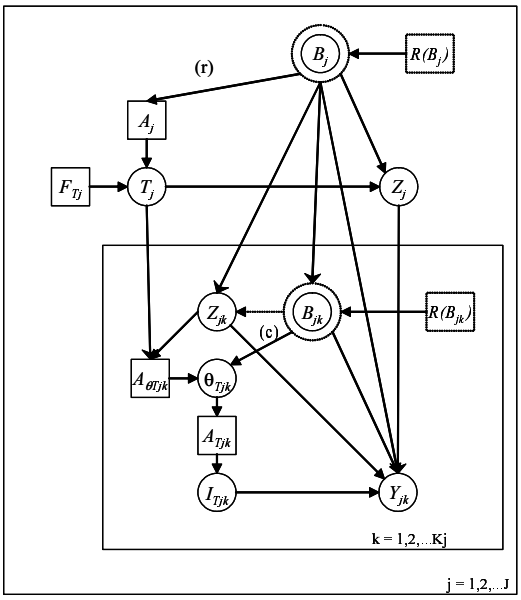


Figure 5: Close-up of individual-level plateau

imenter based on covariates Z_{jk} , then conditional independencies $(T_{jk} \perp\!\!\!\perp B_{jk} \mid Z_{jk})_{d_E}$ are introduced and will hold on the data Δ_{d_E} generated by d_E.

As introduced in Madrigal and Smith (2004), design decision strategy d_E including random allocation of policies (i.e. $A_{\theta_T} = do(\theta_T = \theta_T^*) \in d_E$) might, qualitatively speaking, modify the structure of the data we are to collect (see Appendix A). By allocating the treatments completely at random (i.e. $A_{\theta_{Tj}} = do(\theta_{Tj} = \theta_{Tj}^*)$) we ensure that the treatment received is independent of any background variables B_j that, otherwise, might have an influence on the policy assignment mechanism. Then, when random allocation takes place, arrow (r) from B_j to A_j in Figure 4 disappears and the conditional independence statement $(T_j \perp\!\!\!\perp B_j)_{d_E}$ holds. Again, this probability, θ_{Tj}^*, might depend on possible stratification observed variables. When randomising at cluster level we therefore ensure that the level of treatment that is received by cluster j is independent of the level received by cluster j' (i.e. knowing that cluster j was assigned intervention t^* does not give us any further information about the intervention group at cluster j').

4.3 Identifying cluster-level interventions

The appropriateness and consequences of different design decisions $d_E \subset D_E$ will depend on the goals of the experiment. The case in which the interest is in the effect of a cluster-level future intervention $F_{Tj} = do(T_j = t')$ on a cluster-level response Y_j will degenerate to the one-level case. When the interest lies in an individual-level response Y_{jk}, it can be

seen in Figure 4 that, if d_E^* includes a random cluster allocation procedure and arrow (r) is not present, the conditional independencies $(Y_{jk} \perp\!\!\!\perp F_{Tj} \mid T_j)_{d_E^*}$ hold for all j,k. Thus, once the value of the policy assigned to the cluster T_j is known, learning whether the policy status T_j arose from the future policy implemented $F_{Tj} = do(T_j = t')$ or from the 'original' experimental allocation $A_{Tj} = do(T_j = t^*)$ when $F_{Tj} = \emptyset$, is irrelevant. Therefore, direct identifiability of the effect $F_{Tj} = do(T_j = t')$ on Y_{jk} holds and $p(y_{jk} \mid F_{Tj} = do(T_j = t'))$ can be directly obtained from data $\Delta_{d_E^*}$ available as long as $t' \in \{t^*\}$ such that

$$p(y_{jk} \mid F_{Tj} = do(T_j = t'); \Delta_{d_E^*}) = p(y_{jk} \mid T_j = t'; \Delta_{d_E^*})$$

This does not disregard the fact that individuals belonging to the same cluster will have a positive correlation, which must be taken into account in any model used for the analysis and estimation of the effect on an individual-level response.

Again, when non-random cluster allocation is performed as part of the experimental design d_E^*, $(Y_{jk} \perp\!\!\!\perp F_{Tj} \mid T_j)_{d_E^*}$ does not hold anymore, but the conditional independencies $(Y_{jk} \perp\!\!\!\perp F_{Tj} \mid T_j, B_j)_{d_E^*}$ hold for all j,k. Thus, the identifiability of the effect of a future policy $F_{Tj} = do(T_j = t')$ on Y_{jk} will depend on the recordability of cluster-background variables $R(B_j)$ and the causal effect will need to be obtained through an 'adjustment' procedure such that, as before, using the back-door criteria

$$p(y_{jk} \mid F_{Tj} = do(T_j = t')) = \int p(y_{jk} \mid T_j = t', B_j) p(B_j) dB_j.$$

Unless we are ready to assume some prior distribution for $p(B_j)$, the recording of variables B_j as part of the design $(\{R(B_j) = 1\} \in d_E^*)$ are needed to achieve an 'adjusted identification' of the causal effect.

Different recording mechanisms might assist identification. For instance, if we were ready to assume that cluster background variables did not have a direct effect on the individual response, such that arrow from B_j to Y_{jk} was deleted, then all the influence from B_j would be through individual background variables and the observed cluster-level variables Z_j. In this case, conditioning on T_j, B_{jk} and Z_j would be enough and a design able to record these variables will provide identifiability. Thus, if cluster background variables B_j were not accessible to the experiment, this new set of covariates $\{B_{jk}, Z_j\}$ could assist identification.

4.3.1 Bayesian hierarchical models

In a hierarchical setting, data within each cluster j is assumed to depend on parameters θ_j, which in turn are assumed to be drawn from some population distribution with parameters ψ. In an initial model, the response y_{jk} for individual k in cluster j is assumed to have a Normal distribution, such that

$$\begin{aligned} y_{jk} &\sim N(\mu_{jk}, \sigma^2) \\ \mu_{jk} &= u_j \end{aligned} \tag{1}$$

and cluster-specific random effects (u_j) are assumed to have a Normal distribution with mean ϕ_j and variance σ_u^2, such that

$$\begin{aligned} u_j &\sim N(\phi_j, \sigma_u^2) \\ \phi_j &= \alpha + \beta T_j \end{aligned} \tag{2}$$

where T_j represents the treatment given to the jth cluster. There are many potential elaborations to this basic model (see Spiegelhalter 2001; Turner et al. 2001). The priors that need to be specified for this model are $p(\alpha)$, $p(\beta)$, $p(\sigma^2)$, $p(\sigma_u^2)$. Making causal assumptions and a graphical representation of all influences present in the particular system analysed could assist recognition of possible confounders and thus assist both the experimenter's decisions for control via design and the analyst's decisions for control via analysis.

When cluster allocation is done randomly, if we are ready to assume linear relations, the two-level Bayesian hierarchical model as specified in equations (1) and (2) could be used to estimate the effect of F_{Tj} on Y_{jk}, and coefficient beta can 'safely' be given a causal interpretation as an 'overall' effect. For the non-random case, the analysis will need the conditioning on the 'relevant' background variables. The inclusion of individual-level and cluster-level covariates in the analysis could be done directly by including them in equations (1) and (2) respectively. The conclusions just derived hold for both 'pure' cluster and 'non-pure' cluster allocations. The 'overall' causal effect will correspond to a 'total effect' when a 'pure' cluster allocation ($d_C = 1$) is done. However, this will not be the case for $d_C = 0$, where the 'total effect' cannot be obtained. To make the difference between 'overall' and 'total' effects clearer, the interactions of individuals in a cluster have to be considered, and this is discussed below in Section 4.5.

4.4 Identifying individual interventions from clustered data

Consider the case where the main interest is in obtaining the individual-level causal effect, namely $P(Y_{jk} \mid F_{Tjk} = do(T_{jk} = t'))$ from data that is clustered. If randomised allocation could take place at individual level, then it could be directly identified from the experimental data Δ, as individual random allocation will break the possible influence of cluster background variables on the policy allocated to the individual. Suppose that it is not feasible to randomise at individual level, but to intervene clusters is possible. The design network for this case will basically coincide with that shown in Figure 4, but in this case we assume that the future policy will consist of an individual intervention F_{Tjk}, and that we are interested in identifying effects at the individual level.

From the design network in Figure 6 it can be seen that $(Y_{jk} \perp\!\!\!\perp F_{Tjk} \mid T_{jk})$ does not hold even if arrows (r) and (c) are deleted from the graph. So, the effect of policy intervention $F_{Tjk} = do(T_{jk} = t')$ on Y_{jk} cannot be identified directly from the data and some adjustment will be needed.

When a 'non-pure' cluster allocation takes place in the experiment, individual policy assignment will depend on both the policy allocated T_j and individual background

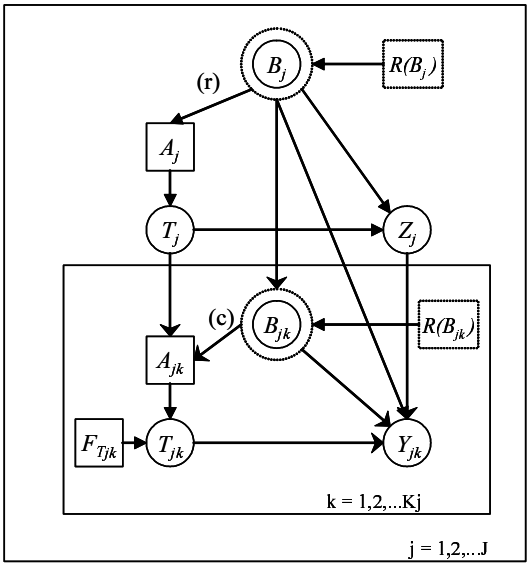


Figure 6: Design Network for cluster allocation and individual-level future intervention F_{Tjk}

variables B_{jk}. Conditioning on this set of variables, the irrelevance of F_{Tjk} is gained such that $(Y_{jk} \perp\!\!\!\perp F_{Tjk} \mid T_{jk}, T_j, B_{jk})$ Thus the recording of B_{jk} is needed in order to obtain adjusted identifiability through

$$p(y_{jk} \mid F_{Tjk} = do(T_{jk} = t'); d_C = 0) = \int p(y_{jk} \mid T_{jk} = t', T_j = t^*, B_{jk}) p(T_j = t^*, B_{jk}) dB_{jk} dT_j \ ,$$

where, if randomisation did not take place at cluster level, T_j and B_{jk} are not independent (both having B_j as an ancestor) and their joint distribution is needed. If recording at individual level for B_{jk} is not undertaken, the causal effect will be unidentifiable.

When 'pure' cluster allocation is done, and as a result arrow (c) is deleted, then there are no individual level confounders and all possible confounders will be at cluster-level. So, 'pure cluster' assignment might improve identification of individual intervention effects, in particular, when randomisation at cluster level is feasible or in the case when cluster-level confounders (B_j) are easier to observe and/or control than individual-level confounders (B_{jk}).

As shown above, the 'overall' effect of a future cluster-level intervention at cluster level F_{Tj} can be identified from the experimental data when policies in the experiment are allocated randomly at cluster level. Something similar happens when the assignment is not carried out randomly, but cluster background variables are recordable and an 'adjustment' measure is needed. Moreover, for $d_C = 1$, the overall effect will coincide with the participants' total effect. Thus, if the indirect effect due to interaction among neighbours is negligible, as could be the case when vitamin supplements are administered to children in Progresa, the total effect measured will be equal to the direct (personal)

effect. Although this might not be always true for social policies, this might be the case for some treatments in a medical setting in which, for example, a drug is supposed to act in an individual regardless of his interaction with other people in the cluster. In this case, a 'pure cluster' design could assist identification of individual interventions. Therefore, a 'good' (randomised or controlled) cluster design might be able to provide more information than a 'bad' (with unobserved confounders) individual design. The distinction of overall, total, direct and indirect effects will be discussed next.

4.5 Overall versus total effects

As Koepsell (1998) states, 'just as infectious agents can be spread from person to person, transmission of attitudes, norms and behaviours among people who are in regular contact can result in similar responses'. So, when people interact or communicate, their response to an intervention can be explained (and partitioned) in terms of direct ('personal') effect and indirect ('neighbours') effect. So, interventions may affect the whole population, not just those who participate (or were subject to interventions).

The fact that all individuals in a group follow the same policy, or are encouraged to take a particular action, has thus an additional 'interaction effect'. This is so as individuals interact with each other, creating a domino effect. In the case of Progresa we have, for example, the fact that mothers talk! Thus, if a mother is encouraged to take children to the health centre for food supplements, besides her possible individual motivation, the fact that other mothers in the village are encouraged as well, creates an additional effect on her (i.e. if everybody is doing it, there is an extra motivation to do it, and being the only one not doing it will be rare and possibly socially penalised).

If in a cluster, not all individuals are allocated the same intervention, then the effects of interventions can be classified, following Hayes et al. (2000)'s definition, according to the 'intervention status of the individual' as participants (treated) or nonparticipants (controls). Those who participate receive both a direct $\left(DE_{(P)}\right)$ and an indirect effect $\left(IE_{(P)}\right)$, which combine to form the total effect $\left(\lambda_{(P)} = DE_{(P)} + IE_{(P)}\right)$. The non-participants receive only an indirect effect, $IE_{(NP)}$, so their total effect contains only those indirect effects $(\lambda_{(NP)} = IE_{(NP)})$. The indirect effects received by participants and non-participants may differ in magnitude, so an index is used to distinguish them:

	Participants (P)	Non-participants (NP)
Total effects	$\lambda_{(P)} = DE_{(P)} + IE_{(P)}$	$\lambda_{(NP)} = IE_{(NP)}$

If we are ready to assume that these effects are equal for all individuals in a cluster, then the overall effect observed in a cluster will correspond to the weighted average of the effects on participants and non-participants such that

$$\text{Overall effect} = w_{(P)}\lambda_{(P)} + w_{(NP)}\lambda_{(NP)} \tag{3}$$

where $w_{(P)}$ and $w_{(NP)}$ are just weights that will be functions of the number of participants and non-participants in the cluster (or in terms of the 'coverage' -% of participants- of the experiment). So, the overall effect will include a combination of direct and indirect

effects. In particular, expression (3) could be extended to be re-written as

$$\begin{aligned}\text{Overall effect} &= w_{(P)}\left(DE_{(P)} + IE_{(P)}\right) + w_{(NP)}IE_{(NP)} \\ &= w_{(P)}DE_{(P)} + \left(w_{(P)}IE_{(P)} + w_{(NP)}IE_{(NP)}\right)\end{aligned}$$

In the case where indirect effects, for both participants and non-participants, are assumed to be negligible such that $IE_{(P)} \approx 0$ and $IE_{(NP)} \approx 0$, then the overall effect will be approximately proportional to the direct effect such that

$$\text{Overall effect} \approx w_{(P)}DE_{(P)}.$$

In the case of a 'pure cluster' experiment, either all individuals are participants (treated cluster) or all are non-participants (control cluster). In this situation, contamination within clusters is completely avoided, and in control clusters no intervention indirect effects are observed (i.e. $IE_{(NP)} = 0$). For the treated clusters, all individuals are participants, and overall intervention effect of the cluster will coincide with the total participant effect, denoted by $\tau_{(P)}$: namely,

$$\begin{aligned}\text{Overall effect (control cluster)} &= 0 \cdot \lambda_{(P)} + 1 \cdot \lambda_{(NP)} = \textit{total effect}_{(NP)} = 0 \\ \text{Overall effect (treated cluster)} &= 1 \cdot \lambda_{(P)} + 0 \cdot \lambda_{(NP)} = \textit{total effect}_{(P)} \quad = \tau_{(P)}\end{aligned}$$

and therefore

$$\text{Overall effect}_{(d_C=1)} = \text{total effect }_{(P)} = \tau_{(P)} = DE_{(P)} + IE_{(P)}$$

Individually randomised trials typically aim to measure the direct effect, $DE_{(P)}$. By contrast, CRTs measure the total effect $\tau_{(P)}$ if all individuals participate, but otherwise they measure the overall effect, which will vary according to intervention coverage and the characteristics of the population.

If individuals are naturally clustered, the magnitude of the indirect effect of an intervention is likely to be important in deciding whether a trial should be individually - or cluster- randomised. Indirect effects, due to interaction, will be included in the outcome measure. As a consequence, if the main interest is in measuring only the direct effect that a possible drug/treatment, say, has on an individual and it is known that indirect effects could be relevant, then CRTs might not be the best option as they will measure the overall effect instead of the direct effect.

In assessing the value of intervention it is important to take into account their indirect as well as direct effects. In some cases it may be better to avoid intervention if the coverage needed to make it beneficial is too high to be realistically achievable. In addition, it may be desirable to separate the overall effect into its direct and indirect components. Methods for measuring direct and indirect effects separately have mostly been developed in the context of vaccination (see Hayes et al. 2000; Longini et al. 1998). Standard CRT designs measure the overall effect of intervention, and this is often the most useful measure for policy makers because it includes all the components, both direct and indirect, which a population would experience if a cluster policy were to be implemented.

It should be clear that the stable-unit-treatment-value-assumption (SUTVA, as labeled in Rubin 1980), which implies that the response of the unit does not depend on which treatment was applied to other units, does not hold when units interact and the indirect neighbours effects are not negligible. However the Bayesian predictive decision-theoretic approach that is followed in this paper does not require this assumption, as would be the case in Rubin's counterfactual approach. The counterfactual model for causal inference could lead to ambiguities and pitfalls, as discussed by Dawid (2000).

5 Progresa effect example using hierarchical models

In this section a hierarchical model analysis based on Spiegelhalter (2001) is performed for the evaluation of cluster- and individual-level interventions based on Progresa data. In the programme, communities were randomly allocated either to a treatment or a control group. The community level interventions G_1 (such as the improvement of health services and educational talks) are received by all households in a 'treatment' community. In addition, all eligible (poor) households that belong to a 'treatment' community receive household interventions, such as financial support, G_2. The data recorded includes a census of eligible and non-eligible households for (treated and control) communities selected for the study.

Let T_j be the cluster treatment indicator, such that $A_{T_j} = do(T_j = 1)$ if community j was allocated to Progresa programme and $A_{Tj} = do(T_j = 0)$ if it was allocated to control, so we have

$$T_j = \begin{cases} 1 & \text{if community Treatment} \\ 0 & \text{if community Control} \end{cases}$$

Let E be an indicator variable denoting eligibility status. Then $E_{jk} = 1$ if household k in community j is eligible and $E_{jk} = 0$ if non-eligible. In Progresa, $E_{jk} = 1$ corresponds to a poor household. Thus,

$$E_{jk} = \begin{cases} 1 & \text{if 'poor' household} \\ 0 & \text{if 'non-poor' household} \end{cases}$$

So, household k in community j will be allocated household-level Progresa interventions T_{jk} (e.g. financial support) through an allocation A_{jk}, in which a household is given extra money if, in addition to belonging to a treatment cluster, the household is 'eligible'. It will not be given extra money if either it is not eligible or if it belongs to a control community. If we denote by P_{jk}, the indicator variable for 'Progresa participant', such that $P_{jk} = 1$ if household k in cluster j receives economical support and $P_{jk} = 0$ if not, then, P_{jk} is defined as

$$P_{jk} = \begin{cases} 1 & \text{if } T_j = 1 \text{ and } E_{jk} = 1 \\ 0 & \text{otherwise} \end{cases}$$

From the general formulation of the Design Network for cluster allocation presented above, a simplified version of Progresa's experimental design containing the main features is shown in Figure 7, where Y_{jk} represents the response of household k in community j for $k = 1, 2, ...K_j$; B_j represents the background variables that are shared by all

individuals in community j. Background variables at individual level were not added to keep the graph simple, but could be easily incorporated. As allocation at cluster level was done randomly in Progresa, no arrow is drawn from B_j to T_j. If G_1 denotes Progresa's cluster-level intervention (action) corresponding to health services and talks (i.e. the 'encouragement' that communities receive to improve their nutrition) and G_2 denotes Progresa's household-level action of giving financial support to poor households, then note that the action $do(T_j = 1)$ will trigger both atomic cluster-level intervention G_1 and contingent (on T_j and E_{jk}) individual-level intervention G_2.

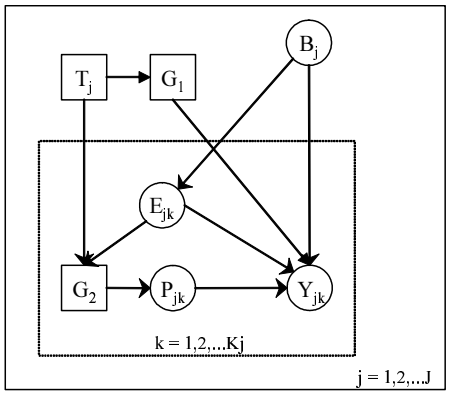


Figure 7: Progresa experimental Design Network for basic nodes

5.1 Cluster-level intervention effect

Imagine that we are interested in the effect of Progresa on the total food consumption y_{jk}, measured in terms of the amount of money spent on food in a household. The complete data set, including poor and non-poor households, consists of 20,589 households in 500 clusters. To begin with, assume we are interested in measuring the overall effect of Progresa intervention $F_{Tj} = do(T_j = t')$. A hierarchical model following the setting presented in Section 4.3 (equations (1) and (2)) is used to estimate this effect. This model was run in BUGS using vague priors following Spiegelhalter (2001) with

$$\begin{aligned} p(\alpha) &\sim Uniform(-10000, 10000) \\ p(\beta^*) &\sim Uniform(-10000, 10000) \\ p(\sigma^{-2}) &\sim Gamma(0.01, 0.01) \\ p(\sigma_u^{-2}) &\sim Gamma(0.01, 0.01) \end{aligned}$$

We encountered no difficulties in convergence of this model. The analysis is based on a sample of 10,000 iterations following a burn-in of 5,000.

The posterior inference (means and 95% intervals) for the parameters involved is presented in Table 1. As can be seen from the table, the posterior mean for beta

Parameter	Mean	95% interval
β^*	39.48	$(20.36, 58.33)$
α	444.9	$(435.2, 454.6)$
σ^{-2}	$2.102E-5$	$(2.061E-5, 2.143E-5)$
σ_u^{-2}	$1.054E-4$	$(1.052E-41, .216E-4)$

Table 1: Posterior distributions of parameters for cluster-level intervention

is $E[\beta^* \mid \Delta] = 39.48$ with a 95% interval of (20.36,58.33). In this case β^* gives the cluster-level total overall effect of Progresa intervention on the food expenditure in a household. So, β^* contains a summary of the effects of the programme (through G_1 and G_2) on Y for all the population in a cluster, by averaging participants (receiving G_1 and G_2) and non-participants (only receiving G_1). Depending on the aims of the study this total overall effect might be the relevant causal effect of interest. In such a case, it could be said that the causal effect of Progresa is to increase, on average, the food expenditure of a household by 39.48 Mexican Pesos. This in relation to the average food expenditure for a household in a control community that will be of $E[\alpha \mid \Delta] = 444.9$ Mexican Pesos.

5.2 Individual-level effect

Now imagine that we are interested in obtaining an estimate of the 'causal' effect of the individual-level intervention $F_{G_2} = do(G_2 = g_2' = q(poor))$ of giving financial support to poor households. The allocation of G_2 depends on the cluster-allocated policy T_j (Treatment/Control) and on the eligibility condition E_{jk} of a household defined as 'poor': both are assumed to have an effect on the household expenditure level and thus act as confounders in this case. So, to identify the individual-level effect of F_{G_2}, it is needed to control by including these two confounders in the analysis.

The hierarchal model used for the cluster-level effect above can be extended to include covariates T_j and E_{jk} at cluster and individual level respectively. The 'Progresa participants' status of a household P_{jk} acts as an indicator variable of the presence of economic support provided by G_2. We include here the household size Z_{jk} as an individual covariate to illustrate the possible inclusion of other covariates in the model. Household size will have an influence on the total expenses of the household Y, and it is neither affected by the policy nor affecting (at least directly) policy allocation. So now this is considered part of the white noise (with respect to T and P) at the recorded individual level. Equations (1) and (2) can be substituted by

$$\begin{aligned} y_{jk} &\sim N(\mu_{jk}, \sigma^2) \\ \mu_{jk} &= u_j + \beta_2 P_{jk} + \delta E_{jk} + \gamma Z_{jk} \end{aligned} \tag{4}$$

$$\begin{aligned} u_j &\sim N(\phi_j, \sigma_u^2) \\ \phi_j &= \alpha + \beta_1 T_j \end{aligned} \tag{5}$$

We chose to use the same priors to estimate this model, as before. Thus all the coefficients (namely $\alpha, \beta_1, \beta_2, \delta$ and γ) were given vague uniform distributions a priori. Again, we encountered no difficulties in obtaining convergence and the sample simulated is the same size as before.

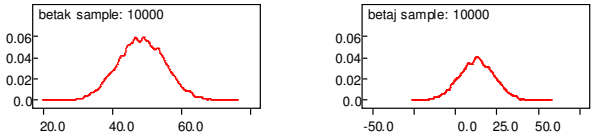
The posterior means and 95% intervals for all the parameters are given in Table 2. The individual-level effect here will be measured by the coefficient of P_{jk}, namely β_2, whose posterior mean is given by $E[\beta_2 \mid \Delta] = 45.71$ with a 95% interval of (33.95, 57.49). In general β_2 will isolate the effect of G_2 (from the effect of G_1) and we could say that the effect of a policy F_{G2} that provides economic support according to the poverty level of a household will increase, on average, the food expenditure of a participant household by 45 Mexican Pesos (regardless of the presence or not of a secondary action G_1). However, β_2 implicitly includes possible indirect effects resulting from the interaction of participant households with non-participant households in a community.

Parameter	Mean	95% interval
α	473.5	(461.7 , 485.5)
β_1	16.7	(-3.848 , 36.63)
β_2	45.71	(33.95 , 57.49)
δ	-34.14	(-43.54 , -24.69)
γ	30.56	(29.54 , 31.6)

Table 2: Posterior distributions of parameters for individual-level intervention

A second reading of this analysis could consider the case in which the total effect of Progresa (defined by G_1 and G_2) is split in its effect on food expenditure, due to the community-level action of educational talks (G_1) and the economic support provided at household level to poor people (G_2). Then, β_1 becomes a parameter of interest containing the direct effect of G_1 (i.e. the effect of Progresa on food expenditure that is not due to economic support) and P_{jk} is regarded as an intermediate variable. In this case, and following the reasoning of path analysis (Bollen 1989; Pearl 2000), we can see that the total overall effect of Progresa β^* could be written as $\beta^* = \beta_1 + \lambda\beta_2^*$ where λ will contain information about the prevalence of participants within a treated community. The total overall effect at household level is here denoted by β_2^* $(= \beta_2 + \delta b_{(E)P})$ where, as before, β_2 represents the direct individual-level effect and $\delta b_{(E)P}$ the confounding effect, which in this case has been controlled via analysis.. In this case it can be seen that, although the posterior mean for β_1 has a value of 16.7, the 95% posterior interval includes the value zero. So we cannot assert that the direct effect of the cluster-level intervention G_1 was different from zero. A more careful analysis, possibly including more 'white noise' covariates at cluster level, might provide narrower intervals for the coefficients. However, given that this response variable is measured in money terms, the main effect of the programme could be expected to be due to the increase of income of the participant households derived from G_2. This might not be true for other response variables.

We can notice that one could be tempted to offer a causal interpretation to the



(a) *(b)*

Figure 8: Posterior distribution for *(a)* β_2 and *(b)* β_1

coefficient of eligibility (δ) that distinguishes between poor and non-poor households. However, the data available was not properly selected in order to isolate the relationship between poorness and food expenditure. The level of poorness in Progresa was obtained through a discriminant function that depends on many household level covariates that could also affect the response, becoming confounders for this coefficient. Therefore, if we are interested in interpreting δ, a data set including these covariates will be needed. In this analysis, household size is known to be part of the variables used to define the poorness of a household. Its inclusion or not in the model (analysis not shown), although 'transparent' for β_1 and β_2 (given E_{jk}), will have an important effect on the posterior mean $E[\delta \mid \Delta]$. Although, in this case the posterior mean seems to be significantly different from zero and has the 'correct' sign (i.e. it should be expected that poor people spend less money than non-poor people), this superficial conclusion might still be subject to unrecorded confounding.

6 Conclusion

The primary contribution of this paper is to expand on Dawid (2002)'s model for causality reasoning within the Bayesian decision-theoretic framework: to 'adapt' it to policy analysis, to include experimental nodes, to allow intervention nodes 'do' parameters nodes, to discuss the relevance (or irrelevance) of experimental design and to include interventions at different levels (clusters) of units. Observational data is considered a degenerate type of experimental data. In addition, there was a need to create some notation to describe the mechanisms derived from choices, such choices as the experimenter might make when choosing the 'lens of the camera' to picture the world. These choices affect the characteristics (units, variables and distributions) of the database. The inspection of influence diagrams and, in particular, the augmented DAGs derived from them, has been shown to be useful to decide if the data available is sufficient for obtaining consistent estimates of the target causal effect of policy intervention $F_T = do(T = t')$. If so, we can derive a closed-form expression for the target quantity in terms of distributions of available quantities. If it is not sufficient, this framework can help suggest a set of observations and experiments that, if performed, would render a consistent estimate feasible. Design Networks expand the IDs framework to address explicitly experimental design and provide the semantics to discuss how design can assist identification, and when and how one can identify causal effects. Incorporating nodes for experimental design decisions is useful in demonstrating their impact on the graphical structure and

on the 'data structure' derived from it. Certain policy assignment mechanisms, such as randomised cluster allocation, will add 'extra' independencies to the ID, defining a new collection of conditional independencies. To make a causal inference of F_T, it is important that we consider the mechanisms producing the data (d_E). Furthermore, we need to differentiate between policy interventions we want to evaluate, $F_T \in D_T$, and experimental allocations, $A_T \in D_E$. The relevance of D_E in assisting the identification and comparison of different mechanisms d_E^* in terms of identifiability can be addressed using DN for diverse types of assignment. Design networks were introduced for cluster allocation, and Spiegelhalter (2001)'s Bayesian hierarchical model was extended to include causal interpretation and used to illustrate a causal analysis of a simplified version of the Progresa programme.

When cluster allocation is done randomly, two-level Bayesian hierarchical models could be used to obtain the effect of F_{Tj} on Y_{jk} and the relevant coefficient can 'safely' be given a causal interpretation as an 'overall' effect. For the non-random case the analysis will need the conditioning on the 'relevant' background variables. Cluster allocation might help identifying individual-policy effects in certain cases.

Design of experiments within the Bayesian decision theoretic approach has been studied broadly in the literature; however, not in terms of causal reasoning and identifiability. In most of the literature on Bayesian experimental design the discussions have been limited to a) a set of options which include the choice of the levels of treatment and the number of repetitions within each level, and b) to a utility function usually defined in terms of minimising the posterior variance of estimates (or maximising entropy), which is an important issue. However, an experiment that overlooks identification could lead to the wrong conclusions if causal analysis is of interest. Causal inference imposes an extra criterion for the evaluation of the designs. This work extends the on-going discussion to a more general setting where the set of options is extended to include decisions about policy allocation and recording mechanisms and where the utility function is allowed to include a measurement of identifiability.

Appendix A. Design Networks: General remarks

By allowing the 'idle' system in Dawid (2002) to refer to any experimental system, the list of propositions in this section could be derived directly or are analogous to the results presented in Dawid (2002).

In general, we will say that the 'causal' effect of T on Y is identifiable *directly* from the available (experimental) data collected under $D_E = d_E$, if learning the value of F_T (i.e. learning if the future policy was set to a value or left to vary 'naturally') does not provide any 'extra' information about the response variable Y given the value of T and experimental conditions $D_E = d_E$ (*i.e.* if $(Y \perp\!\!\!\perp F_T \mid T)_{D_E}$) then $p(y \mid t', F_T = do(T = t'); D_E = d_E) = p(y \mid t', F_T = \emptyset; D_E = d_E)$. Note that this will hold (or not) regardless of $R(B)$.

Definition (Direct identifiability) *The 'causal' effect of T on Y is identifiable* directly

from available (experimental) data collected under $D_E = d_E$, *if* $(Y \perp\!\!\!\perp F_T \mid T)_{d_E}$. *Then*

$$p(y \mid F_T = do(T = t'); D_E = d_E) = p(y \mid t'; D_E = d_E).$$

This will imply that the conditional distribution $p(y \mid t')$ that is extracted from data generated according to d_E can be used directly to estimate the target causal effect $p(y \mid F_T = do(T = t'))$ regardless of the recordability or the actual values of B.

Let $d_{E1} = \{A = a_1; R(B) = r_1\}$ and $d_{E2} = \{A = a_2; R(B) = r_2\}$ be two experimental design interventions.

Proposition *If two experimental DNs under allocations defined by* $A = a_1$ *and* $A = a_2$ *share the same conditional independencies* $S_{a_1} = S_{a_2}$, *and* $(Y \perp\!\!\!\perp F_T \mid T)_a$ *holds for* $a = a_1, a_2$ *then experiments* d_{E1} *and* d_{E2} *share the same 'direct identifiability' status for the causal effect of* T *on* Y *defined by intervention* F_T *for any recording mechanisms* r_1 *and* r_2 *Thus, the choice of assignment mechanism (between* a_1 *and* a_2*) is said to be irrelevant to obtaining direct identification.*

For instance, if a_1= pure random allocation with probability θ_1^* and a_2= pure random allocation with probability θ_2^*, such that $0 < \theta_T^* < 1$ for all t^*, both assignments lead to direct identifiability. Then, regardless of the background variables recorded, the choice between a_1 and a_2 is irrelevant for identification purposes. Both allocations might be different in terms of a balanced sample and the variance and efficacy of the estimates, but this is regarded as a secondary goal of the choice of experiment.

Proposition *If direct identifiability holds for* a_1, *i.e.* $(Y \perp\!\!\!\perp F_T \mid T)_{a1}$, *but not for* a_2, *then the choice between* $d_{E1} = \{a_1, r\}$ *and* $d_{E2} = \{a_2, r\}$ *is not irrelevant for direct identifiability.*

An example of this is when a_1 = pure random allocation and $a_2 = \emptyset$. Although naturally it could be observed that $(T \perp\!\!\!\perp B)_\emptyset$ holds, direct identifiability will usually not hold for $a_2 = \emptyset$. So, the choice between performing a randomised experiment and observing the original mechanism is not irrelevant for the isolation of effects and their direct identification.

Direct identifiability of the causal effect implies assuming $(Y \perp\!\!\!\perp F_T \mid T)_{d_E}$, which is a very strong assumption that usually will not hold when observational studies or imperfect experiments take place. However, we might be ready to assume that for a set $B^* \subseteq B$ where $B^* \perp\!\!\!\perp F_T$, conditional on B^* the learning of F_T is irrelevant for the response, such that $(Y \perp\!\!\!\perp F_T \mid T, B^*)_{d_E}$ and then

$$p(y \mid t', B^*, F_T = do(T = t'); D_E = d_E) = p(y \mid t', B^*, F_T = \emptyset; D_E = d_E)$$

so we could 'substitute' the future intervened probability with the 'natural experimental' distribution available from the data.

Definition (Conditional identifiability) *The 'causal' effect of T on Y* conditional *on B^* is identifiable directly from the available (experimental) data collected under $D_E = d_E$, if $(Y \perp\!\!\!\perp F_T \mid (T, B^*))_{d_E}$. Then*

$$p(y \mid F_T = do(T = t'), B^*; D_E = d_E) = p(y \mid t', B^*; D_E = d_E).$$

Notice that conditional identifiability alone does not imply that procedures like the back-door formula can be used to calculate the overall effect of T on Y, which needs condition $(B^* \perp\!\!\!\perp F_T)_{d_E}$ to hold as well.

Proposition *If direct identifiability does not hold for a_1, $d_{E1} = \{a_1, r_1\}$, then the choice of the recording mechanism $R(B) = r_1$ in the experimental design defined by $d_{E1} = \{a_1, r_1\}$, is relevant for obtaining 'adjusted' identifiability.*

When identifiability cannot be obtained directly from the data defined by D_E, identifiability can still hold for a particular configuration of $R(B)$. Then, we say that the causal effect is identifiable through an '*adjustment*' procedure, and this leads to another definition.

Definition (Adjusted identifiability) *The 'causal' effect of T on Y is identifiable through an* 'adjustment' *procedure if*

$$p(y \mid t', F_T = do(T = t'); D_E = d_E) = h(y, t', B^* \mid D_E = d_E)$$

such that $R(B^*_q) = 1$ for all $B^*_q \in B^* \subseteq B$ and h is a function of known probabilistic distributions of recorded variables under d_E.

If we had a complete picture of the systems, then we could observe all background variables B and their influences and no unobserved or latent variables would exist. Then, $R(B^q) = 1$ would be plausible for all q and we would always be able to find a combination $R(B^*)$ such that $p(y \mid F_T = do(T = t'); R(B^*))$ would be identifiable through an adjustment procedure. However, our vision as experimenters willing to collect data is much narrower and is restricted to a partial view in which not all background variables are accessible and not all settings r are accessible. Nevertheless, we can still choose among different settings of $R(B)$. The design network representation permits us to evaluate identifiability for different choices of the recording mechanism $R(B)$. In consequence, it could assist the experimenter to choose among a possible set of recording settings, r, in order to assist identification of the effect of interest. In a first raw classification, recording mechanisms could be classified into those for which adjusted identifiability holds (h exists) and those for which the effect remains unidentifiable. Different recordings might have further consequences in the inference of causal effect; however, in terms of identifiability, the choice between two recordings that ensure adjusted identifiability is irrelevant. Thus we have,

Proposition *Let $d_{E1} = \{a, r_1\}$ and $d_{E2} = \{a, r_2\}$ be two experimental conditions such that direct identifiability does not hold for the policy assignment mechanism defined by $A = a$, and where r_1 and r_2 represent recording mechanisms in which collections B_1^* and B_2^* are recorded respectively. If functions h_1 and h_2 of known probabilistic distributions can be found for both recordings r_1 and r_2, then d_{E1} and d_{E2} are said to be equivalent for adjusted identifiability and the choice between recording mechanisms r_1 and r_2 is irrelevant for identifiability.*

In the case where neither h_1 nor h_2 can be found, the choice of r_1 and r_2 is also irrelevant, but in this case both recordings produce non-identifiability. However, when h_1 exists, but h_2 does not, then d_{E1} and d_{E2} do not share the same identifiability status, as the target causal effect of future intervention F_T can be obtained through an adjustment procedure for d_{E1} but it is not identifiable under d_{E2}.

When 'adjustment' is needed, some closed-forms for the function h have been given. The 'back-door' criterion (Pearl 1993), the 'front-door' formula (Pearl 1995) and the 'G-computation' formula (Robins 1986) are examples of criteria and formulae that imply the use of background variables to obtain 'adjusted' estimates and are all particular cases of functions h. A broader discussion of these criteria under different approaches can be found in Pearl (2000), Dawid (2002) and Lauritzen (2001).

If we can assume we are in a situation represented by a system in which potential confounders exist, pure random allocation will provide a data-generating mechanism that ensures direct identifiability of the effect of interest. In this case, we are performing *control via design* of the potential confounders that might be affecting the choice of policy, like politicians' preferences to benefit some particular communities. A generating mechanism that can only provide identifiability through an 'adjustment' formulation will correspond to a situation in which potential confounders have to be *controlled via analysis*.

Even if functions h_1 and h_2 can be found for d_{E1} and d_{E2} and adjusted identifiability can be obtained, further considerations are necessary when choosing an experiment. If r_1 and r_2 are such that $B_1 \subseteq B_2$ then d_{E1} will be generally preferred to d_{E2}, as recording a larger data set implies a more costly implementation and storage. In this sense, we would like the set of recorded variables to be minimal, but sufficient for identifiability. Definitions of sufficient sets have been made (see Lauritzen 2001; Dawid 2002; Pearl 2000). Functions h_1 and h_2 might be found for sets $B_1 \neq B_2$ where neither of them is a subset of the other. In any case, functions h_1 and h_2 might not have the same form and particular estimates might not be equally efficient when derived from h_1 than when derived from h_2, reflecting the loss of information associated with our restricted partial views determined by r_1 and r_2. An example of this, for the front-door formulation, can be found in Lauritzen (2001).

When direct or adjusted identifiability holds, the design d_E is ignorable. However, as Rubin (1978) notes, not all ignorable mechanisms can yield data from which inferences for causal effects are insensitive to prior specifications. Direct identifiability gives a situation where effects are insensitive to the specification of prior distributions of the

data. However, this will not hold for adjusted identifiability where the causal effect is dependent on the prior distribution of background variables, $P(B)$.

A.1 The positivity condition

In general, the set $\{t^*\}$ of intervention-values assigned through A_T is not necessarily the same as the set of future-policy-values $\{t'\}$ defined by F_T. In order to be able to evaluate the causal effects of intervention $F_T = do(T = t')$, we need treatment t' to be observed under experimental conditions d_E. So, we need $p(t' \mid B^*, F_T = \emptyset; D_E = d_E) > 0$. This requires that treatment assignment mechanism A_T includes t' as one of its allocated values. In other words, this requires that $t' \in \{t^*\}$. In a prospective study, this condition will usually hold. However, when data has been already collected, we might face the case where $t' \notin \{t^*\}$. In this case, we would only be able to use the data available if we could make some parametric assumptions for $p(Y \mid T, \cdot)$ before the policy effects can be identified. In general, if all the relevant information needed to evaluate the causal effect is encoded in a function $\tau(\eta)$ of η, and the experimental data provides us with information about $\lambda(\eta)$, it will suffice if $\tau(\eta) \subseteq \lambda(\eta)$. In this case, predictively,

$$p(y \mid F_T = do(T = t'); d_E) = \int_{\eta} p(y \mid F_T = do(T = t'), \tau(\eta)) p(\lambda(\eta) \mid \Delta_{d_E}; d_E) d\eta.$$

If different policies represent categorical variables, this could be difficult. In the FS example, imagine the two supplements provided in the experiment through assignment $A_T = do(T = t^*)$ are from different brands, say brand A (t^*_A) and brand B (t^*_B), and the future policy consists of providing a food supplement from brand C (t'_C). The data available, no matter how the actual assignment was made (random or not), will hardly be useful to conclude anything about the effect of food supplement C. However, imagine, that we have a measure in terms of the calorie intake that each supplement provides and that $t^*_A = 100$ kcal and $t^*_B = 300$ kcal, and we know that supplement C has 200 kcal, then if we are ready to assume that η contains a summary of the effect on weight per each increase of one kcal, then we would be able to estimate its effect.

A.2 Choice of experimental design

The problem of choosing an experiment has been set in Bayesian decision theory using decision trees (see Lindley 1971; Bernardo and Smith 1994). A DN could be viewed as its corresponding ID, allowing us to represent influences between decisions and random nodes. Optimality can be defined in various ways, and qualities for the distributions of estimators, such as minimum variance, are desirable (see Chaloner and Verdinelli 1995). Here we focus on the isolation of the target causal effect and thus on its identification. The efficacy of experimental design interventions D_E could then be measured in terms of making the (causal) effects of $F_T = do(T = t')$ identifiable and then two (or more) experiments can be compared in these terms, and among the experimental decisions D_E we choose the one with highest utility. 'Pure' (i.e. non-stratified) individual random allocation contrasted with the 'no experiment' choice (i.e. observational

data) is used to introduce this procedure. When the policy assignment is done through random allocation, two control actions are performed: randomisation and intervention. So treatment t^* is done, $A_T = do(T = t^*)$, according to a probability distribution θ_T^* totally fixed and controlled by the experimenter through $A_{\theta_T} = do(\theta_T = \theta_T^*)$. Node A might be expanded to show explicitly the mechanisms underlying the assignment and the new independencies that might be introduced. This expansion involves parameter and intervention nodes that are included in an *augmented-extended design network*.

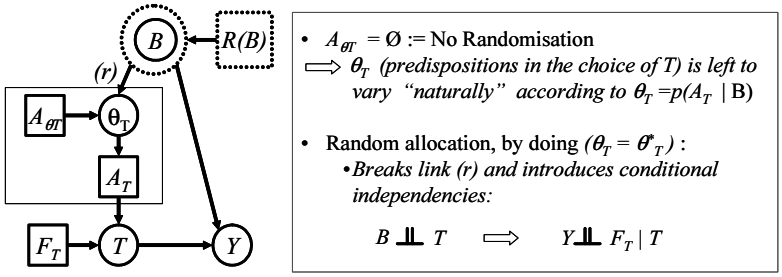


Figure 9: Augmented - Extended DN for random allocation

Experimental actions 'do' parameter nodes. Random allocation breaks the link (r) and therefore two experimental structures arise from this choice. For each design strategy $d_E^* \subset D_E$ taken we can obtain an *experimental DN* from which independencies can be easily read. These experimental DNs define the data structure or data pattern.

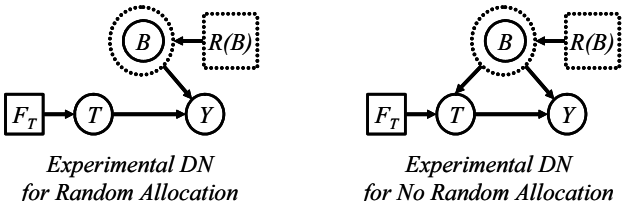


Figure 10: Experimental DN

Imagine we establish that the utilities associated with obtaining direct identifiability, adjusted identifiability and unidentifiability are given by U_D, U_A and U_U respectively. Then for the pure random allocation vs observational case, the four possible combinations of $(A, R(B))$ are shown in Table 3. Both experimental decisions that include random allocation, $A_{\theta_T} = do(\theta_T = \theta_T^*)$, have the same utility associated in terms of identifiability and are equivalent in these terms. However, performing an experiment (randomising and/or recording) will typically involve an associated cost that is not included here. The fact that $U_D \neq U_A$ (and actually we consider $U_D > U_A$) is due to the fact that the recording of B will increase the cost of the experiment and that $p(y \mid F_T = t', d_E = 3)$ is sensitive to the specification of prior distributions of the data. The choice of θ_T^* could have an effect on the efficacy of the estimators as it could af-

fect the balance of the experiment, but the actual value θ_T^* does not affect the graph independence structure and the identifiability status derived from it.

Experimental Decisions			Design Consequence	Utility(d_E)
D_E	A_{θ_T}	$R(B)$	$p(y \mid F_T{=}\, t'; d_E)$	U
1	random	1	direct identifiability	U_D
2	random	0	direct identifiability	U_D
3	$\emptyset$	1	adjusted identifiability	U_A
4	$\emptyset$	0	No identifiable	U_U

Table 3: Choice of experimental design for pure random example

A.3 An influence diagram for policy analysis

Figure 11 shows an influence diagram of the (simplified version of the) complete system for policy analysis. As before, the policy variable is denoted by a decision node T that has been augmented to make explicit policy intervention F_T. When the policy is defined through policy intervention decisions D_T, it can contain a collection of actions G that are triggered when intervention $F_T = do(T = t')$ takes place. Actions G can be contingent on a set of observed variables Z and are children nodes of T. The definition of possible structures and correspondent formulae for the calculation of the overall effect of intervention F_T in Y through actions G have been discussed in Madrigal (2004). The policy assignment mechanisms are contained in decision node A, which could be influenced by some background variables B. Both, the policy assignment mechanisms, A, and the recording mechanisms of B, $R(B)$, are defined as part of the experimental decisions D_E.

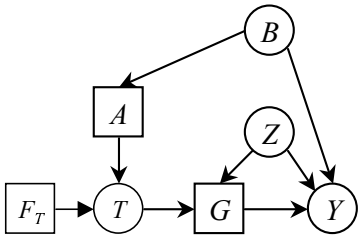


Figure 11: Complete ID for policy analysis

This simple structure shows how the two sets of decisions (namely, policy intervention decisions D_T and experimental decisions D_E) could be represented in the same graph. A more realistic graph should include some possible links between the background variables B and the variables Z in which actions G are contingent on, and possibly some type of influence of Z in the policy assignment mechanism. It is important to use the policy makers' expertise and knowledge to be able to represent in the influence diagram a most accurate version of the 'real' system with all possible influences. This will assist our

causal inferences conclusions and help the choice of actions.

References

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. John Wiley & Sons. 584

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons. 578

Chaloner, K. and Verdinelli, I. (1995). "Bayesian Experimental Design: A Review." *Statistical Science*, 10: 273–304. 562, 584

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems*. New York, NY: Springer-Verlag. 560

Dawid, A. P. (1979). "Conditional Independence in Statistical Theory." *Journal os the Royal Statistical Society (series B)*, 41: 1–31. 561

— (2000). "Causal Inference Without Counterfactuals." *Journal of the American Statistical Association*, 95(450): 407–424 (C/R: p424–448). 559, 575

— (2002). "Influence diagrams for causal modelling and inference." *International Statistical Review*, 70(2): 161–189. 557, 558, 559, 560, 561, 563, 565, 579, 580, 583

Hayes, R. J., Alexander, N. D. E., Bennett, S., and Cousens, S. N. (2000). "Design and Analysis Issues in Cluster-randomized Trials of Interventions against Infectious Diseases." *Statistical Methods in Medical Research*, 9(2): 95–116. 573, 574

Heckerman, D. and Shachter, R. (2003). "Discussion in Pearl, J. Statistics and Causal Inference: A Review." *Test.*, 12(2): 327–331. 560

Howard, R. A. and Matheson, J. E. (1981). "Influence Diagrams." *Principles and Applications of Decision Analysis. Menlo Park, CA*. 559

Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs.*. New York: Springer. 559

Koepsell, T. P. (1998). "Epidemiologic Issues in the Design of Community Intervention Trials." In Brownson, R. C. and Petitti, D. B. (eds.), *Applied Epidemiology, Theory and Practice*. New York: Oxford University Press. 573

Lauritzen, S. L. (2001). "Causal Inference from Graphical Models." In Barndorff-Nielsen, O. E., Cox, D. R., and Klüppelberg, C. (eds.), *Complex stochastic systems*, 63–107. Chapman & Hall Ltd. 558, 561, 583

Lindley, D. V. (1971). *Making Decisions*. Chichester: Wiley-Interscience. 584

Longini, J., Ira M., Sagatelian, K., Rida, W. N., and Halloran, M. E. (1998). "Optimal Vaccine Trial Design When Estimating Vaccine Efficacy for Susceptibility and Infectiousness from Multiple Populations." *Statistics in Medicine*, 17: 1121–1136 (Corr: 1999 V18 p890). 574

Madrigal, A. M. (2004). "Evaluation of Policy Interventions under Experimental Conditions Using Bayesian Influence Diagrams." Ph.D. thesis, University of Warwick,UK. 563, 586

— (2005). "Design Networks, Policy Interventions and Causal Inference." Technical Report 448, University of Warwick,UK. 568

Madrigal, A. M. and Smith, J. Q. (2004). "Causal Identification in Design Networks." In R. Monroy, E. E. (ed.), *MICAI 2004: Advances in Artificial Intelligence, (Conference Proceedings)*, 517–526. LNAI 2972, Springer-Verlag. 557, 559, 569

Oliver, R. M. and Smith, J. Q. (1990). *Influence Diagrams, Belief Nets and Decision Analysis*. John Wiley & Sons. 559

Pearl, J. (1993). "Comment on "Graphical Models"." *Statistical Science*, 8: 266–269. 559, 561, 583

— (1995). "Causal Diagrams for Empirical Research." *Biometrika*, 82: 669–688 (Disc: p688–710). 559, 583

— (2000). *Casuality: Models, Reasoning, and Inference*. Cambridge University Press. 558, 559, 561, 578, 583

Robins, J. (1986). "A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period - Application to Control of the Healthy Worker Survivor Effect." *Mathematical Modeling*, 7: 1393–1512. 559, 583

Rosenbaum, P. (2002). *Observational Studies.*. New York: Springer Verlag. 562

Rubin, D. B. (1978). "Bayesian Inference for Causal Effects: The Role of Randomization." *The Annals of Statistics*, 6: 34–58. 558, 583

— (1980). "Comments on "Randomization Analysis of Experimental Data: The Fisher Randomization Test"." *Journal of the American Statistical Association*, 75: 591–593. 575

Spiegelhalter, D. J. (2001). "Bayesian Methods for Cluster Randomised Trials with Continuous Responses." *Statistics in Medicine*, 20: 435–452. 559, 571, 575, 580

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction and Search*. Cambridge, MA: MIT Press. 559

Turner, R. M., Omar, R. Z., and Thompson, S. G. (2001). "Bayesian Methods of Analysis for Cluster Randomised Trials with Binary Outcome Data." *Statistics in Medicine*, 20: 453–472. 571

Wu, C.-F. and Hamada, M. (2000). *Experiments: Planning, Analysis, and Parameter Design Optimization*. John Wiley & Sons. 562

Acknowledgments

This work was partially funded by CONACYT. The author thanks PROGRESA for giving access to databases for this research.