Research Article

Philip Dawid*

Potential outcomes and decision-theoretic foundations for statistical causality: Response to Richardson and Robins

https://doi.org/10.1515/jci-2023-0058
received October 02, 2023; accepted March 7, 2024

Abstract: I thank Thomas Richardson and James Robins for their discussion of my article, and discuss the similarities and differences between their approach to causal modelling, based on single world intervention graphs, and my own decision-theoretic approach.

Keywords: causal inference, decision analysis, distributional consistency, extended conditional independence, *g*-computation, graphical model, ignorability, intention to treat, potential outcome, single world intervention graph

MSC 2020: 62A01, 62C99

1 Introduction

I am indebted to Richardson and Robins [1], henceforth RR, for their serious and detailed engagement with the ideas and material in my article (Dawid [2], henceforth D21). It is particularly valuable that they highlight the similarities and differences between my decision-theoretic (DT) approach and their own approach based on single world intervention graphs (SWIGs). Indeed, the similarities are manifold and the differences few and largely inconsequential. I will, however, concentrate here on these small differences, in the hope that this will illuminate the differences in our underlying world views.

In Section 2, I address some specific points raised by RR's discussion, and in Section 3, I respond to various critiques they make of D21. Section 4 addresses RR's argument, an alternative to the one I gave in D21, and supplies some corrections to their analysis. Finally, in Section 5, I opine on the relative advantages and disadvantages of SWIG and DT representations.

Note: In the sequel, references to equations in RR are given in the form "equation (R1)," to those in D21 in the form "equation (D1)," and to those in the present article in the form "equation (1)," with a similar convention for other references.

2 Some specific points

(1) RR's introduction says that I

> *aim to develop a graphical framework* for causal models.

* **Corresponding author: Philip Dawid,** Statistical Laboratory, University of Cambridge, Cambridge CB3 0WB, United Kingdom, e-mail: apd@statslab.cam.ac.uk

Not exactly. The fundamental idea of DT is that we can express causal properties by means of extended conditional independence (ECI) assertions, involving both stochastic variables and non-stochastic intervention indicators; my article aimed to develop arguments to support such assertions. These arguments can always be expressed and developed non-graphically, using the purely algebraic theory of ECI. It is true that graphical representations are incredibly useful and near-ubiquitous, which is why I devoted much attention to them in my article; but the underlying theory does not require that we have such a representation (which is, in any case, not always available).

(2) Footnote R1. I am disappointed that RR choose to perpetuate the prevalent but highly misleading terminological confusion between concepts that relate to distinct rungs of the "ladder of causation" [3]. Use of the same term "counterfactual" to denote totally distinct things is particularly dangerous, and I have often found myself confused, in reading the literature, as to which concept is intended. A recent workshop I attended was entitled "Counterfactual Prediction," but was simply about using data to make forecasts for new patients, under various treatments. Nothing about this runs counter to known facts, and it was thus firmly positioned on Rung 2 of the causal ladder, which concerns the effect of a new, actual or hypothetical, intervention on a system – so not involving a contradiction with any known facts, and not meriting the description "counterfactual." On Rung 3, by contrast, we ask genuinely counterfactual questions about what might have happened in a particular case if – in contradiction to the known facts – an action other than the actual one had been performed. The contrast between these tasks is illuminated in Dawid [4], Dawid and Musio [5], where it is shown that different mathematical frameworks are required to formalise them. In particular, while potential outcomes can be used at either of these levels, they are totally inessential for Rung 2 – for all that this accounts for by far the largest share of their current use – but seem unavoidable for Rung 3.

Both my own article and that of RR stand firmly on Rung 2, and involve no genuinely counterfactual considerations. That is why I have been able to dispense entirely with potential outcomes, while still having a theory that – as RR convincingly show – is essentially isomorphic to theirs, where they have opted to employ them.

(3) In §R2, referring D21's "hypothetical distributions," RR say:

> there is no requirement that these distributions live on the same probability space.

The various regime distributions all relate to identical variables, and thus do live on a single space, though admittedly it is not under the control of a single probability measure, so not a probability space. This is the same structure we are familiar with in the context of a parametric statistical model.

In RR's approach, each variable is indexed with one or more actions, leading to a proliferation of variables. These variables can, if so desired, be considered as having an overall joint distribution – so "living on the same probability space"; but it is only margins of this joint distribution, which are just my regime distributions, that are relevant. In particular, the dependence, in the overall joint distribution, between versions of the same variable labelled by different interventions is both unknowable and (fortunately!) irrelevant. So the advantage of having a single probability space is lost on me. A similar approach, if applied for a parametric statistical model for a variable X with parameter θ, would involve constructing an expanded collection of variables $\{X(\theta)\}$, one for each value of θ, all having a joint distribution – of which only the margins are of interest. Why would one ever do such a thing?

(4) Section R3: Labelling issues and individual effects.

RR describe three distinct ways – uniform, temporal, and causal – in which variables in SWIGs may be labelled by actions. They say

> although we may wish to adopt the additional equalities between potential outcomes that are implied by the temporal and/or causal relationships, our results do not require these equalities

– indicating that it really does not make any difference which scheme is employed,

RR opt to work with uniform labelling. Indeed, either of the other schemes would not be representing a single world. Thus, as they point out, Figure R1(d) represents the case that C would take the same value in the distinct worlds corresponding to actions $(a = 0, b = 1)$ and $(a = 1, b = 1)$. While such "absence of individual effects" assumptions may have some intuitive appeal, they add nothing to the analysis.

My personal view is that the very concept of an "individual effect" is not merely unnecessary but metaphysical – and not in a good way [6]. In particular, the emphasis, in the potential outcome approach, on necessarily unknowable[1] individual effects, takes one down a blind alley, which one has to re-emerge from before anything useful can be done.

(5) Footnote R12:

> In Dawid (2021, Figure 15), two conditions are stated as supporting g-computation. The first of these is correct, but the second should be $Y(x_0, x_1) \perp\!\!\!\perp X_0$, not $Z(x_0) \perp\!\!\!\perp X_0$.

I gratefully accept RR's correction, based on Robins [8].[2] This requires the following amendments to D21:

Equation (D78) Replace by:

$$Y(x_0, x_1) \perp\!\!\!\perp X_0.$$

Equation (D81) Replace by:

$$Y \perp\!\!\!\perp X_0^* | (F_0 = x_0, F_1 = x_1)$$

(noting that the dotted arrow from X_0^* to X_0 disappears).

(6) Footnote R20: I thank RR for catching my careless error.

(7) Section R6: The role of fictitious independence.

Note that instead of equation (R57), I had $\perp\!\!\!\perp_{i=1}^{k} F_i$, which is a preferred notation.

I stand corrected by RR's analysis here and am indeed embarrassed that I myself have fallen foul of the very fallacy I identified and analysed in Dawid [10] and Dawid [11, §8.1]. Fortunately, this is easily rectified with an additional assumption such as in §R6.2, or, more straighforwardly (if more restrictively) *variation independence*, as described in Remark 1. In partial deflection of their criticism I point out that (as mentioned in Remark 1; see also footnote 11), RR also rely on just such an implicit assumption.

3 Response to RR's "critique of Dawid's proposal"

3.1 §R4.1

RR argue against my aim of expressing causal properties by means of augmented DAGs (or, more generally, ECI statements) including regime indicators but without intention-to-treat (ITT) variables. They ask "why it is necessary to introduce the ITT variables in the first place?"

The first point they make in this section was previewed in their introduction:

> ITT variables are necessary and important in order to encode the notion of ignorability and the effect of treatment on the treated.

But this is not so.

Ignorability In the DT approach, ignorability is directly encoded by ECI, *e.g.*, $Y \perp\!\!\!\perp F_T | T$, without any need to consider ITT variables. The sole purpose of introducing ITTs in D21 was to support one possible argument that

[1] Because of the "fundamental problem of causal inference" [7].

[2] See Dawid and Didelez [9], §8, for a general DT formulation and analysis, and §10.2 for its relation to the potential response approach of Robins [8].

(when appropriate) might be made in justification of such assertions – and thus to justify (when appropriate) the use of an augmented DAG, without explicit ITT nodes, to represent and manipulate causal relations.

Effect of treatment on the treated (ETT) While consideration of ITTs is one way of thinking about ETT, it is not essential: ETT can be meaningfully and helpfully interpreted in ways that do not involve ITTs at all. Thus, [12, §34.4.1] suppose there is an unobserved sufficient covariate ("confounder") U (so that $U \perp\!\!\!\perp F_T$ and $Y \perp\!\!\!\perp F_T | (U, T)$). The *specific causal effect*, in the subpopulation of individuals having $U = u$, is given by $\mathrm{SCE}(u) = \mathrm{E}(Y | U = u, F_T = 1) - \mathrm{E}(Y | U = u, F_T = 0)$. Then, ETT is defined as $\mathrm{E}\{\mathrm{SCE}(U) | T = 1, F_T = \varnothing)$: the average, in the observational regime, of the specific causal effect for those individuals who, in fact, receive the active treatment. (This generalises the ITT analysis, which is recovered on taking U to be the ITT variable.) A variant approach [12, §34.5.1] contemplates the policy of making the treatment available to a previously untreated population and defines ETT as the average improvement in response, per individual choosing to take the treatment. Both these approaches lead to exactly the same form for ETT as that derived from the ITT analysis.

RR go on to argue, by means of an example, that an augmented DAG, without ITTs, cannot distinguish a "genuine" causal relationship from a "spurious" one. But, as with any model, it is essential to keep in mind the real-world characteristics that the ingredients of the model are intended to represent. In particular, the states of the decision node index the data distributions associated with carefully described hypothesised interventions.

The "spurious" case they discuss is represented[3] by Figure R4(b). This involves particular "fat hand" interventions and describes their effect on the response Y. In the – admittedly implausible – case that I myself was considering undergoing just such fat-hand interventions, this might describe my own decision problem. And it would then, indeed, be the case (assuming I could accept the appropriate exchangeability assumptions) that I could consider the observational distributions of Y given T as germane to that problem, and so alternatively represent the problem as in Figure R4(a). As RR say, correctly, "the causal diagram shown in (a) cannot be refuted." But in this case, in Figure R4(a), the states of F_T would represent the "fat hand" interventions, and, with this interpretation, the problem can still appropriately be called "causal." Such a representation would be distinguishable from what RR consider to be a "genuine" causal case, likewise represented by Figure R4(a), but where the states now represent different, "surgical," interventions.

As mentioned earlier, the introduction of ITT variables in D21 was made to support ignorability assertions, here $Y \perp\!\!\!\perp F_T | T$, in particular kinds of problems. But they are not essential, and all that RR's example demonstrates is that ignorability can hold even when the argument based on ITTs fails. This does not make such a problem any less genuinely causal. Moreover, while the DT description is unproblematic, I do not see how a SWIG approach could represent ignorability in such a problem.

3.2 §R4.2.1

Here, RR argue that I could (should?) have regarded variables such as T^* appearing in different regimes as identical, not merely identically distributed. But as I mentioned earlier when discussing labelling issues, *while "absence of individual effects" assumptions may have some intuitive appeal, they add nothing to the analysis.*

I am bemused by RR's complaint that my DT account "leads to an unnecessary multiplicity of random variables," when in their SWIG approach, even using the relatively lean causal labelling scheme, every single variable is replaced by a host of potential variables (one for each combination of actions that could affect it).

[3] Well, not entirely, since the assumed underlying distribution is unfaithful to the graph, having the property $Y \perp\!\!\!\perp F_T | T$ even though that is not represented in it.

3.3 §R4.3

While, as I argued in D21, it is extremely useful to think about ITT variables when trying to justify ignorability assumptions, I do not agree with RR's preference to retain the intention-to-treat variable T^* in the final augmented DAG representation, while omitting the received treatment variable T. First, as I have explained in Section 3.1, T^* is not needed to "rule out spurious invariance" (which I do not consider spurious); nor is T^* essential for defining the effect of treatment on the treated. (However, I do not rule out that there may be some special cases where it is helpful to retain T^*, as well as T, in the final model – in which case by all means do so.)

Second, again as mentioned in the case of "spurious causation," in some cases, properties such as ignorability can be meaningfully justified, and again expressed by $Y \perp\!\!\!\perp F_T | T$, even when no argument involving ITT variables is available – in such a case T is essential, while T^* is a red herring.

Another case is where causality is understood as a property of invariance across differing contexts [13], as in considerations of transportability and external validity [14]. For example, a medical device may have the same probability of registering a positive result, given whether or not a patient has a certain condition, irrespective of who it is used on, or in which hospital. This can still be encoded as $Y \perp\!\!\!\perp F_T | T$, where Y denotes the response, T denotes the condition, and F_T now labels the context. Considerations of ignorability and intention-to-treat are simply not relevant here, and there is no SWIG representation.[4]

3.4 §R5

At (R36), and again in Figure R7, RR point out that a contextual independence, here $Y \perp\!\!\!\perp F_T | M, F_T \neq \varnothing$, is not implied by the (non-contextualised) conditional independence assumptions that they label A and B. This is so, but I can't see why it is a problem. The full set of required assumptions includes, as well as A and B, the description, in (D33), of how T depends on T^* and F_T. Taken all together, these imply all relevant contextual independencies. Moreover, the inclusion of dashed edges in an augmented DAG allows such properties to be derived directly from the graph, bypassing algebraic manipulations.

4 RR's alternative argument

The major part of RR is devoted to an argument alternative to the one that I presented in Section 2 and Appendix A of D21 – using different assumptions and arguments, but leading to the same conclusion. RR very helpfully conduct this argument twice, first (in Sections R3.2–R3.7) in the language of SWIGs, and again (in Sections R5.1–R5.6) in my own DT language. I like this alternative development and especially appreciate the two parallel descriptions: it is illuminating to compare different ways of looking at the same thing. In particular, the twin analyses demonstrate the close correspondence between our approaches, such differences as there are being largely (though not entirely) notational.

The argument presented by RR appears basically correct, but certain details of it, particularly in its DT version, require clarification and amplification.

4.1 Distributional consistency

Following their introduction of their own version of "distributional consistency" in Definition R2, RR give a variation on this definition in terms of a "dynamic regime" g_i^*, supposed to have the effect of setting an

[4] However [15,16], SWIGs with their ITT variables (which could alternatively be given DT formulations) can be used to model the process by which a study group is selected from a population, and so relate these two contexts.

intervention target, B_i (in my unstarred notation), to agree with its "natural value" (which I conceive of as an "intention-to-treat," ITT, variable, B_i^*). I do not see how this advances the argument. In particular, it generates still further proliferation of potential variables, which now require g_i^* as an additional argument. And in order for this to work at all, a variety of additional conditions, as detailed in footnote R8, are required, detracting considerably from this approach – which is, in any case, totally superfluous.

When RR introduce distributional consistency in the DT context, in Definition R13, they do so solely in terms of g_i^*. But the description of what g_i^* does is indistinguishable from how the idle regime $\varnothing$ operates. So the two states of F^* embody a distinction without a difference and collapse into one – making the interpretation of (R38) problematic,[5] and rendering the argument from (R39) to (R40) decidedly dodgy. In particular, the first equality again requires additional conditions, translations of those in footnote R8, which effectively beg the question. Fortunately, variables such as g_i^* are, again, entirely superfluous and – as RR themselves later acknowledge – DT distributional consistency can perfectly well be defined by the equality of (R39) and (R40), which I phrase as:[6]

Definition 1. (Distributional consistency) This requires that, for $B \in A$, $Y = V \backslash B$,

$$\Pr(Y = y, B = b | F_B = b, F_{A\backslash B}) = \Pr(Y = y, B = b | F_B = \varnothing, F_{A\backslash B}) \tag{1}$$

(where we do not distinguish between a variable and its singleton set).

Definition 1 is the direct DT translation of the SWIG-based Definition R2. Note that, in (1), we could set the value of $F_{A\backslash B}$ as $(F_C = c \neq \varnothing, F_{(A\backslash B)\backslash C} = \varnothing)$, for $C \subseteq A\backslash B$, showing more clearly the equivalence with (R39) and (R40). This simplification will be used without further comment in the sequel.

Remark 1. Note that it is implicitly assumed, here and in the sequel, that knowing the values (fixed or idle) of some intervention indicators (here $F_{A\backslash B}$) does not constrain the possible values of others (here F_B) – the property of *variation independence* [17]. This assumption – or a suitable weaker one, such as in (R64) – is also required throughout RR's arguments in §R3 for SWIGs,[7] as well as their DT §R5.

4.1.1 Relationship with D21's distributional consistency

Equation (1) is equivalent to the pair of properties:

$$\Pr(Y = y | B = b, F_B = b, F_{A\backslash B}) = \Pr(Y = y | B = b, F_B = \varnothing, F_{A\backslash B}), \tag{2}$$

$$\Pr(B = b | F_B = b, F_{A\backslash B}) = \Pr(B = b | F_B = \varnothing, F_{A\backslash B}). \tag{3}$$

Equation (2) is similar to my own definition of distributional consistency (Definition D2), applied under given interventions on some or all of the variables in $A\backslash B$. A difference is that I was implicitly considering Y to comprise "response variables" that could be affected by B, whereas RR also allow variables that are causally prior to B. This seems very reasonable and indeed essential if we have not yet introduced a causal ordering of the variables.

As for (3) (not in itself a "distributional consistency" property): because it involves the same value b for both B and F_B on the left, it is a weaker[8] version of

[5] In any case, (R38) is incomplete as formulated, since it does not specify the values of the regime indicators $F_{A\backslash(C\cup B_i)}$. Presumably, here and elsewhere in RR, any unmentioned regime indicators are implicitly supposed idle.

[6] Henceforth, unless otherwise indicated, I follow RR in dropping the * notation, and consider only the ITT variables associated with intervention targets.

[7] Where variation independence is equivalent to the existence of $p(V(a))$ in Equation (R1), for all $a \in \mathcal{X}_D$, for each subset D of A.

[8] Unless B is binary

$$B \perp\!\!\!\perp F_B | F_{A \setminus B}, \tag{4}$$

which extends (3) to allow $F_B = b' \neq b$ on the left. Condition (4) encodes the intuitively desirable property that (for any interventions on some or all of the other manipulable variables) the distribution of the ITT variable B (which is B^* in my own notation) is not affected by applying any intervention, or none, to its target (my unstarred B). Extending a remark in footnote R6, under the conditions of Lemma R8, the stronger property (4) will, in any case, hold.

4.1.2 Lemma R14

Because of its reliance on F_B^*, Lemma R14 is meaningless as stated. Its statement and proof should be replaced by a DT paraphrase of Lemma R3, as follows:

Lemma 1. *Distributional consistency implies that* (1) *continues to hold for B a general subset of A and $Y \subseteq V \setminus B$.*

Proof. Use induction on the cardinality of B. Write B as a disjoint union $B = D \cup E$, with E a singleton. Then, with $Z = V \setminus B$,

$$\begin{aligned} & \Pr(Z = z, B = b | F_B = b, F_{A \setminus B}) \\ & \quad = \Pr(Z = z, D = d, E = e | F_D = d, F_E = e, F_{A \setminus B}) \\ & \quad = \Pr(Z = z, D = d, E = e | F_D = d, F_E = \varnothing, F_{A \setminus B}) \end{aligned} \tag{5}$$

$$\begin{aligned} & = \Pr(Z = z, D = d, E = e | F_D = \varnothing, F_E = \varnothing, F_{A \setminus B}) \\ & = \Pr(Z = z, B = b | F_B = \varnothing, F_{A \setminus B}). \end{aligned} \tag{6}$$

Here, (5) follows from (1), and (6) by the inductive hypothesis. Finally, marginalise from Z to Y. □

4.2 RR's further argument

While the results in the remainder of §R5 are essentially correct, there are some deficiencies in the arguments employed.

4.2.1 Lemma R15

Again because of its reliance on F_B^*, Lemma R15 is meaningless as stated.[9] A suitable DT translation of Lemma R4 is

Lemma 2. *Let $B \subseteq A$, and $W \subseteq Y = V \setminus B$. Then, under distributional consistency,*

$$\Pr(Y = y | B = b, W, F_B = b, F_{A \setminus B}) = \Pr(Y = y | B = b, W, F_B = \varnothing, F_{A \setminus B}). \tag{7}$$

This follows directly on further conditioning (2) on W.

[9] However, this lemma does not appear to be used by RR in the sequel.

4.2.2 Lemma R16

The introduction of F_B^* in the proof of Lemma R16 is pointless: the passage from line 2 to line 5 is immediate from distributional consistency expressed as the equality of (R39) and (R40). To clarify, I re-express Lemma R16 and its proof as follows, where, by annotating an intervention variable with the check mark ˇ, we understand that it does not take value $\varnothing$.[10]

Lemma 3. *Let $B \subseteq A$ and $W \supseteq B$. Then under distributional consistency*

$$W \perp\!\!\!\perp \check{F}_B | F_{A \setminus B} \;\Rightarrow\; W \perp\!\!\!\perp F_B | F_{A \setminus B}. \tag{8}$$

Proof. Let w be a possible state of W, with projections w', w'' onto $W \setminus B, B$, respectively. Then,

$$\Pr(W = w | \check{F}_B = b, F_{A \setminus B}) \tag{9}$$

$$= \Pr(W \setminus B = w', B = w'' | \check{F}_B = b, F_{A \setminus B}) \tag{10}$$

$$= \Pr(W \setminus B = w', B = w'' | \check{F}_B = w'', F_{A \setminus B}) \tag{11}$$

$$= \Pr(W \setminus B = w', B = w'' | F_B = \varnothing, F_{A \setminus B}) \tag{12}$$

$$= \Pr(W = w | F_B = \varnothing, F_{A \setminus B}). \tag{13}$$

Here, (11) follows from the premise of (8) and (12) from Lemma 1. □

The next result is not explicit in RR, but is useful.

Corollary 1. *Furthermore, let $D \subseteq A$ be disjoint from B. Then,*

$$W \perp\!\!\!\perp (\check{F}_B, F_D) | F_{A \setminus (B \cup D)} \;\Rightarrow\; W \perp\!\!\!\perp (F_B, F_D) | F_{A \setminus (B \cup D)}. \tag{14}$$

Proof. Fix w. The premise of (14) implies that there exists a variable G, measurable with respect to $A \setminus (B \cup D)$, such that, for any value $b \neq \varnothing$ of F_B,

$$\Pr(W = w | F_B = b, F_D, F_{A \setminus (B \cup D)}) = G. \tag{15}$$

Then, the equality of (9) and (13) shows that (15) holds also for $b = \varnothing$, and the result follows. □

Corollary 2. *Lemma* 3 *and Corollary* 1 *continue to hold if some or all of the intervention indicators F_i in $F_{A \setminus B}$ are replaced by their checked versions $\check{F}_i$.*

4.2.3 Lemma R17

Again, the line in the proof of Lemma R17 involving F_B^* is superfluous and should be omitted. Our version of its statement is as follows.

Lemma 4. *Let $B \subseteq A$, and let Y and W be disjoint with $B \subseteq W$. Then, under distributional consistency,*

$$Y \perp\!\!\!\perp \check{F}_B | (W, F_{A \setminus B}) \Rightarrow Y \perp\!\!\!\perp F_B | (W, F_{A \setminus B}). \tag{16}$$

The proof is similar to that of Lemma 3.

[10] In D21, it was the variable itself, rather than its intervention variable, that was so annotated.

4.2.4 Lemma R19

RR’s proof of Lemma R19 is inadequate in a number of ways.

(R48) RR’s argument for equating (R47) and (R48) fails because the property

$$W_{\mathrm{pre}(i)\cup\{i\}} \perp\!\!\!\perp F_{A\setminus\mathrm{pre}(i)} | F_{A\cap\mathrm{pre}(i)}, F_A \neq \emptyset$$

does not satisfy the requirement “$B \subseteq W$” that would support direct application of Lemma R16. Instead, I supply the following argument – a DT analogue (notably missing from RR) of Lemma R8.

Let the ITT nodes be labelled $A_1, \ldots, A_k$ following the topological order, with associated intervention indicator nodes $F_1, \ldots, F_k$, respectively. For $r \leq s$, $F_{r:s}$ will denote the sequence $(F_r, \cdots, F_s)$, *etc.* Define $Z_r = (\mathrm{pre}(A_r), A_r)$.

Lemma 5. *For* $r = 1, \ldots, k$,

$$H_r : Z_r \perp\!\!\!\perp F_{r:k} | \check{F}_{1:r-1}. \tag{17}$$

Proof. We proceed by backward induction.

From Definition R18, we have

$$Z_k \perp\!\!\!\perp \check{F}_k | \check{F}_{1:k-1}. \tag{18}$$

Then, by Lemma 3, we have

$$Z_k \perp\!\!\!\perp F_k | \check{F}_{1:k-1}, \tag{19}$$

So H_k holds.

Now, suppose H_{r+1} holds. Then,

$$Z_r \perp\!\!\!\perp F_{r+1:k} | \check{F}_{1:r}. \tag{20}$$

Also, from Definition R18, we have

$$Z_r \perp\!\!\!\perp \check{F}_r | (\check{F}_{1:r-1}, \check{F}_{r+1:k}). \tag{21}$$

Now, fix a value $\check{f} \neq \emptyset$ of $F_{r+1:k}$ (and thus of $\check{F}_{r+1:k}$). By (20), for any possible values $\check{f}_{1:r}$ of $\check{F}_{1:r}$ and $f_{r+1:k}$ of $F_{r+1:k}$,[11]

$$\Pr(Z_r = z_r | \check{F}_{1:r} = \check{f}_{1:r}, F_{r+1:k} = f_{r+1:k}) = \Pr(Z_r = z_r | \check{F}_{1:r} = \check{f}_{1:r}, \check{F}_{r+1:k} = \check{f}). \tag{22}$$

Since, by (21), the right-hand side of (22) is a function only of $\check{f}_{1:r-1}$, the same holds for the left-hand side, i.e.,

$$Z_r \perp\!\!\!\perp (\check{F}_r, F_{r+1:k}) | \check{F}_{1:r-1}. \tag{23}$$

Then, H_r follows from Corollaries 1 and 2, and the induction is established. □

The following is immediate by marginalisation of (17) (where by F_S when $S \not\subseteq A$ we understand $F_{A\cap S}$):

Corollary 3. *Lemma* 5 *continues to hold if we replace* Z_r *by* $(W_i, W_{\mathrm{pre}(i)})$, *where* W_i *lies between* A_{r-1} *and* A_r *in the topological order. That is to say,*

$$(W_i, W_{\mathrm{pre}(i)}) \perp\!\!\!\perp F_{A\setminus\mathrm{pre}(i)} | \check{F}_{\mathrm{pre}(i)}. \tag{24}$$

Conditioning on $W_{\mathrm{pre}(i)}$ in (24), we deduce the equality of (R47) and (R48).

[11] Here, and in other similar arguments, we rely on Remark 1.

(R49) Contrary to RR’s assertions,

$$W_i \perp\!\!\!\perp F_{(A\cap \mathrm{pre}(i))\backslash \mathrm{pa}(i)} | F_{A\cap \mathrm{pa}(i)}, F_{A\cap \mathrm{pre}(i)} \neq \varnothing \tag{25}$$

does not follow from the local Markov property, and even if it were valid, it would not allow the application of Lemma R17.

Instead, I prove the following result, which implies the equivalence of (R47) and (R49).

Lemma 6.

$$W_i \perp\!\!\!\perp F_{A\backslash \mathrm{pa}(i)} | (W_{\mathrm{pre}(i)}, \check{F}_{\mathrm{pa}(i)}). \tag{26}$$

Proof. It does follow from the local Markov property that

$$W_i \perp\!\!\!\perp \check{F}_{\mathrm{pre}(i)\backslash \mathrm{pa}(i)} | W_{\mathrm{pre}(i)}, \check{F}_{\mathrm{pa}(i)}, \check{F}_{A\backslash \mathrm{pre}(i)}. \tag{27}$$

Also, by (24),

$$W_i \perp\!\!\!\perp F_{A\backslash \mathrm{pre}(i)} | W_{\mathrm{pre}(i)}, \check{F}_{\mathrm{pa}(i)}, \check{F}_{\mathrm{pre}(i)\backslash \mathrm{pa}(i)}. \tag{28}$$

Again, fix a value $\check{f} \neq \varnothing$ of $F_{A\backslash \mathrm{pre}(i)}$ (and so of $\check{F}_{A\backslash \mathrm{pre}(i)}$). By (28),

$$\begin{aligned} &\Pr(W_i = w | W_{\mathrm{pre}(i)}, \check{F}_{\mathrm{pa}(i)}, \check{F}_{\mathrm{pre}(i)\backslash \mathrm{pa}(i)}, F_{A\backslash \mathrm{pre}(i)}) \\ &\quad = \Pr(W_i = w | W_{\mathrm{pre}(i)}, \check{F}_{\mathrm{pa}(i)}, \check{F}_{\mathrm{pre}(i)\backslash \mathrm{pa}(i)}, \check{F}_{A\backslash \mathrm{pre}(i)} = \check{f}). \end{aligned} \tag{29}$$

By (27), the right-hand side of (29) depends only on $(W_{\mathrm{pre}(i)}, \check{F}_{\mathrm{pa}(i)})$. Then, the same holds for the left-hand side, proving (26). □

Similar arguments deliver (R50) and (R51).

5 Discussion

RR have very nicely demonstrated the close connexions between their SWIG approach and my own DT approach, as described in D21. Minor notational issues aside, there are two main differences:

(1) While both approaches introduce “intention-to-treat” variables as a way of justifying ignorability assumptions, once this has been done DT can dispense with them, relying on regime indicator variables to express and manipulate those assumptions. SWIGs, on the other hand, retain these additional ITT variables explicitly. I contend that the DT approach with regime indicators makes for cleaner representation and analysis.

(2) A more substantial difference is that SWIGs explicitly represent potential outcomes, whereas – as demonstrated by RR, as well as in D21 – in a DT analysis, they are not needed.

In many articles over many years, I have argued convincingly (at least to my own satisfaction) that the potential outcome approach to statistical causality is unnecessary and misleading. But I cannot deny that – for some unfathomable reason – it is still regarded as fundamental by most researchers in the field. It is thus a valuable feature of SWIG representations that they engage directly with the large audience of potential outcome enthusiasts. But I hope that RR’s own clear demonstration that there is a cleaner, DT, way of framing the same problems, in which potential responses simply have no place, will help to curb that misplaced enthusiasm.

Acknowledgement: I am grateful to Thomas Richardson and James Robins for continuing valuable discussions and to Andrew Yiu for helpful comments.

Author contribution: The author confirms his sole responsibility for the results presented and manuscript preparation.

Conflict of interest: Prof. Philip Dawid is a member of the Editorial Board of Journal of Causal Inference but was not involved in the review process of this article.

References

[1] Richardson TS, Robins JM. Potential outcomes and decision theoretic foundations for statistical causality. J Causal Inference. 2023;11:20220012. doi: https://doi.org/10.1515/jci-2022-0012.

[2] Dawid AP. Decision-theoretic foundations for statistical causality. J Causal Inference. 2021;9:39–77. doi: http://dx.doi.org/10.1515/jci-2020-0008.

[3] Pearl J, Mackenzie D. The book of why. New York: Basic Books; 2018.

[4] Dawid AP. Counterfactuals, hypotheticals and potential responses: A philosophical examination of statistical causality. In: Russo F, Williamson J, editors. Causality and probability in the sciences. vol. 5 of Texts in Philosophy. London: College Publications; 2007. p. 503–32.

[5] Dawid AP, Musio M. Effects of causes and causes of effects. Ann Rev Stat Appl. 2022;9:261–87. doi: https://doi.org/10.1146/annurevstatistics-070121-061120.

[6] Dawid AP. Causal inference without counterfactuals (with discussion). J Am Stat Assoc. 2000;95:407–48.

[7] Holland PW. Statistics and causal inference (with discussion). J Am Stat Assoc. 1986;81:945–70.

[8] Robins JM. Addendum to “A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy worker survivor effect”. Comput Math Appl. 1987;14:923–45.

[9] Dawid AP, Didelez V. Identifying the consequences of dynamic treatment strategies: a decision-theoretic overview. Stat Surveys. 2010;4:184–231.

[10] Dawid AP. Some misleading arguments involving conditional independence. J R Stat Soc Ser B. 1979;41:249–52.

[11] Dawid AP. Conditional independence for statistical operations. Ann Stat. 1980;8:598–617.

[12] Geneletti SG, Dawid AP. Defining and identifying the effect of treatment on the treated. In: Illari PM, Russo F, Williamson J, editors. Causality in the sciences. Oxford, UK: Oxford University Press; 2011. p. 728–49.

[13] Bühlmann P. Invariance, causality and robustness (with discussion). Stat Sci. 2020;35:404–36.

[14] Pearl J, Bareinboim E. External validity: from do-calculus to transportability across populations. Stat Sci. 2014;29:579–95.

[15] Dahabreh IJ, Robins JM, Haneuse SJPA, Hernán MA. Generalizing causal inferences from randomized trials: counterfactual and graphical identification. 2019. http://arxiv.org/abs/1906.10792.

[16] Kenah E. A potential outcomes approach to selection bias. Epidemiology. 2023;34:865–72.

[17] Dawid AP. Some variations on variation independence. In: Jaakkola T, Richardson TS, editors. Artificial intelligence and statistics. 2001. San Francisco, California: Morgan Kaufmann Publishers; 2001. p. 187–91.