Method Hierarchies in Clinical Epidemiology

Mathew Mercuri
Department of Medicine, McMaster University
mercurmd@mcmaster.ca

Hakob Barseghyan
IHPST, University of Toronto
hakob.barseghyan@utoronto.ca

An earlier version of this paper authored by Mathew Mercuri was awarded the 2016 IHPST Award for the Best Essay on the Theory of Scientific Change.

Keywords
theoretical scientonomy, method hierarchy, method, methodology, clinical epidemiology, randomized controlled trials, evidence-based medicine, employment record

Abstract
Accumulating evidence from diverse fields of inquiry suggests the existence of method hierarchies, where criteria employed by the same epistemic agent constitute a certain preference hierarchy. In this paper, we illustrate the phenomenon of method hierarchy by discussing several prominent studies in clinical epidemiology of coronary artery disease. The current “gold standard” in clinical epidemiology is the randomized controlled trial (RCT) method. Yet, in the absence of studies that satisfy the strict requirement of the RCT method, clinical epidemiologists often relax the requirements of double-blinding, complete follow-up, no treatment switching, and/or randomization. Instead, they sometimes employ less stringent requirements, such as the requirement to account for the potential imbalances between groups through statistical models. This suggests the existence of a certain method hierarchy. However, it is unclear how method hierarchies are to be conceptualized and documented. Specifically, it remains to be seen whether a method hierarchy is best understood as being composed of individual employed methods or as a single composite method with a complex system of if-s and else-s.
Introduction

Accumulating evidence from diverse fields of inquiry seems to suggest the existence of method hierarchies – the phenomenon whereby different criteria employed by the same epistemic agent (individual or communal) constitute a certain preference hierarchy. Consider how practitioners in different fields customarily speak of more or less reliable evidence. For instance, when trying to establish the authenticity of a work of art, the art history community often accepts the opinion of the most credible expert. If for whatever reason that expert fails to give their opinion, the community then refers to the second-best expert, and so on (Loiselle, 2017). The existence of method hierarchies is also confirmed by our recent attempts to explicate the methods currently employed in clinical epidemiology. They suggest that an epistemic agent can have a hierarchy of expectations – from stricter to laxer – where they first look for a theory that satisfies the strictest requirements, and only in the case that such a theory is not found do they settle for a theory that satisfies the laxer requirements in the hierarchy, and so on. In the language of scientonomy, this amounts to saying that methods employed by scientists can, at times, have different weights, i.e. that there can be hierarchies of methods. Despite this mounting evidence, the very existence of the phenomenon of method hierarchy is yet to be accepted in scientonomy.

In this paper, we illustrate the phenomenon of method hierarchy by examples from clinical epidemiology. While clinical epidemiology covers a substantial variety of medical domains and healthcare related phenomena, in the era of the Evidence-Based Medicine (EBM) movement, it is safe to say that the methods for assessing the effectiveness of therapies is generally considered to be applicable to all medical phenomena (Guyatt, Sackett, & Cook, 1993). We focus our discussion on coronary artery disease (CAD), with particular attention to acute coronary syndromes, including myocardial infarction, in order to show the existence of one specific method hierarchy. We examine both therapeutic trials for the treatment of CAD and studies to uncover risk factors for its occurrence. Particular focus on CAD is advantageous for our purpose here for a number of reasons. First, CAD has been among the most studied health states for decades; thus, it has an established scientific community and a significant experience/knowledge base from which we can determine both its accepted methodology and employed method of theory acceptance. Second, by focusing on a single health state, one can avoid considering any possible variations in the community’s implicit expectations brought about by other factors, such as training, our understanding of the nature of the health state, and patient experience. Our major goal here is to show that different requirements employed in clinical epidemiology of CAD constitute a certain hierarchy.

Methodology of Indicators

There is an important methodological question that needs to be settled before we can proceed to the reconstruction of the method hierarchy of clinical epidemiology – that of the respective indicators of theory acceptance. As we know, when trying to reconstruct the mosaic of an epistemic agent it is important to identify the types of markers that indicate if a given theory was, in fact, accepted by that agent. In principle, these indicators can vary across communities, time periods, and fields of inquiry (Fatigati, 2017). Thus, the goal is to determine what indicators can help us establish that the effectiveness of a certain therapy is currently accepted by clinical epidemiologists.

While we do not attempt to compile a comprehensive list of indicators of acceptance in contemporary clinical epidemiology, it is safe to assume that a theory can be considered accepted in clinical epidemiology when it is published in a leading international medical journal and is widely used in practice. It is, of course, possible for a theory to be accepted without being widely used, as there may be many practical impediments to using that theory in practice. For example, a therapy may be accepted as an effective treatment, and yet may not be used in practice due to resource considerations (e.g. costliness, lack of access to equipment, absence of trained practitioners). However, we believe that when a study which demonstrates a therapy’s effectiveness is published in a prestigious journal and the therapy is used in practice, it is a safe indicator of the acceptance of the therapy’s effectiveness by the community. By opting to use this strict indicator we are taking a conservative approach to ensure that the cases we select are uncontroversial: each of the studies discussed below are published in prestigious journals and the respective therapies and techniques are widely used in practice.

Mercuri & Barseghyan
Note that the cases presented here are not exhaustive with respect to all types of studies related to CAD. However, the selected studies do constitute common designs for the relevant therapeutic questions of the CAD clinical research community. Furthermore, each of the presented cases were the result of international collaboration, were published in top tier international general medicine (e.g. *The Lancet*) or cardiac medicine (e.g. *Heart*) journals, as determined by reputation and impact factor, and are among the most highly cited papers in the field. Thus, it is reasonable to believe that the cases reflect the community’s expectations and can be taken as indicative of the methods employed by the community in CAD research.

**Randomized Control Trials and Abstract Clinical Epidemiology Method**

In the second half of the 20th century, the medical community witnessed a dramatic rise in clinical study, in particular the investigation into the effectiveness of therapeutic techniques, including medicine, surgery, and biotechnology. These developments culminated in the establishment of the EBM movement in the early 1990s (Guyatt *et al.*, 1992). Key proponents of EBM define it as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients” (Sackett *et al.*, 1996, p. 71). This movement has been endorsed by major medical journals, such as *The New England Journal of Medicine, The Lancet*, and *The Journal of the American Medical Association*. Research methods and critical appraisal skills advocated by EBM are an important component of undergraduate medicine and postgraduate residency training around the world.

Although EBM as defined above may seem self-evident, it is not the “use of current best evidence” that made EBM important to clinical practice, and subsequently the methodology of clinical epidemiology research; rather it is the specific criteria imposed by EBM concerning the efficacy of therapies. While EBM recognizes that “any empirical observation about the apparent relationship between events constitutes potential evidence” (Guyatt *et al.*, 2000, p. 1292), it also stipulates that unsystematic observations and mechanistic reasoning should be considered less reliable than other sources of evidence, namely, randomized controlled trials and systematic reviews of those trials.

In most cases, the community’s ultimate acceptance of the effectiveness of a particular therapy hinges on its satisfaction of the requirements of the method of randomized controlled trials (RCT). Among other things, the RCT method stipulates that the effectiveness of a therapy must be shown (1) in a prospective trial, whereby (2) group allocation is determined by randomization, (3) one group does not receive the therapy in question, (4) both group participants and clinicians/research staff are blinded with respect to group allocation, (5) every patient receives only that therapy which is assigned to them during randomization (i.e. there is no crossover), and (6) complete data is acquired for all participants for the duration of the trial. The requirements of the RCT method are meant to reduce the threat of bias in the estimate of a causal association between the exposure (e.g. medical therapy, surgical technique, or risk factor) and the outcome (e.g. morbidity, mortality, or onset of disease state) (Guyatt, Sackett, & Cook, 1993).

First, consider the requirement of *prospective* trial, which stipulates that participants should be enrolled in the study and then followed over time (as opposed to retrospective designs that look backwards and examine exposures to suspected risk factor or therapeutic intervention). This requirement allows the investigator to ensure both that the exposure precedes the outcome, crucial for establishing the direction of causality, and that necessary information with respect to the study purpose is measured on all study participants.

Second, group allocation (on the basis of exposure status) by *randomization* improves the chances that the groups examined and compared will be balanced with respect to potential confounding variables, such as those associated with the exposure and causally related to the outcome. For example, in order to correctly estimate the magnitude of the association between exposure to an industrial chemical and the onset of some cancers in a population, one will need to rule out the effect of other known cancer-causing agents that may be present in that population (e.g. cigarette smoking). Although one can potentially adjust (e.g. mathematically, or through use of sampling techniques) the estimate of effect for other cancer-causing agents, such can only be done to the extent
that those agents are both known and measured. Where group allocation by randomization is employed, one might assume that the smokers, for example, have an equal opportunity to be placed in the industrial chemical exposure group and the control group (i.e. no exposure to the chemical in question). The same may be assumed for other cancer-causing agents in the population. As a result, the distribution of these cancer-causing agents is thought to be the same in both groups (i.e. exposure and control), and thus, considered balanced in that the only difference between the groups is the exposure in question (i.e. the industrial chemical).

Third, the use of a control group allows the investigator to minimize the incorrect attribution of any natural maturational effects to the exposure in question. For example, if one were interested in testing a therapy for pain reduction after surgery, it may be that pain will naturally subside over time without the use of the therapy. Only by introducing a control group can one hope to estimate the average incremental benefit on pain reduction (i.e. above what would naturally occur) attributable to the therapy. Again, failure to compare the “therapeutic” effect of pain reduction in a single population without a comparator population may leave the investigator ignorant of what amount of pain reduction was due to the therapy and what could have been achieved without the therapy over a natural course of healing.

Fourth, by blinding the participating patients, research staff, and clinical personnel to group assignment, investigators can minimize the threat of allocation bias, whereby stakeholders can “game” the process so as to selectively allocate a particular patient (or patient type) into one group over the other or selectively provide adjunctive treatment beyond that which is determined by group assignment (e.g., crossing over to active therapy group if assigned placebo, receiving additional therapy), thereby systematically shifting the balance of confounding variables towards one group. Blinding of research staff also ensures that discretionary measurement and analysis decisions do not systematically favour one group over the other. In addition, in conjunction with the use of a control group, blinding minimizes the potential contribution of a “placebo effect” on the estimate of the causal relationship between the exposure and outcome.

Fifth, by ensuring that all patients receive only that treatment which was assigned to them during randomization, investigators account for imbalances that could otherwise be introduced if patients were to switch treatment, i.e. if they were to cross over to the treatment in a group that is different from what they were originally assigned. The presence of crossover can bias the estimate of the effect of the exposure/therapy in question if the groups systematically differed in the rate of and/or reason for failing to adhere to the assigned treatment. Even where the treatment switching is not associated with a systematic difference between groups, its presence may result in a more conservative effect estimate.

Finally, complete follow-up of all participating patients minimizes the threat of bias from systematic differences between groups due to patients dropping out of the study (i.e. attrition bias). For example, one can imagine a case where the therapy under investigation is associated with substantial negative side effects (compared to the placebo), such that patients are motivated to drop out of the study at a higher rate than in the comparison group. Such a situation would potentially inflate the estimated relative benefit of that therapy if those who drop out of the active therapy group were less likely to benefit from the therapy.

These and other requirements of the RCT method are applicable both in the assessment of the effectiveness of therapies and the determination of risk factors on some defined outcome – each a goal of clinical epidemiology. The extent to which the aforementioned “balance” can be achieved through randomization has been challenged by notable philosophers of science (e.g. see Worrall, 2002; Thompson, 2011; Saint-Mont, 2015).

This list of six requirements in our explication of the RCT method is by no means exhaustive. For instance, we have omitted the requirement of plausibility of the hypothesis under scrutiny. If the mechanism of the purported effect of an exposure on an outcome is implausible from a physical, chemical, or biological perspective, clinical epidemiologists will likely be reluctant to accept the existence of that effect even if it is confirmed in a randomized controlled trial. One notable example of this is the randomized controlled trial supposedly showing a positive treatment effect of intercessory prayer (praying for others) (e.g. Harris et al., 1999). Even though it has been demonstrated in a randomized controlled trial, all indicators suggest that the effectiveness of intercessory prayer is not something that clinical epidemiologists would accept these days: it is not typically used in practice, taught in schools, or presented in reputable medical textbooks, and it is not mentioned in any clinical practice guidelines by respected bodies.

Finally, these six requirements of the RCT method can be explicated along the following lines:
The RCT method

A hypothesis about the causal relationship between an exposure and an outcome is acceptable if it is demonstrated in a trial that meets all of the following requirements:

1. It’s prospective,
2. It’s randomized,
3. It’s controlled,
4. It’s double-blind,
5. There is no treatment switching,
6. The follow-up is complete.

The employment of the RCT method can be illustrated on the case of the PCI-CURE study. The PCI-CURE study was the result of an international collaboration to investigate the benefit of pretreatment with clopidogrel, followed by the long-term administration of that therapy in addition to aspirin, on cardiovascular outcomes (composite of cardiovascular death, repeat myocardial infarction, or target vessel revascularization within 30 days) for patients presenting to an emergency department with a diagnosis of non ST-segment elevation myocardial infarction (NSTEMI) and who would undergo subsequent angioplasty (Mehta et al., 2001). The study was prospective in nature. Local investigators approached all patients presenting with NSTEMI to one of the study hospitals and who met the inclusion criteria. Those who gave consent to participate in the study were randomized to either clopidogrel or placebo as determined by a centralized, 24-hour, computerized randomization service located at the project office (off-site). The clopidogrel and placebo were manufactured to appear identical to the patient and clinical staff administering the medication. Patients, clinical staff, and investigators (including statisticians) were blinded with respect to which patients received which therapy (i.e. group assignment). Study investigators were made aware of group assignment only after analysis was complete. Only those patients who subsequently underwent angioplasty were included in the final analysis. Complete follow-up at 30 days post angioplasty was obtained for both the active therapy (i.e. those on clopidogrel) and placebo groups. The study demonstrated a statistically significant reduction in the primary outcome (i.e. death, MI, revascularization) at 30 days post angioplasty for those receiving clopidogrel therapy compared to the placebo group.

This case is an example of a typical prospective randomized controlled double-blind trial with complete follow-up and, thus, its results satisfy the requirements of the RCT method. We know that the results of this study are accepted as they are published in The Lancet, one of the most prestigious journals in the field, and the therapy is also widely used in clinical practice.

The fact that the requirements of the RCT method are currently employed by clinical epidemiologists in evaluating many types of therapies is also evident from the numerous examples which have been traditionally used to show the value of EBM. Consider for instance, extracranial-intracranial bypass for stroke prevention or steroid use to reduce facet-joint back pain – the perceived benefits of such therapies, both in use in clinical practice, were re-evaluated by the employed RCT method (Guyatt, Sackett, & Cook, 1993). In addition, many international consortiums of medical researchers, including The Grading of Recommendations, Assessment, Development and Evaluation working group (GRADE Working Group, 2004), and The Cochrane Collaboration (e.g. Cochrane handbook table 8.5.d; Higgins & Green (Eds.), 2011), advocate the employment of the RCT method in clinical epidemiology. Thus, it is reasonable to consider the RCT method as the standard of evidence for acceptance of effectiveness in clinical epidemiology.

Importantly, because the RCT method has advantages in reducing the threats of bias compared to other methods of clinical epidemiology, agencies entrusted with the approval of medication and other clinical therapy, such as Health Canada and the United States Food and Drug Administration (USFDA), uniformly favour evidence derived from high-quality randomized controlled trials. For instance, both Health Canada and the USFDA have adopted the guidelines developed within the Expert Working Group (Efficacy) of the International
Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Among other things, these guidelines stipulate that randomized controlled trials are generally required before a therapy can be approved for a specified indication (Health Canada, 2011; U. S. Food and Drug Administration, 2001). This confirms that, in the hierarchy of methods employed in clinical epidemiology, the RCT currently occupies the highest position. Specifically, in cases when the connection between an exposure and an outcome has been tested in randomized controlled trials, clinical epidemiologists usually accept the results of such trials over those obtained by other means, other things being equal. But what other means are there?

While the requirements of the RCT method are currently considered the “gold standard” in clinical epidemiology, there are cases when a hypothesis concerning the efficacy of a certain therapy is accepted without any randomized controlled trials (Hatswell et al., 2016). In some cases, this is due to the very nature of the exposure in question that simply precludes the employment of the RCT method. For example, in some instances, blinding the patients and/or researchers proves to be technically impossible or ethically undesirable. In other cases, complete follow-up proves to be impossible. There are also cases, where patients switch from one group to another during the course of the trial. Finally, there are cases when the requirement of randomization cannot be satisfied. We will show that, in such instances, clinical epidemiologists employ alternative methods of evaluation that relax one or more of the requirements of the RCT method. While these alternative methods are clearly treated by clinical epidemiologists as inferior to the RCT method, their results are normally accepted in cases when no randomized controlled trials on that exposure can be or have been performed.

What all these alternative methods have in common with the RCT method is that they all implement the same abstract requirements implemented by the RCT method. Namely, all these methods are implementations of the more abstract requirements of controlling participant and investigator bias. More specifically, these abstract requirements stipulate that the causal relationship between the exposure and the outcome should be confirmed in a study in which the exposure is somehow shown to precede the outcome in question. Additionally, it requires that the groups of patients should be balanced with respect to potential confounding variables, yet it doesn’t specify how exactly this is to be done. It also requires that those patients who are exposed to the therapy or risk factor should be compared to those who are not exposed to it to account for the placebo effect, natural maturation effect, etc.; once again, it is not specified how exactly this is to be achieved. The abstract method also stipulates that the trial should minimize the potential for the patients and clinicians/researchers to bias its outcome. It also requires that any potential bias due to patients receiving therapy that they were not originally assigned to or patients dropping out from the study be accounted for. The abstract method of clinical epidemiology can be summed up along these lines:

**Abstract CE Method**

A hypothesis about the causal relationship between exposure E and outcome O is acceptable if it is demonstrated in a study that meets all of the following requirements:

1. E precedes O,
2. those exposed to E and not exposed to E are similar with respect to potential confounding variables,
3. those exposed to E are compared to those not exposed to E,
4. the potential for stakeholder intervention to bias the relationship between O and E is minimized,
5. the possible bias due to treatment switching is accounted for,
6. the possible attrition bias due to patient dropout is accounted for.

It is easy to see that this more abstract method is applicable not only to therapeutic interventions such as drugs or surgeries, but to all exposures including risk factors for disease.

Because of its abstract nature, this method can be implemented differently, and that is exactly what takes place in contemporary clinical epidemiology. The RCT method is only one of many possible implementations of these
abstract requirements. In what follows, we will review several alternative implementations of this abstract method with the goal of illustrating the existence of a certain hierarchy of methods. While the discussion will focus on CAD, many of the case studies illustrate features that are common in other clinical research domains.

Coronary artery disease, defined as a narrowing of the arteries that supply oxygen-rich blood to the myocardium (i.e., heart muscles) as a result of plaque deposits and inflammation, is among the leading causes of significant morbidity and mortality worldwide (Yusuf et al., 2015). Of particular concern clinically is Acute Coronary Syndrome (ACS), which result from a sudden blockage in a coronary artery and causes myocardial infarction (i.e., heart attack) and unstable angina, among other conditions. The immediate cause of ACS is significant or total occlusion of one or more of the coronary arteries that supply blood to the myocardium. Failure to re-establish blood flow to the myocardium can result in tissue damage, subsequent heart attack, or even death. The onset of CAD can typically be treated with medical therapy, minimally invasive intervention (e.g., angioplasty to open up the occluded blood vessel), surgery (e.g., coronary artery bypass graft surgery to open up new channels of blood flow), or a combination of two or more of these approaches. The purpose of these therapeutic strategies is to alter the progression of the disease, such that morbidity can be reduced, and premature mortality can be avoided.

Outcomes associated with CAD can be optimized in other ways beyond the therapeutic strategies discussed above. For example, participation in a cardiac rehabilitation program, which may include smoking cessation, diet and lifestyle modification, and exercise, can alter the progression of disease, or modify the risk of developing CAD. Likewise, primary prevention strategies can be developed on the basis of intervention to reduce known risk factors. The risks of bias in assessing the benefit of cardiac rehabilitation or determining a risk factor for CAD are the same as those in assessing the benefit of therapeutic strategies, and thus, the standard for acceptance (of the causal relationship between exposure and outcome) in the community should be similar. Thus, we will also examine studies of these types in the current analysis.

Relaxing the Blinding and Follow-up Requirements

It is readily seen that the requirements of the RCT method are not easily applicable in the case of many surgical approaches. For one, the very nature of many surgical treatments makes it difficult to blind patients and physicians; in such cases, the abstract requirement to account for the possible placebo effect and stakeholder “meddling” (that may bias the estimate of effect) is usually implemented without blinding the patients and researchers the way it is done in double-blind trials. While sham surgeries have been used in clinical trials to blind the patients, such surgeries are often impractical and ethically questionable (Hostiuc et al., 2016 and references therein). Thus, the community resorts to a more lenient requirement to minimize the potential risk of stakeholder intervention to bias the outcome. In addition, in many cases, the strict requirement of complete follow-up also seems to be replaced with a laxer requirement to account for any possible attrition bias (due to patient dropout) by means of a statistical analysis.

The willingness to accept results from studies where the requirements are relaxed can be illustrated by the case of the SYNTAX trial that investigated the effectiveness of coronary artery bypass graft surgery (CABG) compared to angioplasty in patients with left main or three-vessel CAD in reducing the occurrence of major cardiac events, including death, myocardial infarction, stroke, and repeat revascularization (CABG or angioplasty) (Mohr et al., 2013). This prospective study followed 1800 patients randomly allocated to receive either CABG or angioplasty for 5 years post intervention. Patients were recruited from among those diagnosed with left main or three-vessel CAD at one of 85 participating centers in the United States or Europe. Those patients meeting inclusion criteria and providing written consent were enrolled. Patients were randomly assigned to one of the two therapies using a centralized voice response system. Patients and physicians were aware of which therapy the patient was assigned after randomization occurred. The study did not report whether or not the research and analysis personnel were blinded with respect to which group of patients received which therapy. Prior to analysis, 50 patients in the CABG group withdrew consent, and an additional 41 patients were lost to follow-up,
compared to 11 and 21 in the angioplasty group, respectively. The study demonstrated a significant reduction in major cardiac events at 5 years for patients in the CABG group, compared to those in the angioplasty group. This was confirmed with a sensitivity analysis (i.e. a statistical adjustment) that accounted for potential bias in the estimate due to patient dropout and loss to follow-up.

It is safe to assume that the results of this study are accepted by the medical community. First, this study is published in *The Lancet*. Second, the technique examined in the study is widely used in practice (i.e. CABG is the preferred strategy for patients with significant CAD, compared to angioplasty). Thus, we can take the case as indicative of the expectations of the community of clinical epidemiologists concerning surgeries.

This and other similar trials reveal that, in most instances, the estimation of the effectiveness of a surgery is determined by a method somewhat different from the RCT method. Specifically, there is no requirement of performing a double-blind trial (which would rarely be a viable requirement in case of a surgery), just as there is no strict requirement of complete follow-up. Instead, the method stipulates accounting for the possible attrition bias by means of a statistical adjustment. Thus, we can formulate the requirements of this method as follows:

```
Method with Relaxed Blinding and Relaxed Follow-Up

- A hypothesis about the causal relationship between an exposure and an outcome is acceptable if it is demonstrated in a trial that meets all of the following requirements:
  1. It’s prospective,
  2. It’s randomized,
  3. It’s controlled,
  4. The potential for stakeholder intervention to bias the outcome is minimized,
  5. There is no treatment switching,
  6. The statistical analysis accounts for potential imbalances between groups due to patient dropout.
```

Note that this method is applicable when evaluating the effectiveness of any surgery compared to the effectiveness of other surgeries or minimally invasive interventions.

Of course, if it were possible to test the effectiveness of a surgery in randomized controlled trials (i.e. in double-blind trials with complete follow-up), the community would readily accept the outcome of those trials. Importantly, the findings of such trials would very likely trump the findings of studies where patients and/or researchers were not blinded or where there was no complete follow-up. Thus, the results of studies such as SYNTAX are accepted not because they are thought to be superior to results of other studies that meet the stricter requirements of the RCT method, but because there have been no other studies on these topics that would satisfy the requirements of the RCT method. This goes to confirm the existence of a certain hierarchy of methods in clinical epidemiology.

**Relaxing the No Treatment Switching Requirement**

Another indication of the existence of method hierarchies is the usual expectations of clinical epidemiologists concerning the effectiveness of different rehabilitation programs. When evaluating the effectiveness of cardiac rehabilitation programs, epidemiologists often employ requirements which are somewhat different from those of the RCT method or those employed to evaluate surgeries. Similar to many surgical interventions, the nature of cardiac rehabilitation programs often precludes blinding of group assignment. Whereas a medication can be made to look similar to another medication or a placebo, rehabilitation programs are rather obviously different. Because patients and health care providers are acutely aware of the activity and lifestyle modifications that comprise the therapy, the requirement of double-blinding is not viable in this case. In addition, in many rehabilitation effectiveness studies, patients are in a position to switch to from one treatment program to another, if they find that the originally assigned treatment program does not work for them. Although such a switch is discouraged, researchers recognize that it can happen and must account for it. To account for patient treatment switching and minimize the potential imbalance between the groups, these studies often perform an analysis using the *intention*
to treat protocol, i.e. these studies consider patients to belong to their original group at randomization, rather than based on what treatment they ultimately received. This suggests that clinical epidemiologists are prepared to accept the results of a study not only without double-blinding but also with treatment switching. We contend that this happens in the absence of other studies concerning the same causal relationship between the exposure in question and its purported outcome that would meet the strict requirements of the RCT method.

This can be illustrated in the example of the RAMIT trial, a multi-centre study in England and Wales designed to determine the benefit of a formalized and comprehensive cardiac rehabilitation program on reducing mortality, morbidity, and health care service use, improving quality of life, and prevalence and control of cardiac risk factors in patients who suffered an acute myocardial infarction (West, Jones, & Henderson, 2011). The study prospectively followed approximately 1800 patients. Mortality (the primary outcome) was examined at 2 years post enrollment, and again at 7-9 years, whereas other outcomes were examined at 1 year. Patients meeting inclusion criteria and consenting to participation in the study were randomized to receive either a formal weekly or bi-weekly session of exercise training, health education, and counseling (20 hours total over 6-8 weeks) with an allied health care professional, or “usual care” (as determined by the patient in consultation with their physician). Randomization was carried out using a pre-set protocol executed by a centralized agent (i.e. not determined by the centre where the patient was enrolled). The extent to which patients, clinicians, or research staff were aware of which patients were allocated to which groups is not discussed in the study. However, it is reasonable to believe, given the nature of the intervention, that stakeholders were aware of group assignment. Complete follow-up with respect to the primary outcome (mortality) was achieved only for all participants at 2 years and 7-9 years, whereas, data regarding secondary outcomes was incomplete at 1 year for 108 and 99 patients in the intervention and control groups respectively (this was due to patient death in approximately half of these cases). In order to minimize the potential imbalance between groups caused by the patient treatment switching, the intention to treat approach was used: patients were analysed depending on the treatment to which they were originally randomised. The analysis yielded no significant difference between groups in any of the primary or secondary outcomes. Thus, the study suggested that formal cardiac rehabilitation programs were not better than “usual care”.

It is safe to say that the community has accepted the results of this study, as it is published in Heart, the official journal of the British Cardiovascular Society. Heart is a cardiology specific journal published by the British Medical Journal, a leading publisher in clinical medicine. Furthermore, while cardiac rehabilitation is a standard of care post heart attack, specific recommendations regarding the structure of such programs are not presented in guidelines, consistent with the findings of this study which found that formal rehabilitation programs were not better than “usual care”.

This suggests that in the absence of studies that would meet the no treatment switching requirement, clinical epidemiologists are prepared to accept the results of studies which meet a more lenient requirement of accounting for potential imbalances due to patient treatment switching by means of the intention to treat protocol. Thus, in such cases, they employ a method that could be explicated as follows:

**Method with Relaxed Blinding, Treatment Switching, and Follow-up**

A hypothesis about the causal relationship between an exposure and an outcome is acceptable if it is demonstrated in a trial that meets all of the following requirements:

1. it’s prospective,
2. it’s randomized,
3. it’s controlled,
4. the potential for stakeholder intervention to bias the outcome is minimized,
5. the analysis by means of the intention to treat protocol accounts for any potential imbalances between groups due to treatment switching,
6. the statistical analysis accounts for potential imbalances between groups due to patient dropout.
A brief examination of cardiac rehabilitation trials for CAD shows consistency with the RAMIT trial with respect to both group allocation procedures (randomization, control group), and concealment of therapy (blinding of stakeholders) (e.g. O’Connor et al., 1989; Lawler, Filion, & Eisenberg, 2011). Based on this, it is reasonable to believe that the employed method is also consistent with other studies in this domain.

It can be shown that similar requirements are employed not only in the cases of many rehabilitation programs, but also in evaluating many diagnostic procedures, surgeries, and surgical techniques. One illustration for the employment of this method in evaluating a technique used in a minimally invasive cardiac diagnostic procedure is provided by the RIVAL study, an international, multicentre trial examining the comparative benefit of radial versus femoral access for patients undergoing diagnostic angiography after presentation to hospital with ACS (Jolly et al., 2011). The primary outcome was a composite of death, myocardial infarction, stroke, or major bleed at 30 days post intervention. The study prospectively followed over 7000 patients presenting to one of 158 hospitals in 32 countries. Those patients who met inclusion criteria and consented to participation in the study were randomized to angiography via either radial or femoral access. Group allocation was determined by a computerized and automated voice response system located at the project office (off site). Patients and clinical staff were aware of which technique was used for the procedure; however, group assignment was concealed from the centrally located adjudication committee that reviewed and determined patient outcomes at 30 days. Following group assignment, 28 patients in the radial access group and 46 patients in the femoral access group withdrew from the study. A further 245 patients in the radial group crossed over to the femoral group on account of failure to achieve access by the operators. This was the case for 32 patients in the femoral group. Analysis was based on intention to treat from the point of randomization. The study demonstrated a reduction in vascular site complications with use of radial access, but no statistically significant differences with respect to other outcomes.

Two of the six pre-specified (i.e. prior to data collection) subgroup analyses favoured radial over femoral access on reducing the occurrence of the composite outcome.

Since the RIVAL trial was published in The Lancet, the use of radial access has continued to grow at the expense of femoral access. Based on the continued uptake of the technique, in conjunction with statements by key professional organization attesting to its use, it is safe to assume that the community has accepted the results of this study.

Once again, the results of the study were accepted, having met requirements that are in some respects laxer than those of the RCT method. Just as in the case of the RAMIT trial, the RIVAL trial did not involve double-blinding due to both practical and ethical considerations. Similar to what we observed in case of the RAMIT trial, this case implies that it may be acceptable to rely on intention to treat approach when dealing with potential imbalances between groups due to participant treatment crossover.

Relaxing the Randomization Requirement

There is evidence suggesting that similar method hierarchies are common in clinical epidemiology and concern many different requirements. While the requirements of the RCT method are considered the “gold standard” for determining if a particular exposure (e.g. patient characteristic, behaviour, clinical intervention, etc.) is a risk factor for a disease, it is the case that the nature of many exposures makes them unnamable to randomized trials. When the standard randomization is not possible, clinical epidemiologists are willing to relax the requirement and employ the somewhat more lenient requirement that any differences between those in the exposure group and those in the non-exposure group are accounted for statistically. This requirement is employed when the nature of

Interventions and surgeries for CAD might not only differ in type but might also differ in technique within a type. For example, diagnostic angiography and angioplasty can be performed via access through the femoral artery (i.e. leg), or alternatively through the radial artery (i.e. wrist), and yet be otherwise similar in all other aspects of the procedure.
the exposure under scrutiny prevents randomization or when no alternative study of that exposure with randomization is available.

The SHARE study examines one such exposure for CAD (Anand et al., 2000). Among other purposes, the SHARE study sought to determine if ethnicity (specifically, South Asian, European, or Chinese) is a risk factor for CAD. Individuals meeting the inclusion criteria, living in one of 3 major cities in Canada, and stratified according to ethnicity were randomly selected and contacted regarding their interest in participating in this study. Due to the nature of the variables of interest in this study, both allocation to group by random assignment and blinding of participants and researchers/clinicians could not be carried out (i.e. participants cannot be assigned an ethnicity or a risk factor for CAD, nor can a participant or the researcher collecting data be blinded as to her or his status on those variables). Data regarding known cardiac risk factors and cardiovascular disease status, as derived from questionnaires and laboratory tests, was prospectively collected from those consenting to participation. Data collection was complete for all those who participated in the study. Statistical models were used to determine which factors (from among the data collected) were independently associated with CAD (i.e. after adjustment for known or potential confounding variables). The study found ethnicity to be a significant predictor (i.e. risk factor) for CAD.

It is safe to say that this finding became accepted by the community, as indicated both by the fact of its publication in *The Lancet* and by the fact that the ethnicity of patients is often taken into account by medical practitioners when determining risk of disease.

The study was cross-sectional in nature, meaning that data on both the potential risk factors and CAD status was collected at the same time. Thus, factors beyond study design were required to ascertain if the exposure preceded the outcome. While no particular criteria were noted with respect to how one could determine the temporality of the exposure-outcome relationship, it is certainly reasonable to believe that ethnicity precedes the onset of CAD. Noteworthy, this study also reveals that the scientific community seems willing to relax the requirement for experimenter controlled random allocation to groups if other means can be used to achieve balance between groups with respect to potential confounding variables. Here, this is believed to be achieved through use of statistical means (i.e. random sampling from the population, and regression models that incorporate confounding variables in question). That is, the investigators used a statistical model to balance the groups with respect to known confounding factors, and in doing so, hoped to approximate what can be achieved using randomization. Examination of this study suggests that in some instances clinical epidemiologists are prepared to accept the results of a study that satisfies requirements that are more lenient than those of the RCT method. These more relaxed requirements can be formulated thus:

### Method with Relaxed Randomization and Relaxed Blinding

A hypothesis about the causal relationship between an exposure and an outcome is acceptable if it is demonstrated in a study that meets all of the following requirements:

1. it’s prospective,
2. the similarity between groups with respect to potential confounding variables is accounted for statistically,
3. it’s controlled,
4. the potential for stakeholder intervention to bias the outcome is minimized,
5. there is no treatment switching,
6. the follow-up is complete.

It is worth repeating that in clinical epidemiology the results obtained in strict randomized controlled trials normally trump the results obtained in trials with more lenient requirements. The case of hormone replacement therapy in post menopausal women, often cited by key members of the Evidence-Based Medicine movement (e.g. Sackett, 2006), illustrates that clinical epidemiologists normally prefer the results of randomized studies over those of non-randomized cohort studies. In that case, clinical practice was informed by results derived from non-randomized cohort studies, which suggested that hormone replacement therapy offers a therapeutic benefit to the
target population. However, the hypothesis of the therapeutic benefit of hormone replacement therapy was thrown into question when it was observed in a randomized controlled trial that the “overall health risks exceeded benefits from use of combined estrogen plus progestin for an average 5.2-year follow-up among health postmenopausal US women”, leading the researchers to conclude that “the results indicate that this regimen should not be initiated or continued for primary prevention of CHD [coronary heart disease]” (Writing Group for the Women’s Health Initiative Investigators, 2002, p. 321). Suffice it to say that hormone replacement therapy is not prescribed for primary prevention of CHD in women and is a questionable line of therapy for postmenopausal women.

What this analysis aims to show is that there are method hierarchies, i.e. there are cases when hypotheses that satisfy more stringent requirements are preferred to those that satisfy less stringent requirements. Thus, when a study manages to meet the requirements of the RCT method, its outcome normally overrides the outcomes of studies which did not meet the requirements of the RCT method. However, in cases when no such studies have been done because of technical, financial, ethical, or other considerations, clinical epidemiologists are prepared to accept the results of studies where one or more of these requirements is not strictly satisfied. Studies such as SYNTAX, RIVAL, RAMIT, and SHARE illustrate precisely this point.

Now, since all of these concrete methods within the hierarchy are implementations of the same abstract method of clinical epidemiology, one may wonder if these methods are sufficiently different to constitute any meaningful hierarchy. We agree that the concrete methods discussed here are similar in many respects, as most of their requirements overlap. Yet, importantly, even when one of the requirements of a method is different, we are effectively dealing with a different method. The fact that at least some of the requirements of the concrete methods discussed above are different provides enough grounds to differentiate them.

**Conceptualizing Method Hierarchies**

As we are suggesting the admission of a new type of phenomena into the scientonomic ontology, it would also make sense if we provided some working definition of the term:

| Method Hierarchy |
|----------------------------------|
| A set of methods is said to constitute a hierarchy iff theories that satisfy the requirements of methods that are higher in the hierarchy are preferred to theories that satisfy the requirements of methods that are lower in the hierarchy. |

Once the existence of method hierarchies is accepted, we will have to tackle an important practical question of how observational scientonomists are to document actual historical hierarchies of methods employed by different epistemic agents in encyclopedias and databases. Specifically, what would an employment record of such a complex method look like? Suppose we have been studying the expectations of a certain epistemic agent and discovered that those expectations constitute a certain hierarchy. Are we to ascribe separate employment records for each individual method in the agent’s method hierarchy, or are we to ascribe one employment record to the whole method hierarchy of the agent? Regardless of which of these two scenarios we opt for, there seem to be clear practical obstacles. While, in this paper, we will not be suggesting any means of overcoming these obstacles,

For example, the Committee on Gynecologic Practice of The American College of Obstetricians and Gynecologists in Committee Opinion No.565, which was reaffirmed in 2018, state that “Menopausal hormone therapy should not be used for the primary or secondary prevention of coronary heart disease at the present time” (Committee on Gynecologic Practice, 2013, p. 1407). See also the Mayo Clinic Staff take on prescribing the therapy (Mayo Clinic, 2018).

Note that the notion of method hierarchy discussed in this paper concerns hierarchies of expectations and should not be conflated with other hierarchies of methods. For instance, one may consider a hierarchy where methods range from more abstract to more concrete, i.e. “where the requirements of more concrete methods specify the requirements of more abstract methods; by satisfying more concrete requirements, a theory also satisfies some more abstract requirements” (Barseghyan, 2015, pp. 7-8). In contrast, methods that constitute a hierarchy of expectations are normally at the same level of abstractness/concreteness, as they implement the same abstract requirements.
we nevertheless think it’s important to highlight them and pose an open question for future research. Let us consider each of the two scenarios in turn.

Scenario 1: Suppose we decide to document the employment of each of the individual methods within the agent’s method hierarchy individually. In this scenario, the respective employment records would look as follows:

One advantage of this approach is that it would help us trace the historical fate of different methods and, possibly, even individual requirements within these methods. However, in this scenario, there is no simple way of knowing the exact place of each of these methods within the agent’s method hierarchy. All it would allow us to say is that such and such a method was employed by the agent during a certain time period, but we would not have any records of how this method was related to other methods in the agent’s hierarchy.

Scenario 2: Alternatively, we can decide to treat the whole hierarchy of the agent’s methods as one giant method with a complex tapestry of if-s and else-s, and-s and or-s. Since, in this scenario, the agent’s method hierarchy would be captured within the formulation of one giant method, this method would require simple employment records, such as this:

However, there are obvious practical issues with this approach. For one, such complex methods would be virtually unmanageable. Here is what a tiny fragment of such a giant method would look like:

A hypothesis about the causal relationship between an exposure and an outcome is acceptable if [other requirements] and it is demonstrated in a study
(a) with complete follow-up,
or
(b) where the statistical analysis accounts for any potential attrition bias and no other study with complete follow-up shows otherwise,
or
(c) [another requirement] and no other study with complete follow-up or statistical analysis of potential imbalances between groups shows otherwise.
In addition, each time there was the slightest change in the agent’s method hierarchy or across methods of two different agents, the observational scientonomist would have to add another giant method to the respective database together with its respective employment record. That would saturate our historical database with a plethora of unmanageable giant methods, each unique to one or few epistemic agents employed at very limited time periods. Finally, and perhaps most importantly, such an approach would make it impossible to trace the historical fate of individual requirements that constitute these giant methods. After all, as observational scientonomists, we may want to know when and which agents have employed, say, the requirement of double-blinding. Since, in scenario 2, the historical fate of these individual requirements would not be traced, we would have no simple way of querying for agents who employed these requirements.

What we are facing here is an important question: how should method hierarchies be captured and documented? While this question is of great practical importance to observational scientonomy, it also has an important theoretical aspect. It concerns not only how method hierarchies are to be documented in databases of intellectual history, but also – and even primarily – how method hierarchies are to be conceptualized and understood. Are they to be understood as being composed of individual methods, each with its place in the hierarchy, or are they to be understood as being composed of individual requirements, or are they to be understood as giant methods with complex decision trees? We hope that future research will shed light on this important question.

Note that, for now, we suggest no changes to the current definition of method. The current notion of method, understood as a set of criteria for theory evaluation (Barseghyan, 2018), is inclusive enough to also capture hierarchies of criteria, so no new definition is needed at this stage. One may be needed once we have a better understanding of how method hierarchies are to be conceptualized.

**Conclusion**

While previous discussions in theoretical scientonomy have hinted at the existence of hierarchies among methods employed by an epistemic agent (Loiselle, 2017), the phenomenon itself is yet to be properly understood and its existence its yet to be accepted. As we have demonstrated, there are clearly cases suggesting that method hierarchies do exist. We discussed a number of prominent studies in clinical epidemiology of coronary artery disease that reveal a variety of different requirements — from more stringent to more lenient. The randomized controlled trial method is the “gold standard” in clinical epidemiology: if a study meets its requirements, the study’s outcome is normally accepted, as was illustrated by the case of the PCI-CURE study. However, in the absence of studies that satisfy the strict requirement of the RCT method, clinical epidemiologists are prepared to accept studies that satisfy less stringent requirements. In some cases, the requirement of double-blinding is relaxed and is replaced with a more lenient requirement to minimize the potential for stakeholder intervention to bias the outcome. This was illustrated by the case of the SYNTAX studies. This study also showed that the requirement of complete follow-up is at times replaced with a more lenient requirement to account for any potential attrition bias by means of a statistical analysis. Similarly, clinical epidemiologists are sometimes prepared to relax the requirement of no treatment switching. As our discussion of the RAMIT and RIVAL studies revealed, the community is prepared to accept the results of a study with patient crossover, if potential imbalances between groups are accounted for by the intention to treat protocol, or perhaps other statistical methods. In principle, any of the standard requirements of the RCT method can be relaxed provided that the respective abstract requirement is satisfied; the case of the SHARE study is an illustration of the relaxation of the randomization requirement, and its replacement with the requirement to collect data from a random sample of participants and account for important differences between groups through statistical models. Importantly, our examination suggests that these...
more lenient requirements are employed only in the absence of studies satisfying the more stringent requirements, which suggests the existence of a certain hierarchy of expectations.

Once the existence of method hierarchies is accepted, a question emerges concerning the nature of these hierarchies. The question is how method hierarchies are to be understood and documented. Specifically, should we conceive of an agent’s method hierarchy as being composed of individual employed methods (or even individual requirements), or should we think of the agent’s hierarchy as one composite method with a complex system of if-s and else-s, and-s and or-s? This question is of both practical and theoretical importance and we hope that it will receive its proper treatment in due course.

Acknowledgements

We would like to thank Paul Patton and Juan Serrano for their helpful comments on earlier drafts of this paper.

Suggested Modification

We suggest the following modification:

[Sciento-2019-0013]

Accept the existence of method hierarchies.

Accept the following definition of method hierarchy:

- **Method Hierarchy** ≡ a set of methods is said to constitute a hierarchy iff theories that satisfy the requirements of methods that are higher in the hierarchy are preferred to theories that satisfy the requirements of methods that are lower in the hierarchy.

Accept the following question as a legitimate topic of scientonomic inquiry:

- **Conceptualizing Method Hierarchies**: should we conceive of a method hierarchy as being composed of individual employed methods/requirements, or should we think of it as constituting one composite method with a system of if-s and else-s, and-s and or-s?

Bibliography

Anand, S. S.; Yusuf, S.; Vuksan, V.; Devanesen, S.; Teo, K. K.; Montague, P. A.; Kelemen, L.; Yi, C.; Lonn, E.; Gerstein, H.; Hegele, R. A.; & McQueen, M. (2000). Differences in Risk Factors, Atherosclerosis, and Cardiovascular Disease Between Ethnic Groups in Canada: The Study of Health Assessment and Risk in Ethnic groups (SHARE). *Lancer* 356, pp. 279-284.

Barseghyan, H. (2015). *The Laws of Scientific Change*. Springer.

Barseghyan, H. (2018). Redrafting the Ontology of Scientific Change. *Scientonomy* 2, pp. 13-38. Retrieved from https://scientojournal.com/index.php/scientonomy/article/view/31032.

Committee on Gynecologic Practice. (2013). Hormone Therapy and Heart Disease. Committee Opinion No.565. *American College of Obstetricians and Gynecologists. Obstetrics and Gynecology* 121(6), pp. 1407-1410. Retrieved from: https://www.acog.org/Clinical-Guidance-and-Publications/Committee-Opinions/Committee-on-Gynecologic-Practice/Hormone-Therapy-and-Heart-Disease.

Fatigati, M. (2017). A Method for Reconstructing the Medieval Arabic Scientific Mosaic. *Scientonomy* 1, 19-28. Retrieved from https://www.scientojournal.com/index.php/scientonomy/article/view/27761.
Morden, J. P.; Lambert, P. C.; Latimer, N. R.; Abrams, K. R.; & Wailoo, A. J. (2011). Assessing Methods for Dealing with Treatment Switching in Randomised Controlled Trials: A Simulation Study. *BMC Medical Research Methodology* 11(1), article 4. Retrieved from https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-11-4.

O’Connor, G. T.; Buring, J. E.; Yusuf, S.; Goldhaber, S. Z.; Olmstead, E. M.; Paffenbarger Jr. R. S.; & Hennekens, C. H. (1989). An Overview of Randomized Trials of Rehabilitation with Exercise After Myocardial Infarction. *Circulation* 80(2), pp. 234-244.

O’Connor, G. T.; Buring, J. E.; Yusuf, S.; Goldhaber, S. Z.; Olmstead, E. M.; Paffenbarger Jr. R. S.; & Hennekens, C. H. (1989). An Overview of Randomized Trials of Rehabilitation with Exercise After Myocardial Infarction. *Circulation* 80(2), pp. 234-244.

Sackett, D. L.; Rosenberg, W. M. C.; Muir Gray, J. A.; Haynes, R. B.; & Richardson, W. S. (1996). Evidence-Based Medicine: What It Is and What It Isn’t. *British Medical Journal* 312(7023), pp. 71-72.

Saint-Mont, U. (2015). Randomization Does Not Help Much, Comparability Does. *PLoS ONE* 10(7), article e0132102.

Thompson, R. P. (2011). Causality, Theories and Medicine. In Illari, Russo, & Williamson (Eds.) (2011), pp. 25-44.

U. S. Food and Drug Administration. (2001). *Guidance for Industry—E 10 Choice of Control Group and Related Issues in Clinical Trials*. Retrieved from https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073139.pdf.

West, R. R.; Jones, D. A.; & Henderson, A.H. (2012). Rehabilitation After Myocardial Infarction Trial (RAMIT): Multi-Centre Randomised Controlled Trial of Comprehensive Cardiac Rehabilitation in Patients Following Acute Myocardial Infarction. *Heart* 98, pp. 637-644.

Worrall, J. (2002). What Evidence in Evidence-Based Medicine? *Philosophy of Science* 69(S3), pp. S316-S330.

Writing Group for the Women’s Health Initiative Investigators. (2002). Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results from the Women’s Health Initiative Randomized Controlled Trial. *Journal of the American Medical Association* 288(3), pp. 321-333.

Yusuf, S.; Wood, D.; Ralston, J.; & Reddy, K. S. (2015). The World Heart Federation’s Vision for Worldwide Cardiovascular Disease Prevention. *Lancet* 386(9991), pp. 399-402.