Evidence of biological mechanisms and health predictions: an insight into clinical reasoning

Saul Perez Gonzalez[1] and Elena Rocca[2]

1 Warsaw University of Technology, saul.perezgonzalez@pw.edu.pl

2 Norwegian University of life Sciences, elena.rocca@nmbu.no

Abstract

Traditionally, understanding biological mechanisms has had a central role in clinical reasoning. With the raise of the evidence-based paradigm, however, such role has been under debate. On the one hand, evidence of pathophysiological mechanisms has been de-emphasised in clinical guidelines. This is often motivated by the unreliability of our understanding of complex biological mechanisms. On the other hand, evidence of mechanisms has been defended by some scholars as key to clinical practice. Here, we assess the relevance of evidence of biological mechanisms in two types of clinical predictions: predictions about efficacy and predictions about safety of a certain intervention for the particular patient. Further on, for each type of prediction, we analyse separately two roles that evidence of mechanisms might have, confirming and disconfirming, depending on whether or not it supports that certain epidemiological results apply to the single patient. We argue that the 'unreliability because of incompleteness' argument against the emphasis on mechanistic clinical thinking only applies to some of the considered cases. We conclude by offering a model for a more granular view of the role of mechanistic thinking in clinical practice.

1. Introduction

In evidence-based medicine (EBM), there is a disagreement about the type of evidence required to make reliable causal predictions. Contra the dominant EBM paradigm (Howick 2011a), which priories evidence of difference-making, many authors have argued that evidence of mechanisms

is crucial to make accurate predictions from population studies to a target population or an individual patient (Clarke et al. 2013; 2014; Parkkinen et al. 2018; Rocca 2017; Russo and Williamson 2007). Evidence of mechanisms is understood as evidence of either the existence or features of mechanisms in the domain of inquiry (Illari 2011). In that framework, a broad view of mechanisms is usually adopted. A mechanism can be a complex system (Illari and Williamson 2012), a causal process (Salmon 1998), or some combination of both. Since it is beyond our scope to advocate one of the existing definitions of mechanism, we are going to follow that broad approach to the notion of mechanism.

In this paper, we aim at offering an analysis of which precise roles evidence of mechanisms should have in clinical reasoning. In particular, we will look at the significance of mechanistic evidence for the cases in which a clinician needs to evaluate the relevance of epidemiological results for the single patient. Previous analyses have highlighted the importance of evidence of mechanisms for the clinical practice (Andersen 2012; Tonelli and Williamson 2020). They show that, in certain contexts, evidence of mechanisms can even offer sufficient basis for reliable causal predictions. Nevertheless, it has not been specified *how* relevant is evidence of mechanisms for the clinical practice and *when* it offers sufficient bases for reliable predictions. In order to address those questions, we will diversify the general issue into different scenarios. Each scenario is characterised by a different role played by mechanistic evidence. In this sense, we will follow the approach adopted by Pérez-González and Iranzo (Forthcoming) in their discussion on mechanism-based causal extrapolation from a study population to a target population of interest. In their analysis, the authors distinguish between a 'positive', or confirming, and a 'negative', or disconfirming, role of evidence of mechanisms, depending on whether it supports or undermines a causal extrapolation. In the positive, or confirming, scenario, the relevant mechanisms at work in the study and the target population are similar in the relevant aspects. The extrapolation of the causal claim is thus justified. In the negative, or disconfirming, scenario, the relevant mechanisms at work in the study and the target population differ in relevant aspects, making the extrapolation of the causal claim not justified. The authors take this distinction as reference and argue that evidence of mechanisms is not equally relevant in both scenarios. While the disconfirming role

2

of evidence of mechanisms is highly reliable, the confirming role faces important difficulties and additional evidence is required to support the extrapolation of a causal claim. Here, we extend that approach to the clinical setting, where extrapolation needs to be made from epidemiological data (experiments, observational studies, other patients) to one single patient.

The structure of the paper is as follows. Section 2 introduces the main causal predictions in the clinical practice: prediction about efficacy and prediction about safety. Predictions about specific side effects are presented as a particular case of prediction about efficacy. Section 3 discusses mechanism-based predictions about efficacy. It is argued that evidence of mechanisms is more reliable for determining that an intervention would not produce a specific effect in a patient than for establishing that it would produce the effect. Section 4 analyses the difference between mechanism-based predictions about efficacy for targeted effects and for untargeted effects. Section 5 discusses mechanism-based predictions about safety. It is considered that, while evidence of mechanisms faces difficulties for supporting the extrapolation of a safety claim, it can reliably establish that an intervention would probably produce side effects and is not safe for a patient. Finally, section 6 concludes.

2. Two kinds of causal prediction

Wise clinical choices must involve two kinds of causal prediction: prediction about efficacy and prediction about safety. Clinicians, indeed, are not only interested in whether the intervention will produce the desired (targeted) effect in a particular patient. They must also consider whether it will cause relevant side (untargeted) effects. Here, we are going to treat predictions about efficacy and safety of interventions separately for two main reasons.

First, the target effect is one (or few) and known, while untargeted effects are many and often unknown. This asymmetry demands different standards of evidence for efficacy and safety of interventions (Osimani 2013).

Second, there is a difference regarding the specificity of the research question in predictions about efficacy and in predictions about safety. When predicting efficacy, the question is usually whether the intervention is likely to produce a *specific* effect in the patient. In this case, it is often

sufficient that the patient presents a single *relevant*[1] difference from the population average, for the confidence in the efficacy to drop. For example, we might wonder whether a proton pump inhibitor (PPI) will relieve symptoms of stomach acidity in the patient, given that it worked in a certain experimental group. Suppose that the patient has visceral sensitivity, while patients with this condition were excluded from the experiment. This piece of information alone would decrease significantly our expectations of getting effects in the patient comparable to the population average, since decreasing acidity might not be sufficient to relieve the symptoms in an extra-sensitive patient.

On the contrary, when predicting safety, the question is less specific: will the intervention produce *any* undesired effect in the patient? In the previous example, suppose that the clinical trial for efficacy of PPI included a subpopulation of older patients, and that some of these patients in the experimental group developed pneumonia. Say that this is interpreted as a side-effect of the PPI, since reduced stomach acidity leads to reduced defence to bacterial infections. If we do not find in our patient any evidence for a mechanism of vulnerable immune system (older age, chronic illness, use of immunosuppressant medicines), we will probably not predict a high risk of—specifically—pneumonia (or other serious infections). However, risk of milder bacterial infection, especially for long-term therapies, should not be excluded. In addition, we still need to evaluate whether reduced stomach acidity could have other consequences in this patient in the long run, such as fractures because of reduced calcium uptake, rebound symptoms after discontinuation, or even other unexpected effects. We see then that, given the presence of hazardous mechanisms in the experimental population, identifying one or few relevant differences in the patient is not sufficient for concluding the absence of side effects. Those differences could modify the side effects, but they would probably still be undesired effects. Knowing *one or few* differences might at best tell us that the predisposition of the patient to experience *one or few of all the possible* untargeted effects is different than average.

There is one more clarification we need to make. One might observe that, sometimes, clinicians need to make mechanism-based predictions also about *specific* side effects. Returning to the previous example, if the patient using PPI is a post-menopausal woman, we should

4

specifically worry about the decreased calcium uptake and increased risk of fractures. This is because the patient has a propensity to calcium loss and osteoporosis to begin with, given the hormonal phase she undergoes. In this and similar cases, we are not primarily interested in whether the intervention is safe *in general*, but in whether the intervention would produce certain *specific* untargeted effect on the patient. In this sense, mechanism-based predictions about one precise side effect fall in the same *biological* category of predictions about efficacy. The difference with paradigmatic examples of predictions about efficacy is that the considered effect is undesired. Nonetheless, as we shall see in section 4, the 'desired-undesired' distinction is crucial for defining the role of evidence of mechanism in clinical predictions.

Predictions about specific side effects and predictions about safety are of course related. If, given the presence of a mechanism through which an intervention produces a specific side effect, we assess that the intervention might produce that side effect in the patient, we must conclude that the intervention is not safe for her. The prediction of a side effect, thus, is equal to the absence of safety. Nevertheless, the opposite does not hold. Assessing that the intervention will not produce a specific side effect in the patient does not imply its safety. In the above example, even if the risk of reduced calcium uptake is identified and monitored, this does not necessarily mean that PPI are safe for the patient, in a general sense. The question about safety is a broader one, since it asks whether the patient will be unhurt, or not hurt in a way that is considered serious or relevant. This is a complex question, since there is a nearly infinite potential of what a substance could do, depending on the nearly infinite combinations of other entities that it could encounter (Rocca, Anjum, and Mumford 2020; Ruthenberg 2016).

3. Predictions about efficacy

In mechanism-based causal predictions about efficacy from a population study to a single patient, the situation seems to be similar to that faced in causal extrapolation to a target population (see section 1). Evidence of mechanisms is more reliable for determining that a causal relation does not hold in the patient (disconfirming role) than for establishing that a causal relation holds

(confirming role). In this section, we will separately address both roles of evidence of mechanisms in predictions about efficacy.

3.1. Confirming predictions about efficacy

Evidence of mechanisms may support the extrapolation of an efficacy claim from a population study to a particular patient. The confirming role of evidence of mechanisms in predictions about efficacy could be characterised as follows. If the relevant mechanisms at work in the study population and the patient are highly similar in the relevant aspects, we can conclude that the intervention would produce the targeted effect in the patient.

The confirming role of evidence of mechanisms in predictions about efficacy faces important difficulties.

i. There is no unproblematic procedure for identifying the degree of similarity between the relevant mechanisms. Comparative process tracing and alternative procedures for comparing the relevant mechanisms have been highly criticized (Howick et al. 2013; van Eersel et al. 2019). It has been argued that they are usually unfeasible. In order to establish the degree of similarity, those procedures require information about the relevant mechanisms (e.g., detailed information about their components) that is rarely available.

ii. Our knowledge about the relevant mechanisms at work in the study population and the particular patient is usually very fragmentary (Howick et al. 2013; Reiss 2010; van Eersel et al. 2019). Even when a mechanism is identified, many of their components remain unknown. Furthermore, it is not always the case that careful studies aimed to identifying the relevant mechanisms at work in the patient can be conducted (e.g., emergency surgeries).

iii. In the individual patient, there may be unknown interfering mechanisms that influence the outcome (Clarke et al. 2014; van Eersel, Koppenol-Gonzalez, and Reiss 2019). Even if mechanisms similar to the relevant mechanisms identified in the study population are present in the individual patient, there may be other

relevant mechanisms too. Those interfering mechanisms could even interfere with the identified mechanisms and mask or modify their own contribution to the outcome.

iv. Similar mechanisms may be present in the study population and the particular patient but not behave in a similar way (Howick et al. 2010; Howick et al. 2013; van Eersel et al. 2019). The behaviour of mechanisms may change depending on the context. Even if the relevant mechanisms at work in the study population and the patient are highly similar, they may have unanticipated and paradoxical effects in her.

Nevertheless, within this general problematic scenario, it is possible to differentiate types of situations. Positive mechanism-based predictions about efficacy are more reliable in some cases than in others. For instance, a mechanism-based prediction on whether a general anaesthetic will work in a patient will be more reliable than a mechanism-based prediction on whether a certain anti-proliferative drug will work against her cancer. The reason is that some mechanisms are 'general' enough to make problems (i) and (iv) little relevant. In order to illustrate that idea, let us discuss an example about inhalation anaesthetic (IA). Suppose that we have to evaluate whether certain IA will work on a specific patient, who is not well represented by any subpopulation of the available trials. Say, for instance, that she is an obese and diabetic patient. How reliable would be a mechanism-based prediction about the efficacy of the intervention?

IAs work by reducing neuronal and synaptic transmission through the interference with ion channels in the neuronal membrane. Such interference provokes a hyper-polarisation of the neuronal membrane and therefore inhibits post-synaptic neuronal excitability (Khan, Hayes, and Buggy 2014). The mechanism by which ion channels regulate the polarisation of the neuronal membrane and the transmission of the electric signal through the synapsis is an evolutionary conserved one. This means not only that it is general to human beings and mammals, but also to the majority of other animals. Moreover, it behaves in a similar way in all of them. This is different from the case of most anti-proliferative dugs, which target specific mechanisms for aberrant proliferation in certain types of cancers. Given that IAs interfere with such a universal (and

foundational) mechanism, issues about dissimilarity or irregularity would not be very worrying. It could then be inferred with a margin of safety that they will work at least to some extent in the patient.

Problems (ii) and (iii), however, would still be relevant. The prediction, indeed, is based on just a part of the mechanism of action (problem ii). IAs are gasses and their interference at neuronal level demands that they are first dissolved in the blood and distributed. An obese patient has a larger fat compartment and IA is highly absorbed and slowly released from the fat tissue. This means that the same dose of IA would work differently in the patient in question than in a patient with normal weight. So, although by knowing a part of the mechanism one can predict that the anaesthetic will work, predicting more specifically how it will work (e.g., for how long) requires knowledge about other aspects of the mechanism of action. In the same way, there might be some other mechanisms at place in this patient, which are still unknown and influence the way IA works in her case (problem iii).

The reliability of a confirming mechanism-based prediction about efficacy depends on how complete our mechanistic knowledge is and how accurate the comparison between the relevant mechanisms is. This is the case even when some involved mechanisms are general or conserved. Nevertheless, there is never certainty about how complete our mechanistic knowledge is. In sum, even when there is clinical evidence that a certain intervention works for a certain patient group, and the specific patient shares (part of) relevant mechanisms with the patient group, clinicians still should stay alert.[2] In order to establish with confidence the efficacy of a treatment in a patient, evidence of mechanisms should be complemented by other kinds of evidence. This could be, for instance, evidence about particular aspects of the specific patient and her context, as well as additional evidence of difference-making from similar patients.

3.2. *Disconfirming predictions about efficacy*

Evidence of mechanisms may also disprove the extrapolation of an efficacy claim from a population study to a particular patient. The disconfirming role of evidence of mechanisms in predictions about efficacy could be characterised as follows. If the relevant mechanisms at work

in the study population and the patient differ in relevant aspects, it can be concluded that the intervention would not produce the targeted effect in the patient

The difficulties faced by the confirming role of evidence of mechanisms in predictions about efficacy are less threatening for the disconfirming role.

    i.   In the disconfirming scenario it is not necessary to specify the degree of similarity between the relevant mechanisms in the study population and the patient. It is only required to identify (at least) one relevant difference between them. This is a more feasible task, which can generally be carried on with available procedures. Consider, for instance, comparative process tracing (Steel 2008): the procedure of carefully analysing the relevant mechanisms in the study population and, subsequently, comparing them with mechanisms at work in the target in (some of) those stages in which they are likely to differ. Adopting comparative process tracing may result in the identification of relevant differences between the relevant mechanisms at place in the study population and in the patient.

    ii.   Even if our knowledge about the relevant mechanisms in the study population and the patient is fragmentary, it is possible to identify a relevant difference between them. Note that the identification of all the relevant differences is not required in disconfirming mechanism-based predictions.

    iii.   Although unknown interfering mechanisms may be present in the patient, it is unlikely the case that the causal relation holds (despite the identified difference(s)) because of them. The presence of interfering mechanisms in the patient would only enable the causal relation if they operated so that they exactly compensated the identified difference—i.e., if they 'restored' the similarity. However, given the complexity of most biological mechanisms, that exact counterbalance is highly unlikely (Andersen 2012; Howick 2011b; Howick et al. 2010). The interfering mechanisms would probably modify the effect of the intervention, but not exactly compensating the identified difference.

iv.   Mechanisms may not behave in the patient as in the study population, but mechanisms' absence of regularity would hardly produce that the causal relation holds (despite the identified difference(s)). The irregular behaviour of a mechanism would only enable the causal relation if it operated so that it exactly compensated the identified difference. Nevertheless, as in the case of masking, that exact counterbalance is highly unlikely given the complexity of biological mechanisms. The irregular behaviour would probably just result in a different untargeted effect.

The disconfirming role of evidence of mechanisms is not undermined by the problems usually faced by mechanism-based predictions about efficacy. Therefore, once we identify a relevant difference between the mechanisms at place in the patient and the mechanism of action by which the intervention works in the study population, we have a solid ground to predict that the intervention's efficacy will be hindered in the patient.

In order to illustrate the disconfirming role of evidence of mechanisms in predictions about efficacy, consider the following example about botulinum. Botulinum toxins cause flaccid paralysis (by interfering with vesicle fusion and neurotransmitter release in the neuronal cells) and are used to treat many conditions (Chen 2012). Although they interfere with an evolutionary conserved mechanism, many other factors influence the therapeutic action, such as age and type and stage of the illness, so it is usually difficult to predict whether a particular patient will respond and which doses will work for her (see subsection 3.1) (Misra et al. 2012). Nevertheless, in cases of long-term therapy, some patients develop neutralising antibodies against the protein, which diminish or counteract the therapeutic effect (Torres et al. 2014). The presence of neutralising antibodies is evidence for a relevant difference between the mechanism at place in the study population (where botulinum toxins produce the effect) and the mechanism at place locally. Consequently, although many components of the mechanisms at work in a particular patient are always unknown, and although there is variation in the effects of the neutralising antibodies, once they are found in a patient's blood, it is justified to expect a reduced (or neutralised) efficacy of the therapy.

4.   Predictions about specific side effects

In section 2, we have seen that, since predictions about efficacy concern specific effects, they should be treated differently from predictions about safety, which normally include a range of possible undesired effects. What, then, when we are wondering about the risk of a specific side effect? At biological level, the question 'will this specific effect happen?' seems to be the same, regardless from whether the effect is a desired or undesired one. Nevertheless, the considerations we made in the previous section do not seem to completely apply in case of undesired effects. Let us illustrate this by an example.

The antiviral abacavir can provoke violent, life-threatening allergic reactions. Population studies have correlated this undesired effect to a certain point mutation of the HLA-B protein (Mallal et al. 2002). Furthermore, the mechanism underlying this correlation has been elucidated: abacavir activates antigen-presenting cells in genetically susceptible individuals, potentially initiating the pathological hypersensitive response (Martin et al. 2007). At present, there are available genetic tests to screen patients and identify if they carry the mutated version of HLA-B and, consequently, are susceptible to the undesired reaction.

Consider now each of the two scenarios presented above in relation to this undesired allergic reaction. First, if the genetic test shows that the patient does not carry the genetic mutation, the pathological mechanism underlying the population data is missing a key element. We have then a case of disconfirming role of evidence of mechanisms. In this case, the clinician has good grounds to believe that this *specific* hypersensitivity reaction will not happen. So far, the reasoning is the same as with regular predictions about efficacy.

Second, say that the patient tests positive to the genetic screening: this element of the mechanism (the mutated HLA-B) is the same in the study population and in the patient. This would give us a case of confirming role: since relevant elements of the mechanisms are at place, we can predict that the allergic reaction will happen in the patient. In principle, because of the problems discussed in subsection 3.1 (e.g., mechanisms' absence of regularity), we would be much less confident about that prediction and insecure about the intensity of the reaction. Here, however, there is a significant difference with respect to the cases in which predictions are about

11

targeted effects. In the present case, the clinician will normally *not* take the chance of trying abacavir in the patient in the presence of positive genetic screening. On the contrary, this result will be considered an excellent ground to avoid abacavir and look for alternative therapy.

The main reason of this divergence between targeted and untargeted effects is that, in general, evidence of mechanisms is given more weight when predicting undesired effects than when predicting efficacy for desired effects (Osimani 2013). This is due to both epistemological and ethical considerations.

In predictions about particular side effects of treatments, the available evidence is usually scarce. Firstly, randomized controlled trials (RCTs) cannot be designed for testing undesired effects, primarily for ethical reasons, but also for other limitations. For instance, the limited time spam of experiments cannot pick up long-term side effects. And secondly, evidence about potential harm from large population studies, such as cohorts, is often unavailable, especially for new or relatively new treatments. What clinicians often have available, as evidence of potential harm, are case-reports, case-series, case-control studies, and evidence of statistical disproportionality in the databases of spontaneous reports of side-effects (Norén, Hopstadius, and Bate 2013). Therefore, evidence of mechanisms has a crucial role in predictions about side effects and a significant weight is given to it.

In addition to those general epistemological considerations, there are also some important value choices in place. When a possible treatment for a patient is evaluated, depending on the type and magnitude of the effects and symptoms, clinicians might be more concerned about predictions about untargeted effects than about targeted effects, or vice versa. When clinicians are concerned about avoiding a lethal side effect (and less concerned about producing the targeted effect), they are likely to give more weight to evidence pointing to the existence of mechanisms for producing the side effect (and to be more demanding with evidence in support of the existence of mechanisms producing the targeted effect).

Consider the abacavir example again. Identifying important similarities regarding the relevant mechanisms for a potentially fatal side effect between the study population and the patient is enough evidence for the clinician to avoid that treatment, regardless of the fact that

mechanisms could be masked or behave irregularly. Although mechanism-based confirming predictions about a specific effect have a bigger margin of uncertainty that the disconfirming ones, this uncertainty may be counterbalanced by the magnitude of risk at stake or the relevance of the targeted effect.

5. Predictions about safety

In this section, we will analyse in detail the confirming and the disconfirming role of evidence of mechanisms in predictions about safety. The confirming role refers to cases in which mechanistic knowledge supports predictions about the presence of side effects, while the disconfirming role refers to cases in which mechanistic knowledge supports predictions about the absence of side effects.

When we predict the safety of an intervention from a population study to the single patient and we use mechanistic knowledge to help the prediction, we seem to face the opposite situation that in predictions about efficacy. In the case of efficacy, the disconfirming scenario is less demanding (i.e., it requires less information about the relevant mechanisms). It only requires identifying at least one relevant difference between the relevant mechanisms at work in the study population and the patient in order to lose confidence in the positive outcome. The confirming role, instead, requires establishing a high degree of similarity (i.e., the absence of any relevant difference) between the mechanisms at place. On the contrary, in predictions about side effects, the less demanding scenario is the confirming one. It only requires identifying in the patient at least one mechanism through which the intervention produces side effects in the study population to confirm lack of safety. Instead, the disconfirming case requires corroborating the absence of all the mechanisms through which the intervention produces side effects in the study population.

We will now consider the two scenarios in detail. As in section 3, we will consider first the more problematic case.

5.1. Disconfirming predictions about safety

Evidence of mechanisms may support the extrapolation of a safety claim—i.e., absence of relevant side effects—from a population study (be it an observational study, a case series, or even a single case report) to a particular patient. The disconfirming role of evidence of mechanisms in predictions about safety could be characterised as follows. If the mechanism(s) through which certain intervention produces side effects in the study population is absent in the patient, it can be concluded that the intervention would not produce side effects and is safe for that patient.

The disconfirming role of evidence of mechanisms in predictions about safety has important problems.

i. It is sometimes impossible to corroborate the absence of certain mechanisms (due to ethical or technical reasons). Consequently, it can be the case that the presence of some known mechanisms through which an intervention produces side effects in the study population cannot be ruled out. For instance, consider psychological mechanisms, which can lead to the abuse of some types of drugs and to addiction. Psychological mechanisms cannot be mapped biomedically; their identification relies exclusively on clinical dialogue, collaboration and compliance of the patient, and clinician's skills and time availability. Therefore, when these resources are absent, it cannot be ruled out whether relevant psychological mechanisms are at place in the patient.

ii. In a case in which, in principle, the absence of every known relevant mechanism can be tested (no problem i), their amount or the required resources could make the testing unfeasible in the relevant contexts. The latest vaccine against the virus of Dengue, for instance, can paradoxically provoke a deadly Dengue infection if it is given to patients who have never been infected by a sub-family of the Dengue virus before (the first Dengue infection is generally light and unnoticed). This is due to a mechanism called 'antibody-dependent enhancement'. In order to make sure that this mechanism is not possible in the patient, it is necessary to verify before the vaccination that she has antibodies against one of the four subtypes of Dengue virus in her blood. However, Dengue is endemic in developing countries, where the

technical and economic resources for such a blood test in every child are unavailable and almost unthinkable (Sridhar et al. 2018).

iii.   There may be unknown mechanisms through which an intervention produces side effects in the study population. It means that, even if it is corroborated that all the known relevant mechanisms are absent in the patient (no problems i and ii), it can be the case that mechanisms through which the intervention produces side effects in the study population are present in her. Consequently, we can hardly be sure that the premise of the disconfirming-scenario prediction is met.

Consider the following example about the drug warfarin. Warfarin is an anticoagulant drug (used to prevent blood cloth) which must be used with extreme attention because an excessive dose can provoke gastrointestinal bleeding and death. One problem related with the use of warfarin is that many other drugs, foods, and drinks can influence how much and how quickly it is absorbed in the intestine and, therefore, produce a deadly interaction. In a specific patient, one can make sure that all the hazardous known interactions are avoided. However, the same side effect (altered absorbing—excess in the blood—internal bleeding and death) can be given by interactions that are so far unknown. This may be the case because this patient has access to a type of food, drink, or spice not commonly used in the observed population. Or because the effect and underlying mechanism of the interaction, despite such interaction being common, have gone unnoticed. Certain common interactions may remain unknown because of over-determination of the effect by multiple causes. For instance, the fact that grapefruit juice interferes with the intake of many drugs (including warfarin) through inhibition of the enzyme cytochrome P450 3A4 (CYP3A4) was unknown until its serendipitous discovery in 1989 (Bailey et al. 1998).

iv.   In the case that the patient is not well enough represented by the study population, there may be certain mechanisms through which the intervention produces side effects in that specific patient that are absent in the study population. This means

that, although all the mechanisms through which the intervention produces side effects in the study population were identified (no problem iii) and it was corroborated that none of them is present in the patient (no problems i and ii), other hazardous mechanisms could be present in her. As a consequence, even if the premise of the disconfirming-scenario prediction was met, the intervention could produce relevant side effects.

The possibility of additional hazardous mechanisms is the reason why practitioners are typically reluctant to make predictions about safety from a population study to pregnant women, multi-morbid patients, older patients, and other patient groups usually excluded from clinical studies. Regarding these patient groups, we often have enough knowledge to predict in advance that, given some specific mechanisms at place in them, results on safety cannot be directly applied from population studies. Nonetheless, these precautions are not always possible, since some interventions can provoke rare and idiosyncratic reactions in some patients, where it is not easy to say how the patient differs from the rest of the population. For example, some children show a rare and fatal liver reaction to the anti-epileptic drug valproic acid. However, from the 1970s, when it was first marketed, to 2010, because the mechanism of interaction in those specific children was unknown, it was not possible to predict which children had the propensity to be fatally injured by the drug (Price et al. 2011). Only after discovering that the drug interacts with a mutated form of the mithocondrial protein POLG, safety predictions could be done from the frequency of the side effect in populations to the single patient (Sitarz et al. 2014). This and similar examples, once again, show how lack of mechanistic knowledge can seriously undermine safe use of drugs.

The disconfirming role of evidence of mechanisms in predictions about safety faces important difficulties. It means that a safety (or absence of side effects) claim can hardly be inferred from a population study to a particular patient exclusively on the basis of evidence of mechanisms. In order to reasonably establish the safety of a treatment for a patient, evidence of

mechanisms should be complemented by other kinds of evidence. Such evidence could, for instance, come from a thorough mapping and understanding of the patient's specific context.

5.2. Confirming predictions about safety

Evidence of mechanisms may also support the extrapolation of an unsafety claim—i.e., presence of relevant side effects—from a population study to a particular patient. The confirming role of evidence of mechanisms in predictions about safety could be characterised as follows. If (one or more) mechanisms through which certain intervention produces side effects in the study population are present in the patient, it can be concluded that the intervention would produce side effects and is not safe for the patient.

The confirming role of evidence of mechanisms in predictions about safety is not severely undermined by the problems faced by the disconfirming role.

i. In cases where it is not possible to check the presence of certain mechanisms through which an intervention produces side effects in the study population, it is often possible to corroborate the presence of other relevant mechanisms in a particular patient. In predictions about safety, we are not concerned about a specific side effect (or hazardous mechanism), but about side effects (or hazardous mechanisms) in general. Therefore, the presence of any relevant mechanism would undermine the safety of the intervention.

ii. Even if, given their large number or the limited resources available, it is unfeasible to check the presence of all the known mechanisms through which an intervention produces side effects in the study population, the presence of some of them in a single patient could be corroborated. It should be noted that often not all the relevant mechanisms are equally expensive or difficult to detect. The identification of some relevant mechanisms would be enough for considering that an intervention is unsafe. The presence of all the mechanisms (and the potential occurrence of all the side effects) is not required for unsafety.

iii. Although only some mechanisms through which an intervention produces side effects in the study population are known, some of the known hazardous mechanisms could be identified in the single patient. In those cases, despite the fragmentary knowledge about the study population, absence of safety for the patient can be inferred on the bases of the common relevant mechanisms.

iv. Even if only part of the mechanisms through which an intervention produces side effects are present in the study population, the study population can provide a reference for corroborating the unsafety of the intervention for a particular patient. Identifying in the patient some of the hazardous mechanisms present in the study population is enough for considering that the intervention would produce side effects and is not safe for her.

Masking and mechanisms' absence of regularity, which undermine the confirming role of evidence of mechanisms in causal prediction about efficacy (see subsection 3.1), would not be very problematic in mechanism-based predictions about side effects. Firstly, although unnoticed interfering mechanisms present in the patient could influence in the effects and/or the identified hazardous mechanisms, it is unlikely that they completely mask the identified mechanisms and prevent the side effects. They would probably just modify some side effects. And secondly, the identified mechanisms may not behave in the patient as in the study population, but that absence of regularity would hardly completely prevent the side effects. As in the case of masking, the irregularity would probably just modify some side effects.

The confirming role of evidence of mechanisms in predictions about safety is not undermined by the problems usually faced by mechanism-based predictions. This means that, once some mechanisms through which certain intervention produces relevant side effects in the study population are identified in a particular patient, it can be claimed that the intervention would probably produce side effects and is not safe for the patient. Further investigations on other hazardous mechanisms or complementary evidence are welcome, but they are not required for making a reliable prediction about lack of safety.

Consider, for instance, the following example. It might be found out that a patient is allergic against one of the components of certain drug formulation, either from previous history or from serological analysis. In that case, it would be justified to conclude that the drug would probably produce an allergic reaction and is not safe for the patient. Obviously, there might be other mechanisms for other side effects, and it could be unfeasible to check for everything. Furthermore, unknown interfering mechanisms or irregular behaviour of known mechanisms could influence the allergic reaction. Nonetheless, the knowledge already available is enough to infer that the patient is likely to experience an allergic reaction against the drug.

6. Conclusion

Evidence of mechanisms contribute significantly to decision-making in the clinical practice. Nevertheless, its reliability and the support it gives to causal prediction may vary considerably. In order to clarify the contribution of evidence of mechanisms to clinical reasoning, we have identified and analysed the different roles that it can play in the involved causal predictions: predictions about efficacy and predictions about safety. Regarding predictions about efficacy, we have argued that evidence of mechanisms is more reliable for determining that an intervention would not produce a specific (target or untargeted) effect in a patient than for establishing that it would produce the effect. With regard to predictions about safety, on the contrary, we have argued that evidence of mechanisms is more reliable for establishing that an intervention would produce side effects in a patient and is not safe than for determining its safety.

Our analysis is intended as a help for critical reasoning and evaluation of evidence-based choices in the clinical setting, and it has important practical implications. It shows that, generally, evidence of mechanisms is more decisive for discarding inadequate treatments than for identifying suitable ones. Evidence of mechanisms by itself can hardly support that an intervention, which was effective or safe in a study population, will be effective or safe for a particular patient. This, of course, does not mean that in these cases evidence of mechanisms should be discarded as useless. Rather, predictions based on this type of evidence should seek further support or, at least, be taken with caution. On the other hand, according to our analysis,

evidence of mechanisms can offer sufficient basis for predicting that an intervention that has been

shown to be effective or safe in a study population might not be so for the single patient.

References

Andersen, H. 2012. "Mechanisms: What Are They Evidence for in Evidence-based Medicine?" *J Eval Clin Pract* 18 (5): 992–99. https://doi.org/10.1111/j.1365-2753.2012.01906.x.

Bailey, D. G., et al. 1998. "Grapefruit Juice–Drug Interactions." *Br J Clin Pharmacol* 46 (2): 101–10. https://doi.org/10.1046/j.1365-2125.1998.00764.x.

Chen, S. 2012. "Clinical Uses of Botulinum Neurotoxins: Current Indications, Limitations and Future Developments." *Toxins* 4 (10): 913–39. https://doi.org/10.3390/toxins4100913.

Clarke, B., et al. 2013. "The Evidence That Evidence-Based Medicine Omits." *Prev Med* 57 (6): 745–47. https://doi.org/10.1016/j.ypmed.2012.10.020.

———. 2014. "Mechanisms and the Evidence Hierarchy." *Topoi* 33 (2): 339–60. https://doi.org/10.1007/s11245-013-9220-9.

Howick, J. 2011a. *The Philosophy of Evidence-Based Medicine*. Oxford, UK: Wiley-Blackwell.

———. 2011b. "Exposing the Vanities—and a Qualified Defense—of Mechanistic Reasoning in Health Care Decision Making." *Philos Sci* 78 (5): 926–40. https://doi.org/10.1086/662561.

Howick, J., P. Glasziou, and J. K Aronson. 2010. "Evidence-Based Mechanistic Reasoning." *J R Soc Med* 103 (11): 433–41. https://doi.org/10.1258/jrsm.2010.100146.

———. 2013. "Problems with Using Mechanisms to Solve the Problem of Extrapolation." *Theor Med Bioeth* 34 (4): 275–91. https://doi.org/10.1007/s11017-013-9266-0.

Illari, P. M. 2011. "Mechanistic Evidence: Disambiguating the Russo–Williamson Thesis." *International Studies in the Philosophy of Science* 25 (2): 139–57. https://doi.org/10.1080/02698595.2011.574856.

Illari, P. M., and J. Williamson. 2012. "What Is a Mechanism? Thinking about Mechanisms across the Sciences." *Eur J Philos Sci* 2 (1): 119–35. https://doi.org/10.1007/s13194-011-0038-2.

Khan, K. S., I. Hayes, and D. J. Buggy. 2014. "Pharmacology of Anaesthetic Agents II: Inhalation Anaesthetic Agents." *Continuing Education in Anaesthesia, Critical Care & Pain* 14 (3): 106–11. https://doi.org/10.1093/bjaceaccp/mkt038.

Mallal, S., et al. 2002. "Association between Presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and Hypersensitivity to HIV-1 Reverse-Transcriptase Inhibitor Abacavir." *The Lancet* 359 (9308): 727–32. https://doi.org/10.1016/S0140-6736(02)07873-X.

Martin, A. M., et al. 2007. "Immune Responses to Abacavir in Antigen-Presenting Cells from Hypersensitive Patients." *AIDS* 21 (10): 1233–44. https://doi.org/10.1097/QAD.0b013e3280119579.

Misra, V. P., et al. 2012. "Factors Influencing Response to Botulinum Toxin Type A in Patients with Idiopathic Cervical Dystonia: Results from an International Observational Study." *BMJ Open* 2 (3): e000881. http://dx.doi.org/10.1136/bmjopen-2012-000881.

Norén, G N., J. Hopstadius, and A. Bate. 2013. "Shrinkage Observed-to-Expected Ratios for Robust and Transparent Large-Scale Pattern Discovery." *Stat Methods Med Res* 22 (1): 57–69. https://doi.org/10.1177/0962280211403604.

Osimani, B. 2013. "Until RCT Proven? On the Asymmetry of Evidence Requirements for Risk Assessment." *J Eval Clin Pract* 19 (3): 454–62. https://doi.org/10.1111/jep.12039.

Parkkinen, V., et al. 2018. *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*. Cham, Switzerland: Springer.

Pérez-González, S., and V. Iranzo. Forthcoming. "Assessing the Role of Evidence of Mechanisms in Causal Extrapolation." *THEORIA. An International Journal for Theory, History and Foundations of Science*. https://doi.org/10.1387/theoria.21642.

Price, K. E., et al. 2011. "Effects of Valproic Acid on Organic Acid Metabolism in Children: A Metabolic Profiling Study." *Clin Pharmacol Ther* 89 (6): 867–74. https://doi.org/10.1038/clpt.2011.47.

Reiss, J. 2010. "Across the Boundaries: Extrapolation in Biology and Social Science, Daniel P. Steel. Oxford University Press, 2007. Xi + 241 Pages." *Econ Philos* 26 (3): 382–90. https://doi.org/10.1017/S0266267110000325.

Rocca, E. 2017. "Bridging the Boundaries between Scientists and Clinicians—Mechanistic Hypotheses and Patient Stories in Risk Assessment of Drugs." *J Eval Clin Pract* 23 (1): 114–20. https://doi.org/10.1111/jep.12622.

Rocca, E., R. L. Anjum, and S. Mumford. 2020. "Causal Insights from Failure: Post-Marketing Risk Assessment of Drugs as a Way to Uncover Causal Mechanisms." In *Uncertainty in Pharmacology: Epistemology, Methods, and Decisions*, eds. A. LaCaze and B. Osimani, 39–57. Cham, Switzerland: Springer.

Russo, F., and J. Williamson. 2007. "Interpreting Causality in the Health Sciences." *International Studies in the Philosophy of Science* 21 (2): 157–70. https://doi.org/10.1080/02698590701498084.

Ruthenberg, K. 2016. "About the Futile Dream of an Entirely Riskless and Fully Effective Remedy: Thalidomide." *HYLE–International Journal for Philosophy of Chemistry* 22: 55–77. https://doi.org/10.1142/9789811233548_0006.

Salmon, W. C. 1998. *Causality and Explanation*. Oxford, UK: Oxford University Press.

Sitarz, K. S., et al. 2014. "Valproic Acid Triggers Increased Mitochondrial Biogenesis in POLG-Deficient Fibroblasts." *Mol Genet Metab* 112 (1): 57–63. https://doi.org/10.1016/j.ymgme.2014.03.006.

Sridhar, S., et al. 2018. "Effect of Dengue Serostatus on Dengue Vaccine Safety and Efficacy." *N Engl J Med* 379 (4): 327–40. https://doi.org/10.1056/NEJMoa1800820.

Steel, D. 2008. *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford, UK: Oxford University Press.

Tonelli, M. R., and J. Williamson. 2020. "Mechanisms in Clinical Practice: Use and Justification." *Med Health Care Philos* 23 (1): 115–24. https://doi.org/10.1007/s11019-019-09915-5.

Torres, S., et al. 2014. "Neutralizing Antibodies to Botulinum Neurotoxin Type A in Aesthetic Medicine: Five Case Reports." *Clin Cosmet Investig Dermatol* 7: 11. https://doi.org/10.2147/CCID.S51938.

van Eersel, G. G., G. V. Koppenol-Gonzalez, and J. Reiss. 2019. "Extrapolation of Experimental Results through Analogical Reasoning from Latent Classes." *Philos Sci* 86 (2): 219–35. https://doi.org/10.1086/701956.

Endnotes

[1] Note that by 'relevant' here we mean 'relevant given the known mechanism of action'.

[2] It should be noted that this is not necessarily an argument to discourage the use of evidence of mechanisms in clinical practice. Rather, it can be understood as an argument for an increased alert and effort to constantly improve and expand mechanistic understanding. Notice that, in the cases in which we make a wrong prediction and the intervention does not work as expected, we have the chance to investigate the reason of failure and improve our mechanistic knowledge (Rocca 2017; Rocca, Anjum, and Mumford 2020).