Can understanding mechanisms solve the problem of extrapolating from study to target populations (the problem of 'external validity')?

J Howick[1] • P Glasziou[2] • JK Aronson[1]

[1]Department of Primary Care Health Sciences, University of Oxford, Oxford OX2 6GG, UK
[2]Bond University Centre for Research in Evidence-Based Practice (CREBP), Robina QLD 4226, Australia
Correspondence to: J Howick. Email: jeremy.howick@phc.ox.ac.uk

The problem of extrapolation

The problem of extrapolation arises because average study results do not always apply to target populations. Three examples illustrate different aspects of the problem.

In the European Atrial Fibrillation Trial (EAFT) the risk of intracranial haemorrhage in patients with atrial fibrillation taking warfarin was minimal.[1] However, when warfarin was given to certain patients in clinical practice (those in sinus rhythm) it increased the risk of haemorrhage.[2] Treatments that benefit study patients may be useless or even harmful for some patients in the potential target population.

In another study carotid endarterectomy appeared to increase overall mortality by 0.5%. But in a subgroup analysis of patients with severe carotid stenosis the procedure had a clear benefit.[3,4] This example illustrates why the terms 'external validity' and 'generalizability' can be misleading: the problem of extrapolation also arises when applying average study results to individuals or subpopulations within the study (particularizing).

Study and target contexts can also differ in ways that make extrapolation problematic. For example, the Tamil Nadu Integrated Nutrition Programs improved the nutritional status of Tamil children, by using a package of care that included food supplements, medical care and (importantly) nutrition education for pregnant mothers. Unfortunately, when the same intervention was used in Bangladesh it did not have any measurable benefit.[5] This was because Bangladeshi mothers-in-law and husbands were in charge of shopping so educating mothers in that context had little effect.[5] Even when study

and target populations are similar, relevant contextual factors must be shared for extrapolation to be justified.

One strategy for dealing with the problem of extrapolation is simple induction. For instance, the Consolidated Standards of Reporting Trials (CONSORT) checklist[6] includes an item about 'generalizability' that involves asking whether target populations would have met the inclusion criteria for a trial. But as we illustrate below, even if a patient would not have met the inclusion criteria it is nevertheless sometimes safe to implement the study results, while in other cases although the patient would have met the inclusion criteria it is unsafe to do so.

Critics and proponents of Evidence-Based Medicine (EBM) suggest that knowledge of pathophysiological mechanisms can solve the problem of extrapolation.[7,8] Guyatt et al.[8] suggest 'a good understanding of pathophysiology is necessary … for appropriate interpretation of evidence (especially in deciding on its generalizability)'. However, no explicit guidelines for implementing this suggestion are provided. Paul Glasziou and Gordon Guyatt (in conversation) offer the following example to illustrate the EBM proposal. If a trial of an intervention excludes everyone over the age of 60, they claim the intervention is likely to work for a 61-year-old but not for a 90-year-old. Presumably they take it that the success of the intervention depends on the operation of pathophysiological mechanisms that change only slowly beyond 60. So the mechanisms would not have changed substantially in most 61-year-olds but would be highly likely to have changed by 90. This example is useful as a starting point, and here we elaborate on how to use mechanistic knowledge to justify extrapolation.

Mechanisms of action and why they do not always help

A mechanism of action is the causal chain or web linking the intervention with the clinical outcome via pathophysiologic mechanisms[9] (see Figure 1, middle). If we know the mechanism of action in the study population, and we know it is shared with the target population, then extrapolation is more likely to be justified. Unfortunately, we rarely have sufficiently complete knowledge of the mechanisms of action in the study to form any basis for comparison (and justifying extrapolation).

For example, widespread use of antiarrhythmic drugs for (supposedly) reducing mortality was based on an incomplete understanding of the mechanism(s) of action of the drugs. Several mechanisms (swallowing, gastric emptying, metabolism, circulatory and binding mechanisms) were involved in getting the drug to its pharmacological targets. These mechanisms are often well understood and are referred to as ADME (mechanisms for absorption, distribution, metabolism and excretion). Having reached their cellular targets, antiarrhythmic drugs were believed to reduce the frequency of ventricular extra beats (VEBs) by modifying the heart's electrochemical mechanism. Finally, a reduction in VEBs should (allegedly) reduce the risk of sudden death, presumably by reducing the risks associated with insufficient blood flow to vital organs. But subsequently a randomized trial of two antiarrhythmic drugs, encainide and flecainide, suggested that they increased mortality (see Figure 1).[10] The evidence for the mechanism of action linking antiarrhythmic drugs with a reduced frequency of VEBs was strong, but the evidence for the mechanism linking reduced VEBs with reduced mortality was weak. It turns out that antiarrhythmic drugs (and indeed most interventions) activate unsuspected mechanisms, making their effects on patient-relevant outcomes difficult to predict. Sometimes the unsuspected pathways can lead to a paradoxical effect;[11] for example antiarrhythmic drugs are proarrhythmic in about 7% of patients (see Figure 1, right-hand side). Many other recent examples show how incomplete knowledge of relevant mechanisms led to adoption of harmful treatments.[9] If our knowledge of mechanisms is flawed or incomplete, then any comparison of such mechanisms

Figure 1
Comparative clinical studies, mechanisms of action
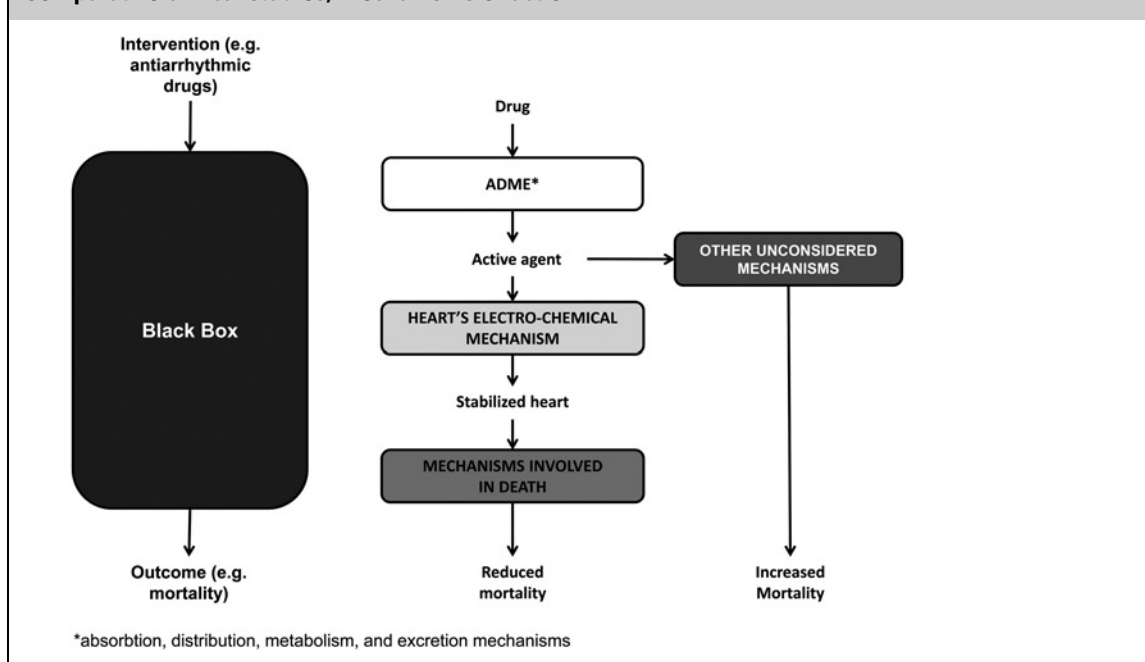
in study and target populations will not lead to justifiable extrapolation.

To make things worse, the actions of many mechanisms are discovered in tightly controlled laboratory experiments, eliminating potentially interfering variables, but, *in vitro* and *in vivo* mechanisms can behave differently. Knowledge gained in these conditions is inevitably oversimplified when compared with real clinical cases, often rendering inferences from *in vitro* and *in vivo* effects unwarranted.

How knowledge of mechanisms of action can be used to mitigate the problem of extrapolation

The National Institute for Health and Clinical Excellence (NICE) in the UK recommends macrogols for the treatment of chronic constipation in children under two years of age.[12] However, this is an unlicensed indication since the available evidence in children relates to those aged over two years.[13] Appealing to CONSORT would not support such use, because children under two years did not meet the inclusion criteria of published trials. Nevertheless, we know that the mechanism of action for how macrogols reduce constipation (by drawing fluid into the gut osmotically) is independent of age. Hence, the NICE recommendations are acceptable.

Conversely, the clinical evidence that metformin causes lactic acidosis comes from many anecdotal reports, and trial evidence has failed to reveal an increased risk. This is probably because the trials were not sufficiently powered to detect this rare complication. Nevertheless, current guidelines recommend that metformin should not be used in people with renal impairment.[14] Again following the CONSORT strategy would lead us to believe that metformin was acceptable for patients with renal impairment. However, we know that the mechanisms of action in patients with renal impairment differ from other patients; hence, extrapolation between results in these two groups is unjustified.[15]

Besides telling us when extrapolation is justified, knowledge of the mechanistic of action is also necessary for guiding adequate descriptions of interventions.[16,17] A description of metformin therapy would specify the dose, the dosage interval, the duration of therapy and so on, but not presumed irrelevant factors (such as clothing and planet position). Factors can be deemed irrelevant if they do not affect the mechanisms involved in the intervention's action.

A guide to using mechanisms to help solve the problem of extrapolation

While our understanding of mechanisms of action can be flawed, it would be unwise to ignore the cases in which our understanding is more complete than others.[9] Our proposed method for using mechanistic knowledge to mitigate the problem of extrapolation involves three steps and the use of a mechanistic taxonomy.

Step 1. Establish whether the problems with obtaining and interpreting mechanisms of action have been overcome. Sometimes we have correctly identified the mechanisms, and we have evidence that the links in the mechanistic chain are well established. For example, the proximate causes of stroke have been known for centuries.[18,19] A burst artery in the brain causes a haemorrhagic stroke, while an ischaemic stroke is caused by a blockage of an artery, by either thrombosis or embolism. Aspirin benefits patients who have had an ischaemic stroke, but may harm those who have had a haemorrhagic stroke. The cause of the stroke (identification of the mechanism that has been disturbed) can be discovered by a computed tomography scan. In this case identifying the mechanism helps us to identify patients who are likely to benefit or not from aspirin.

Box 1

What is known about this topic and what this study adds

What is known about this topic
- Average study results are sometimes not applicable to target populations.

What this study adds
- The problem of extrapolation is two problems: particularizing and generalizing.
- Appealing to inclusion/exclusion criteria is important but insufficient.
- Cautious use of mechanistic knowledge can justify extrapolation of study results in target populations.

Table 1

A taxonomy of the types of mechanisms that can inform the wisdom or otherwise of extrapolating from trial results to another population[*]

Type of mechanism	Examples[*]	Comments
Genetic	Patients whose colorectal tumours carry the wild-type KRAS genotype respond better to monoclonal antibodies that target epidermal growth factor receptors (EGFR)	In both cases, discovering the genetic components of the mechanisms responsible for therapeutic benefit (colorectal tumours) or an adverse drug reaction (abacavir) helps maximize benefit and reduce harm from these therapies; extrapolation from trial data to individual patients is guided by these genetic mechanisms
	The adverse cutaneous effects of abacavir can be avoided by screening for human leukocyte antigen (HLA) B*5701	
Age	Benoxaprofen proved efficacious in trials, but killed some elderly patients in routine practice	In all three cases extrapolation was/is dangerous because of age-related differences: *Benoxaprofen*: Benoxaprofen is metabolized in the liver; its pharmacokinetics is different in elderly patients with reduced liver function from the patients who were studied in trials
	The dosage of growth hormone for adults with growth hormone deficiency is 25% that of children, in spite of greater body mass	*Growth hormone*: The degree of growth required is much greater in children
	Antihypertensive drugs reduce total mortality in middle-aged patients but may not in elderly ones, or not to the same extent	*Antihypertensive drugs*: Many physiological mechanisms change with age, altering responses to these drugs
Sex	The risk of drug-induced lupus-like syndrome is greater in women than in men	The mechanisms that cause idiopathic systemic lupus erythematosus, which is more common in women, make them more susceptible to the drug-induced form; extrapolation from men to women may be dangerous
Physiological variants	Obese patients require different weight-related drug dosages	The dosage of a drug may be affected by its distribution into body fat; knowing which drugs are thus affected informs extrapolation from lean to obese individuals
Co-morbidities	Renal impairment, hepatic impairment	Knowledge of pharmacokinetic mechanisms can predict outcomes for drugs eliminated by the kidneys or liver; extrapolation from individuals with normal renal and hepatic function to those with impaired function requires pharmacokinetic knowledge
	Digoxin normally relieves heart failure, but paradoxically worsens it if there is also hypertrophic cardiomyopathy (HCM)	Increased cardiac contraction leads to reduced cardiac output in the face of obstruction to flow caused by the cardiomyopathy; one cannot extrapolate from, say, ischaemic heart disease to HCM when treating heart failure
Drug–drug interactions	Parkinsonism due to antipsychotic drugs should be treated with anticholinergic drugs, not levodopa	Antipsychotic drugs cause parkinsonism by inhibiting the action of dopamine at its receptors, making levodopa ineffective; one cannot extrapolate from idiopathic parkinsonism to the drug-induced form – the mechanisms are different
Sociological factors	Integrated Nutrition Programs succeeded in Tamil Nadu, but failed when implemented in Bangladesh	The mechanisms for improving the access of malnourished children to nutrition were not the same in the two countries (see text)

[*]Includes examples of both similar (generalizable) and dissimilar (non-generalizable) mechanisms

Table 2

A checklist for using mechanisms to justify extrapolation of study results to target populations

Question	Y/N
Is the mechanism of action supported by evidence?	
Are the mechanisms of action in study and target populations shared?	
Even if shared, are the mechanisms of action likely to behave differently (paradoxically)?	
Have we established that the mechanisms of action in the study and target populations are relevantly similar?	

If the answer to all these questions is 'Yes', then extrapolation may be justified. Otherwise, extrapolation may be risky

Step 2. Establish whether the mechanisms of action in the study and target populations are similar. Having specified the mechanistic chain linking the intervention and the outcome in the study and target populations, we can compare them. If mechanisms are shared (as they were in the macrogol example), extrapolation is more likely to be safe and beneficial; otherwise (as was the case in the metformin example), we should proceed cautiously.

Step 3. Establish that the study and target context are sufficiently similar. Finally, we need to ensure that contextual factors influencing whether relevant mechanisms are activated are shared. Failure to do this (as in the TINP example) can lead to unsuccessful extrapolation (Box 1).

Table 1 shows a taxonomy on which decisions about shared mechanisms of action can be based in cases of pharmacological interventions. It classifies mechanisms as determined by genetic factors, age- and sex-related factors, physiological variants, co-morbidities, drug–drug interactions and sociological factors. Examples are given in each case. Using this taxonomy, one can compare mechanisms of action in study and target populations.

Table 2 provides a checklist that can be used to decide whether average study results are applicable to target populations. Judgement will always

be required, and the problem of extrapolation will never be completely resolved. At the same time, extrapolation to target populations is sometimes better justified than at other times. If the answer to all the questions in Table 1 is 'Yes', then extrapolation is justified, while too many 'No' answers suggest reasons to be cautious.

Conclusions

We all hope that study results will benefit target populations. However, experience suggests that extrapolation based on simple induction can be problematic. Here we have argued that examining mechanisms of action in study and target populations can help justify successful extrapolation and also guide adequate descriptions of interventions. Our proposed method involves (i) establishing that common problems with obtaining knowledge of mechanisms of action have been overcome; (ii) verifying that relevant mechanisms in the study and target populations are similar; and (iii) establishing that study and target contexts are relevantly similar.

References

1 EAFT (European Atrial Fibrillation Trial) Study Group. Secondary prevention in non-rheumatic atrial fibrillation after transient ischaemic attack or minor stroke. *Lancet* 1993;**342**:1255–62

2 The Stroke Prevention in Reversible Ischemia Trial (SPIRIT) Study Group. A randomized trial of anticoagulants versus aspirin after cerebral ischemia of presumed arterial origin. *Ann Neurol* 1997;**42**:857–65

3 Rothwell PM. External validity of randomised controlled trials: 'to whom do the results of this trial apply?'. *Lancet* 2005;**365**:82–93

4 OCEBM Levels of Evidence Working Group. OCEBM 2011 Levels of Evidence: Oxford Centre for Evidence-Based Medicine, 2011

5 Nutrition NIo. *Endline Evaluation of Tamil Nadu Integrated Nutrition Project II.* Hyderabad: Indian Council of Medical Research, 1998

6 Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med* 2010;**7**:e1000251

7 Cartwright N. Will this policy work for you? Predicting effectiveness better: how philosophy helps (Presidential Address). *Philos Sci* 2012;**79**:973–89

8 Evidence Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992;**268**:2420–25

9 Howick J, Glasziou P, Aronson JK. Evidence-based mechanistic reasoning. *J R Soc Med* 2010;**103**:433–41

10 Investigators TCASTC. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial

of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989;**321**:406–12

11 Hauben M, Aronson JK. Paradoxical reactions: under-recognized adverse effects of drugs. *Drug Saf* 2006;**29**:970

12 National Collaborating Centre for Women's and Children's Health. In: NHS, ed. *Constipation in Children and Young People: Diagnosis and Management of Idiopathic Childhood Constipation in Primary and Secondary Care*. London: Royal College of Obstetricians and Gynaecologists, 2010

13 Candy D, Belsey J. Macrogol (polyethylene glycol) laxatives in children with functional constipation and faecal impaction: a systematic review. *Archiv Dis Childhood* 2009;**94**:156–60

14 PRODIGY. *Diabetes-Type 2 – Prescribing Information*. Newcastle-upon-Tyne: PRODIGY, 2012

15 Hanley JA, Lippman-Hand A. If nothing goes wrong, is everything all right? Interpreting zero numerators. *JAMA* 1983;**249**:1743–5

16 Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ* 2008;**336**:1472–4

17 Golomb BA, Erickson LC, Koperski S, Sack D, Enkin M, Howick J. What's in placebos: who knows? Analysis of randomized, controlled trials. *Ann Intern Med* 2010; **153**:532–5

18 Thompson JE. The evolution of surgery for the treatment and prevention of stroke. The Willis Lecture. *Stroke* 1996;**27**:1427–34

19 National Institute of Neurological Disorders and Stroke (NINDS). *Stroke: Hope Through Prevention*. Bethesda: National Institutes of Health, 1999