THEORY AND METHODS

A landmark for popperian epidemiology: refutation of the randomised Aldactone evaluation study

Elard Koch, Alvaro Otarola, Aida Kirschbaum

In 1999 a great multi-site clinical trial known as the randomised Aldactone evaluation study (RALES) showed that the use of spironolactone importantly reduced complications attributable to chronic heart failure without major negative side effects. Recently, RALES has been questioned by a large scale observational study in the Ontario population. In contrast with predictions, the complications and mortality increased dramatically because of hyperkalaemia, reaching dimensions that from a public health perspective are comparable to an epidemic. This review analyses both researches in the light of Karl Popper's science theory applying the *modus tollens* syllogism to the reality proposed by the main empirical enunciations that ensue from its epidemiological designs. RALES is deductively refuted because of the non-fulfillment of auxiliary assumptions that would act as reciprocal potential falsifiers in both studies, taking the logical form of a bi-conditional argument of the type: (a) *P-then-Q* and (b) *Q-if-X_P*, X_P being a set of potential falsifiers of Q as part of the explicit falsity content of P. From this popperian model, implications for clinical research are discussed.

See end of article for authors' affiliations

Correspondence to:
Professor E Koch, Division of Epidemiology, School of Public Health, University of Chile, 939 Independencia, Santiago, Chile; elard@123mail.cl

Accepted for publication 7 June 2005

A well known parable teaches us that every seed must die to give its fruit. With the thought of Sir Karl R Popper one decade from his death, exactly that is occurring. Popperian critical rationalism starts to give its fruit in the applied sciences and the future of the practice of epidemiology seems to fit in with the popperian approach of what we call, since Carol Buck's article in 1975, popperian epidemiology.[1] Ever since the decade of the 1980s numerous epidemiologists have approached the subject in a rich and intense philosophical and methodological debate.[2–13] Although refutationism is one among a number of critical tools for empirical sciences,[1–5 11–13] popperian epidemiology has often been presented as an opposed approach to a paradigm of inductive inference,[6–10 12] which represents the presently dominant current in epidemiology.[11–18] Nevertheless, the recent publication of a populational study carried out in Ontario, Canada,[19] which questions a great multi-site clinical trial[20] such as the randomised Aldactone evaluation study (RALES) revives the discussion about the difficulties in inductively transferring the results of experimental designs to the "real empirical world" of clinical practice and the way in which physicians make use of scientific evidence in the paradigm of evidence based medicine (EBM), nowadays influenced by a positivised critical rationalism.[21] In this article we analyse the results of these studies in the light of Popper's science theory as far as scientific evidence is concerned and we evaluate the applicability of the falsifiability principle to the main empirical enunciations that emerge from their epidemiological designs.

AN APPROACH TO KARL POPPER'S SCIENCE THEORY

Neopositivism argues in favour of the inductive method by means of the process of justification or verification through repetition (addition of multiple verified cases) formulating the principle of positive verification as a significance criterion that can lead to a degree of certainty or truth, a type of inductive "support".[22 23] Popper is opposed to all types of verificationism.[24] The inductive process is not demonstrable in a logical way, as there exists an asymmetry between the part observed and the whole. The problem of induction consists in logically justifying the inductive inferences. Accepting Hume's arguments,[25] Popper considers that this inductive step is not justified, as there will always exist unobserved cases, no matter how large the number of individual observations may be: thus, we will only reach probable conclusions, and following this road, we come to a situation where there are many theories propounded, all of them probable, but leading to no advance in knowledge. For real progress in knowledge, we must increase empirical content of theories refuting them and correcting their errors, and the rebuttal of theories is conditioned by deductive reasoning that, unlike induction, allows to reach a sure or necessary conclusion.[24]

Opposing the intent of refuting his own theories to the normal intent of verifying them, Popper deduces the falsifiability principle, no longer as a criterion of empirical significance, but as a manner of delimiting what science is and what it is not.[23] A theory is scientific if it can be falsified by means of experience or by means of its internal inconsistency. Falsification is the corroboration that an enunciation is false for not having resisted the falsification test. This procedure does not require a process of induction that is logically impossible, but a simple logical

deduction. This deduction is based on the syllogism called *modus tollens*, which can be formulated as:

P → Q[If P-then-Q]

~ Q [and the contrary result is obtained "no-Q"]

~ P [then "no-P"]

On the basis of the *modus tollens* Popper points to the asymmetry between verifiability and falsifiability by means of which, although scientific theories can never be empirically verifiable, they can be falsifiable by the non-observed cases, and thus the *modus tollens* turns into the logical rule of the empirical sciences.[23][24] The induction principle is rejected by Popper because an inductive method, as a logically valid process of contrasting, is inadequate, as theory always precedes observation; it is even necessary for choosing among the innumerable observation objects that are often mentioned as a technological problem for experimental design.[26]

On the basis of scientific deductive methodology, if a series of consequences derives from one theory and at the same time we are able to set forth a series of contradictory statements with those consequences, we have at our disposal a series of potential falsifiers of the theory. For a theory to be falsifiable it has to preclude, as a minimum, one empirical happening.[23] For example, it is what we would usually call inclusion and exclusion criteria in the design of an epidemiological experiment or clinical trial. If the number of possible falsifiers of a theory is greater than that corresponding to another theory in competence, the first theory will have more occasions to be refuted by the experience; therefore, when compared with the second theory, we can say that it is "falsifiable to a larger degree". It follows that a scientific theory has a greater degree of corroboration when it has resisted more criticisms and has been subject to more severe contrasts and not when it has been more verified. Thus, honesty and scientific objectivity reveal themselves in the formulation of falsifiable theories, which are tested without recurring to stratagems of immunisation against error. Scientific theories are such because of their explanatory, informative, and predictive capacity, not because of their capacity of adapting actual happenings to their conceptual formulation, which can be described as subjective and unjustified defence of a theory, a situation to which methodological inductivism would lead in a verificationist approach.[27] In Popper's approach a hypothesis will be accepted as provisionally true (corroborated) only when it continues to explain observed data after repeated attempts to falsify it have failed.[28]

EVIDENCE BASED MEDICINE AND RANDOMISED CLINICAL TRIALS

EBM has recently been described as the "integration of best research evidence with clinical expertise and patient values", and a new model of "based on evidence decision-making" has been proposed.[29][30] Despite the fact that the model considers evidence proceeding from research as one more component of the process, its weight will be greater in clinical decision making with the progress of medical knowledge, thus diminishing the uncertainty and heuristic component of EBM, seeking maximal safety and efficacy in therapeutic interventions.[31][32] The classification of evidence reliability into hierarchical levels, according to the scientific rigour of methodological designs, has placed randomised clinical trials (RCTs) and systematic reviews at the highest level, followed by controlled observational studies, and finally by non-controlled studies and expert opinions.[30–34] Special attention is granted to the evidence provided by large multi-site RCTs over small RCTs. Their main characteristic is that the patients composing the sample are recruited from different parts of

the world.[35] The weight of evidence proceeding from these RCTs is often considered decisive for accepting or rejecting new therapeutic interventions, as they provide a greater inductive "support" for generalising their results to the population and can considerably influence the prescription of a drug on clinical practice.[35][36]

Nevertheless, the hierarchised consideration of evidence can induce a somewhat dogmatic critical rationalism in researchers and clinicians and underrate evidence of smaller RCTs and other epidemiological designs.[21][34][35][37][38] Despite the fact that RCT is considered to be the epidemiological design of highest scientific rigour on an experimental scale,[29–39] it has certain limitations. The main presumption of this methodology is that disease variations can be sufficiently quantified as to allow the real probability of differences between active treatment and placebo to have occurred randomly. As long as there is a sufficiently large size of the sample and a high rate of events, the validity of the study will not be compromised and the results will have clinical value.[31][35][37] Nevertheless, with strict inclusion and exclusion criteria, numerous patients are often excluded from their designs and social determinants are often not included.[38] The greater the complexity of the RCT, more careful is the sample selection, more rigorous and limited the inclusion of sick people in the study, and more controlled the intervention on them. Thus, both the population studied and the environment in which the experimental intervention is carried out are different from those found in clinical practice.[40] In fact, these restrictive and standardised conditions increase the internal validity of the conclusions but they negatively affect the external validity, namely the possibility to transfer the results to patients with characteristics different from those of the selected sample.[35][37][38][40] From this perspective, both clinical epidemiology and EBM in practice essentially respond to a model of hypothesis verification and its subsequent generalisation by the inductive method.

THE RANDOMISED ALDACTONE EVALUATION STUDY (RALES)

A good example of a great multi-site RCT is the RALES research conducted by Pitt *et al*[20] involving 195 centres in 16 countries (table 1). The study was designed to evaluate the effectiveness of spironolactone (Searle's Aldactone), an aldosterone inhibitor in chronic heart failure (CHF). Its main hypotheses were that the use of low doses of this drug combined with a standard therapy of angiotensin converting enzyme (ACE) inhibitors, would importantly reduce the mortality and complications in severe CHF without major negative side effects.

In the RALES study, 1663 patients with severe CHF and no history of renal insufficiency, diabetes, or hyperkalaemia were included. A total of 822 patients were randomly assigned to receive 25 mg of spironolactone daily, and 841 to receive placebo. The study found that spironolactone added to ACE inhibitors and diuretics in patients with CHF attributable to systolic left ventricular dysfunction significantly decreased all cause mortality by 30%, sudden death by 29%, and hospitalisation attributable to advancement of CHF by 35% over two years compared with placebo. The incidence of serious hyperkalaemia was minimal (below 2%) in both groups of patients. The conclusion of its authors was that the block of aldosterone receptors by spironolactone in addition to standard therapy, substantially reduced the risk of both morbidity and death among patients with severe CHF with minimal negative side effects.

This RCT was stopped early (November 1998) and received great attention because of the beneficial effects seen during monitoring.[41] There was a rapid diffusion of its results even before the original article by Pitt *et al* was available on line in

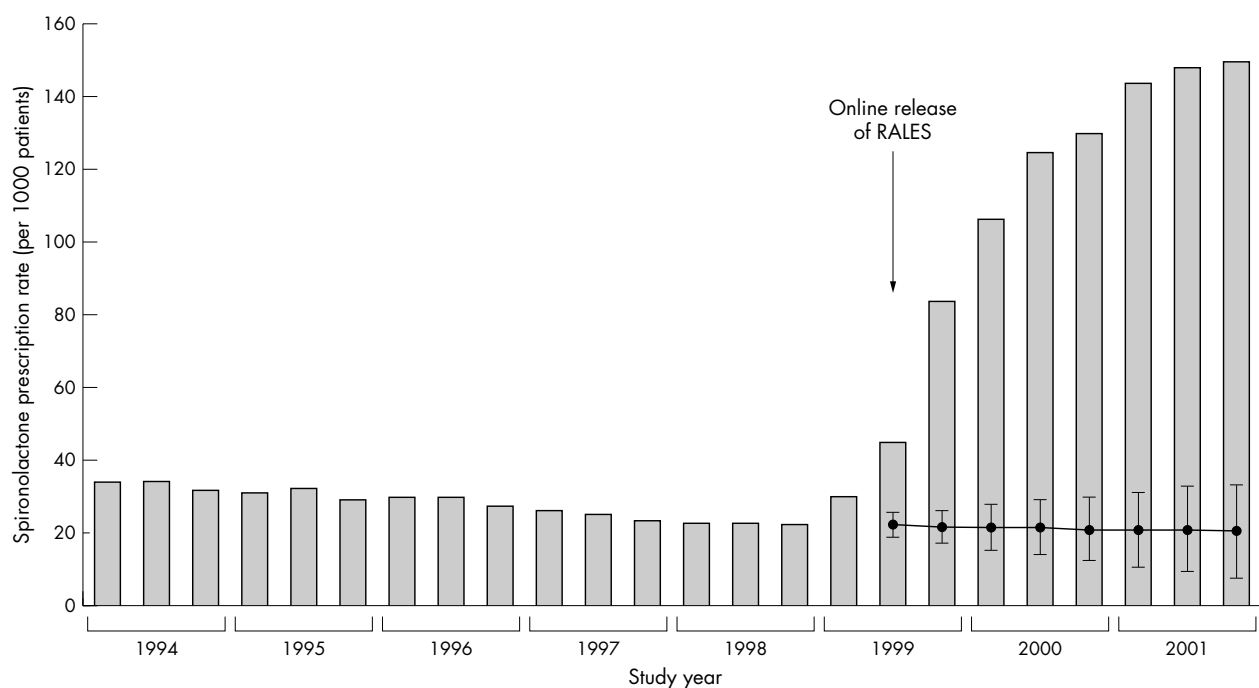**Table 1** The randomised Aldactone evaluation study (RALES)[20]

Purpose	To evaluate whether treatment with 25 mg spironolactone per day can diminish total mortality in patients with severe cardiac insufficiency secondary to systolic dysfunction who are under standard treatment including angiotensin converting enzyme (ACE) inhibitor when the patient tolerated it.
Design	A multi-site, double blind randomised clinical trial
Countries	Belgium, Brazil, Canada, France, Germany, Japan, Mexico, Netherlands, New Zealand, South Africa, Spain, Switzerland, United Kingdom, USA, and Venezuela.
Population	1663 Patients with severe chronic heart failure (CHF). Four countries enrolled more than 100 patients with a 68% of total sample (n = 1138). Twelve countries enrolled less than 100 patients with a 32% of total sample (n = 525).
Treatment being studied	Patients were randomised to 25 mg spironolactone (n = 822) or placebo (n = 841). After an eight week follow up the dose could be increased to 50 mg/day if progression of the cardiac insufficiency was observed without evidence of hyperkalaemia (K>6 mmol/l). If hyperkalaemia appeared, the dose of spironolactone could be reduced to 25 mg/48 h, although previous adjustment of the dose of other pharmaceuticals was recommended. Monthly evaluations were carried out during the first three months, three monthly evaluations during the first year, and after that, every six months.
Primary objective	Death due to any cause. Secondary objectives: death of cardiovascular origin, hospitalisation for cardiovascular cause, combination of the two previous causes. Changes in the NYHA class.
Analysis	For the intention of trying. Average follow up, two years. The study was detained when efficacy in the spironolactone group was shown.
Results	Reduction of mortality by 30% (RR = 0.70; 95% CI = 0.60 to 0.82; p<0.001). Reduction of hospitalisation for heart failure by 35% (RR = 0.65; 95% CI = 0.54 to 0.77; p<0.001)

June 1999. The use of spironolactone blossomed after its publication with greater hopes for the medical community and CHF patients.[42]

POPULATIONAL REFUTATION IN THE ONTARIO STUDY

This study carried out by Juurlink *et al*[19] was a population based, time series analysis of health care databases in Ontario, Canada, from 1 January 1994 to 31 December 2001. During this period, Ontario had a population of about 12.3 million inhabitants, of which about 1.3 million were 65 years of age or older. The number of patients who were treated with an ACE inhibitor after hospitalisation for CHF rose gradually over time, from 20 820 in early 1994 to 32 283 by late 2001. Before the publication of RALES, 4539 patients

hospitalised for CHF were given spironolactone, and after RALES, 12 422 patients. The spironolactone prescription rate remained comparatively constant from early 1994 (34 per 1000 patients) until early 1999 (30 per 1000 patients). After the publication of RALES, the rate of prescriptions increased by a factor of about five, to 149 per 1000 by late 2001 (fig 1). The median dose of spironolactone was 25 mg per day. The increased spironolactone use was accompanied by a pronounced parallel increase in hospitalisations for hyperkalaemia from 2.4 per 1000 patients in 1994 to 11.0 per 1000 patients in 2001 (fig 2), and the associated mortality rose from 0.3 to 2.0 per 1000 patients. According to the results of this population based study, every 1000 additional prescriptions for spironolactone issued after RALES led to 50 additional admissions for hyperkalaemia. The authors' conclusion was that the publication of RALES coincided

Figure 1 Rate of prescriptions for spironolactone among patients hospitalised for CHF in Ontario, Canada, 1994 to 2001. Each bar shows the rate of hospital admission for hyperkalaemia per 1000 patients during one four-month interval. The horizontal line beginning in the second interval of 1999 shows projected prescription rates derived from interventional autoregressive integrated moving average (ARIMA) models, with vertical lines representing the 95% confidence intervals. (Source: Juurlink DN, Mamdani M, Pharm D, *et al*. Rates of hyperkalemia after publication of the randomized aldactone evaluation study. *N Engl J Med* 2004;**351**:543–51. Copyright 2004 Massachusetts Medical Society. All rights reserved.)
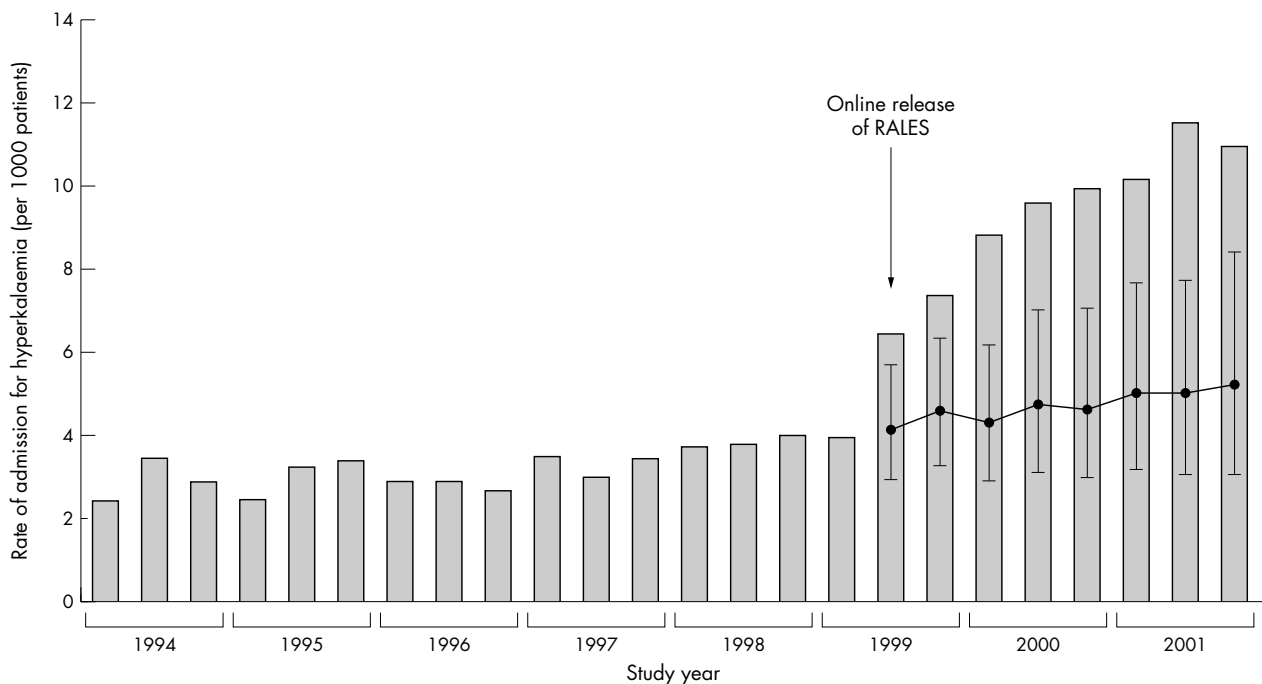
with the abrupt increment in the prescription rates of spironolactone and, at the same time, with an increment in the morbidity and mortality from hyperkalaemia, recommending the more judicious use of spironolactone and more strict laboratory controls to prevent the complications seen.

One year before the publication of the Ontario study, smaller clinical studies described complications arising from the inappropriate use of spironolactone.[43–45] From these reports, it was obvious that many physicians were giving spironolactone to patients who did not meet the RALES entry and exclusion criteria. MacMurray and O'Meara,[46] point out the incongruence of the inclusion and exclusion criteria of the large RCTs with the "real world" of clinical practice. In this case, the evidence of RALES became manifest in a generalisation of its results beyond the scope of inclusion and exclusion criteria of the study. On the other hand, the concept of inductive "support" leads to a false sensation of safety that would provoke an injudicious use of the drug and subsequent adverse outcomes.[45] Besides, the inclusion and exclusion criteria of the RCTs, which respond to a positive verification model, can partly be considered strategies of immunisation against error, which implies that their generalisation to the population be limited and subsequently hazardous. Finally Goldfarb,[47] as a counterpart, speculates with an ecological fallacy in the Ontario experience due to added uncontrolled morbid conditions.

The Ontario study questions, at a populational level, one of the main conclusions of RALES with respect to the complications seen attributable to hyperkalaemia. According to Popper, all empirical sciences share the same hypothesis deductive method[27 48] and the *modus tollens* logical rule, a matter that we will presently analyse in the RALES and Ontario research experiences.

LOGICAL ARGUMENTATION

According to Lakatos[49] and other falsificationist philosophers,[50 51] no hypothesis is outright falsified, but only a hypothesis along with an unspecified number of auxiliary assumptions. Keeping this in mind, RALES and the Ontario study follow a time dependent sequence of events forming a complete logical analysis unit and *modus tollens* in its simplest form *(P-then-Q; no-Q; then no-P)* it can be expressed as: "spironolactone is associated to minimal complications by hyperkalaemia" represented by the symbol P, and its logical consequence as: "spironolactone in Ontario will be associated to minimal complications by hyperkalaemia" represented by the symbol Q. In contrast with what was anticipated, use of spironolactone was associated to an increase in complications by hyperkalaemia. In this case the observational predicate "increase in complications" is represented by the symbol *no-Q* and therefore the assertion of P "minimal complications" is false *(no-P)*. Nevertheless, the corroboration of RALES is more complex and assumes the form of a bi-conditional proposition[51] of the type (i) *P-then-Q* and (ii) *Q-if-X_P*, the symbol X_P representing the exclusion and inclusion criteria of the study: Let X_P be the series of observational predicates X_1, X_2, X_3....X_n as explicit falsity-content of P and therefore, a series of potential falsifiers of Q. Expressed as a canonical logical argumentation: (i) If *P-then-Q*, and (ii) *Q-if-X_P*, X_P being (iii) the series X_1, X_2, X_3....X_n and (iv) *no-X_n* then (v) *no-X_P*, (vi) *no-Q* then (vii) *no-P*. Deductively *no-X_P* can be any criterion of inclusion or exclusion or conditional criterion of RALES that was not fulfilled in the Ontario population (table 2), for example, let X_3 be "patients with a fraction of ejection of the left ventricle <35%", X_4 "diabetic patient", X_{14} "close laboratory monitoring", etc. In this sense, it would suffice for one of these auxiliary assumptions contained in X_P to be false so that Q were not fulfilled and P would result refuted. From the beginning, RALES included numerous

Figure 2 Rate of hospital admission for hyperkalaemia among patients for CHF in Ontario, Canada, 1994 to 2001. Each bar shows the observed spironolactone prescription rate per 1000 patients during one four-month interval. The horizontal line beginning in the second interval of 1999 shows projected admission rates for hyperkalaemia derived from interventional ARIMA models, with vertical lines representing the 95% confidence intervals. (Source: Juurlink DN, Mamdani M, Pharm D, *et al.* Rates of hyperkalemia after publication of the randomized aldactone evaluation study. *N Engl J Med* 2004;**351**:543–51. Copyright 2004 Massachusetts Medical Society. All rights reserved.)

Table 2 Inclusion criteria, exclusion criteria, and conditional criteria as part of the explicit falsity content (X_P) of RALES experimental design. Adapted from original article published by Pitt et al[20]

X_P arguments	Observational predicates	Criteria
X_1	Patients having suffered chronic heart failure IV degree of the New York Heart Association during the six previous months.	Inclusion
X_2	Functional class III or IV at the moment of inclusion, diagnosed as cardiac insufficiency at least six weeks earlier.	Inclusion
X_3	Patients with an ejection fraction of the left ventricle below 35% during the past six months.	Inclusion
X_4	Diabetes	Exclusion
X_5	Renal insufficiency or creatinine above 2.5 mg/dl	Exclusion
X_6	Congenital cardiopathy	Exclusion
X_7	Unstable angina	Exclusion
X_8	Patients with potassium >5 mm/l	Exclusion
X_9	Hyperkalaemia history	Exclusion
X_{10}	Operable valvular disease	Exclusion
X_{11}	Primary hepatic failure	Exclusion
X_{12}	Cardiac transplant indication	Exclusion
X_{13}	Active cancer or any life threatening disease	Conditional
X_{14}	Close laboratory monitoring during treatment	Conditional (research process)
X_{15}	Potassium supplements were not recommended, except when potassium concentrations in serum were below 3.5 mmol/l	Conditional (research process)
X_{16}	Patients had to be in treatment with a loop diuretic and an angiotensin converting enzyme inhibitor if it were tolerated	Conditional (research process)
X_{17}	Treatment with digitalis and vasodilators was permitted but not with potassium sparing diuretics	Conditional (research process)

exclusion criteria and the inclusion of patients was very careful, its empirical results being ''highly falsifiable'' and its extrapolation to clinical use ''highly restricted''.

ECOLOGICAL FALLACY

Although the Ontario experience does not isolate one or several of the auxiliary assumptions contained in X_P that are guilty of the increase in complications, the basic population may contain other potential falsifiers not included in the RALES criteria, giving place to a possible ecological fallacy.[47 52] Conjectures set forth by different authors[19 43–46] that explain these contradictory results are that spironolactone was systematically used beyond the RALES inclusion and exclusion criteria, this means that the conditional argument X_P of Q results to be false. The ecological design of Ontario admits the submittal of a fundamental hypothesis to falsification so as to corroborate this hypothesis. Let H be the proposition ''…RALES increases the prescriptions for spironolactone'' and its logical consequence P ''…RALES increases the prescriptions for spironolactone in Ontario''. Thus H-then-P results corroborated because this phenomenon was observed in Ontario after the publication of RALES. Consequently an increase in hyperkalaemia complications was seen (no-Q) that contradicts the RALES prediction (Q), because both studies predict their contradictory observational predicates (Q and no-Q) starting from the logical consequence of the proposition P, whose subject is ''prescription of spironolactone''. A possible ecological fallacy can be set forth on a speculative scale, but it remains included in the falsity content of X_P, that is, everything that is not RALES. Juurlink et al conclude that the relation established is temporally compelling, biologically plausible, and consistent with existing evidence.[19] It is also logically consistent at a populational level of analysis and reciprocally we could speculate an individualistic fallacy[52] in the RALES design. We can set forth the following set of enunciations:

- H = ''… RALES increases the prescriptions of spironolactone''

- O = ''…RALES increases the prescriptions of spironolactone in Ontario''

- P = ''…spironolactone increases hyperkalaemia complications''

- Q = ''…spironolactone increases hyperkalaemia in Ontario''

The corroboration that took place in Ontario is: (i) H-then-O; (ii) O-then-P and (iii) P-then-Q, so that the conclusion predicted in the empirical scale is made by logical synthesis (iv) H-then-Q, a logically valid problem for Ontario's public health which remains unfalsified.

SOLVING THE H-THEN-Q PROBLEM

One of the contributions of EBM is, no doubt, the generation of clinical guides that permit an effective and safe use of the evidence of new available therapies.[30] In previous logical argumentation the challenge is presented to EBM of solving the problem H-then-Q that is the symbolic synthesis of the enunciation ''… RALES increases the prescriptions of spironolactone and hyperkalaemia complications in Ontario'', that is, trying to falsify it. As we discussed previously, RALES has in its design inclusion and exclusion criteria (potential falsifiers) symbolised by X_P, so that X_P can form the content of a new clinical guide for the use of spironolactone in patients affected by CHF. According to Watkins and Miller, modus tollens must, by definition, be amplifiable and applied to false theories.[50 51] Let X_P be the auxiliary enunciation ''utilisation of clinical guides for the use of spironolactone''. We can then define H as a false proposition and its prediction symbolised by another false proposition in the form of a bi-conditional argument Q-if-X_P, Q being false if X_P is true and vice versa if X_P is false. This means that when not using clinical guides (X_P is false) the enunciation H will be true because of its corresponding logical consequence Q-if-X_P and, on by contrast, when using clinical guides (X_P is true) the enunciation H will be really a false conjecture by the contradiction in Q-if-X_P. In other words, the hyperkalaemia complications secondary to the use of spironolactone will be minimal, because the RALES results will be rationally applied. Thus, inclusion and exclusion criteria in RALES act as reciprocal potential falsifiers in both studies.

DISCUSSION

Two criticisms, important for epidemiological science have been proposed to Popper's original falsificationist theory. Firstly, no hypothesis is outright falsified, but only a

What this paper adds

- This paper presents two studies with contradictory results that are in a synchronic relation, forming a complete logical analysis unit: one hypothesis corroborated experimentally in an RCT confronted with a populational reality of an observational study in which its predictions are questioned.

- A basic logical structure is presented for a refutationist analysis of RCTs using bi-conditional *modus tollens* arguments in the form of (i) *P-then-Q* and (ii) *Q-if-X_P*, X_P being a set of potential falsifiers of Q (predictions in the population) as part of the explicit falsity content of P (corroborated hypothesis).

- It is shown, following a deductive, canonical, logical argumentation, that the falsifiability of an RCT is determined by explicit auxiliary assumptions in the empirical content of X_P (criteria of inclusion, exclusion, and conditional criteria) restricting its external validity.

- This logical model constitutes the basis for the development of quantitative methods allowing a probabilistic approach to the falsifiability degree (probability of *Q-if-X_P*) and generalisability (probability of X_P) of experimentally corroborated hypotheses in specific populations.

- Both RALES and the Ontario study have allowed to hypothesise the popperian model presented in this article. In practice, not to follow a deductive reasoning, transferring the results of RALES to Ontario, is a logical error with negative public health consequences.

- Popperian epidemiology is an important critical tool that should be incorporated by the EBM authors in the analysis and discussion of the external validity of multinational RCT papers that will permit an effective and safe transfer of the evidence to clinical practice.

hypothesis along with an unspecified number of auxiliary assumptions.[49] Secondly, corroboration statements would have no predictive content. Although they motivate and justify our preference for some theory over another, generalisations involved in drawing predictive conclusions should be rationally supported,[53] so that the induction cannot be totally disproved in empirical sciences. Nevertheless, this pragmatic justification of induction[13 53] has been controversial in epidemiology, as it would also justify to be illogical, quick, imprecise, and uncritical in analysing data and writing papers.[54] In contrast, epidemiological scientists should be meticulous, precise, critical observers who are rigorous in their use of logic.[3 6 12 54] Epidemiological analysis is not only a problem of empirical observation and pragmatism but a process of logical and theoretical construction.[11 18] Thus, the refutationist approach is an important critical tool for epidemiology[28] and there are many examples in medical research that support the utility of Popper's philosophical principles.[1 55–59]

The falsifiability of scientific hypotheses justifies the need for experimentation, a demarcative criterion widely recognised by different medical scientists.[1–7 11–15 48 60–66] Recently, Hyams[67] illustrates how non-falsifiable hypotheses are insufficient to advance in medical knowledge, even when there is an abundance of inductively supported empirical data. If popperian epidemiology is incorporated to clinical research, hypotheses will be ranked as scientific when they can be tested and falsified. In our study, both RCT and

ecological designs are in good accord with the principle of falsifiability of their scientific enunciations and their logical rule, the *modus tollens*, although in practice the evidence appears as being treated in an exclusive inductive model. Researchers in multinational RCTs assume that the clinical effects of the therapies under study are homogeneous from one country to another, centring the discussion of their reports more on the internal validity than on the external validity of the results.[68] Thus, a critical rationalism sustained exclusively on the principle of positive hypotheses verification and of inductive ''support'' provided by large multi-site RCTs can lead physicians to a false sensation of safety and iatrogenic use of evidence with unfavourable consequences for public health as has been the case with RALES.

Canonical logical analyses applied to epidemiological studies are not frequent in the literature and they appear with popperian epidemiology.[1 2 6 12 55–57] In the present falsificationist science metatheory, a theory's content is the totality of its logical consequences; its true consequences constitute its truth content and its false consequences, if any, its falsity content.[50 51] This means that, when transferring the experimental results to the ''real world'', all the empirical content of the tested hypotheses is transferred. From this perspective, the ecological design of Ontario, based on a population, represents a part of the ''real world'', and RALES, a corroborated hypothesis that has been moved to this ''real world'', so that both studies are in synchronous relation, forming a complete logical analysis unit. In Ontario, RALES would be refuted for the non-fulfilment of its auxiliary assumptions represented by its inclusion and exclusion criteria, which act as reciprocal potential falsifiers in both studies, taking the logical form of a bi-conditional argument of the type (i) *P-then-Q* and (ii) *Q-if-X_P*, X_P being a set of potential falsifiers of *Q* (predictions in the population) as part of the explicit falsity-content of *P* (corroborated hypothesis). In this model ecological fallacy and confounders remain included in the falsity content of X_P and falsifiability of RCTs is logically determined by the empirical content of the argument *Q-if-X_P*. Thus, an estimation of the generalisability degree is possible knowing the probability of X_P in specific populations. In the same way, the probability of *Q-if-X_P* would be an estimator of the falsifiability degree, so that Bayesian models[59] starting from bi-conditional *modus tollens* arguments can be adapted to evaluate falsifiability of RCTs from observational designs, and identify other potential falsifiers, thus allowing to raise new testable hypotheses.

To summarise, no epidemiological study logically contributes more than what is contained in its design. In their turn, auxiliary assumptions on RCTs act as potential falsifiers of the empirical reality they propose, that is, RCTs contain, in their experimental designs, the limits of their applications. In practice, not to follow a deductive reasoning by moving RALES results to Ontario, is to say (i) *P-then-Q* and (ii) *Q-if-X_P*, is a logical error with unfavourable consequences for public health. It being impossible to logically justify induction, pragmatic justification is the only alternative, nevertheless, pragmatism can also justify mistakes and epidemiological imprecisions. Lastly, popperian epidemiology would be a beneficial critical tool for EBM facing fundamental issues in multinational RCT papers, especially, the largely unresolved problem of external validity.[68]

ACKNOWLEDGEMENTS

···················
Authors' affiliations
E Koch, A Otarola, A Kirschbaum, Division of Epidemiology, School of Public Health, Faculty of Medicine, University of Chile
E Koch, A Otarola, School of Health, Duoc-UC Institute, Pontifical Catholic University of Chile

REFERENCES

1 **Buck C**. Popper's philosophy for epidemiologists. *Int J Epidemiol* 1975;**4**:159–68.
2 **MaClure M**. Popperian refutation in epidemiology. *Am J Epidemiol* 1985;**121**:343–50.
3 **Weed D**. On the logic of causal inference. *Am J Epidemiol* 1986;**123**:965–79.
4 **Senn SJ**. Falsificationism and the clinical trials. *Stat Med* 1991;**10**:1679–92.
5 **Renton A**. Epidemiology and causation: a realist view. *J Epidemiol Community Health* 1994;**48**:79–85.
6 **Poole C**. Induction does not exist in epidemiology, either. In: Rothman KJ, ed. *Causal inference*. Chestnut Hill, MA: ERI, 1988.
7 **Banegas JR**, Rodriguez F, Del Rey Calero J. Popper and the problem of induction in epidemiology. *Rev Esp Salud Publica* 2000;**74**:327–39.
8 **Jacobsen M**. Against popperized epidemiology. *Int J Epidemiol* 1976;**5**:9–11.
9 **Karhausen L**. The poverty of popperian epidemiology. *Int J Epidemiol* 1995;**24**:869–74.
10 **Pearce NE**. Crawford-Brown DJ. Critical discussion in epidemiology: problems with the Popperian approach. *J Clin Epidemiol* 1989;**42**:177–84.
11 **Cole SR**. Atheoretical science. A response to the poverty of Popperian epidemiology. *Int J Epidemiol* 1996;**25**:899–900.
12 **Greenland S**. Induction versus Popper: substance versus semantics. *Int J Epidemiol* 1998;**27**:543–8.
13 **Susser M**. What is a cause and how do we know one? A grammar for pragmatic epidemiology. *Am J Epidemiol* 1991;**133**:635–48.
14 **Rothman KJ**. *Epidemiology. An introduction*. New York: Oxford University Press, 2002.
15 **Goodman SN**. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med* 1999;**130**:995–1004.
16 **Greenland S**. Probability logic and probabilistic induction. *Epidemiology* 1998;**9**:322–32.
17 **Poole C**. Positivized epidemiology and the model of sufficient and component causes. *Int J Epidemiol* 2001;**30**:707–9.
18 **Weed D**. Theory and practice in epidemiology. *Ann NY Acad Sci* 2001;**954**:52–62.
19 **Juurlink DN**, Mamdani M, Pharm D, et al. Rates of hyperkalemia after publication of the randomized aldactone evaluation study. *N Engl J Med* 2004;**351**:543–51.
20 **Pitt B**, Zannad F, Remme WJ, et al. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med* 1999;**341**:709–17.
21 **Shahar E**. A Popperian perspective of the term 'evidence-based medicine'. *J Eval Clin Pract* 1997;**3**:109–16.
22 **Suppe F**. *The structure of scientific theories*. Urbana, IL: University of Illinois Press, 1977.
23 **Popper KR**. *The logic of scientific discovery*. London: Hutchinson, 1997.
24 **Popper KR**. *Conjectures and refutations: the growth of scientific knowledge*. London: Routledge, 1992.
25 **Hume D**. *An inquiry concerning human understanding*. Oxford: Oxford University Press, 1974.
26 **Popper KR**. *The poverty of historicism*. London: Routledge, 1961.
27 **Popper KR**. Replies to my critics. In: Schilpp PA, ed. *The philosophy of Karl Popper*. Illinois: Open Court, 1974.
28 **Maclure M**. Karl Popper and his unending quest: an epidemiologic interpretation. *Epidemiology* 1995;**6**:331–4.
29 **Haynes RB**, Devereaux PJ, Guyatt GH. Clinical expertise in the area of evidence- based medicine and patient choice. *Evid Based Med* 2002;**7**:36–8.
30 **Sackett DL**, Richardson WS, Rosemberg W, et al. *Evidence-based medicine: how to practice and teach EBM*. Edinburgh: Churchill Livingstone, 2000.
31 **Guyatt GH**, Sackett DL, Cook DJ, et al. for the Evidence Based Medicine Working Group. How to use an article about therapy or prevention. *JAMA* 1993;**270**:2598–601.
32 **Sehon SR**, Stanley DE. A philosophical analysis of evidence-based medicine debate. *BMC Health Serv Res* 2003;**3**:14.
33 **Cook DJ**, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;**126**:376–80.
34 **Tonelli MR**. In defense of expert opinion. *Acad Med* 1999;**74**:1187–92.
35 **Charlton BG**. Fundamental deficiencies in the megatrial methodology. *Curr Control Trials Cardiovasc Med* 2001;**2**:2–7.
36 **Ghali WA**, Cornuz J. Early uptake of research finding after fast-track publication. *Lancet* 2000;**355**:579–80.
37 **Benson K**, Hartz AJ. A comparison of observational studies and randomized controlled trials. *N Engl J Med* 2000;**342**:1878–86.
38 **Cohen AM**, Stavri PZ, Hersh WR. A categorization and analysis of the criticisms of Evidence-Based Medicine. *Int J Med Inform* 2004;**73**:35–43.
39 **MaClure M**. Taxonomic axes of epidemiology study designs: a refutationist perspective. *J Clin Epidemiol* 1991;**44**:1045–53.
40 **Naylor CD**. Grey zones of clinical practice: some limits to evidence based medicine. *Lancet* 1995;**345**:840–2.
41 **Kulbertus H**. Study of the month. The RALES study (randomized aldactone evaluation study). *Rev Med Liege* 1999;**54**:770–2.
42 **Majumdar SR**, McAlister FA, Soumerai S. Synergy between publication and promotion: comparing adoption of new evidence in Canada and the United States. *Am J Med* 2003;**115**:467–72.
43 **Bozkurt B**, Agoston I, Knowlton A. Complications of inappropriate use of spironolactone in heart failure: when an old medicine spirals out of new guidelines. *J Am Coll Cardiol* 2003;**41**:211–14.
44 **Masoudi FA**, Havranek EP, Wolfe P, et al. Most hospitalized older persons do not meet the enrolment criteria for clinical trials in heart failure. *Am Heart J* 2003;**146**:250–7.
45 **Tang WH**, Francis GS. Spironolactone in chronic heart failure: all's well that ends well. *J Am Coll Cardiol* 2003;**41**:215–16.
46 **MacMurray J**, O'Meara E. Treatment of heart failure with spironolactone. Trial and tribulations. *N Engl J Med* 2004;**351**:526–8.
47 **Goldfarb DS**. Hyperkalemia after publication of RALES. *N Engl J Med* 2004;**351**:2448–50.
48 **Antiseri D**. Philosophers, physicists and physicians in defense of the unity of the scientific method. *Clin Ter* 1998;**149**:429–33.
49 **Lakatos I**. *The methodology of scientific research programmes*. London: Cambridge University Press, 1978.
50 **Miller D**. *Critical rationalism: a restatement and defense*. Chicago: Open Court, 1994.
51 **Watkins J**. Popperian ideas on progress and rationality in science. *The Critical Rationalist*. 1997;**2**:1–11. (http://www.eeng.dcu.ie/~tkpw/tcr/volume-02/number-02/v02n02.pdf).
52 **Pearce N**. The ecological fallacy strikes backs. *J Epidemiol Community Health* 2000;**54**:326–7.
53 **Salmon W**. The pragmatic justification of induction. In: Swinburne R, ed. *The justification of induction*. Oxford: Oxford University Press, 1974.
54 **Maclure M**. On the logic and practice of epidemiology. *Am J Epidemiol* 1987;**126**:554.
55 **Weed D**. An epidemiological application of Popper's method. *J Epidemiol Community Health* 1985;**39**:277–85.
56 **Maclure M**. Multivariate refutation of aetiological hypotheses in non-experimental epidemiology. *Int J Epidemiol* 1990;**19**:782–7.
57 **Maclure M**. Demonstration of deductive meta-analysis: ethanol intake and risk of myocardial infarction. *Epidemiol Rev* 1993;**15**:328–51.
58 **Maclure M**. Inventing the AIDS virus hypothesis: an illustration of scientific vs unscientific induction. *Epidemiology* 1998;**9**:467–473.
59 **Malavasi A**, Realdi G. Logical structure of the problem oriented clinical record. A combined Popper-Bayes approach. *Ann Ital Med Int* 2002;**17**:21–30.
60 **Boissel JP**. Note on the epistemology of clinical pharmacology: comparison with the approach of Karl Popper. *Therapie* 1999;**54**:67–73.
61 **Holmberg L**, Baum M, Adami HO. On the scientific inference from clinical trials. *J Eval Clin Pract* 1999;**5**:157–62.
62 **de Dombal FT**. Epidemiology of Crohn's disease of the colon. *Ann Gastroenterol Hepatol (Paris)* 1985;**21**:191–200.
63 **Ayres JR**. The problem of true knowledge in epidemiology. *Rev Saude Publica* 1992;**26**:206–14.
64 **Havlicek J**. Popper's critical rationalism and the biomedical sciences. *Cas Lek Cesk* 1993;**132**:641–4.
65 **Hogel J**, Gaus W. The procedure of new drug application and the philosophy of critical rationalism or the limits of quality assurance with good clinical practice. *Control Clin Trials* 1999;**20**:511–18.
66 **Allmark P**. Popper and nursing theory. *Nurs Philos* 2003;**4**:4–16.
67 **Hyams KC**. The investigation of chronic fatigue syndrome: a case-study of the limitations of inductive inferences and non-falsifiable hypotheses in medical research. *Med Hypotheses* 2003;**60**:760–6.
68 **Rothwell PM**. External validity of randomised controlled trials: ''To whom do the results of this trial apply?''. *Lancet* 2005;**365**:82–93.