

Erik Weber
Leen De Vreese

THE CAUSES AND CURES OF SCURVY

How Modern was James Lind’s Methodology?*

1. Introduction

The Scottish physician James Lind is the most celebrated name in the history of research into the causes and cures of scurvy. This is due to the famous experiment he conducted in 1747 on H.M.S. *Salisbury* in order to compare the efficiency of six popular treatments for scurvy. This experiment is generally regarded as the first controlled trial in clinical science (see e.g. Carpenter 1986, p. 52).

The experiment relates to the possible cures of scurvy, not to its causes. The traditional view is that Lind was much less innovative in his research on the causes of scurvy. In 1867 the French naval surgeon H. Rey, in his analysis and critique of Lind’s *Treatise of the Scurvy*, complained that, while Lind relied only on experiments in matters of therapeutics, his theory on the causes of scurvy was based on “ideas that are far removed from any kind of scientific precision”. According to Rey, Lind accepted some ideas on the

*The research for this paper benefited from a bilateral scientific exchange project funded by the Ministry of the Flemish Community (project BIL01/80) and the Polish State Committee for Scientific Research. We thank Steffen Ducheyne and Robrecht Vanderbeeken for their comments on earlier drafts of this paper.

causes on scurvy too readily, because in that domain he did not rely on experimental results. (see Carpenter 1986, p. 57).

The aim of this paper is to qualify both the idea of Lind as a modern therapist and of Lind as old fashioned in his investigation of the causes of diseases. In Section 2 we will explain how contemporary biomedical scientists investigate causes and cures of diseases. In Section 3 we argue that there are important differences between Lind's experiment and contemporary experimental practice. In Section 4 we argue that Lind's methodology for investigating the causes of scurvy was more modern than Rey suggests.

2. Contemporary Experimental Practice

2.1. The Concept of Causation

A detailed description of current experimental practice in the biomedical and social sciences can be found in chapter 7 and 8 of Giere (1997). Chapter 7 deals with the *concept* of causation that is used, chapter 8 with experimental methods. Most of what we say in 2.1–2.3 is based on these chapters of Giere's book.

Let us start with some examples. Suppose we do the following experiment (cf. Giere 1997, p. 210 ff.). A random sample of laboratory rats is randomly divided into an experimental group and a control group. The experimental group is fed a diet that contains 5% saccharin. The control group receives the same diet, minus the saccharin. When a rat dies, we check whether it has developed bladder cancer. After two years, the remaining rats are killed and examined. We observe that the fraction of animals with bladder cancer is higher in the experimental group. The difference is statistically significant, so we conclude that saccharin consumption is positively causally relevant for bladder cancer in populations of laboratory rats. What does this claim mean? It means that, in a population where all rats would consume saccharin (this hypothetical population is represented in the experiment by the experimental group) there would be more bladder cancers than in a population without saccharin consumption.

If we claim that smoking is positively causally relevant for lung cancer in the Belgian population, we make a similar counterfactual claim. In Belgium, some people smoke, others don't. Positive causal relevance means that, if every Belgian would smoke (while their other characteristics remained the same) there would be more lung cancers than in the hypothetical population where no Belgian smokes (while their other characteristics are preserved).

Giere proposes the following definitions for causal relevance in populations (**C** is a variable with two values (C and Not-C); the same for **E** (values E and Not-E)):

C is a *positive causal factor* for **E** in the population **U** whenever $\mathbf{P}_X(\mathbf{E})$ is greater than $\mathbf{P}_K(\mathbf{E})$.

C is a *negative causal factor* for **E** in the population **U** whenever $\mathbf{P}_X(\mathbf{E})$ is less than $\mathbf{P}_K(\mathbf{E})$.

C is *causally irrelevant* for **E** in the population **U** whenever $\mathbf{P}_X(\mathbf{E})$ is equal to $\mathbf{P}_K(\mathbf{E})$. Giere 1997, p. 204

X is the hypothetical population which is identical to **U**, except that each individual exhibits the value **C** of the causal variable **C**. **K** is the analogous hypothetical population in which all individuals exhibits $\neg\mathbf{C}$.

Claims about causal relations in populations are typically probabilistic. For instance, not all rats that consume saccharin develop bladder cancer. However, these probabilistic claims about populations are compatible with determinism at the individual level. For instance, we can assume that there is at least one set of properties **S** such that all rats which have **S** and consume saccharin develop bladder cancer while rats with **S** that don't consume saccharin do not get bladder cancer. Giere defines causal factors in individuals as follows:

C is a *positive causal factor* (deterministic) for **E** in an individual, **I**, characterized by residual state, **S**, if in **I**, **C** produces **E** and Not-**C** produces Not-**E**.

C is a *negative causal factor* (deterministic) for **E** in an individual, **I**, characterized by residual state, **S**, if in **I**, **C** produces Not-**E** and Not-**C** produces **E**. Giere 1997, p. 200

If **C** is neither a positive nor a negative causal factor for **E** in **I**, given **S**, then we say that the variable **C** is causally irrelevant for **E** in **I**, given **S**. Giere 1997, p. 201

The residual state **S** refers to all other characteristics of the individual besides the cause and effect variable.

In order to clarify the relation between the two types of claims, it is important to see that Giere's population claims "always average over individuals and, therefore ignore what might be important differences among individuals" (pp. 204–205). Giere discusses only one specific case: if in **U** there are some individuals for which **C** is a positive causal factor for **E**, and

an equal number for which C is a negative causal factor, C is causally irrelevant for E in U . More generally, Giere's definitions leave open the possibility that a population contains individuals for which C is a positive causal factor for E , as well as individuals for which C is a negative causal factor for E . If the first subpopulation is larger, C will be a positive causal factor for E . If the second subpopulation is larger, C will be a negative causal factor.

We will now discuss the experimental methods that scientists use in order to establish claims about causal relations in populations. We first look at experimental design (2.2) and then at the interpretation of experimental results (2.3).

2.2. Experimental design

How do we test a causal hypothesis? The most direct way is by randomly selecting a sample from the real population and then randomly dividing it into two groups (cf. the rat example in 2.1). The individuals in the first group are manipulated by the experimenter so that they have the putative cause (in the example: saccharin consumption). This group, the experimental group is just as if it had been sampled from the hypothetical population X . The other group, the control group, is manipulated so that it does not have the putative cause (it is just as if it had been sampled from the other hypothetical population K).

Random sampling is one of the conditions that any contemporary experiment must meet in order to be taken seriously by the scientific community. There are two other important requirements: the experiment should exclude bias due to the placebo effect, and exclude observer bias. Suppose that we want to test the efficacy of a new drug. The fact that people know that they are in the experimental group, combined with their belief that the drug will work, is sometimes sufficient for ameliorating the patient's condition. The classical solution for this problem consists in administering a placebo drug (a product of which we know that it is inert) to the control group. This is called a blind experiment, because the subjects cannot know any more whether they are in the experimental or control group. Placebo effects are not excluded in this way, but we are sure that they occur in both groups with comparable frequency. In order to exclude observer bias (distortion of the results due to the beliefs of the experimenter) the experiment has to be made double-blind. This means that the person who observes the results must not know which subjects are in the experimental group, and which in the control group. In case of the test of a drug, this can be accomplished by

having two experimenters: one that administers the test-drug or the placebo to the subjects, and one that follows up the condition of the patients.

In many cases, random division into experimental and control group cannot be realised for ethical reasons. In the words of John Dupré:

A decisive test of whether smoking causes heart disease, then, would be to take a large sample of human infants randomly selected from the human population, divide them into two equal groups, and force one group to smoke for the rest of their—no doubt abbreviated—lives.

Dupré 1993, pp. 202–203

In many cases, *prospective designs* offer a way out. Like in the examples above (which are instances of *randomized experiments*), we start from a random sample of individuals. However, the individuals put themselves into the experimental or control group by the way they act. For instance, people that decided to smoke end up in the experimental group, non-smokers in the control group. Because of this non-random selection, there may be disturbing factors. For instance, if there are more heart diseases among the smokers, this may be due to the fact that both smoking and heart disease are positively influenced by coffee drinking. Randomized experiments avoid this problem by the random division into experimental and control group.

Prospective designs (and the other alternative, the backward-looking *retrospective* design) presuppose that the purported cause-variable is spontaneously divided among the population. So prospective and retrospective designs do not always offer a way out. Another very popular way to deal with ethical worries is the use of laboratory animals as surrogates for humans.

2.3. Interpretation of results

Let us now look at the interpretation of results. The first thing we do, after we have executed the appropriate experimental manipulations, is to check whether there is a difference with respect to the presence of the putative effect factor between the two groups. If there is a difference (e.g. if the rats that consume saccharin develop more bladder cancers than the rats that don't consume saccharin), we have to show that it is statistically significant, i.e. so large that the chance that they are due to mere chance fluctuations is rather small. A widespread scientific tradition is to take results of experiments seriously if they are significant at the 0.05 level (less than 5% probability that the difference is due to chance).

3. How modern was Lind as a therapist?

3.1. Description of Lind's experiment

From 1746 till he left the Navy in 1748, James Lind was full surgeon on H.M.S. *Salisbury*. In the summer of 1746, during a cruise in the English Channel, there was a first outbreak of scurvy, involving 80 men (of a crew of 350). During a second outbreak in 1747, Lind conducted his experiment. His aim was

to relate the effects of several medicines tried at sea in this disease, on purpose to discover what might promise the most certain protection against it upon that element. Stewart & Guthrie 1953, p. 144

The design of the experiment has been described by Lind as follows:

On the 20th of May 1747, I took twelve patients in the scurvy, on board the *Salisbury* at sea. Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of their knees. They lay together in one place, being a proper apartment for the sick in the fore-hold; and had one diet common to all, [...]. Two of these were ordered each a quart of cider a-day. Two others took twenty-five gutts of elixir vitriol three times a-day, upon an empty stomach; using a gargle strongly acidulated with it for their mouths. Two others took two spoonfuls of vinegar three times a-day, upon an empty stomach; having their gruels and their other food well acidulated with it, as also the gargle for their mouth. Two of the worst patients, with the tendons in the ham rigid, (a symptom none of the rest had), were put under a course of sea-water. Of this they drank half a pint every day, and sometimes more or less as it operated, by way of gentle physic. Two others had each two oranges and one lemon given them every day. These they eat with greediness, at different times, upon an empty stomach. They continued but six days under this course, having consumed the quantity that could be spared. The two remaining patients, took the bigness of a nutmeg three times a-day, of an electuary recommended by an hospital-surgeon [...]; using for common drink [a tamarind decoction].

Stewart & Guthrie 1953, pp. 145–146

The electuary (medicinal paste) contained (among other things) garlic, mustard seed and gum myrrh. What were the observed effects after a period of two weeks?

the most sudden and visible good effects were perceived from the use of the oranges and lemons; one of those who had taken them, being at the

end of six days fit for duty. [...] The other was the best recovered of any in his condition [...]. Next to the oranges, I thought the cider had the best effects. It was indeed not very sound,... However, those who had taken it, were in a fairer way of recovery than the others at the end of the fortnight [...]. As to the elixir of vitriol, I observed that the mouths of those who had used it by way of gargarism, were in a much cleaner and better condition than many of the rest, especially those who used the vinegar; but perceived otherwise no good effects from its internal use upon the other symptoms [...]. There was no remarkable alteration upon those who took the electuary and tamarind decoction, the sea-water, or vinegar, upon comparing their condition, at the end of the fortnight, with others who had taken nothing but a little lenitive electuary and *cremor tartar*, at times, in order to keep their belly open; or a gentle pectoral in the evening, for relief of their breast.

Stewart & Guthrie 1953, pp. 146–148

3.2. Analysis

It might be useful to point out that, though he does not mention them in the quotation above, Lind had a control group: all the other patients on board of his ship. These patients did not get anything that might cure their disease: all they got was a pain-killing paste (“lenitive electuary”), a laxative (*cremor tartar*) and/or a cough syrup (“pectoral”). It is clear that these products can have an effect on the symptoms (pain, constipation) but will not cure the disease. The right way to describe Lind’s design is that he had six small experimental groups (each containing 2 patients) and a large common control group.

Because he takes the non-treated patients as controls, Lind’s experiment involves all the patients on board of the ship. This does not mean that his experiment satisfies the criteria of random sampling. First, he conducted his experiments on board of one ship, in a rather limited area: the home waters of the British Empire. Lind left the Navy in 1748 to write his *Treatise*. From a contemporary perspective that is a premature step: he should have stayed in the Navy to conduct similar experiments in the Mediterranean, the West Indies, and other areas where the Navy was active (the war with Spain ended in 1748; Lind might have done this if the war lasted longer).

A second problem with Lind’s experiment is that it is unclear whether the 12 patients in the experimental groups were selected randomly. There are indications (e.g. the fact that he took patients “as similar as he could have them”; and the fact that the two patients that were in the worst condition were both treated with seawater) that this was not the case.

A third problem is that no measures have been taken to exclude observer bias. It is obvious that Lind knew who got what when we registered the effects. It is also obvious (this is the fourth problem) that there were no placebo treatments.

Because of these four problems, we think it is a bit optimistic and/or liberal to call Linds' experiment the first controlled trial in clinical science. Moreover, the way in which Lind argues from his data to his conclusion is completely different from what contemporary scientists do. Contemporary biomedical scientists would try to show that the differences are statistically significant (and therefore would have bigger experimental groups). In Linds' era, inferential statistics did not yet exist. So he had to use a different argumentation strategy. This is probably one of the most interesting aspects of his experiment. In his book on scurvy, Kenneth Carpenter writes:

It is interesting that the two subjects in worst condition both received the seawater treatment. Was this chance, or had Lind perhaps been a believer in it and expected that they would give a dramatic response from "worst" to "best"?

Carpenter 1986, p. 53

It is possible that Lind felt that he had to exclude chance in some way. Carpenter suggests that, lacking the statistical methods we have now, Lind tried a rhetorical trick: he wanted to show that seawater worked by a spectacular improvement of the two patients. His trick did not work: the patients that were given lemons and oranges were in the best condition after two weeks. This evolution (from "bad but not worst" to "best") is not as spectacular as the one Lind had hoped for.

We think that Lind's experiment is much overrated: there are serious problems with the design and statistical reasoning is completely absent. Later in his career Lind performed better experiments. In 1758 he became physician at Haslar, a Navy hospital near Portsmouth. In one of his experiments there he put 130 scorbutic patients on a course of wort during two weeks (Carpenter 1986, p. 65). This experiment (and others he performed at Haslar) is better than the original one in two respects: the experimental group is much larger, and the patients come from different ships. These experiments are less famous because they did not have any positive results.

4. Lind and the causes of scurvy

4.1. The “blocked perspiration” theory

At the beginning of the second part of his *Treatise of the Scurvy*, Lind claims that the primary cause of scurvy is a humid atmosphere. The mechanism by which a humid atmosphere is supposed to have its unwanted effect is based on the idea of “blocked perspiration”. Corruble & Ganascia have summarised the blocked perspiration theory as follows:

The body is made mainly of solid tissues and fluids. The fluids naturally tend to become corrupted. An important function of all the excretions, and especially of perspiration, is to evacuate these corrupted fluids from the body to keep only some healthy fluids inside. If the perspiration is blocked, the corrupted fluids act as a poison and produce some diseases.

Corruble & Ganascia 1997, p. 217

This theory was very popular around 1750 and was used to explain a.o. cholera and various fevers. Lind claims that moisture blocks perspiration and that blocked perspiration leads to scurvy. For the first causal link, he suggests two mechanisms (see Carpenter 1986, pp. 57–59). The first is that moisture constricts the pores in the skin. The second is that moist air is simply unfit for salutary purposes.

Lind believed that one can fight against the poisonous effect of the corrupted fluids by eating lemons, oranges and green vegetables. The acids (present in the fresh fruit/vegetable or formed during fermentation) restores the balance in the chyle, by acting as a detergent. Note that this theory explains the results of Lind’s experiment: the lemons and oranges block the causal mechanism that leads from moist air to scurvy.

4.2. The rivals

A lot of alternative views on the causes of scurvy were circulating during the 18th century. First of all, there is John Bachstrom who claims that it is a pure nutritional disease, resulting from limited food variety:

Cependant c’est à Bachström (1734: *Observationes circa scorbutum ejusdemque indolem*, etc.) qu’il faut rapporter l’insigne honneur d’avoir appelé et fixé l’attention des médecins sur la cause principale du scorbut cherchée dans l’alimentation [...] cet esprit sagace prouva que l’abstinence totale des aliments végétaux frais est la seule, la véritable et la première cause du scorbut[.]

Mahé 1880, p. 100

Lind made it clear that he did not agree with this. He explicitly denies that

the constitution of the human body, [is such] that health and life cannot be preserved long, without the use of green herbage, vegetables, and fruits; and that a long abstinence from these, is alone the cause of the scurvy. Stewart & Guthrie 1953, p. 73

We will discuss the arguments of Bachstrom and Lind in 4.3 and 4.4. Less popular ideas on the cause of scurvy were restricted food quantity and bad hygiene:

On sera donc porté à conclure que la diminution de quantité des aliments, l'insuffisance, ou pour parler net, la disette et l'inanition peuvent bien servir occasionnellement la cause du scorbut, mais qu'elles sont impuissantes à le produire par elles-mêmes. Mahé 1880 p. 101

Si les équipages de Cook furent préservés du scorbut, dans une mesure relativement grande eu égard au temps et à l'époque, on a pensé que ces magnifiques résultats étaient précisément l'heureuse conséquence de soins de propreté et d'assèchement des navires, soins dirigés habilement et poursuivis sans relâche. Mahé 1880, p. 89

We will not discuss these last two possibilities further.

4.3. Bachstrom's arguments

Bachstrom's main argument is the blockade of Thorn (Torun, Poland) which took place in the summer (from May till the beginning of October) of 1703. While the besieging Swedes remained healthy, many thousands inhabitants and soldiers inside the city developed scurvy. Carpenter remarks that Bachstrom's argument seems straightforward for the modern reader (1986, p. 44). We think that there is a good reason for that: though he does not use the term, Bachstrom treats the siege as a prospective design in which citizens and soldiers of Thorn constitute the experimental group, and the Swedes the control group. Of course the design has serious flaws: the sample is not representative and there can be observer bias (cf. the problems with Lind's design discussed in Section 3.2). And of course there is no statistical analysis of the results. But the basic structure of the argument by which he argues that lack of food variety is a cause of scurvy, is modern: the basis is an observed difference between an experimental and a control group. Bachstrom did not use the terms "experimental group" and "control group", but he treats the inhabitants of Thorn and the Swedes in this way.

In order to assess the power of Bachstrom's argument it is important to notice that, while the siege can be the basis of an argument in favour of the claim that food variety is causally relevant for the development of scurvy, there is no way in which one can argue from this case that moisture is causally irrelevant: the only conclusion is that moisture is not a necessary cause. Lind regarded moisture as a necessary cause. This idea is refuted by Bachstrom. However, this does not mean that moisture is causally irrelevant. More general, Bachstrom does not have an argument for the view that food variety is the *only* causally relevant factor.

4.4. Lind's arguments

A first argument offered by Lind against Bachstrom's view that food variety is the only cause is this:

if this were truly the case, we must have had the scurvy very accurately described by the ancients, whose chief study seems to have been the art of war; and whose manner of besieging towns was generally by a blockade, till they had forced a surrender by famine. Now, as they held out many months, sometimes years, without a supply of vegetables; we should, no doubt, have heard of many dying of the scurvy, long before the magazines of dry provisions were exhausted. The continuance of those sieges far exceeded most of our modern ones; even the five months blockade of Thorn, upon which Bachström has founded this supposition.

Stewart & Guthrie 1953, p. 73–74

This argument is not very convincing for us, because it relies on the historical accuracy of a literary tradition. However, Lind offers a more convincing argument, drawing on his own experience:

I have been three months on a cruise, during which time none of the seamen tasted vegetables or greens of any sort; and although for a great part of that time, from want of fresh water, their beef and pork were boiled in the sea-water, yet we returned into port without one scorbutical complaint [...]. But it was remarkable that in the two cruises afterwards to be mentioned, in his Majesty's ship the Salisbury, where I had an opportunity of making observations on this disease, that it began to rage upon being less than six weeks at sea.

Stewart & Guthrie 1953, pp. 73–74

Now, it was observable, that both these cruises were in the months of April, May, and June; when we had, especially at the beginning of them, a continuance of cold, rainy and thick Channel weather, as it is called; whereas in our other cruises, we had generally very fine

weather; [...]. Nor could I assign any other reason for the frequency of this disease in these two cruises, and our exemption from it at other times, but the influence of the weather; the circumstances of the men, ship, and provisions, being in all other respects alike.

Stewart 1953, p. 84

In this passage Lind treats his own experience as a prospective study: the crew of the first cruise are the control group, the crew of the two other cruises the experimental group. Of course there are serious shortcomings: lack of representativeness (crews from only three ships), no attempt to exclude observer bias, no argument for the claim that the circumstances of men and ships were the same (that the provisions were the same is plausible). However, we have shown that Lind's celebrated experiment on the cures of scurvy faces similar problems. So we conclude that Lind was as modern in his investigation of the causes of scurvy as he was in his investigation of the cures.

5. Conclusion

In this paper we have assumed that Lind implicitly shared the concept of causation as explained in Section 2.1 and investigated to what extent his research practice is modern. It goes without saying that Lind did not conceptualise his own practice in the modern way: he never heard of our modern concepts like "randomised trial", "prospective design", "control group", etc. So we focussed on his practice, on the methodology he actually used. Our conclusion is that what he does when investigating cures is not more modern than what he does when investigating causes. The crucial modern feature in his practice is the comparative nature of his methodology: in both contexts, he compares (what we would call) an experimental group with (what we would call) a control group. Lind did not use these terms, but quite obviously was convinced that causal relations have to be investigated by comparative methods. Therefore, after all, he deserves being regarded as the founding father of modern biomedical science.

References

- Carpenter, K. J. (1986), *The History of Scurvy and Vitamin C*. Cambridge: Cambridge University Press.
- Corruble, V., and J.-G. Ganascia (1997), "Induction and the Discovery of the Causes of Scurvy", *Artificial Intelligence* **91**, 205–223.

- Dupré, J. (1993), *The Disorder of Things*. Cambridge and London: Harvard University Press.
- Giere, R. (1997), *Understanding Scientific Reasoning* (4th edition). Fort Worth: Harcourt College Publishers.
- Mahé J. (1880), “Le Scorbut”, in *Dictionnaire Encyclopédique des Sciences Médicales. Série 3. Tome 8*. Paris: Masson, pp. 35–257.
- Stewart, C. P., and D. Guthrie (eds.) (1953), *Lind’s Treatise on Scurvy. A Bicentenary Volume Containing a Reprint of the First Edition of “A Treatise of the Scurvy” by James Lind*. Edinburgh: Edinburgh University Press.

ERIK WEBER, LEEN DE VREESE
Centre for Logic and Philosophy of Science
Ghent University
Blandijnberg 2
B-9000 Gent, Belgium
{Erik.Weber,Leen.DeVreese}@UGent.be