Keekok Lee
www.keekoklee.org
3 September 2021

A Diagnosis and a Prognosis of RCT as Gold Standard and Pillar of EBM

Introduction

This paper attempts to show:
1. In spite of its obvious perceived methodological strengths, a more detailed scrutiny of the historical and theoretical/philosophical background and presuppositions of RCT may reveal at least one fundamental flaw.
2. The flaw is identified as the **Axiom of Homogeneity**.[1]
3. Its accompanying causal model, though clear and simple to understand and to apply, may also be implicated.
4. Recent developments in Systems Biology show that Biomedicine may increasingly rest on the **Axiom of Heterogeneity** with a causal model which is more complex, and an orientation in diagnosis and treatment which recognizes individual differences between people in general and patients in particular.
5. Such arguments imply that RCT's future as Gold Standard and as a pillar of EBM may not be as secure as its present commanding position in the Pyramid of Evidence leads one to believe.

RCT

History

Although Archie Cochrane (a Scottish doctor, 1909-1988) cannot be credited with pioneering the concept of RCT itself, he could be credited with having successfully spread its mission abroad. In 1972, he published *Effectiveness and Efficiency: Random Reflections on Health Services*, in which he championed using RCTs in order to make medicine more effective. A step in that direction is:

> … to measure the effect of a particular medical action in altering the natural history of a particular disease for the better. Since the introduction of the randomized controlled trial (RCT) our knowledge in this sphere has greatly increased but is still sadly limited. It is in this sense that I sue the word 'effective' in this book, and I use it in relation to research results, as opposed to the results obtained when a therapy is applied in routine clinical practice in a defined community. (1972, 2)

Cochrane praised Austin Bradford Hill for having introduced the techniques of the RCT in his study of whether streptomycin was effective in the treatment of pulmonary tuberculosis (see Daniels and Hill 1952).[2] He said:

> … from the point of view of the NHS it enabled Bradford Hill … to introduce to the medical world the techniques of the RCT which added the experimental approach to medical research. Its importance cannot be exaggerated. It opened up a new world of evaluation and control which will, I think, be the key to a rational health service. (1972,11)

To make a pedantic point: the honour thus accorded is not warranted as Cochrane appeared to have overlooked that Marshall 1948 had already published an RCT study on streptomycin in the treatment of tuberculosis, also in the *British Medical Journal*, four years earlier. Furthermore, according to Jefferson 2007, the Medical Research Council (MRC) of the UK had pioneered Randomisation in the 1940s in the trial of whooping cough vaccines.

Cochrane continued:

> The basic idea, like most good things, it is very simple. The RCT approaches the problem of the comparability of the two groups the other way round. *The idea is not to worry about the characteristics of the patients*, but to be sure that the division of the patients into two groups is done by some method independent of human choice, i.e., by allocating according to some simple numerical device such as the order in which the patients come under treatment, or, more safely, by the use of random numbers. In this way *the characteristics of the patients are randomized between the two groups*, and it is possible to test the hypothesis that one treatment is better than

[1] For further discussion of the **Axioms of Homogeneity and Heterogeneity** as well as the distinction raised later between Statistical Relevance and Clinical Relevance, see Lee 2017.

[2] This is ironical. Bradford Hill is not generally celebrated for advocating the notion of RCT, but for his contribution with Richard Doll (1912-2005) in Epidemiology, elevating it to a proper scientific footing. Their research convinced enough of the scientific community and governments worldwide of the causal link between smoking and lung cancer. He had also formulated 9 criteria (what he calls "viewpoints" to determine how confident one can be that such a kind of link epidemiologically established could be a causal one–see Applying the Bradford Hill Criteria in the 21st century 2015. The Bradford Hill and Doll's study is a case-control study which in the Pyramid of Evidence is classed Level ll-2.

> another and express the results in the form of the probability of the differences found being due to chance or not. (1972, 22)

Note simply for the moment that this author has italicized certain phrases (not in the original) to alert the reader that they have raised issues in which the paper is particularly interested.

The first issue to notice is that the RCT as presented through Cochrane 1972 focuses on Randomisation as a technique. Randomisation presupposes at least two groups as it means allocating patient-participants to the experimental arm or the control arm of the trial with the aim of excluding bias. This is done by (a) using double-blinding (patient-participants and staff-participants involved with the trial in any way would not know which patient-participant belonged to the experimental and which to the control arm); (b) the allocation to the two arms should be done by a numerical/mechanical device of some sort[3]–today, one assumes it to be done by a computer software, whose algorithm would be written by computer experts who are not privy to the precise use of the software.

This model of RCT (call it Model C) is a very simple technique as Cochrane himself put it. Its simplicity lies in the way in which Randomisation alone is expected to do all the heavy lifting to ensure that the result of the trial, whatever it might be, is not due to chance and/or contamination by the placebo effect.

Its simplicity apart, Model C also chimed in well with the spirit of the age of "the magic bullet" in Clinical Medicine. This age took a long time in coming. Theoretically, it was ushered in with the germ theory of disease causation established by Pasteur (1822-1895) and Koch (1843-1910) and, in particular of Koch's discovery in 1882 of the tubercle bacilli as the causal agent of tuberculosis for which he was awarded the Nobel Prize in 1905. However, this radical theoretical discovery which put medicine at last on a scientific footing took a long time in delivering any effective treatments for the victims of germ infection, whether it be cholera, diphtheria or tetanus. Venesection and leeching (blood-letting) continued to be used in spite of the fact that they were shown to be not effective for treating pneumonia in 1835.[4] The concept of the magic bullet was enunciated by Paul Ehrlich (1854-1915) in 1900 but it was not till 1909 that he (and Sahachiro Hata 1873-1938) produced the first magic bullet by demonstrating that Salvarsan, an arsenical compound, could kill the spirochete of syphilis without such drastic side effects as killing the patient. The age of the magic bullet was only finally established with the mass manufacture of penicillin (which Alexander Fleming (1881-1955) in 1928 had discovered by chance) through the efforts of Howard Florey and Ernst Chain (1989-1968 and 1906-1979) after WWII when the age of antibiotics made a dramatic appearance in Clinical Medicine. Streptomycin appeared in 1948 as a cure for tuberculosis.

In 1906, Theodore Roosevelt, President of the USA passed the first Food and Drug Act in 1906. This provided an important framework eventually a half century later for the pharmaceutical manufacturers to become what today some people call "Big Pharma" with its deep pockets and immense lobbying power after the end of WWII with the emergence of the golden age of antibiotics. In Phase III of a drug trial as conducted by a pharmaceutical company, the new drug is given to some patients with the disease for which the treatment is targeting, using what this paper called RCT-Model C (to be made clearer a little later)–should the drug perform better (in an acceptable statistical sense of that term), then the FDA (Food and Drug Administration) would license the company to sell it. Phase III preceded by Phase I and II are said between them to warrant sufficient degree of safety and effectiveness.

Philosophical roots of Model C

The British philosopher, John Stuart Mill (1806-1873) published his *System of Logic* in 1843.[5] In Book III, Chapters 8-10, he set out five methods for determining the cause of a phenomenon–agreement, difference, agreement and difference, concomitant variations, residues.[6] Model C as used in drug trials conducted today appear to rely, in the main, on the method of difference. Mill put the method of difference thus:

> If an instance in which the phenomenon under investigation occurs and an instance in which it does not occur, have every circumstance in common save one, that one occurring in the former; the circumstance in which the two instances differ is the effect, or the cause, or are indispensable part of the cause of the phenomenon.

In what sort of context would Mill's account have clear purchase and hold true without reservation? One can think immediately of two: (a) in engineering and (b) in a chemistry lab. In the former, engineering products are

[3] Historically, double-blinding appeared before Randomisation–according to Garner and Thomas 2010 but also see Jefferson 2007 for pointing out the importance of Randomisation from the methodological standpoint.

For the purpose of his paper, it may be appropriate to be pedantic and distinguish between double-blinding from Randomisation, even if in practice in RCTs they go together.

[4] This was shown by Pierre Charles Alexandre Louis in Paris–see Silverman 1985.

[5] An electronic version may be found at http://www.gutenberg.org/etext/27942 .

[6] For an overall account of Mill's thoughts, see Macleod 2020; for an account of Mill's canons for medicine and biology, see Schaffner 1993, Chapter 4.6, 142-46.

designed to be homogeneous and as near identical as they can be. When engineers test their products by setting up an experiment, their products can be divided into two identical groups; they then perform the intervention on only the experimental group (subjecting it, say, to a stress test), thereby making sure that this group differs from the control group only in one "circumstance". The Millean idea (call it the Millean Requirement), therefore, sits well in the machine context, where every "circumstance" of manufacture and production is within the control of the production team. As engineered products are paradigmatically artefacts and as artefacts, *ex hypothesi*, are things intentionally designed and executed by us humans, it follows that Aristotle's four causes (material, efficient, final and formal) are all within human control. In principle, we can choose whatever material we want to manufacture the artefact; we can choose whatever tools or techniques we please; we can choose whatever design, shape or form we please. As for the final cause, it all depends on our intention. If we want to celebrate Alexander the Great, just a statue of our hero would do; if we want to celebrate our hero on his famous horse, Bucephalus, then we must design a statue of Alexander riding Bucephalus. In practice, everything depends on our limited resources in time, money, effort as well as on the state of our knowledge and on the level of our technological development.

The second context is a chemistry lab. Imagine the following scenario: the teacher hands out identical test tubes (same material, same size, from the same batch made by the same factory in the same country), two to each student. He ensures that all the beakers involved in the experiment satisfy similar criteria of identity. He tells the student to pour out 5ml of the liquid in Beaker A into their two test tubes (Test Tube 1 and Test Tube ll); he next tells them to pour out 5ml of the liquid in Beaker B into Test Tube ll only. He then says to observe the result (R) of the experiment. If he were also teaching them philosophy of science at the same time, he would use the experiment to talk about Mill's Method of Difference, that the validity of the experiment depends on it satisfying the Millean Requirement of similarity in all circumstances "save one". This is the methodology of a controlled experiment.

Note that the teacher cannot use the chemistry demonstration to teach the class about the RCT for the simple reason that it is a chemistry lesson and not a lesson in Clinical Medicine –test tubes and beakers do not need to be randomised in their allocation to the experimental or the control arm, as test tubes, unlike human beings have no consciousness or beliefs about the outcome of the experiment which could conceivably affect the outcome positively or negatively.

In Clinical Trials, humans with their peculiar consciousness are involved. Randomisation has to be introduced to avoid any bias which could conceivably arise on the part of the patient-participants or the staff-participants. Hence, the need to have double-blinding.

In other words, RCT-Model C is addressing a cluster of issues which collectively amounts to overcoming what may be called *Allocation Bias*.[7] However, more than one kind of bias may exist. *Selection Bias* appears also to exist to undermine a trial; this form of bias is related to the representative or unrepresentative nature of the patient-participants chosen to enter the trial. Two absurd trials could be cited to illustrate the difference between the two types of bias–see Jackson 2019, 296 and Dusenbery 2018, 25. The first conducted in the early 1960s in the USA to investigate whether supplementation with the hormone oestrogen would help women faced with problems after their menopause, absurdly enrolled 8,341 men but no women in such a trial. The second was a pilot study mounted by Rockefeller University and supported by the National Institutes of Health to study how obesity affected breast and uterine cancer; nevertheless, it did not recruit a single woman. We could grant that these two absurd cases adhered scrupulously to RCT-Model C; true, such trials avoided Allocation Bias but not Selection Bias.

Any society normally includes people who differ in sex, age, life-style/diet, religion, ethnicity, genetic make-up, physiological, immunological functioning, and so on. Hence the recruitment criteria must include whatever groupings are regarded as relevant to the treatment under test.

We have already observed that Model C finds favour with pharmaceutical companies whose mind is mainly focussed on getting the new drug to market when the FDA sees fit to bestow a license on them to do so. Big Pharma would sub-consciously, if not consciously, have found Mill's Method of Difference agreeable in the sense it is easy to implement in the context of testing a new drug especially in the form of a pill. Dummy pills, identical in size, shape, weight, colour to the experimental pill, could easily and cheaply be manufactured; randomise the patient-participants either to the experimental or the control arm; under double-blinding staff-participants would also not know which is the one and which the other, who is getting the real pill and who the dummy. Perfect - as far as controlling Allocation Bias and the placebo effect are concerned, Big Pharma gets full marks.[8]

---

[7] Caution: usage is not uniform.

For in-depth exploration between Allocation and Selection Bias, see Lee 2017.

[8] Remember too that it is not often that Big Pharma, through its R and D, discover a truly new drug. While working on such a goal, the bread-and-butter business of drug testing goes on but the drugs tested are "me too" drugs, that is, drugs which are already in their archive but which they could tweak, introducing some variations to their molecular structure. Every variation, however, requires a new trial. So, their labs and scientists are kept busy and their profits, hopefully, are steadily augmenting. See Angell 2004.

However, medical scientists who do not work for Big Pharma are interested in other issues (not that of getting a new drug to market) but say, in comparing the effectiveness (and safety) of the new drug/treatment with the standard/extant treatment, in determining whether "off label" drugs could be "recycled" to treat a new medical condition/disease.[9] For instance, they could be interested in whether a long standing practice does have the good effect claimed for it and which no-one has challenged but simply taken for granted.[10]

Revamping Model C: Model S[11]

With the above considerations in mind, medical scientists start to refine Model C, enrich it in such a way as to overcome bias other than Allocation Bias, and hence, to maximise useful knowledge. Simplistically put, the main methodological guidelines are as follows:

1. Assuming that the new treatment will make a difference, estimate the difference it would make between the experimental and control groups – this difference is called the "effect size".[12]
2. The ability of the trial to pick out the effect size is called the "power of the trial" which would be greater, the larger the number of patient-participants (the sample size) in the trial.[13]
3. Statistical formulae are available to help calculate the size of the trial, depending on the effect size one is expecting to observe and on the type of analysis one is engaged in doing (for instance, is the trial interested in two outcomes only which are mutually exclusive, such as dead or alive or in variables which form a continuous range, say weight gain where some patient-participants may gain x kg, others x+1 kg, yet others x +2 kg and so on?)
4. Generally, trials are designed to have minimally a power of 80%, which means that 4 out of 5 repeated trials ought to detect the smallest predicted difference in the effect size.
5. Upon completion of the trial, statistical analysis will determine whether the outcome satisfies the level of statistical significance which may be defined as: "less than a 5% chance of observing the difference in outcome, if the treatment really has no effect at all and the difference is due to the play of chance".

The key thing to bear in mind when all the statistical finessing has been done is that the sixty-four-thousand-dollar question still remains: will the treatment under study make a real difference to the health and well-being of real patients out there in the world? To answer this question, we need to introduce another distinction, between *Statistical Relevance* and *Clinical Relevance.*[14]

RCT-Model S, like Model C, undoubtedly satisfies Statistical Relevance, but doubts ought still to exist with regard to Clinical Relevance. Some of these doubts include:

(a) Clinical trials use exclusion criteria in their recruitment policy. For instance, for good reasons, pregnant women are generally excluded. For that reason, the thalidomide trials failed to uncover that thalidomide could cause defects to a developing foetus. This discovery did not come to light until long after the release of the drug for general clinical use.
(b) Some trials do not so much deliberately exclude but fail unintentionally to include. For instance, for practical reasons, trials (say, in the UK) may find it easier to recruit from more educated, middle-class people who may turn out to be predominantly white than from less educated, lower-income people,[15] whether white or non-white in the society in which the RCT is being conducted.

Decreasing such doubts to decrease Selection Bias is key to the problem of ensuring that the patient-participants are not unrepresentative of the population of real people out there in the world with the medical

[9] Their objectives chime in admirably with those of Cochrane 1972. Cochrane was interested in both effectiveness and efficacy. A new drug on the market released by Big Pharma is bound to be very expensive as the company has not only to recoup what it has spent on its R and D over the years nursing it to fruition but also to keep its shareholders happy. When its patent expires, then generics can come on stream, which would then keep the cost of medical treatment at a lower level.
[10] Doctors up to 1994 had prescribed bed rest for patients with lower back pain; yet after reviewing all the available evidence, the Clinical Standards Advisory Group in the UK concluded that not only was bed rest not doing patients any good, it was actually doing them harm when compared with those patients who were allowed to continue with their normal daily activities – see Garner and Thomas 2010.
[11] C stands for Classical as well as Cochrane; S for a revamped C finessed to overcome Selection Bias, using tools of Statistical Analysis.
[12] See Sullivan and Feinn 2012.
[13] See Campbell 2012.
[14] For more detailed exploration, see Lee 2017.
[15] This should not be taken to imply that non-white people in the UK are not found amongst the educated and the higher-paid jobs.

condition under study in the trial. Suppose you are conducting a trial for a new treatment for diabetes (Type 2) in England with an estimated of just under 3.5 million cases. Suppose you know that the disease is more common in men, 9.6% compared with 7.6% of women. So, you recruit proportionately more men than women for your trial. However, at the time of designing and carrying out the trial, there might have been no figures available regarding south Asian[16] and black ethnic groups and so you would not know that people from these groups are nearly twice as likely to have the disease compared with people from white, mixed or other ethnic groups, 15.2% compared to 8.0%. Your RCT may conform impeccably with Model C and you think also with Model S. All the same, the results of your trial, whatever they may be, have, at best, limited Clinical Relevance. It suffers from Selection Bias, as your sample size is predominantly, though not exclusively, white people.

Another strategy to promote the same end of ensuring that the patient-participants included would not be unrepresentative of the population at large with the medical condition in question is, as already seen, to increase the sample size, the number of patient-participants. As a statistical ploy, it is impeccable, as it is obvious that the larger the sample size, the greater is the probability of ensuring that the patient-participants chosen would not be too unrepresentative of the population at large. However, a larger sample size is only a necessary condition but not a necessary and sufficient condition for controlling Selection Bias and for ensuring Clinical Relevance in all contexts.

To conclude: Model S has solved only to a certain extent the problem of ensuring that the patient-participants are truly representative of the population in the real world at large with the condition which the RCT is putting to the test for effectiveness and safety.

Given this limitation, it is ironic to realize that Model C may be said to be more successful than Model S for the simple reason that it does what it says on the tin: RCTs constitute the "Gold Standard" of Clinical Evidence provided the results are reached in a trial which abides by Randomisation as well as double-blinding. By these means, Model C avoids Allocation Bias; as it has nothing to say about avoiding Selection Bias, it seems unfair to judge it deficient in terms of a goal it has never set itself. In contrast, Model S does have this more ambitious aim. Hence, judged by its own goal, it seems to have been at best only partially successful.

Furthermore, Model C, as Cochrane had put it: "The idea is not to worry about the characteristics of the patients". In other words, its default axiom is the **Axiom of Homogeneity** – patients are deemed[17] to be uniform and homogeneous so that one can get on with the business of mounting RCTs. As already pointed out, unless patient-participants are deemed to be homogeneous, how can one logically satisfy the Millean Requirement behind his Method of Difference that they be similar in all circumstances "save one"? Without this deemed homogeneity, how can one conclude that the circumstance in which the two arms (the experimental and the control) of the RCT differ (that is, that the former gets the real pill or treatment and the other dummy pill or some other treatment) is the cause of the observed effect?

Table 1.1 below shows the place of RCTs in the traditional Pyramid of Evidence.

Level l The Gold Standard	Evidence from one or more Randomised Controlled Trials (RCTs) Model C: Conforming to **Axiom of Homogeneity** & satisfying the Millean Requirement
Level ll	ll.1 Evidence from Controlled Trials without Randomisation
	ll.2 Evidence from Cohort or Case-Control Analytic Studies
	ll.3 Evidence from multiple Time Series Observational Studies
Level lll	Opinions of respected authorities built upon vast clinical experience, descriptive studies, reports of expert committees
Level lV	Anecdotal evidence based on one person's experience only

Table 1.1 Table of Clinical Evidence (more familiarly presented as The Pyramid of Evidence) invoked today
(If a higher or the highest level (Level I) is not available, then move down to the next best level)

[16] According to the language used in demographic discourse today in the UK, "south Asian" generally refers to people whose geographical provenance or ancestral origin can be traced to Pakistan, Bangladesh and India. "Black ethnic groups" refer generally to people of black Caribbean or black African origin.

[17] "Deeming" is a term borrowed from jurisprudence which involves a legal fiction. For instance, a corporation for legal purposes may be deemed to be a person, although we all know that corporations are not flesh and blood individuals or persons, like you and I.

Model C is perfectly coherent and logically consistent, whatever other limitations critics may say it possesses. Can the same be said of Model S? The short answer is no–it is not perfectly coherent and logically consistent. The reasons for this claim are presented in the next section.

Model S: incoherent?

Model S attempts to replace the **Axiom of Homogeneity** with the **Axiom of Heterogeneity** (at least an approximation of it). It is obvious, after all, that human beings share commonalities but also significant differences – we all have a heart, but your heart beats at a slightly different rate from mine (indeed, our hearts beat differently depending on the context we find ourselves in, stressed or relaxed or, indeed, at different moments of the day). Our metabolic rate varies from person to person which accounts for why some people can consume more calories and yet not gain weight or as much weight as others. For every physiological function, for every anatomical structure, for every genetic feature we can think of, we differ from one another.

If Model C appears not to recognize this fundamental fact, then obviously Model S which does is superior. However, the defect of Model C notwithstanding, its main methodological virtue rests in satisfying the critical Millean Requirement. That is why in the Pyramid of Evidence, it enjoys Level l status, as the Gold Standard. It embodies the model of Monofactorial, Linear Causation: one identifiable and clearly identified factor/variable as the cause generating its one clearly identified effect. Model S fails to satisfy the Millean Requirement: with two or more variables (such as age, ethnicity, income-level) playing a role in leading to an effect, the casual chain is more complex and more "confused" in a manner of speaking. So, it looks as if that while avoidance of Allocation Bias alone is logically compatible with the Millean Requirement, the goal of avoiding Selection Bias as well is not.

An alternative way of coping is to split Level 1 into two sub-levels: Level la (Model C) and Level lb (Model S) for the following reasons: Model C is primarily about Statistical Relevance and Model S is more about Clinical Relevance. If you are in hospital management and administration, you could be more interested in the former which will then be your Level la while the latter becomes your Level lb. However, if you are the "field doctor"/clinician, you would want to elevate the kind of trial which strives to minimise Selection Bias and you would elevate Model S to Level la instead. The clinician (in a certain region of the UK) is faced with a particular patient who is female, of Bangladeshi origin, 75, speaks very limited English, housewife, husband in low-paid job, with co-morbidities such as Type 2 diabetes. Which is the most effective/appropriate treatment I (her GP) can prescribe for such a patient?

Cochrane 1972 spoke of effectiveness as well as efficiency; although they are related, yet they are distinguishable and distinguished. The first is about whether the treatment works and to what extent; the second is primarily about its economic costs in terms of the price of the drug, time and personnel. If you work for Nice (National Institute of Health and Care Excellence, UK), you would recommend generic drugs which cost much less but are just as effective as the very expensive equivalents not yet out of patent. You might then elevate Model S to Level la instead as you would want to know how well a generic drug would work across the sick population which differs in terms of sex, age, ethnicity, income and so forth – you then prioritise those RCTs which minimise Selection Bias.

Context of Use/Application	Levels of Evidence
Statistical Relevance (Cochrane's Efficacy & Effectiveness for managers of medical resources at the level of hospital and national levels)	la: Model C designed to avoid Allocation Bias lb: Model S designed to reduce Selection Bias
Clinical Relevance (the "field" doctor treating individual patients with their own specificities)	la: Model S; Level lla in Table 1.1 1b: Model C

Table 1.2 Showing changes in the Levels of Evidence depending on the context of use

Contemporary Medicine develops fast. So, we must also peer a little into the near future to see where the notion of RCT itself might stand. At the moment, it stands at Level 1, constituting the Gold Standard in Clinical Medicine in Table 1.1 or at Level la in Table 1.2 in the context of Statistical Relevance.

The Supersession of RCT?

The Human Genome Project, following the discovery of the double-helix structure of DNA in 1953 was completed in 2004; it revealed to an astonished world that the human genome contains ~21,000 genes (at most 30,000 genes). How could such a number account for the actual behaviour displayed by us humans, as organisms? What exactly

is the relationship between the so-called genotype and the phenotype? The study of genetics, whether Mendelian or Molecular, has prompted two responses to the question: on the one hand, genetic determinism and on the other, its rejection.[18] These dichotomous stances are easily grasped in the context of human diseases. The former is impressed by the fact that some diseases are caused by single genes, such as Down Syndrome,[19] or cystic fibrosis.[20] Other single gene or monogenic disorders include sickle cell disease, muscular dystrophy, Huntingdon disease and Fragile X syndrome. However, such disorders are said not to be common, in spite of the fact that several thousands are known to exist. However, although disorders may be caused by single faulty genes, it does not mean that the disease would necessarily manifest itself – Phenylketonuria (PKU) is one such instance.[21] In many countries, infants when born are tested to see if they have inherited this condition and in the rare cases where they have, their parents are immediately advised to put the child on a diet (probably lasting a life time) to avoid foods which are rich in phenylalanine, as the inherited defective gene means that the body cannot make an enzyme called phenylalanine hydroxylase (PAH, for short), an enzyme necessary for converting the amino acid phenylalanine into other substances the body needs. The defective PAH gene then leads to the accumulation of the amino acid in the blood, causing a whole range of conditions (intellectual disability, delayed development, behavioural and social problems, psychiatric disorders and so forth), from mild to severe. In other words, the disease will manifest itself or the gene would only be "activated"/expressed in the presence of certain adverse environmental factors.

This understanding implies the rejection of Monofactorial Causation, opting instead for the Multifactorial Causal Model, in this case of inheritance (or polygenic) disorders, such as heart disease, diabetes, obesity and most cancers. These are caused by a combination of small inherited variations in genes but acting in conjunction with certain adverse environmental factors.

Genes are embedded in cells; we also know that the internal as well as the external environment of a cell can affect which genes are "turned on", so to speak. For instance, hormones can "tell" a cell to activate a specific gene (internal environmental factor);[22] outside temperature can change the fur colour in rabbits (external environmental factor).[23]

The complexities above are reflected in the figure below:

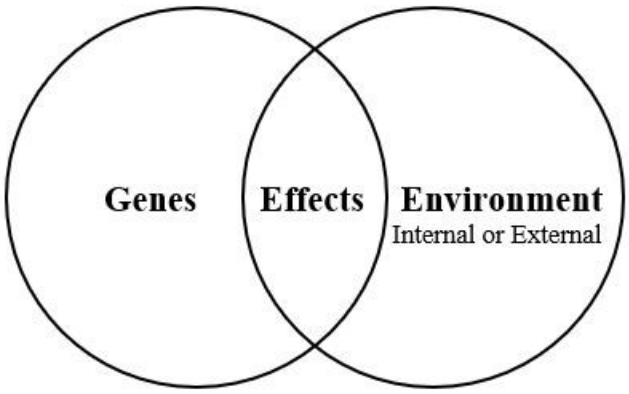

Figure 1.1 Genes interacting with environmental factors, internal or external

[18] In political discourse, the dichotomous stance appears to be starkly adhered to – it is sometimes referred to as the Nature versus Culture debate, especially over the characteristic of intelligence. Those who uphold the Nature stance claim or at least imply that intelligence is something one is born with, as part of one's natal genetic endowment. Those who disagree, opting for the other extreme, maintain that intelligence has nothing to do with genes but only with education, upbringing and the environment in general in which people are brought up. The former stance was/is normally associated with the right of the political spectrum and the latter with the left. The former response is also sometimes associated with expressions of racism.

[19] This is caused by an error in the process of cell division during foetal development, resulting in an extra full or partial copy of chromosome 21. This type of Down syndrome is called trisomy 21, accounting for 95% of cases.

[20] It is a condition caused by a faulty gene which affects the movement of salt and water in and out of cells, resulting in the body's tubes and passageways (lungs and digestive organs) being filled up with thick, sticky mucus. There is no cure but the disease can be managed.

[21] PKU is what is called an autosomal recessive metabolic genetic disorder. As such it requires two PKU alleles (one from each parent) to be present in an individual before s/he would experience symptoms of the disease.

[22] See Ing 2005 for a review of the literature on how steroid hormones regulate gene expression.

[23] An experiment used to demonstrate this: an ice pack was strapped to the back of a rabbit with white body fur. At the end of the experiment, when the ice pack was removed, one would find that the rabbit had a patch of much darker fur on its back in exactly the place where the ice pack was. See Role of Environmental Factors in Gene Expression 2012.

Not only do genes interact with the external larger environment or with the internal environment of the body in producing certain phenotypical changes and characteristics, they are also related with one another in a complex manner, as shown in the figure below.

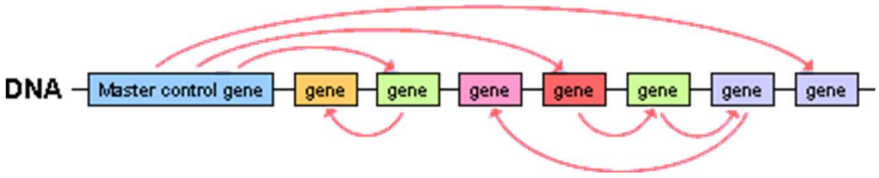

Figure 1.2 Complex inter-relationships between genes[24]

This discovery led to the emergence of several new disciplines in Human Biology. There is time and space for a brief discussion of only one of them, Metabolomics (a particular instantiation of what is sometimes called Systems Biology) and its metabolites. A few words, then, are in order about the term "Systems Biology". The term itself came into use at the turn of this century (even if the concept itself may not be new) in Modern Western Science; it can be found in Kitano 2001, 2002 and in the account of systems biology at the website of the Institute of System Biology (ISB), established in 2000 by Leroy Hood and others.[25] Since then, it looks as if the term itself is much associated with the particular vision of the subject and its application to disease and well-being propagated by the ISB; for this reason, it seems wise to refer to and explore that cluster of ontological and causal concepts which this author has called the Integrated Systems Approach to Human Biology,[26] leaving out the version of the ISB, to which we will now say only a few brief words.

The ISB version is what is called Big-data Science, a multi-disciplinary project (drawing in physics, mathematics, computer science, computer engineering and technology and, of course, biology) composing of three main inputs, namely, Biology, Technology and Computation. Such Big-data Science has spawned four "omics" fields of study: genomics, proteomics, transcriptomics[27] and metabolomics. Here, we will focus on metabolomics:

> Metabolomics is the large-scale study of small molecules (within a mass range of 50- 1500 daltons), commonly known as metabolites, within cells, biofluids, tissues or organisms. Collectively, these small molecules and their interactions within a biological system are known as the metabolome. … Metabolomics is a powerful approach because metabolites and their concentrations, unlike other "omics" measures, directly reflect the underlying biochemical activity and state of cells / tissues. Thus metabolomics best represents the molecular phenotype.
> …
> The metabolome is the complete set of metabolites within a cell, tissue or biological sample at any given time point. The metabolome is inherently very dynamic; small molecules are continuously absorbed, synthesised, degraded and interact with other molecules, both within and between biological systems, and with the environment.[28]

These small molecules have been discovered using two analytical techniques, nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS).[29] From the two quotations cited above, one could see why metabolites have been considered as such exciting "finds". The reasons include the following:

1. Their non-invasive nature, as they can be extracted from any sample of body fluids such as urine, saliva or blood (although this last is already relatively more invasive and stressful than the other two) not to mention gut fluid.

---

[24] This figure has been downloaded from: https://evolution.berkeley.edu/evolibrary/images/evo/control_gene.gif.

[25] What is Systems Biology? ISB, 2020. See also Hood 2018; Hood and Rowen 2013.

[26] Note that Noble 2006 does not use the term either; the index includes terms like "systems-level interaction", "systems-level properties", "systems-level theory", "systems-level view". Admittedly, the book is intended as a publication in popular science.

[27] We have already looked at genes (genomics), *en passant* at genes synthesising proteins (proteomics). Transcriptomics is concerned with how organisms store genetic information (both as DNA and RNA) for transmission, involving a class of proteins which regulate gene expression. See Horgan *et al*. 2011.

[28] EMBL-EBI of the ELIXR Training Platform; URL = https://www.ebi.ac.uk/training-beta/online/courses/metabolomics-introduction/what-is/#:~:text= Metabolomics%20is%20the%20large%2Dscale,are%20known%20as% 20the% 20metabolome.

What appears between round brackets has been copied from ELIXR on another page and inserted here by this author.

[29] See MS vs. NMR: Which One Is a Better Pick for Biofluids Analysis? 2020. Today, it includes a larger suite of techniques: liquid chromatography-mass spectrometry (LC-MS), gas chromatography-mass spectrometry (GC-MS), nuclear magnetic resonance (NMR) and enzyme assays render metabolite measurements possible. This is not to say that they are methodologically impeccable either–for an assessment, see Lu *et al*. 2017.

2. As they are produced at the levels of cells, tissues, organs, they can be said to be indicative of a Wholist/Integrated Systems View of the human organism.
3. They reflect dynamic processes of change at work within the human organism.
4. They act as ready biomarkers (a) of the effects of pharmaceutical interventions, (b) indicating the presence of certain risk potentials in patients, (c) alerting the medical profession to the fact that different people may react differently to particular treatments.

   Biomarkers are defined as follows: "objective indications of medical state observed from outside the patient–which can be measured and accurately and reproducibly." (See EMBL-EBI.) They are signs (objective, observable, measured by third parties such as doctors and scientists), not symptoms (subjective/reports by the patient of how s/he feels) and so pass the test of Scientificity.

   As a concept in medicine, it is not new. For instance, the humble blood pressure machine was invented in 1881 by Karl Samuel Ritter von Basch, an Austrian physician. A high blood pressure reading is regarded as a biomarker of cardiovascular disease risk. This biomarker in the medical language of today is called "a surrogate endpoint for cardiovascular disease". Biomarkers can also be used in a retrodictive way to help understand the spread of disease in a population (in Epidemiology) – the Chinese Center for Disease Control and Prevention released data (29/12/2020) showing that nearly half a million people in Wuhan would have had Covid-19, ten times the earlier figure of just over 50,000 people, based on the testing for antibodies in blood serum samples from 34,000 people in Wuhan and neighbouring cities.[30]
5. They act in general as highly useful tools in two major industries, the pharmaceutical and the bio-agricultural.[31]

The implication of the ISB vision of Big-data Science for medicine may be seen in Hood and Friend 2011 and Hood *et al*. 2012 in which the notion of P4 Medicine is set out as an implication of Systems Biology-ISB. This refers to Predictive, Preventive, Personalized and Participatory Medicine. In the light of what has been said so far, it is easy to see what is meant by the first three Ps; the last P, Participatory Medicine may be less familiar but it refers to the possibility of taking healthcare away from hospitals and GP clinics into workplaces, even schools but certainly into homes, as individuals can do self-monitoring about weight and calorie intake against a background of information and insights gathered from new technologies, analytic tools and forms of care.[32]

The existence and discovery of biomarkers (extant and new) may have the potential of dispensing with Randomisation altogether as the latter's rationale is to decrease/eliminate Allocation Bias.[33] The Randomised bit of the RCT would disappear, leaving it as CT (Controlled Trial). A further stranger consequence may follow – such a new-fangled trial may only require one patient-participant at a time. Imagine the following scenario: a patient turns up who is then diagnosed as suffering from condition C, with a biomarker reading of X (ascertained by analysing a sample of the patient's saliva, say.) The clinician after taking into account the details specific to this patient (age, gender, co-morbidities and so forth), drawing upon his/her clinical experience and having done an intensive literature search would decide that a certain drug, at a certain dosage, would be the right treatment. S/he could explicitly or sub-consciously construct an argument and a procedure looking like this:

According to respected authorities, other patients similar to this patient in my consulting room, in the relevant aspects – possessing a biomarker reading X for condition C – are reported to have benefited from treatment Y at dosage Z.
My own clinical experience also inclines me to the above.
This patient of mine will, therefore, be likely to benefit from treatment Y at dosage Z.
Explain above and discuss with patient.
Patient agrees and consents to treatment.
Take biomarker reading just before patient begins the treatment. Record this datum.
Give the patient the treatment.
Take another reading and measurement of the biomarker after an appropriate stipulated interval following treatment.
If second reading and measurement of the biomarker show decrease in a certain anticipated manner, record this as success; if second reading shows no change or an increase, record as failure. Outcome (whether success or failure)

[30] See Gan 2020.
[31] Research in agriculture from this perspective is expected to increase crop protection and engineering, offer better control of pesticides and in food industries, to identify potentially harmful bacterial strains.
[32] It is none too clear how such a proactive, data-driven type of healthcare would take shape in reality in societies with very different economic and cultural outlooks and practices. A critical voice has been raised – see Fiala *et al*. 2019 who have proposed O4 Medicine (Overtesting, Overdiagnosing, Overtreating and Overcharging) as a reminder that P4 Medicine may not be the rosy vision it appears to uphold.
[33] The glorious future of biomarkers in medicine is not guaranteed. The jury is still out: can they provide "final proof" of effectiveness and safety long-term? It is simply used here to pose a scenario in which it might make sense to postulate the supersession of RCT as Gold Standard.

would be sent on to a central collecting point for data of this kind for further study and analysis in the fashion of today's Systematic Reviews of RCTs in Evidence-based Medicine.

This may be said to be an instance of Personalised or Individualised Medicine in the history of Biomedicine which rests on the possibility of correlating signs with symptoms:

1.The patient before treatment tells the doctor that s/he feels pain in named parts of the body/feels depressed and so forth (symptoms of illness).
2. The doctor measures the metabolites in a sample of body fluid (sign of illness).
3. Biomedical understanding assumes a meaningful correlation between 1 and 2.
4. Doctor treats patient with a certain drug at a certain dosage.
5. Doctor measures the metabolites in another sample of the same kind of body fluid after an interval of time and finds a change in the number of metabolites in it (a sign of change in patient's condition).
6. Patient reports an improvement in condition (a symptom of improvement in patient's condition).
7. Biomedical understanding assumes a meaningful correlation between 1, 2 on the one hand and 4, 5, 6 on the other as well as between signs and symptoms.
8. The correlations established at 7 at the level of specificities are not expected to hold necessarily for other patients, as in the judgment of the doctor, the actual drug prescribed and the actual dosage applied would/could vary from individual patient of individual patient.

Reductionist and Integrated Systems Approaches to Human Biology: their respective metaphysics and models of causality

The differences between the two approaches are presented in the two Textboxes below.

The Reductionist Approach
Metaphysics of Reductionist Biology

A whole is no more than the sum of its component parts – call this "whole"/ "wholism"

The metaphysics of wholism goes hand in hand with an over-arching Reductionist Methodology: higher levels of organisation (structure and/or function) must be reduced with no remainder to the sum of the structure and/or function of component parts at lower levels of organisation. Let this be represented by a downward pointing arrow: ↓

There are no emergent properties

Body/Organism
↓Organs
↓Tissues
↓Cells
↓Molecules (DNA – deoxyribonucleic acid)
↓Atoms (phosphorus, nitrogen, oxygen, carbon)

Monofactorial, Linear Model of Causality: one cause leading to one effect
Cause → Effect

Textbox 1.1a The Reductionist Approach

The downward pointing arrow means that happenings at the atomic level (the lowest level) constitutes both the necessary and sufficient conditions to explain the behaviour of the Organism (the highest level)
Linear Model of Causation: causal arrow → is unidirectional
Unidirectional arrow also accounts for the "manufacture" of different body components: atoms and molecules of chemicals in genes make specific Proteins, which in turn make Cells, and so forth, leading to the One Gene, One phenotypical Effect postulate
See also Noble 2006, Figure 1 for an analogous though different way of presenting the same concept of Reductionism.

The Integrated Systems Approach
Metaphysics of Systems Biology

A Whole is different from or more than the sum of its component parts – call this "Whole"/"Wholism"

The metaphysics of Wholism goes hand in hand with a Non-reductionist Methodology: higher levels of organisation (structure and function) cannot be reduced without remainder to the sum of the structure and function of component parts at lower levels of organisation. Let this be represented by the bi-directional arrow:

↕

There are emergent properties[a] arising from the complex causal inter-relationships between component parts of the Whole (synergism, reciprocity, feedback mechanisms),[b] invoking a Non-linear Multifactorial Model of Causality

Body (Network of Networks)
↕Populations also necessarily embedded in ecological environments
↕Populations necessarily embedded within social/cultural environments
↕Bodies/Organisms (each body containing trillions of cells) forming a population
↕Organs
↕Tissues
↕Cells (Genes in nucleus of each cell)
↕Molecules (DNA in chromosomes, genes)
↕Atoms (phosphorus, nitrogen, oxygen, carbon in DNA's deoxyribonucleic acid)

Textbox 1.1b The Integrated Systems Approach

[a] Emergence: examples of emergent properties include (human) consciousness and memory (see Lee forthcoming, Chapter 8); in non-human organisation, an ant colony; at the level of atoms in chemistry, H_2O or Sodium Chloride (common salt, NaCl,) formed by combining hydrochloric acid (HCl) with Sodium Hydroxide (NaOH), two very dangerous chemicals for humans each separately in their own rights, yet their combined products, salt and water, are harmless and even necessary for human survival.

[b] Synergism: if two relevant causal variables obtain, their total causal effects is greater than the sum of each acting in isolation

Reciprocity: where the causal arrows between two relevant variables (x & y) are bi-directional, with x and y causally impacting on each other

Feedback loops: negative feedback mechanisms return the system to equilibrium upon perturbation; positive feedback mechanisms, upon perturbation, move system to a new/different level of existence and operation

See Lee 1989a/2020, 63-70.

The term "Network of Networks" is borrowed from "What is Systems Biology?", Institute of System Biology (ISB), 2020

Conclusion

In this futuristic scenario, the **Axiom of Homogeneity** (which after all was only deemed to obtain) would be dispensed with; the **Axiom of Heterogeneity** would come into full play. **Personalised Medicine** would be the order of the day. Mass Medicine based on RCTs would be a thing of the past. Not only would Model C be superseded, so would Model S. When that day dawns, the Pyramid of Evidence would become redundant. Clinical experience of respected authorities now ranked at Level lll in that Pyramid and the clinical experience of the field doctor at Level lV (see Table 1.1) would form an integral part of and play a key role in the clinician's decision-making regarding his/her patient in this futuristic order.

Heterogeneity and Personalised Medicine are given another boost by genomics, which can now concentrate on those DNA variations which differentiate one individual from another, and not only on those sequences which we have in common, to shed light on why this person and not another succumb to the pathogen given similar exposure to it.

RCTs, resting on the **Axiom of Homogeneity**, from this vantage point, would be regarded as crude, simplistic, belonging to a past age of theory and practice. Its status of Gold Standard would be rendered superfluous. Randomisation as its methodological high-point would be rendered irrelevant, as biomarkers, which are objective signs of improving, degenerating or remaining unchanged following a treatment would remove the fear that the efficacy of a treatment could be contaminated by subjectivities either on the part of the patient or of the doctors administering the treatment. Today under the RCT imperium, the placebo effect is the major bugbear behind the impulse to randomize clinical trials, to engage in double-blinding or even triple-blinding. However, when

biomarkers (which are objectively determinable) become available to evaluate the clinical outcome of a treatment on any one specific patient, then it makes no sense to have angst about the placebo or nocebo effect – on the contrary, the former can be co-opted to help improve the degree of efficacy of a treatment. The subjective beliefs and feelings of the patients as well as of those who effect the treatment are not the main consideration – what truly matters is the outcome of the patients receiving the treatment, whether their condition has objectively improved. In other words, subjectivities can be harnessed to help produce a better clinical outcome.

References and Select Bibliography

Angell, Marcia. 2004. *The Truth About the Drug Companies: How They Deceive Us and What To Do About It.* New York: Random House.

Anjum, Rani, Samantha Copeland, and Elena Rocca. 2020. Eds. *Rethinking Causality, Complexity and Evidence for the Unique Patient*. URL = https://link.springer.com/book/10.1007%2F978-3-030-41239-5 .

Applying the Bradford Hill Criteria in the 21st century. 2015. URL - https://www.ncbi. nlm.nih.gov /pmc/articles/PMC4589117/ .

Campbell, Michael J. 2012. Doing clinical trials large enough to achieve adequate reductions in uncertainties about treatment effects. The James Lind Library. URL = https://www. jameslindlibrary.org/articles/doing-clinical-trials-large-enough-to-achieve-adequate-reductions-in-uncertainties-about-treatment-effects/ .

Cochrane, Archibald. 1972. *Effectiveness and Efficiency: Random Reflections on Health Services*. London: Nuffield Trust.

Daniels, Marc, and A. Bradford. Hill. 1952. Chemotherapy of pulmonary tuberculosis in young adults. An analysis of the combined results of three Medical Research Council trials. *British Medical Journal*, 1, 1162.

Desai, Mehul, Norman Stockbridge, and Robert Temple. 2006. Blood pressure as an example of a biomarker that functions as a surrogate. *Journal of the American Association of Pharmaceutical Scientists*, 8(1): E146-E152.

Dusenbery, Maya. (2018). *Doing harm: The Truth About How Bad Medicine and Lazy Science Leave Women Dismissed, Misdiagnosed, and Sick.* New York: HarperCollins Publishers Ltd.

Evans, David. 2002. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing.* URL = https://onlinelibrary.wiley.com/ doi/10.1046/j.1365-2702.2003.00662.x.

Fiala, Clara, Jennifer Taher, and Eleftherios Diamandis. 2019. P4 Medicine or O4 Medicine? Hippocrates Provides the Answer. *The Journal of Applied Laboratory Medicine*, Vol. 4, Issue 1, 108-119.

Gan, Nectar. 2020. CNN. URL = https://edition.cnn.com/2020/12/29/asia/china-coronavirus-seroprevalence-study-intl-hnk/index.html .

Garner, Sarah, and Rachel Thomas. 2010. Evaluating a Medical Treatment–how do you know it works. Plus. URL = https://plus.maths.org/content/evaluating-medical-treatment-how-do-you-know-it-works .

Hood, Leroy. 2018. Lessons Learned as President of the Institute for Systems Biology (2000-2018). *Science*, Vol 16, Issue 1, Genomics, Proteomics & Bioinformatics.

Hood, Leroy, and Lee Rowen. 2013. The Human Genome Project: big science transforms biology and medicine. *Genome Medicine* 5, 79 (2013).

Ing, Nancy H. 2005. Steroid hormones regulate gene expression posttranscriptionally by altering the stabilities of messenger RNAs. *Biology of Reproduction*, June 72(6): 1290-6.

Jackson, Gabrielle. (2019). *Pain and Prejudice: A call to arms for women and their bodies.* London: Piatkus.

Jefferson, Tom. 2007. Why the MRC randomized trials of whooping cough (pertussis) vaccines remain important more than half a century after they were done. *Journal of the Royal Society of Medicine*, 100(7): 343-345. URL = https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC1905874/ .

Kitano, Hiroaki. 2001. Editor. *Foundations of Systems Biology.* Cambridge, MA: MIT Press.

—. 2002. Systems biology: a brief overview. *Science*, Vol. 295, 1662-1664. URL = https://science.sciencemag.org/content/295/5560/1662 .

Lee, Keekok. 1989/2020. *Social Philosophy and Ecological Scarcity*. London: Routledge.

—. 2017. RCTs & EBM as Gold Standards: A Trojan Horse for Biomedicine?' in *Knowing and Acting in Medicine*. Ed. Robyn Bluhm, Rowman & Littlefield International, Lanham.

—. Forthcoming. *A Case-study of Intercultural Philosophy of Medicine: Biomedicine, Chinese Medicine and the Colonial Mind-set*

Lu, Wenyun, Xiaoyang Su, Matthias Klein, Ian Lewis, Oliver Fiehn, and Joshua Rabinowitz. 2017. Metabolite Measurements: Pitfalls to Avoid and Practices to Follow. *Annual Review Biochemistry*, 86: 277-304.

MacGill, Markus. 2018. What is a randomized controlled trial? *Medical News Today*. URL = https://www.medicalnewstoday.com/articles/280574#what-is-a-randomized-controlled-trial .

Macleod, Christopher. 2020. John Stuart Mill. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = https://plato.stanford.edu/entries/mill/ .

Marshall, Geoffrey. 1948. Chairman of the British Medical Council of the Streptomycin Trials Committee. Streptomycin Treatment of Pulmonary Tuberculosis–A Medical Research Council Investigation. *British Medical Journal*, 2, 79-782.

Mill, John Stuart. *A System of Logic*. (Toronto and London: University of Toronto Press, Routledge and Kegan Paul, 1974): http://www.gutenberg.org/etext/27942 .

MS vs. NMR: Which One Is a Better Pick for Biofluids Analysis? 2020. URL = https://www.nanalysis.com/nmready-blog/2020/1/2/ms-vs-nmr-which-one-is-a-better-pick-for-bio-fluids-analysis .

Noble, Denis. 2006. *The Music of life: Biology Beyond Genes*. Oxford: Oxford University Press.

Salek, Rezak, Laura Emery, and Stephan Beisken. 2020. Metabolomics: An Introduction. URL = https://www.ebi.ac.uk/training/online/course/introduction-metabolomics/what-metabo- lomics.

Schaffner, Kenneth F. 1993. *Discovery and Explanation in Biology and Medicine.* Chicago and London: Chicago University Press.

Silverman, WA. 1985. *Human Experimentation: a Guided Step into the Unknown.* Oxford: Oxford University Press.

Sullivan, Gail M., and Richard Feinn. 2012. Using Effect Size—or Why the P Value is not Enough. *Journal of Graduate Medical Education*, 4(3): 279-282. URL = https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3444174/ .

What is Systems Biology? 2020. *Institute of Systems Biology*. URL = https:// systems biology.org/about/what-is-systems-biology/.

End

First edition: 3 September 2021
How to cite this article in your bibliography:
Keekok Lee. 3 September 2021. First Edition. A Diagnosis and a Prognosis of RCT as Gold Standard and Pillar of EBM. URL = www.keekoklee.org.