# Treatment Effect vs. Treatment Response

Lev S Sverdlov, M.D., Ph.D.[1]
Redmond Analytics, LLC

Abstract. The treatment response is designated as the individual ways of responding to a defined treatment. The major concepts of the treatment response approach are the aggregation, treatment-outcome complex, sensitivity to the treatment, and capacity for spontaneous recovery. Contrary to the treatment effect (defined in the frame of RCT as risk reduction in compared cohorts), the treatment response approach derives from the assumptions of the uniqueness of each member of the population, relatedness of the co-occurring events, and possibility of making valid causal inferences from single cases. The data driven deterministic hypotheses of non-random aggregation of the elements, events, and characteristics are subject to contrasting with the hypotheses of their random gathering. Complementing the traditional approach towards analysis of the treatment effect can have most substantial implications in the following areas: 1) identification and prediction of the individual effect of treatment; 2) identification and prediction of spontaneous recovery and sensitivity to the treatment, and 3) generalizing the results of RCT and predicting the individual results of treatment in the general population

Key words: treatment response, treatment effect, aggregation, treatment-outcome complex, spontaneous recovery, sensitivity to treatment

## Introduction

This paper considers approaches toward analysis of the treatment process through two complementary aspects, the *effect of treatment* and the *response of a subject to the treatment*.

Statistical analysis of the *treatment effect* has a century-long tradition deriving from the pioneering works by Jerzy Neyman[i] and Ronald Fisher.[ii,iii] The major tool for studying the *treatment effect* is the Randomized Controlled Trial (RCT), which is a statistical experiment examining a hypothesis, typically that treatment $A$ is more effective than $B$.

A common consensus is that RCT provides a unform, quantitative scientific approach towards the assessment of the efficacy of treatment. These undisputed advantages of the RCT approach make it a gold standard for the assessment of the effect of drugs, and an integral part of the system of drug development. Numerous variants and modifications of the trial design developed over decades share the same fundamental assumptions and principles.

The efficacy of the experimental treatment, i.e., the effect of the treatment in the RCT environment, is routinely defined as risk reduction for a negative outcome. The index of risk reduction refers to the entire trial population, and additional studies of heterogeneity of the *treatment effect* such as subgroup analysis and subgroup identification procedures are required to explore the variability of risk across the clinically significant segments of the trial population, and statistically assess an

[1] lev@redmondanalytics.com

association of these variations with co-variates (conditions) of the study. The accomplishments and limitations of this approach are thoroughly investigated.[iv, v]

One of the major trends in modern medicine is the individualization of treatment. Some important aspects of the RCT related to individualization remain problematic in the frame of the *treatment effect* approach. Here we discuss some of these:

- Can be a result of the RCT translated to an individual patient?
- In a statistical context, is there a way to distinguish a recovery induced by treatment from spontaneous recovery?
- To what extent can the efficacy-effectiveness gap be overcome, and can individualized predictions be made in the general population?

In our opinion, further advances on each of these problems require not only improved and new methods, but a changing in the angle from which we consider the problem.

We propose the *treatment response* approach as it is described in our previous reports,[vi,vii] which hereafter will be referred to as [*] and [**] respectively, and which will refer to our earlier related publications. Contrary to the *treatment effect* approach, which refers to the entire treatment population, the *treatment response* approach focuses on an individual patient. The *treatment response* approach starts with the observations that 1) when exposed to the same treatment, various individuals can produce different outcomes; and 2) in response to different treatments, various individuals may produce the same outcome or different outcomes.

The objectively observed co-occurrences of the treatment, outcome, and co-variates in individual patients and small groups are the starting point for logical and statistical analysis. In this sense, the *treatment response* approach is complementary to the *treatment effect* approach.

We reuse the notation and set of definitions for describing the *effect of treatment* and *response of the subject to the treatment* used in our previous reports [*, **]. While this paper refers to binary models throughout (a positive outcome vs. negative one, i.e., recovery vs. death, or having a heart attack vs. not having it), generalization to continuous outcomes would be a natural next step and follows the same principles.

# 1. Treatment effect

## 1.2. Assumptions in treatment effect analysis

A classic randomized controlled trial (RCT) design implicitly assumes the following.

- *Subjects are anonymous and interchangeable*. Personal information may move from one group to another for the reasons not related to analysis. Exclusion of any single subject from the study, moving him or her from one group to another, and/or trading single subjects between groups does not substantially affect the result of the study.
- *Two or more events co-occur by chance unless the contrary is proven.* Statistical models inherently involve randomness. The hypothesis of random co-occurrence of the events (null hypothesis) is accepted or rejected because of its correspondence (or a lack of correspondence respectively) to the observed relationships.
- *Numerous subjects are required for making valid inferences.* Any clinical research manual has a chapter devoted to choosing the sample size before data collection to obtain a desired level of power.

## 1.2. Individualization

In RCTs and observational studies, the *treatment effect* measures the impact of the treatment on the treatment population. It is defined as the absolute or relative risk reduction of a negative outcome (or, symmetrically, as increased chance for recovery), where risk is the probability of the negative outcome, the proportion of the negative outcome in the treatment population.

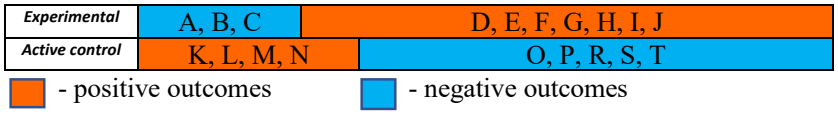
Thus, the object of the RCT is a population, not an individual. This has important ramifications.

*Case 1*

Consider a classic randomized placebo-controlled trial, and assume that sample sizes and other trial parameters are sufficient for statistical significance. On diagrams (Fig. 1, 2), the color-coded bars show positive and negative outcome.

Fig. 1


Fig. 2




The capital letters within the bars are designating conditional names of the individuals participating in the trials. The experimental cohort includes 7 patients $(A, B, C, D, E, F, G)$ with a positive and 3 patients $(H, I, J)$ with a negative outcome (Fig. 1). In the control cohort, there were 4 positive $(K, L, M, N)$ and 6 negative outcomes $(O, P, Q, R, S, T)$. The absolute risk reduction, therefore, is $30\%$.

Now suppose the outcome of the trial is as shown on Fig 2. Individuals $H, I, J$ had the positive outcome, and instead $A, B, C$ had the negative outcome. This would not matter for the assessment of the efficacy of the experimental treatment, the absolute risk reduction is the same at 30%. For individuals $A, B, C$ and $H, I, J$, however, who are the human beings (not gadgets, not bolts, not peas, not drosophilae, not molecules, etc.), it would literally be a matter of life and death.

The individual outcome is of existential concern for each individual member of the trial. Certainly, this problem looks quantitatively less substantial in trials claiming a high (e.g. 95%) product efficacy, but the human essence of the problem does not disappear even in such cases.

This is the fundamental difference of the clinical trial from the statistical experiment in agriculture, physics, engineering, etc., and this aspect of the RCT should be given a priority status in the further development of the methodology of the clinical trials.

The clinical trial is designed to assess the *treatment effect*, i.e., to identify risk and change in risk for the *trial population*; this effect cannot be directly translated to each individual member of the trial. The information on the *treatment effect* obtained from the RCT can be valuable for public health purposes. However, applying this information to the individual requires, the very least, substantial reservations and additional investigations beyond the scope of the classical RCT design. Substantial efforts undertaken to overcome this limitation will be discussed later in the paper.

## 1.3 Individual treatment effect

D. Rubin defined *causal effect of treatment* as follows: "Intuitively, the causal effect of one treatment, $E$, over another, $C$, for a particular unit and an interval of time from $t_1$ to $t_2$ is the difference between what would have happened at time $t_2$ if the unit had been exposed to $E$ initiated at $t_1$ and what would have happened at $t_2$ if the unit had been exposed to $C$ initiated at $t_1$: 'If an hour ago I had taken two aspirins instead of just a glass of water, my headache would now be gone,' or 'because an hour ago I took two aspirins instead of just a glass of water, my headache is now gone.' Our definition of the causal effect of the $E$ versus $C$ treatment will reflect this intuitive meaning." [viii]

This intuitive definition considering the *treatment effect* at the individual level contains at least two strong assumptions. It is assumed that if aspirin had been taken instead of water, the headache would be gone, which is not necessarily the case. Also, it is assumed that "headache is now gone" *because* aspirin was taken, which is not necessarily true either. Therefore, this definition contains substantial uncertainty. This is acceptable for statistical analysis of the *treatment effect* though, where the individual cases are grouped into a population, but when it is applied to an individual case, the question about the effect of aspirin remains unanswered.

The positive outcome of treatment (the "headache is gone" state) is then said to be *caused* by aspirin only if 1) this positive outcome has occurred after taking aspirin, and 2) we have a reason to believe that it would not have occurred if aspirin were not taken.

Otherwise, the positive outcome ("headache is gone") could have occurred without taking aspirin. It could possibly have occurred after taking aspirin, but not because of the effect of aspirin, regardless of aspirin, or in spite of aspirin. The difference between these cases is the difference between 'reasons to believe…' in condition 2.

## 1.4. Individual risk and population

Thus, the definition of the individual *treatment effect* is not strict, and it leaves open the possibilities for mutually exclusive inferences. This makes bridging the individual and the population *treatment effect* difficult. In the frame of the prevailing methodology, we circumvent these difficulties by not operating with the individual *treatment effects*. Rather we operate with the concept of risk, and in RCT, the *treatment effect* is defined as risk reduction via comparing risk in the trial cohorts.

In clinical research and epidemiology, risk is understood as a probability of a negative outcome estimated as a proportion of this outcome in the studied population; this definition implicitly refers to a population, not individual. A source of numerous misunderstandings is that often the index of risk computed for one of the arms of the RCT is intuitively ascribed to each individual member of this arm and is treated as his/her virtual property. In this notion, it is assumed that if, for instance, a proportion of the negative outcome among male patients in the cohort is 0.6, then the risk for each male patient of this cohort also is 0.6. Retrospectively this can be verified for the subgroup only as the proportion of negative outcomes.

However, for a single patient randomly selected from this subgroup this risk cannot be verified. Retrospectively, we always find that in each individual patient with a negative outcome the risk was underestimated; and in each one with a positive outcome was overestimated. In this sense we never find out what the individual risk really was.

The inability to assess the individual risk and to measure the individual risk reduction under the exposure to treatment inspires strong generalizations. In the context

of "population thinking," "the ubiquitous presence of individual-level variability makes it impossible to study individual-level causal effects." [ix] A widespread view among statisticians is that the assessment of the effect of treatment in an individual case is impossible. Yu Xie presents it as a paradox in social science: "Whereas there is always variability at the individual level, causal inference always requires statistical analysis at an aggregate level overlooking individual-level variability."

Richard von Mises does not leave any doubts on this matter: "When we speak of the "probability of death," the exact meaning of this expression can be defined in the following way only. We must not think of an individual, but of a certain class, e.g., "all insured men forty-one years old living in a given country and not engaged in certain dangerous occupations." A probability of death is attached to the class of men or to another class that can be defined in a similar way. We can say nothing about the probability of death of an individual even if we know his condition of life and health in detail. The phrase "probability of death," when it refers to single person, has no meaning for us at all."[x]

Richard von Mises introduced the concept of "collective" to emphasize that probability does not deal with individual cases. (An) "… example of a collective is a whole class of insured men and women whose ages at death have been registered by an insurance office." … "The definition of probability which we shall give is only concerned with 'the probability of a certain attribute of this collective'."

Summarizing the experience of numerous clinical studies, McEvoy, et al., (2014) conclude that (1) predictions of risk are accurate at the level of populations but do not translate directly to patients, and (2) perfect accuracy of individual risk estimation is unobtainable even with the addition of multiple novel risk factors.[xi]

In summary, the *treatment effect* is defined on the target population. The interpolation of the effect onto the individual within the population or extrapolation of the result to the individual member of another population is always problematic. A common notion among statisticians is that making valid causal inferences regarding the effect of treatment in a single case is impossible. [xii] Some of them even state that "identifying individual causal effects is generally not possible, or even does not make sense."[xiii]

## 1.5. Approaching the problem

Contrary to the declared and thoroughly justified senselessness of identifying individual causal effects, there exists a brunch of applied statistics aiming to predicting individual *treatment effects*. Generally, the problem is approached by considering the *treatment effect* as a function of conditions (co-variates).

*1.5.1. Subgroup analysis.* The simplest method for addressing the heterogeneity of the *treatment effect* is subgroup analysis, in which risk of the negative outcome is conditional on the variables selected based on clinical or theoretical considerations. The intention of the approach is to identify clinically significant groups of patients with elevated or low risk. Substantial limitations of subgroup analysis include false positives due to multiple comparisons, false negatives due to inadequate power, and simultaneous variation in multiple patient characteristics. Together these lead to a limited ability to inform individual treatment decisions. [iv, v]

*1.5.2.* Subgroup identification. Substantial progress in identifying clinically significant high-risk subgroups is related to the use of modern statistical and computational methods. A "subgroup identification" approach[xiv, xv] associates the high-risk groups with a combination of several co-variates. Using deep learning, Ran Chen et al., performed identification of the variability of the treatment effect across clinically relevant subgroups

associated with clusters of variables.[xvi] Unlike subgroup analysis, which is selecting variables based on clinical or theoretical considerations, the "subgroup identification" approach is creating data driven hypotheses. One of the issues requiring methodological elaboration is the relationships between the power of the entire study and the power of the procedures of the identifying the subgroups.

*1.5.3.* Identification of the predicted individual treatment effect. This approach[xvii] involves treating of the effect of treatment in individuals as a problem in prediction, and a high degree of predictive accuracy has been reached using modern statistical methods. We see the limitation of this approach in that it does not predict the individual *treatment* effect, but rather identifies and predicts the *individual outcome.* This is not just a terminological difference.

*Case 2*

Consider a typical RCT design schematically shown on Fig. 3.

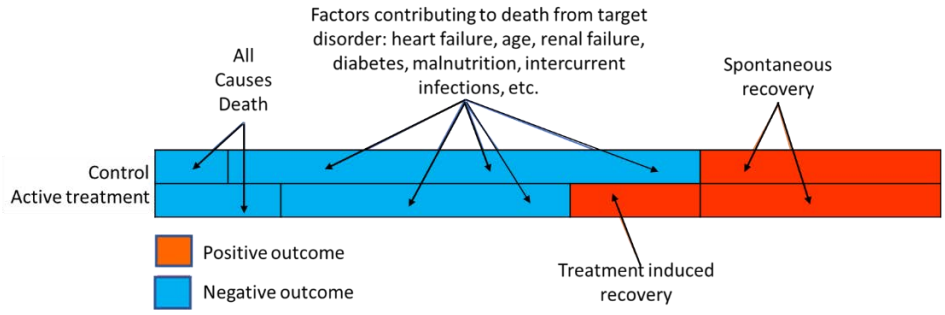
Fig. 3. Diagram of Placebo-Controlled Randomized Trial


Note, that since the placebo is biologically inactive, the patients with a positive outcome in the control cohort have recovered with no treatment, i.e., spontaneously. For simplicity, we do not consider the "placebo effect." [xviii] In the placebo control, there is a fraction of spontaneously recovered patients. If the trial was randomized 1:1, it should be expected that in the experimental cohort, there should be an amount of spontaneously recovered patients (*Spontaneous*) approximately equal to that in the placebo control. The recovery can be positively attributed to the experimental treatment only in the rest of patients with positive outcome (*Treatment-induced*). These relationships are inherent to the randomized placebo-controlled design, but they are not included in the concept and practice of the assessment of the *treatment effect* by apparent reasons: The *Spontaneous* and *Treatment-induced* cases are not clinically distinguishable. Their identification and differentiation require special analysis described in a greater detail in our report.[xix]

*Case 3*

A separate issue is the heterogeneity of the group of patients with negative outcome. While some patients could have died from the target disorder, the others could have died from other causes. Once again, the outcome itself, death, is not distinguishable in these patients, and to establish the differences in the causes of death is the prerogative of the clinical, pathology, etc. analyses.

Fig. 4. Diagram of Randomized Placebo-Controlled Trial



Altogether, in terms of causality, the RCT trial data present a complex mosaic (Fig. 4) not sufficiently accounted for in the analysis of the *treatment effect* and more thoroughly described in [*].

The very definition of the individual causal effect (see above) makes it impossible to tell whether a positive outcome, recovery, should be attributed to the effect of treatment, or it has developed spontaneously, regardless of the treatment. Also, some of the negative outcomes (deaths) were not caused by the target disorder. Clinical and pathology analysis would be required to conclude whether a specific negative outcome was caused by the target disorder, or this death was due to other cause; thus, likewise, the effect of the treatment in such cases could not be identified (See below).

Predicting the outcome is important, yet it is different from predicting the *effect of treatment*. Having the outcome in the individual predicted still leaves the question regarding the individual effect of the treatment wide open. A related limitation of the predictive approach is the difficulty of replicating the prediction from the trial population to the general population. This generalization problem applies both in terms of proportion of negative outcomes, and the identification of the individual outcomes.

## 2. Efficacy-effectiveness gap

The strong results of the "identification of the predicted individual outcome" approach are necessarily obtained within the trial population. This implies the problem of extrapolating from "in-sample" to "out-of-sample." This problem is analogous to that encountered in numerous studies using classical methods of automated classification, image recognition, etc., which were popular in 1960s-70s. The extrapolation of these results beyond the experimental population (training set vs. test set, or "in-sample" vs. "out-of-sample") to another population, or to a general population, lead, as a rule, to reducing the accuracy of the prediction. This applies both to the projection of the *treatment effect* established for the trial population to the general population or its sub-populations, and to the prediction of the *treatment effect* in the individual members of the general population.

In contemporary drug clinical trials, a positive result of a comparative trial ("$E$ is more efficacious than $C$"), once made public, can lead to an unlimited preference of drug $E$ in physicians' prescription practices and patients' preferences, effectively eliminating the less efficacious drug $C$. The demonstrated superiority of drug $E$ over $C$ often leads to indiscriminate prescribing of the more efficacious drug $E$. This is a not an optimal way of generalizing the result of the RCT.

The efficacy-effectiveness gap[xx] means that the proportion of patients with a negative outcome computed in the trial population is not necessarily the same in the general population. The indices of risk computed for the trial population and for the general population can differ, as can the *treatment effect* in the clinically significant groups of the trial population. The "individual *treatment effect*" can be predicted with accuracy next to 100% within the trial population, yet not accurately predicted in the general population.

While the result of an RCT remains an important factor in making a clinical decision, "physicians should base treatment decisions on their knowledge of the pathophysiology of the disease, the mechanism of action of the proposed treatment, and the clinical characteristics of the individual patient while informing their decision with a critical understanding of the results of relevant trials."[xxi] The appeal towards professional knowledge, experience, and intuition in the interpretation of the results of the RCT in

their reference to clinical practice is traditional and common because multiple factors - environmental, organizational, logistic, and statistical - are involved in forming the efficacy-effectiveness gap but the attempts identify the major factors forming the efficacy-effectiveness gap so far have a limited success.[xxii]

There are numerous reasons for the discrepancy between the treatment effect in RCT and general population. From a purely statistical viewpoint, in RCT, the indices of risk used for the assessment of the efficacy are computed for only a small subset of the general population selected by the inclusion/exclusion criteria which may not be representative of the distribution of the factors in the general population and described with limited number of co-variates.

The problem of determining the size and the content of the gap between populations has no satisfactory solution at one point in time. Moreover, with the passage of time, the distribution of known and unknown factors might change, and new factors might come into play, changing heterogeneity of risk and heterogeneity of effect, as well as choice of therapy [**]. Under these circumstances it would be naïve to expect that the individual *treatment effect* perfectly predicted for the out-sample of the RCT population would be projected for the general population with the same level of accuracy. Prediction of the individual outcome or the effect of treatment using regression or classification models built on the RCT sample have limited accuracy for the same reason.

## 3. Paradox

The effectiveness and productivity of the methodology of analysis of the *treatment effect* is proven with a multitude of practical results. Advances in the methodology and the use of the modern computational methods [xxiii] brought a notable progress in studying the heterogeneity of the *treatment effect* and identifying and predicting individual outcomes. Further progress can be expected with the use of the modern statistical and computational approaches (e.g., “deep learning”) which are capable of “educating” themselves and find those parameters that have a greater impact on the result.

However, so far, practical achievements with respect to the individualization of treatment are not as impressive as the theoretical and technological advances.

Consider the analysis of the *treatment effect.* A key requirement of traditional RCT methodology is that the process of the statistical analysis of the *treatment effect* can proceed only after data collection is complete, i.e., after having the entire set of individual characteristics (variables) uniformly describing each member of the population put in the record.

With the set of individual data records complete, we can measure characteristics of the studied population. The individual characteristics of each subject are interpreted as the random variations of the general characteristics of the population. The distribution of these variations is described with a set of the parameters, providing a generalized and de-individualized description of the population. This lets us use bivariate or multivariate methods to compute risks in the comparison cohorts. If our objective was limited to computing the risk and change in the risk within the trial population, this is sufficient. If our objective is to explore the heterogeneity of the *treatment effect*, or the “identification of the predicted individual *treatment effect*,” we undertake additional, more complex procedures, rooted in probability theory, requiring modern statistical methods, computational technology, time, and intellectual resources.

Thus, first, we compose the population from individuals. Then, we get rid of the individual characteristics. Then we follow a roundabout path, transforming the data to go back to the individual and his/her relationships with his/her individual characteristics.

Even after this, the results cannot be applied as-is to individuals in the general population. The actual application is guided not by the rigorous and mechanical application of the scientific method, but rather by the clinical judgement, experience, and intuition of physicians. Put this way the situation seems absurd, and it is a wonder the present author and reader have survived to this day in spite of numerous visits to physicians over their lives. The diagnosis and treatment decisions during these visits have not necessarily, if ever, involved classic or modern statistics.

At this point, having considered the analysis of the *treatment effect* from some distance, we should ask the following naïve questions.

- Is it possible that, by taking this roundabout road to study the results of treatment and to project them to the individual patient, we lose some important information or an opportunity for analysis along the way?
- Are we spending too much time and resources assessing the association between treatment, outcome and "generating conditions," which possibly could be obtained by other, more direct ways?
- Is it possible that considering the treatment process from a different angle, we could explore some aspects of the process which cannot be explored from our traditional position?
- Should we take a better look at the capacity of clinical thinking, and at least to borrow some of the ideas explicitly or implicitly realized in clinical experience and intuition?

## 4. Treatment response

Exploring causality in the treatment process requires establishing relationships between three categories of factors: treatment, outcome, and conditions (co-variates).

When exposed to the same treatment, various individuals can produce different outcomes. In response to two different treatments, various individuals can produce the same outcome or different outcomes. Thus, the outcome of the treatment depends not only on the treatment itself, but also on the way each individual patient responds to the treatment.

We define *treatment response* as the *individual way of responding to the* treatment, in contrast with *treatment effect.* An analogy illustrates the distinction:

Planting a seed into watered soil with the necessary nutritional components, results in vegetation; planting it into dry sand will have different results. Similarly, different results should be expected from hammering a nail into a wooden wall and into a granite boulder. In other words, for achieving the expected result, there should be a *correspondence* between 1) the treatment (e.g., a seed, or the nail and hummer,) and 2) the *properties* of the object to which this treatment has been applied (fertile soil *vs*. dry sand; wooden wall *vs*. granite boulder).

In this dualistic approach, the *treatment effect* and the *treatment response* characterize the same treatment process, but from two different, complementary positions. The *treatment response* approach sees the trial population as intrinsically heterogeneous regarding the factors determining the result of treatment.

## 4.1. Variability of variability

Per Jerzy Neyman, at a logical level, a population is defined as "categories of entities satisfying certain definitions but varying in their individual properties." (ref. [ix], p.1). Populations can differ in *type of variability* of individual and group properties. In the simplest case, height can vary across individuals and across groups in a certain range. Other populations can be conveniently described with hierarchical (multilevel, spatial) models, e.g., the indices of quality of educations can vary at individual, class, school, district, and so on levels. Other populations have complex heterogeneity where some of their subgroups differ from one another by *clusters of properties*. For instance, consider the staff of a large corporation as a population. It consists of various divisions with a different structure and function, with the individuals and subgroups differing by their skills, experience, cultural background, area of expertise, objectives of their work, their position in the industrial process, etc., etc. It can be a military population, with the apparent differences, for instance, between the air force, infantry, and navy, as well as substantial variability within the branches. It can be the population of a large city, or a population of the companies trading on a stock market, and numerous other heterogenous populations.

A distinctive example of heterogeneity can be borrowed from the theory of origin of life. Per A.I. Oparin, the infant Earth was surrounded by a reducing atmosphere, containing molecules of methane, ammonia, hydrogen, and water vapor, which were the raw materials for the evolution of life. The random encounters of the diverse, multiple, chaotically moving molecules, in accordance with their properties were followed by arrangement of the molecules into molecular structures of increasing complexity, with new properties, size, and forms.[xxiv] At some point of evolution, this "primitive ocean" became a mix of freely moving molecules of gas, water, mineral and organic solutions, and organic compounds of various size, form, complexity, described by J.B.S. Haldane as a "soup," [xxv] from which the first reproducing entities were created.

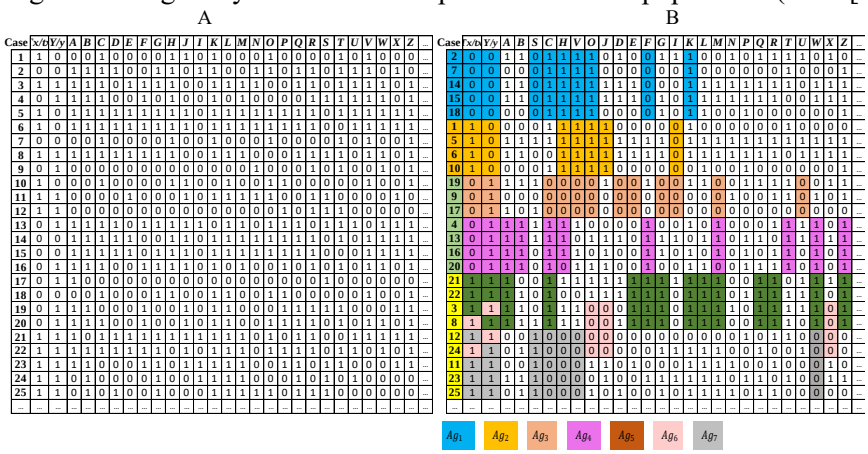
This description provides two distinctive pictures. One of them presents randomly distributed and free moving molecules, which were the raw materials for the evolution of life. In the other, there are still chaotically moving molecules, but also their various aggregations, a class of components radically different from free moving elements because of the relationships between the elements of these structures. In other words, this is a mix of the randomly distributed elements and deterministically related components composed from the same elements. These random and non-random extremes, together with a diversity of intermediate variants, form the range of possibilities of variability. The choice of analytical method for a population must be adequate to its variability type.

## 4.2. Data matrix

A data matrix is the framework for analysis of the RCT. The object of analysis, i.e., the trial population, can be described in two complementary ways, presenting two different images. An example of the matrix with binary data described by a set of variables is shown on Fig. 5 (A,B).

On Fig. 5(A), before sorting, a configuration of the elements of the matrix is visibly chaotic. On Fig. 5(B) the same data have been sorted to demonstrate the object of our interest, which are the group of color-coded subsets of variables (including treatment and outcome) identical in the subsets of cases (individuals).

Fig 5. Heterogeneity of treatment response in the trial population (from [*])




The mounting data on heterogeneity of the *treatment effect* make evident that in various individual patients and subgroups of patients, the treatment and outcome are related to various combinations of the variables, but a "physical" structure of the relatedness is not known, and it requires special investigation. Thus, the dataset of the heterogenous population presents a mix of elements, either distributed by chance, or related to each other in various combinations.

The assumption of randomness has strong implications: while part of the objects may chaotically behave like molecules of gas, others can randomly create assemblages, and one more part of the objects can present aggregations, i.e., subsets of these "molecules" related deterministically, and the data matrix can be described as a mix of randomly distributed and deterministically related elements ("soup").

When an unprejudiced investigator observes two co-occurring events, he/she must form a hypothesis as to whether these events have co-occurred by chance or were related to each other. Initially, both assumptions are equally legitimate, leading further analysis in different directions. Ultimately the results of the two analysis paths are compared and contrasted, and the initial hypothesis of each analysis should be accepted or rejected.

## 4.3. Assumptions in *treatment response* approach

The traditional analysis of the *treatment effect* derives from the assumption of randomness. If the population is complex, in the sense that it contains a mix of random and deterministically related influential factors, the assumption that the observed co-occurred events are related to each other is also valid. It coexists as a parallel hypothesis along with the assumption of randomness, which is natural for statistical analysis, and we must assume it is a legitimate possibility unless proven otherwise using logic, computations, and yet to be developed conventional criteria.

The assumptions of the *treatment response* approach, in the context of analysis of the relationships between treatment, outcome, and conditions within the experiment of treatment (RCT), are as follows:

- *Each subject is a unique individual.* Changing, removing, and/or adding individuals might completely change the subject and result of analysis. The individual outcome of treatment is of the existential value for each individual patient.
- *Two or more co-occurring events are related unless the contrary is proven.* This

assumption is the major motivating factor for analysis, and it can be accepted if the probability of co-occurrence of the events by chance is small, and it should be rejected when this probability is large.

- *Valid inferences potentially can be made from single cases and small number of cases.*

We emphasize that these are not positive assertions, but rather the assumptions in a genuine meaning of this term, designating starting positions of the approach, which is exploratory by its nature. The inference deriving from analysis based on these assumptions is relevant for a particular instance (case or cases) only. A conclusion in a particular instance can then be an element of further analysis. This can mean comparisons with analogous elements, producing inductive and deductive inferences, and comparing inferences with other hypotheses.

## 4.4. Aggregation

The RCT data matrix is represented as a rectangular space populated with sparsely distributed defined elements, with some relatively crowded areas, and packs, in which identical clusters of variables are repeated in several cases. Cases with identical clusters can be observed in matched rows and columns of the list of cases. By sorting, these cases can be brought next to each other. Similarly, variables identical in two or more individuals can be located apart in the list of variables without any impact on the content of the matrix and statistical analysis. Using relevant column sorting, they also can be brought together, creating a rectangular subset in which each line consists of the identical sub-set of variables, and each column is presented by the identical value of the relevant variable (Fig. 6., highlighted blue)

Some variables can be clustered in some subsets of individuals and differently grouped in the others. The grouping can create rectangular subsets of individuals with identical subsets of variables (highlighted blue).

Fig. 6. Scheme of Aggregation

| $I_j \backslash V_i$ | … | A | B | C | D | E | F | G | H | I | … |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $I_1$ | … | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | … |
| $I_2$ | … | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | … |
| $I_3$ | … | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | … |
| $I_4$ | … | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | … |
| $I_5$ | … | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | … |

We designate the described subset, the elements which have gathered presumptively non-randomly, with the term “aggregation,” and we use the term “case-variable association” for such assembly of elements gathered by chance [*]. The difference between the case-variable associations and aggregations [Fig. 5(B)] can be determined statistically [*].

Via purposeful sorting, or using special algorithms, we can identify and visualize one particular aggregation. If we then sort to visualize another aggregation or case-variable association, the visualization of the first aggregation can be destroyed. This process can continue, and at each step, the visualization of a new aggregation might destroy the previous ones. Still, the change in the order of variables in the matrix, as well as the order of cases (individuals), does not change its content. We only change the visualization of the multidimensional relationships between the variables and individuals existing in various virtual groups of the population.

By this procedure, we see that the data matrix is a mix of sparsely distributed elements, case-variable associations, and aggregations of various size, form, content, and

function. These components can exist solo or joined with each other, exist separately or create hierarchy, etc. The entire picture can change in time [**].

#### *4.4.1. Correlation and aggregation*

The correlation index, a measure of association between two variables, characterizes the population, and it does not refer to any individual member of the population. For instance, a positive correlation between two binary variables (designating different characteristics) means that a proportion of the pairs concordant by these characteristics (i.e., both 1, or both 0) is larger in this population than the proportion of discordant pairs (one 0, another 1). However, the correlation coefficient does not indicate which pairs are concordant and which are not.

In contrast, the aggregation is a set of the observed evens having co-occurred in certain identified individuals, irrespective of the rest of the population.

#### *4.4.2. Quantification*

Each aggregation implies a deterministic hypothesis of relatedness between its elements. This hypothesis can be and should be contrasted with an alternative hypothesis of random association of the elements. In our report [*], a set of indices including the size of the aggregation, the probability of random gathering of the elements of the aggregation, "density" of the aggregation, and other indices, was introduced. Using the report's notation, we summarize here some key quantitative features of case-variable association/aggregations.

After having the case-variable assemblage identified, the next objective is establishing its random or non-random character. The alternative to the deterministic hypothesis ("null hypothesis") above is the hypothesis of random gathering of the elements comprising the association/aggregation:

$$Pr(Ag_1') = [Pr(C_+) * Pr(D_-) * Pr(D_-) * Pr(E_+) * Pr(F_+)]^{|I_{2,3.4}|}.$$

The size ($Sz$) of the aggregation expected under the condition of random gathering of the elements comprising the association/aggregation:

$$Sz_{Ag_1'} = n * Pr(Ag_1');$$

where $n$ is the size of the matrix.

The observed size of the aggregation is a product of the cardinalities of the subset of variables and subset of individuals comprising the association/aggregation:

$$Sz_{Ag_1} = |I_{2,3.4}| * |\boldsymbol{V}_{C,D,E,F}|.$$

A comparison of the expected and observed size of the aggregation can be a basis for examining the hypothesis of random gathering of the elements comprising the association/aggregation.

In the analysis of the aggregation, relatedness between the treatment, outcome, and co-variates is inferred not via comparison of risk in various groups or subpopulation. Rather, it is defined as the relatedness of these elements *within* the aggregation, which makes it the instrument of choice for studying heterogeneity.

#### *4.4.3. Interpretation of aggregations*

So far, we've considered sets of variables constituting aggregations from a quantitative angle. Such a set might be 1) a cluster of symbols without an interpretable meaning, or 2) it can have an identifiable connotation and be a readable "word," or 3) be a combination of meaningful and "meaningless" subsets. Combining symbols into "words" in this sense can reduce the dimensionality of the description of the population. The interpretation of the "words" reveals the meaning and content of the aggregations. Ultimately, "words" can help explain the structure of heterogeneity and mechanisms of

the *treatment response*.[xxvi] Aggregation is an exploration tool, and interpretation of the aggregation is the search for new useful factors or properties.

The process of the interpretation of the "words" is similar, but not identical, to the process we described for the interpretation of principal components.[xxvii] The major difference is that with principal components we deal with a set of correlations, which requires comparing a "positive presentation" and "negative presentation." Within the aggregation, the "word" is identical in all individuals. The meaning of the "word" is to be surmised directly from the combination of the events, things, or properties denoted by the "letters" comprising the "word." This inference is a key "human-machine interaction" stage in the analysis of complex relationships and mechanisms of the *treatment effect* and *treatment response*.

## 4.5. Determinants of treatment response

In the context of *treatment response*, the numerous factors determining the result of treatment can be aggregated into two, not mutually exclusive, categories, namely the capability for spontaneous recovery, *"Spontaneous,"* ($Sp$) and sensitivity to a specified treatment *"Sensitive,"* ($St$). Under each of these categories, we mean a relevant property (attribute, quality, characteristic, ability, or trait) of a patient, determined by a single factor inherent to a relevant category of patients (e.g., polymorphism), or by a confluence of multiple internal or external factors, either prevalent or rare. The presence or absence of one or both of these properties determines the outcome of treatment. [*]

### *4.5.1. Property $Sp$*

Recovery, as one of possible outcomes in the natural course of most disorders, has been described by physicians long before the age of modern medicine. Spontaneous recovery (or remission, or intermission) is a prevalent phenomenon observed in most known human disorders. A history of this phenomenon, its epidemiological and experimental aspects, and its theoretical and practical implications are discussed in our report.[xix] We understand the capability for spontaneous recovery as a property reflecting an existence of an evolutionarily developed protective system in response to a specific hazard.

A capability to recover spontaneously, designated by the category, *"Spontaneous,"* ($Sp$), is the capacity (propensity, predisposition, readiness) to recover spontaneously, i.e., without, or regardless of treatment. We call a patient *"Spontaneous"* ($Sp_+$) if a positive outcome has developed without treatment, or there are reasons to believe that in the patient exposed to the treatment, recovery would have occurred if the treatment were not applied.

### *4.5.2. Property $St$*

The category designating sensitivity to the treatment, *"Sensitive"* ($St$), is the capacity of a subject to respond positively to the defined treatment. We call a patient *"Sensitive"* ($St_+$) to a specified treatment if there are reasons to believe that this treatment has imposed the outcome of interest, or it would have imposed it if applied. Sensitivity should be considered only in reference to the specified treatment.

Logically, the presence of the factors covered by the categories *"Spontaneous"* ($Sp_+$) and/or *"Sensitive"* ($St_+$) is the necessary condition for the development of the positive outcome in the patient exposed to the specified treatment.

### *4.5.3. Treatment-outcome complex*

In the frame of our model, there are four possible combinations of the treatment and outcome ($Tx_+Y_+, Tx_+Y_-, Tx_-Y_+, Tx_-Y_-$ ). We call this combination the "treatment-

outcome complex."[xxviii] It derives from the classification of patients by treatment and outcome, which creates a $2 \times 2$ table (Fig. 7) typically used as a framework for the assessment of the *treatment effect*.

Fig. 7. Combinations of Treatment and Outcomes

| | $Tx_+$ | $Tx_-$ |
|---|---|---|
| $Y_+$ | $Tx_+\ Y_+$ | $Tx_-\ Y_+$ |
| $Y_-$ | $Tx_+\ Y_-$ | $Tx_-\ Y_-$ |

These measures deal with objectively observed individual events of the treatment and outcome, unlike the individual treatment effect, which, as shown above, cannot be determined with certainty. Importantly, the category of the treatment-outcome complex is logically related to the categories of "sensitive" and "spontaneous," which makes a framework for possible inferences regarding treatment response. In analysis of the *treatment response*, we use the variables of treatment and outcome jointly as a unit for the identification of the individuals possessing the properties *"Spontaneous"* and/or *"Sensitive."*

## 4.6 Logic of identifying properties $Sp$ and $St$

The concept of the Treatment-Outcome Complex, logically related to the categories *"Spontaneous"* and *"Sensitive"* [*, **], allows to address the following problems:

1) identifying the properties that *determined* the outcome in the patients in the trial population

2) identifying the patients possessing relevant properties in the defined target population

3) applying this inference to prediction for patients

Problem 2 requires special consideration and is discussed below in Section "Treatment response in general population." Problems 1 and 3 can be approached through establishing logical relationships between the categories of treatment ($T_+Y_-$), outcome ($Y_+, Y_-$), and conditions ($Sp_+, Sp_-;\ St_+\ , St_-$), schematically shown in Fig. 8.

Fig. 8. Relationships between the categories Treatment, Outcome, "*Sensitive*" and "*Spontaneous*"

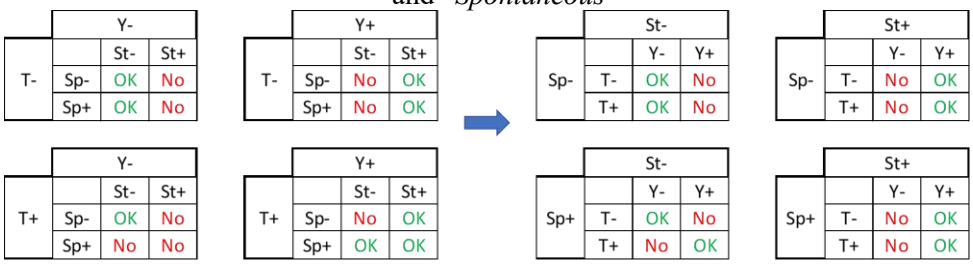



Fig. 8 shows logic relationships between the categories of treatment, outcome, "Sensitive" and "Spontaneous." It shows the possibilities of determining (retrospectively) the presence or absence of the properties "Sensitive" and "Spontaneous" in patients. Also, it shows the possibilities of predicting the outcome in a patient with all possible combination of the properties "Sensitive" and "Spontaneous" exposed and not exposed to the treatment. For instance, if we have reason to believe that the patient is "$Sensitive,$" ($St_+$) to the treatment $Tx_+$ or/and that he/she is capable for spontaneous recovery ["$Spontaneous$" ($Sp_+$)], it would be equivalent to predicting a positive outcome in this patient under the exposure to the treatment $Tx_+$.

## 4.7 Aggregations including treatment-outcome complex

Thus, the presence or absence of the properties $St$ and $Sp$ in the individual is inferred from the combination of treatment and outcome (Fig. 8), i.e., treatment-outcome complex. If the aggregation includes the treatment-outcome complex, it can be hypothesized that *this* treatment, *this* outcome, and *this* subset of the variables are related to each other *in this subset of patients,* i.e., in the individuals comprising this subset possess relevant property or properties.

The set of aggregations which include the treatment-outcome complexes represents the *heterogeneity of the treatment response* in the defined population.

It is not necessary that the aggregation should include a substantial number of individuals. Relatedness of a certain subset of variables with the treatment and outcome can be established even in a single case or a small group of cases having rare variables in their description. Borrowing an example from pharmacovigilance, two or three cases of Steven-Johnson syndrome observed in the patients with rare symptoms or conditions could be a ground for the inference of the relatedness between the treatment, these symptoms, and this serious adverse event.

Above, we have considered the logic of identification of patients with the properties $Sp$ and $St$. The presence or absence of these categories can be logically established in each individual case with the level of assurance and limitations discussed in [*], [**], even when typically a single causal factor cannot be identified. The statistical approach towards the identification of the properties $Sp$ and $St$ we describe [xix] is limited by a strong assumption of the independence of the $Sp$ and $St$ properties.

## 4.8 Mapping "Sensitive" and "Spontaneous"

The properties "sensitive" and "spontaneous" can thus be attributed to individual patients. The following step is the mapping of the factors belonging to the categories "sensitive" and "spontaneous" to variables describing the population. A procedure of the mapping described in [*] is built on the logic of pairwise comparison of individuals or groups of individuals within and across the subpopulations delimited by the treatment-outcome complexes.

There are 10 unique types of pairs, each with a distinct pattern of inferences regarding the mapping. For each type of the comparison, there is a specific set of logical operations leading towards mapping the properties "Spontaneous" and "Sensitive," i.e., ascribing a status of potential carriers of these properties to some subsets of variables. [*]

## 4.7. Implications

Identification of the "*Sensitive*" and "*Spontaneous*" individuals and mapping these properties to subsets of the variables have important implications. Here is a schematic example.

## 4.8 Accounting for $St$

*Case 4*

Consider a RCT comparing the experimental ($E$) and active control drugs ($C$) (Fig. 9). For simplicity, all patients with positive outcome have recovered only due to the treatments, and none recovered spontaneously. The color code scheme is as on Fig. 1, 2. A routinely computed result of the RCT is that the experimental drug $E$ has reduced risk of death (ARR) by 30% compared to the active control drug $C$. Setting aside problems with generalization, we say that treating the target disorder with the more efficacious drug $E$ means 70% of the patients survive, and 30% die.

At least 8 (40%) patients from the active control cohort (1, 2, 3 ,4 ,5, 6, 7, 8) should be sensitive to the drug ($C$).

Fig. 9

| | | |
|---|---|---|
| *Experimental* | A, B, C, D, E, F, G, H, I, J, K, L, M, N | O, P, Q, R, S, T |
| *Active Control* | 1, 2, 3 ,4 ,5, 6, 7, 8 | 9,10,11,12,13,14,15,16,17,18,19,20 |

Because of randomization, it means that, 40% of patients of the experimental cohort also should be sensitive to the drug ($C$). It is true for the entire experimental cohort including the segment of with a negative outcome (patients *O, P, Q, R, S, T*). If the drug ($C$) was prescribed to the patients $O, P, Q, R, S, T$, who are not sensitive to the experimental drug $E$, we should expect a reduction of the deaths rate in this subgroup by ~ 40%.

This means that the drug $C$, which has been qualified as less effective than the experimental drug ($E$), in fact can be effective in cases in which the experimental drug happened to be ineffective. Likewise, it is possible that while the trial does not reveal superiority of drug $E$ over $C$, the patients sensitive to the drug $C$ can be found in the experimental cohort, as well as patients sensitive to $E$ in the control cohort, with relevant implications upon the individualized prescription of both $E$ and $C$.

The inferences above are true in the case that drugs $E$ and $C$ belong to different classes, and the sensitivity of a patient toward the experimental drug $E$ is not related to (independent of) sensitivity toward drug $C$.

### 4.9 Accounting for $Sp$

For simplicity of presentation, capacity for spontaneous recovery ($Sp$) was not included in the Case 4, but implications of accounting for this capacity can be demonstrated on the scheme of Case 2 above, where substantial part of the recoveries in the experimental cohort could be theoretically achieved without even being exposed to the experimental treatment $E$.

Analysis incorporating the capacity for spontaneous recovery and sensitivity to treatment under the assumption of independence of $Sp_+$, $St_+$ is described in our paper. [xxix] In lieu of so strong an assumption, the identification of these properties in a heterogenous population should be performed using the sequence of the steps described in the previous sections (See: “Determinants..).

### 4.10 Potential benefits

Clinical reality is more complex, but ideally individualized prescription of treatments upon the identification of $Sp_+$, $St_+$ creates the possibility of increasing the effect of treatment and more rational use of resources, specifically reducing the number of negative outcomes, reducing risk for adverse reactions, and reducing expenditure via the elimination of unnecessary prescriptions and via prescribing a less expensive but still individually effective drug to those with a relevant indications.

## 5. Treatment response in general population

In the *treatment response* analysis we assume that the treatment population is intrinsically heterogeneous. In its description (data matrix), it presents as a mix of aggregations, case variable associations, and randomly scattered elements. The data matrix can refer to the general population or to the RCT population, which is a sample of the former.

An aggregation is, by definition, a subset of the deterministically related variables in a subset of individuals. There is no reason to believe that the process that formed this structure is specific to the environment of the RCT. If this subset is identified in members of the general population, the variables of this subset are deterministically related in the same way. The proportion of the aggregation of this specific type may vary in various samples of the population, but as identified it exists irrespective of the proportions of the comprising elements in the population, possessing the same properties and functions.

In the RCT data, the property, e.g., $St_+$, can be marked by several subsets of variables belonging to respective aggregations. In other words, the property can be deterministically related to various combinations of factors, i.e., be heterogenous. Once deterministic relatedness of the subset of variables (necessarily including the variables of treatment and outcome) in this aggregation is established, it can be interpreted as a marker for the property of our interest. Using these markers, we can identify the individuals with at least one of the relevant markers, and who, therefore, are in possession of the property $St_+$. The same is true for $St_-$, $Sp_+$, $Sp_-$, or their feasible combinations.

Subsequently the above logic (Fig. 8) leads to a prediction of the outcome in the individuals and relevant subgroups. A prediction by this method is only possible for those cases possessing relevant information in their description; for other cases we must default to traditional predictive modeling based on the *treatment effect* approach.

The *treatment response* approach leads to correct predictions on a part of the population for which information is sufficient and the property inferences are correct. Likewise, predictive modeling leads to correct (true positive or true negative) predictions for part of the population, and false positive or false negative for the rest. The projections made using *treatment effect* and *treatment response* approaches not necessarily should be identical. One approach can confirm or challenge another one, which is the meaning and purpose of complementing the *treatment effect* and *treatment response* approaches.

The frame of the relationships between the results of the *treatment response* and *treatment effect* analysis is shown in [*]. Comparative analysis of these relationships is beyond the objectives of this report.

## 6. Limitations

Clarity about the limitations of the complementary *treatment effect* and *treatment response* approaches should create the frame for combining them. The advantages of one of them might compensate for the limitations of another one, and vice versa.

### 6.1 *Treatment effect*

Above, we have demonstrated the efforts to overcome the limitations of the *treatment effect* approach. As discussed above, the *treatment effect* refers to the population, not individual. Translating the results of analysis of the treatment effect in RCT to the "real world" population is difficult; predicting the individual *treatment effect* even more so. In the real world, the task of individual case decision making is performed primarily based on clinical experience and intuition, while at the same time the *treatment effect*, i.e., risk reduction, is often misinterpreted as the generalized expectation of the positive outcome.

The identification and prediction of the *treatment effect* requires working with the distribution of the variables in the RCT. The difference between sample statistics and the parameters of the distribution in the general population requires adjustments in translating

the population and individual *treatment effect* from RCT to the "real world" population. Additional factors not included in the set of covariates, and factors that emerge during and after the completion of the trial [**], obfuscate filling the "efficacy-effectiveness gap" and predicting the individual *treatment effect*.

The identification and prediction of the *treatment effect* in sub-groups and individuals requires additional analyses beyond the design of the RCT. There is progress in this area, but as shown above, this progress is limited primarily to the experimental populations.

### 6.2 *Treatment response*

While some of the limitations of the *treatment response* approach, as well as its capacities, are already known, others are yet to be learned. At this point the most substantial limitation of exploring the *treatment response* is that, similarly to predictive modeling based on the *treatment effect* approach, it can leave some of the individuals unidentified. The *treatment response* approach does not claim that it should identify the properties determining the response to the treatment in *all* the individuals comprising the RCT and general population. Rather it claims that only individuals whose description contains relevant information can be identified with the level of assurance determined by the conventional criterion of non-randomness of the aggregation. It acknowledges the impossibility of identifying and predicting the *treatment response* in those with insufficient information. But it claims the invariance of the structure of the aggregation across various samples of the general population. In this it differs from the *treatment effect* approach, which diffuses the concept of risk onto all members of the trial population. In this sense, the limitation of the *treatment response* approach can be interpreted as an advantage: it delineates a boundary of its capacity without making an impression of certainty in situations in which there is no reason for a confidence.

## 7. Conclusion

Thus, the questions asked in the Introduction can be answered at this point as follows.

### 7.1 *Translation of the result of RCT to an individual patient*

A *treatment effect* as measured by RCT refers to the trial population and cannot be directly translated to the individual patient. Additional studies, beyond the original design of the RCT, should be done to move towards individualization. Using data on treatment, outcome, and co-variates recorded in the data matrix, it is possible to re-create the results of the trial. In the frame of this re-creation, it is possible to identify and predict the outcome in each individual participant with high accuracy, [xvii] but this result cannot be reproduced with the same level of accuracy in prediction for individuals in the general population.

The proportion of negative outcomes in the general population is not necessarily equal to this proportion in the RCT, and predictions of the individual outcome beyond the trial population are usually accurate for a part of the population. In some predictive models, this part can be substantial, but for the rest the of population the individual predictions are either false positive or false negative. The difference between the trial and general populations in the proportion of the negative (or positive) outcomes, along with the discrepancy between the predicted and observed individual outcomes, constitutes the efficacy-effectiveness gap. The predictive approach identifies the *individual outcome*, not the *individual treatment effect*, because it does not identify the individuals whose recovery was induced by the treatment, and those whose recovery was spontaneous.

### 7.2 Sensitivity to treatment $(St)$ and capacity for spontaneous recovery $(St)$

The very design of RCT shows that spontaneous recovery is a prevalent phenomenon, a notion supported with historical, clinical, epidemiological, and experimental data. [xix] Analysis of the *treatment effect* in RCT leaves unknown whether the positive outcome (recovery) was induced by the treatment, or developed spontaneously.

The *treatment response*, defined as an individual's way of responding to the treatment, is determined by multiple factors forming two, not mutually exclusive, categories, the sensitivity to the treatment $(St)$ and capacity for spontaneous recovery $(Sp)$.

There are difficulties in defining the *individual treatment effect* in the statistical context. In contrast, the determinants of the *treatment response* ($Sp$ and $St$) can be logically defined and identified (with well-defined limitations) from observation of the outcome in the patient exposed to the treatment. The logic of identifying the subsets of variables deterministically related to the properties $Sp$ and $St$ was shown in [*], [**]. These subsets can function as markers for these properties in individuals and subgroups in both RCT and general populations.

### 7.3 Tackling the efficacy-effectiveness gap

The identification of individuals with treatment-induced and spontaneous recovery is of great importance for studying disease and treatment mechanisms, to the individualization of treatment, and to reducing the efficacy-effectiveness gap.

For reasons described above, it is not likely that the efficacy-effectiveness can be eliminated in the frame of the *treatment effect* approach. Nor is it likely to be eliminated using the *treatment response* approach, nor by complementing these approaches; but there are reasons to believe that the combined approach has the potential to reduce the gap to some extent.

In the *treatment effect* approach, depending on the population and on the specific predictive model, the individual predictions are correct (true positive or true negative) for some proportion of the population, with the goal of methodology to improve this proportion. The *treatment response* approach derives from another set of assumptions, and it relies on statistically confirmed inferences about deterministic relationships of the descriptive characteristics of the aggregation [*].

The following considerations suggest that the individual predictions of the *treatment response* used in parallel to various methods of predictive modeling projecting the results of *treatment effect* analysis can lead to reducing the efficacy-effectiveness gap.

The RCT population is a sample of the general population selected using special inclusion/exclusion criteria. The RCT data matrix is a mix of the deterministically related components (aggregations of various types) and randomly scattered elements. The aggregations and randomly scattered elements are contained in various proportions in various samples of the general population, but the structure of the of the aggregations, i.e., the subsets of the respective deterministically related variables, should be identical respectively in various samples of the population, and does not depend on the proportions of the aggregations and randomly scattered elements in these samples.

The method refers to deterministic relatedness (i.e., $Pr = 1$), as distinct from probabilistic relatedness. This is a strong assumption in a statistical sense, but in a "physical" sense (or per Kolmogorov,[xxix] in the "actual world of experiment" as opposes to the "the purely mathematical development of the theory") it refers to the relatedness of the atoms creating a molecule upon a random encounter, where the probability $Pr =$

$0.99(9)$ of the random encounter means the possibility of the random encounter of the atoms, while $Pr = 1$ means that the encounter has occurred.

As a rule, a heterogeneous treatment population contains multiple aggregations comprised of distinctive subsets of variables including treatment, outcome (the treatment-outcome complex), and a subset of co-variates. Using the logic shown on Fig. 8, the *treatment response* approach identifies individuals and groups possessing the properties of sensitivity to the defined treatment $(St)$ and capacity for spontaneous recovery $(Sp)$ within the RCT population. This in turn makes it possible to identify the subsets of variables deterministically related to the property, which can play the role of markers for these properties. In turn these markers make possible the identification of individuals with these properties in the general population. This then makes possible prediction of outcomes in individuals in the general population.

In the predictive modeling, the predictions are correct (true positive or true negative) for part of the population. Similarly, the *treatment response* approach claims correct prediction for the part of the cases, specifically those possessing relevant information in their description, and in which the status of relevant aggregations has been correctly established descriptively and statistically. Importantly, the correct/incorrect predictions made using two different approaches are not necessarily identical in each individual member of the target population. The predictions made using one approach can confirm or challenge the other, which is the meaning and purpose of complementing these approaches.

The contribution of the combined application of the two complementary approaches depends on the heterogeneity of the treatment population, the set of covariates used for analysis, and multiple other factors.

## 7.4. Possible developments

The *treatment response* approach is a tool for the analysis of heterogenous populations, which can be described as a mix of deterministically related and randomly scattered elements.

The assumptions to the *treatment response* approach relate to traditional qualitative, clinical thinking. Applying this approach towards quantitative data intends to combine the flexibility and explorative capacity of qualitative, clinical thinking with the advantages of quantification, measurement, and computation. The inevitable trade-off is that flexibility must be restricted by the hard requirements intrinsic to the quantitative methods.

We are at the very initial steps of the approach. At this point, only basic statistical procedures can be advised in the context of the *treatment response* [*]. More advanced quantitative methods specifically focusing on analysis of the *treatment response* are yet to be explored or developed. Further development will allow the application of this analysis to more structurally complex variable associations.

With necessary adjustments this approach can be applied to the exploration and analysis of treatment using observational, messy, and incomplete data, analysis of single cases and small groups. In this context, a status of the aggregation can be confirmed or rejected using valid and reliable epidemiological data on relevant characteristics of the studied population.

In the context of the modern advanced approaches, the *treatment response* approach can provide a conceptual guidance framework for “self-educating,” and in particular “deep learning,” analysis of the effects of treatment.

The key impactful application of the *treatment response* approach should be expected in the area of individualization of treatment. Further impact may be sought beyond the boundaries of medicine and public health, in a broad class of areas and populations characterized by heterogeneity of response to treatment, where understanding and addressing the members of the population in an individualized manner can render better results for a smaller expenditure.

## Literature


[i] Neyman, Jerzy. 1923 [1990]. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." Statistical Science 5 (4): 465–472. Trans. Dorota M. Dabrowska and Terence P. Speed.

[ii] Fisher, Ronald A. (1971) [1935]. The Design of Experiments (9th ed.). Macmillan.

[iii] Box, JF (February 1980). "R. A. Fisher and the Design of Experiments, 1922–1926". The American Statistician. 34 (1): 1–7.

[iv]Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. BMJ. 2018 Dec 10; 363

[v]Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. BMJ. 2015 Nov 4;351

[vi] Sverdlov, L.S. 2019. Exploring heterogeneity of treatment response: assumptions, logic, algorithm, computations. In JSM Proceedings, Biopharmaceutical Section. Alexandria, VA: American Statistical Association. 3096-3119

[vii] Sverdlov, L.S. 2020. Exploring heterogeneity of treatment response: Longitudinal aspect. In JSM Proceedings, Biopharmaceutical Section. Alexandria, VA: American Statistical Association. 2083-2099

[viii] Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. J. Educ. Psychol. 66 (5): 688–701

[ix] Yu Xie. Population heterogeneity and causal inference. PNAS April 16, 2013, 110 (16) 6262-6268

[x] Mises, Richard von. Probability, Statistics and Truth. 2nd Revised ed. New York: Dover, 1981 [1957] p. 11

[xi] McEvoy JW, Diamond GA, Detrano RC, Kaul S, Blaha MJ, Blumenthal RS, Jones SR. Risk and the physics of clinical prediction. Am J Cardiol. 2014 Apr 15;113(8):1429-35.

[xii] Mises, Richard von. Probability, Statistics and Truth. 2nd Revised ed. New York: Dover, 1981 [1957] p. 11

[xiii]Hernán, MA, Robins, JM. Causal Inference. March 5, 2017 https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2017/03/hernanrobins_v1.10.32.pdf (Accessed March 8, 2017)

[xiv] Lipkovich I., Dmitrienko A., Denne J., Enas G. A recursive partitioning method for establishing response to treatment in patient subpopulations. Statistics in medicine. 2011, 30 (21), 2601-2621.

[xv] Lipkovich I, Dmitrienko A, D'Agostino RB. Sr. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials.2016

[xvi] Ran Chen, Hanzhong Liu. Heterogeneous Treatment Effect Estimation through Deep Learning. arXiv:1810.11010v1 [stat.ME] 25 Oct 2018. https://arxiv.org/pdf/1810.11010.pdf (Accessed 3/2/2019)

[xvii] Lamont A, Lyons MD, Jaki T, Stuart E, Feaster DJ, Tharmaratnam K, Oberski D, Ishwaran H, Wilson DK, Van Horn ML. Identification of predicted individual treatment effects in randomized clinical trials. Stat Methods Med Res. 2018 Jan; 27(1):142-157.

[xviii] Enck P, Klosterhalfen S. Placebos and the Placebo Effect in Drug Trials. Handb Exp Pharmacol. 2019; 260:399-431

[xix] Sverdlov, L.S. 2015. Increased Chance for Recovery vs. Risk Reduction. In: Proceedings of the American Statistical Association, Section on Health Policy Statistics. Alexandria, VA, pp.3316-3330

[xx] Ankarfeldt MZ, Adalsteinsson E, Groenwold RH, Ali MS, Klungel OH. A systematic literature review on the efficacy-effectiveness gap: comparison of randomized controlled trials and observational studies of glucose-lowering drugs. Clin Epidemiol. 2017

[xxi] Rubins HB. From clinical trials to clinical practice: generalizing from participant to patient. Control Clin Trials. 1994 Feb;15(1):7-10

[xxii] Nordon C, Karcher H, Groenwold RH, Ankarfeldt MZ, Pichler F, Chevrou-Severac H, Rossignol M, Abbe A, Abenhaim L; GetReal consortium. The "Efficacy-Effectiveness Gap": Historical Background and Current Conceptualization. Value Health. 2016 Jan;19(1):75-81

[xxiii] Holmes S, Huber W, Modern Statistics for Modern Biology. 2019. Cambridge University Press

[xxiv] Oparin, A. I. The Origin of Life, 1st ed., New York: Macmillan, 1938 [Oparin, A. I. Proiskhozhdenie zhizni. Moscow: Izd. Moskovskii Rabochii, 1924]

[xxv] Haldane J.B.S. "The origin of life" in the Rationalist Annual (1929). https://www.uv.es/~orilife/textos/Haldane.pdf (Accessed 01.15.2021)

[xxvi] Sverdlov, L., Interpretation of Principal Components: Methodological Aspect. 2005 Proceedings of the American Statistical Association, Section on Nonparametric Statistics, Alexandria, VA, pp 1639-1644.

[xxvii] *Ibid.*

[xxviii] Sverdlov, L.S. 2011. Treatment-Outcome Complex and Analysis of Observational Data. In: Proceedings of the American Statistical Association, Section on Statistics in Epidemiology. Alexandria, VA, pp. 5415–5430.

[xxix] Kolmogorov, A.N. "Foundations of the theory of probability." 2nd English Edition. Chelsea Publishing Company. New York. 1956.