Exploring heterogeneity of treatment response: Longitudinal aspect

Lev S Sverdlov, M.D., Ph.D.[1]

Redmond Analytics, LLC

Introduction

This introductory section briefly summarizes our previous report[i], to set a starting point for discussing the longitudinal aspect of heterogeneity of the treatment process. In this report we considered the heterogeneity of the treatment process in terms of two similar yet distinct concepts - the treatment *effect* and treatment *response*. Hereafter, we will refer to this report as [*]. Below, we will use the same randomized controlled trial model and same definitions and same notation with a minimum of additions.

Treatment effect. In the cited paper, a randomized controlled trial is considered as *a statistical experiment examining the hypothesis* that treatment *A* is more effective than *B*. The treatment effect is defined as the absolute or relative risk for the reduction of a negative outcome. Usually, the hypothesis of the study is generated based on preliminary observations, theoretical and experimental data. It is examined in accordance with laws of statistics, following an experimental design, under assumptions as follows:

1. *Subjects are anonymous and interchangeable*.
2. *Two or more events co-occur by chance unless the contrary is proven.*
3. *Numerous subjects are required for making valid inferences.*

The indices of risk reduction refer to the entire trial population. It may or may not be possible to extrapolate these to a general population and a set of its subpopulations.

Heterogeneity of the *treatment effect* is understood as "the nonrandom, explainable variability in the direction and magnitude of treatment effects for individuals within a population."[ii] Subgroup analysis and more sophisticated methods of subgroup identification *are procedures to analyze the relationships between the variables within the set of the data obtained in this experiment*, e.g., to identify significant subgroups of patients with substantially higher or lower risk of the target disorder. Thus these methods explore the *heterogeneity* of the treatment effect.

Treatment response. We define the *treatment response* as an *individual reaction* to a specified treatment. Depending on the factors determining this response, in response 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. In the reference to a population, *heterogeneity of treatment response* can be defined as the qualitative and quantitative diversity of individual responses, as well as responses of small groups of patients, to specified treatments.

Exploring *the treatment response* is not the equivalent to validating a pre-specified statistical hypothesis. Rather it is the *development of the data driven hypotheses* about the individual ways of reacting to a defined treatment. It is performed by identifying associations between the treatment, outcome, and conditions observed in individual patients and in small groups. Identification, analysis, and interpretation of these

[1] lev@redmondanalytics.com

associations (aggregations) stems from the set of assumptions designating the starting position of the explorations.

Another way of generating hypotheses is exploring the data in the spirit of exploratory data analysis (EDA) suggested by John Tukey[iii] in order to

- Suggest hypotheses about the causes of observed phenomena
- Assess assumptions on which statistical inference will be based
- Support the selection of appropriate statistical tools and techniques
- Provide a basis for further data collection[iv]

We apply this approach to explore a longitudinal aspect of the treatment response, using a virtual dataset of the randomized clinical trial as the model.

Analysis of the treatment process in humans has specific features and requirements vis a vis agricultural or mechanical objects. It motivated us to revise the assumptions from which analysis of the treatment response should derive as opposed to analysis of the treatment effect, which is a purely statistical concept in the context of the clinical trial.

I. *Each subject is a unique individual.*
II. *Two or more co-occurring events are related unless the contrary is proven.*
III. *Valid inferences potentially can be made from single cases and small number of cases.*

In the cited paper [*] we discussed fundamental concepts and the methodology of analysis of the treatment response.

Heterogeneity of the *treatment effect* is a rapidly developing area, but most studies in this area are focused on the cross-sectional aspect, while treatment of chronic diseases is usually a long-term process. In contrast, the *treatment response* as we define it, i.e., individual ways of reacting to treatment, is not yet well studied. Treatment response analysis promises to be instrumental in complementing studies of the treatment effect in the area of their limitations. Below we consider the *longitudinal* aspect of heterogeneity in treatment response analysis.

We consider a trial in which the outcome of interest in each of the compared cohorts over time is a summation of the binary individual outcomes (e.g., a case of MI, or first diagnosis of Alzheimer's disease, or death from the target disorder, etc.). The individual outcome of interest ($Y = 1$) can occur at any moment of time (t_k) in the frame of the model. Analyzing other temporal patterns requires relevant adjustments.

Longitudinal analysis of treatment effect: Some problems

Below we refer to long-term clinical trials on age-related disorders, which are a large family of disorders including but not limited to hypertension, type 2 diabetes, cardiovascular and cerebrovascular disorders, COPD and interstitial lung disorders, cancer, Parkinson's disease, Alzheimer's disease, acute renal failure, rheumatoid arthritis, osteoporotic fracture, venous thromboembolism, cataract, glaucoma, and many more. Some of these disorders are among the leading causes of mortality. All these disorders debilitate those affected, worsen their quality of life, and create a severe burden on their families and on the nation's health care system. These disorders affect primarily, although not exclusively, elderly people.

Unless otherwise indicated, our speculations below refer to a longitudinal trial on the effect of treatment Tx on the incidence of the age-related disorder (ARD), meaning absolute and relative risk reduction. However, with necessary adjustments, our

speculations apply to any stage of human life including childhood, *pubertas*, young adulthood, adulthood, so on.

Statistical methods employed for analysis of the treatment effect in longitudinal studies have been comprehensively reviewed by Peduzzi et al. (2003),[v] described in numerous books[vi,vii ,viii,ix,x,xi] and articles.

Randomization

Randomization is a fundamental concept for the RCT design. Usually, a population from which the patients suffering with the target disorder are recruited, is heterogeneous in many respects. The commonly recognized goal of randomization is to minimize allocation bias in the assignment of treatments, balancing both known and unknown prognostic factors. In the process of randomization, study participants are assigned randomly (by chance) to trial arms to minimize the differences between the distribution of patients with particular characteristics (conditions, co-variates). It is assumed that after baseline randomization, two (or more) groups of subjects are followed up in exactly the same way, with treatment being the only difference between the cohorts.

Events, changes and developments

The longitudinal analyses consider treatment as a process unfolding over time. Every component of the treatment process - the treatment, outcome, and co-variates - may undergo some events, changes, and developments.

Treatment. Treatment may vary measurably in its intensity, or various discrete modalities of treatment are considered. Though treatment is intended to be the most standardized component of the clinical trial, the treatment might undergo substantial changes. Usually, a treatment protocol includes the possibility of some planned and some random changes in the treatment over time. It might foresee some changes in dosage, augmentation, omissions, etc., associated or not associated with the concomitant events, intercurrent disorders, adverse effects, etc. Also, changes in treatment might occur beyond the frame of a protocol, including poor adherence, dropout, accidents, and other influential events. Also, during the process, the potency of treatment might change for some physical or biological reasons, as well as sensitivity of a patient to the treatment, and/or the patient's attitude towards the treatment.

Outcome. Our speculations refer to the individual outcome of interest ($Y = 1$), which can occur at any moment of time (t_k). The outcome for each of the compared cohorts is a summation of the individual outcomes (e.g., a number of cases case of MI, or first diagnoses of Alzheimer's disease, or deaths from the target disorder, etc.). In numerous studies of the treatment effect, it is shown that the outcome can be empirically distributed unevenly across the comparison cohorts (which is the nature of the clinical trial) and across various clinically significant subgroups (which is the essence of the heterogeneity of the treatment effect). Also, randomly, and/or non-randomly, the outcome can occur at various moments of the treatment process, and is not necessarily distributed evenly over time.

Pretreatment Measurements or Baseline Characteristics (co-variates). These variables represent conditions (co-variates) of the study. Treatment is preceded by numerous events and factors potentially affecting an outcome of treatment. These events and factors are also unevenly distributed across various segments of the population. The very idea of randomization stems from acknowledging this heterogeneity. *Importantly, the set of co-variates under consideration is usually comprised of characteristics about which there is prior presumptive knowledge, empirical or based on theory, that they can potentially influence the outcome.* In practical analysis, with rare exemptions, the temporal variation of baseline characteristics and pretreatment measurements is ignored. That is, it is either implicitly assumed that, during treatment and the follow up period, they either remain constant, or it is assumed that the changes are rare or insignificant, or the association between the events and outcome is being assessed *in toto*, without considering the events and changes that may have occurred during the treatment and follow up.

Only a few variables, such as blood type or genomic data, etc., are constant during the longitudinal trial. Most of the variables change over time, and these changes can be substantial. They can involve a quantitative expression of the covariates, e.g., during a long-term trial the patients are aging, duration of the disorder is increasing. Clinical, diagnostic, and laboratory data, demographic characteristics, the indices of a family and socio-economic status, or characteristics of environments, etc., can change over time – systematically (e.g., age, height) or occasionally, in either one or another direction (e.g., change in weight, severity of a condition, laboratory data), at various periods and moments of time. Even race and sex/gender, which are commonly used in RCTs as exemplary constant variables, might change due to the change in reporting of race/ethnicity and in cases of sex and gender identity change.

Thus, in fact, most of the pretreatment characteristics included in the set of co-variates undergo changes during the treatment and follow up period. There is no reason to believe that these events and changes occur synchronously, or that they are evenly distributed across the comparison cohorts. In the studies several years long, the developing imbalances between the cohorts might have substantial impact on the outcome.

Time series

The co-variates can have various distributional characteristics over time. Together they form a set of time series, both stationary and non-stationary, in which variables can associate with each other in different combinations over time. Analysis of multivariate time series appears to be an adequate approach to studying the treatment effect, although applying it to the format of the randomized controlled trial requires adjustments. We should note that multivariate time series analysis necessarily deals with *known* and measurable variables included in the set of co-variates.[xii]

Beyond the set of co-variates. In real time, the duration of a clinical trial is longer than the designated duration of treatment and/or follow up. The time from randomization of the first subject to data lock is longer than the defined duration of the trial: the subjects do not enter the trial simultaneously, and for a large trial, or a trial on a rare disorder, this process can take years. Therefore, a trial with planned observation of 5-8-10 years, can in fact be several years longer, which essentially comprises an epoch in a human life. In part, this evolution can be reflected with the dynamics of the covariates. In addition, during a long-term trial, material events and developments can occur which are either not related to these co-variates, or whose effect is beyond our prior or contemporaneous knowledge.

During the long-term treatment and/or follow up, the processes of human development, the progression of the disorder, environmental and social processes all progress following their own laws. Compared to the beginning of the trial, this can lead to the development of qualitatively new factors affecting the outcome. Also, the patients may experience potentially influential events, e.g., intercurrent disorders, emotionally significant events, accidents, adverse reactions to treatment, etc., as well as changes and developments in their physical, biological, social and medical status. They can undergo changes in their individual and family life associated with aging (e.g., change in marital and family status, employment, retirement, etc.), as well the changes related to a climax, menopause, changes in patients' immunological status, their sensitivity to medicines and propensity to adverse reactions. The patients may experience consequences of concomitant and intercurrent disorders and their complications, debilitation, and changes in a style and quality of life. All these changes can be accompanied with the increase in age-related morbidity, severity of disorders, and mortality.

This list of the factors potentially influencing the outcome of a longitudinal trial can on the extended indefinitely. A typology of the Influencing factors and possible changes during the trial is shown in Table 1.

Table 1. Influencing factors and possible changes

Level	Global, regional, local, group, individual
Locus of changes	Internal and external environment
Sphere of changes	Physical, biological, social, economic, political
Character of changes	Developmental changes, crises, disasters
Type of changes	Systematic, random, accidental
Structure of population	Deviation from initial balanced structure achieved via randomization
Distribution	Uneven across cohorts and along the time of trial
Heterogeneity	Cross-sectional, longitudinal

In the frame of the trial, it is not possible to compute and formally assess the influence of the factors acting beyond the scope of co-variates, but intuitively, it can be assumed that the influence of at least some of them on the outcome can be substantial.

Attrition

During the treatment and/or follow up, the trial population undergoes quantitative changes due to attrition, i.e., reducing the size of the trial population and potentially eliminating or reducing some segments of the heterogenous population (mortality, stop treatment, dropout by medicine, administrative, or behavioral reasons). [xiii, xiv, xv] Within a stochastic model assuming random elimination of the subject from the population, the attrition does not affect the outcome of the trial, [xvi] but in real trials, the attrition is not necessarily at random. Among other factors, it can depend on the perceived or real effectiveness of treatment as well as on side effects, meaningfully biasing outcome metrics. Indeed any form of imbalanced attrition might influence the outcome of the study.

Competing risks.

The factors of the attrition (mortality for concomitant or intercurrent disorder, stop treatment, dropout, etc.) can compete with the risk for the mortality from the target disorder. The consequences of effective treatment, or contrary severe side effects, can lead differences between the comparison cohorts in dropout, and therefore to a biased estimate of the treatment effect.

Summarizing, a long-term clinical trial takes a significant period of any individual's life, during which the patients are exposed to numerous influencing factors, and to changes in their medical, biological and psychological status, as well as external social, economic, and environmental factors, such that

- each patient undergoes changes following natural laws of the human development and involution
- the target disorder undergoes a process of recovery, or it evolves in severity, or the patient dies because of the target disorder or from not directly related causes
- during a trial, the treatment might change, or the potency of the treatment might change, or the reaction of the patient to the treatment might change
- age-related morbidity increases with numerous potential negative consequences
- broadly, factors of non-random, random, and/or accidental nature in individual life continue to occur
- these evolutions progress in the context of, or in (direct or indirect) relation to, local and global environmental and social processes and events (epidemics, disasters, crises, etc.) differently affecting various individuals.

The factors can exercise their influence on the result of the trial in two ways:

- direct or indirect influence differently affecting various individuals
- changes in quantitative characteristics of the population (dropout, deaths due to the target disorder or ACD, stop treatment by medical or administrative reasons).

The impact of some of these factors on the result of the trial can be synergetic, the others can cancel out each other, *but there is no reason to expect that they are distributed evenly across the comparison cohorts and along the treatment process and follow up. There is no reason to expect that the influencing factors create a new balanced structure, with the treatment being the only difference between the cohorts*. It is more likely that they be distributed unevenly across the trial arms and along the duration of the trial, which has important implications.

"Intent-to-treat"

Per R. Fisher,[xvii] a randomization procedure must be the last one before the initiation of treatment because the influential factor coming *after* the randomization might destroy the randomized structure of the experiment and, therefore, invalidate its result. In early randomized controlled trials, the treatment was considered during a single, aggregate segment of time[xviii] (Appendix I), and randomization was considered be sufficient to warrant an unbiased estimate of the treatment effect.

Gradually, the problems related to the influencing factors entering the treatment and follow up process became apparent and led to the development of the concept of the "intent-to-treat."[xix, xx,xxi] For instance, a portion of the patients randomized to the active arm of the trial refused, or for some reasons have not received, the treatment, or dropped out. It creates a problem with comparing the active and control arms and computing the treatment effect for the randomized population: if we account only for the patients who in fact have received treatment ("per-protocol"), the balance between the comparison cohort is broken. A result of the trial does not reflect the effect of the treatment on the population for which the trial was initially designed. Rather it reflects the effect of the treatment on *the trial population which has been modified (reduced) by the new (or modified) factors.*

If, to the contrary, we compute the effect of the treatment in all patients randomized to the cohorts then, in fact, we inflate the denominator in the calculation of treatment effect by including the patients who in fact did not receive the treatment.

The intent-to-treat approach acknowledges that known and unknown factors enter the process after the randomization and that dealing with this issue requires stepping beyond the frame of the narrowly defined concept of the treatment effect. The intent-to-treat approach declares that the trial is assessing not the treatment effect *per se*, but rather the effect of the implementation of the policy providing treatment.

Therefore, the intent-to-treat interpretation is a pragmatic compromise between the planned experiment and the result of its real-life implementation. It is an effort to explain the emerging paradoxical uncertainty with a theory acceptable from a position of common sense, and it should be acknowledged that to a certain extent it works. However, with the increasing duration of the trial, the amount of the potentially influential factors grows along with increasing uncertainty, whether such factors are known or unknown, non-random, random, or accidental, considering they are unevenly distributed across the cohorts and along the timeline of the trial.

For trials exceeding a certain duration we should find that the compromise ceases to be acceptable. As far as additional, influential factors enter the context of the trial, the boundaries of the alleged policy of the implementation of the treatment gradually become less certain, until they get washed out. From the beginning of the trial, the convention was that the result of the trail, specifically, the index of the treatment effect, does not assess the treatment effect, it assesses the policy. Beyond some moment, the policy cannot be satisfactory delineated.

If we are true to the principles of experiment, we must answer the following questions: is there any reason to believe that this trial/policy could be replicated under identical or similar diversity of the conditions? Can the result of this trial be reproduced? Using terms like reproducibility broadly/generically, bypassing specialized definitions, [xxii] we must admit that it is not likely that conditions of the long-term trial can be repeated, replicated, and reproduced close enough to the condition of the original trial. In fact, the replica of the long-term trial, if possible, would be carried out in substantially different conditions. Under such conditions, we cannot expect the results of the trial to automatically generalize to the population. The assessment of the result of the trial then moves from the domain of the probability theory and statistics to the area of experience, intuition, and common sense.

The problem is that there is no formal way to determine the cut-off point. This problem arises even if we try to estimate a cut-off point by comparing treatment effect measured at consecutive time segments of the completed trial. Most directly, we observe the number of outcome events for each segment, but the denominator (the number of treated cases) in each segment will be different because of attrition, which makes the segments not comparable.

Longitudinal trials are of paramount necessity for medicine. Therefore, in planning new trials, the duration of the trial should be well justified, and new analytical approaches should be considered. In a long-term longitudinal trial, we achieve a *global* assessment of the treatment effect, referring to the entire trial population during the entire time of treatment and/or follow up. Beyond the global assessment, we seek to assess changes in the treatment process over time. As noted above, there are difficulties in such an assessment.

We have argued that analysis of the treatment effect, as defined in clinical trials, should be complemented with an approach allowing a more granular assessment of the results of treatment and factors involved at all stages of the treatment process. Below, we will explore the longitudinal aspect of the treatment response. This approach is primarily focused on an individual patient (or small groups) during short consecutive segments of time, under changing conditions. A summation of these elements leads to the integration of the picture of the entire process and to the assessment of the process, as well as the character and timing of the changes taking place during the process of treatment and follow up.

Exploring treatment response

Cross-sectionally, a population of a clinical trial is heterogeneous in terms of treatment effect and treatment response. During a long-term trial, the outcome events are, typically, distributed unevenly along the time axis, i.e., number of the events varies during various segments of time. Above we have argued that, at various moments/segments of time, the trial population, and each of its subgroups, can be exposed to diverse combinations of the potentially influential factors, both internal and external. As shown above, the potentially influencing factors also are distributed unevenly across the segments of the trial population and along the time of treatment and follow up. In the data, this uneven distribution gives rise to 'nodes,' or aggregations, created by the associated variables of the co-variates, treatment, and outcome, observable at some points in time.

In our previous works we presented the analytical approach stemming from the assumption (not empirical assertion) that these nodes correspond to deterministic relationships between the of treatment, outcome, and conditions in subsets of the individuals. This approach naturally focuses on the individual way of reacting to the treatment, which we have defined as the treatment response, as distinct to the statistical concept of the treatment effect.

We present here an approach to the longitudinal aspect of heterogeneity of the treatment response. The definitions of the major concepts of analysis of the treatment response of the concepts "sensitivity to treatment," "capacity for spontaneous recovery," "treatment-outcome complex," as well as a notation are in our previous reports [*]. The necessary additions are in the text below. Here we treat the simplest temporal pattern, assuming co-occurrence of the treatment, condition and outcome within one cycle.

Treatment response in a longitudinal context

In our model, in any individual, the event of the outcome of interest can occur at any moment of time during the trial. Most commonly, timing of the occurrence of the outcome is thought as a random process as studied in survival analysis. In Cox's proportional hazard regression analysis, the timing of the occurrence of the outcome is a random variable conditioned on treatment and on a set of pre-treatment measurements and baseline characteristics as they are described above in the relevant section.

The treatment response is defined as an individual's way of reacting to the treatment. It was shown above that treatment in each individual patient can change during the trial, but for illustration of the approach towards analysis we assume it constant.

Directly, we can consider only the factors represented by the set of covariates. Apparently, one can expect that the distribution of these factors both across the trial

subpopulations and along the treatment and follow up process will demonstrate heterogeneity of the treatment process in both these aspects.

We are interested in exploring the factors potentially related to the type of the outcome, to the type of reaction of the patient to the treatment, as well as the factors related to the timing of the event of the outcome. Specifically, we are interested in learning why the outcome Y_+ in the patient I_g treated with treatment T_+ occurred at cycle t_k. Having explored these factors, we will better understand the nature of reaction of the individuals to the treatment, with the ultimate goal of finding opportunities to influence survival time.

Sets and sequences of cycles

Unlike the treatment effect, which should be considered only in the context of the entire study population (or in the context of the entire subgroup in subgroup analysis), analysis of the treatment response, i.e., the individual way of reacting to the treatment, focuses preferentially on

- a single case
- a small group
- defined trial subpopulations
- entire trial population

All these variants can be considered during

- a single or aggregate segment of time
- a sequence of equally spaced segments of time t_i (or "cycles"[xxiii,xxiv]), e.g., analysis of the selected or uncompleted set of the trial data
- a set of cycles during the entire duration of the trial (analysis of the completed set of the trial data)
- time periods from the population data on the incidence of the outcome of interest.

The combination of a segment of the population with a segment of time creates a context determining the boundaries of possible inferences.

Synchronization

Typically, while studying the treatment effect, the clinical trial data are synchronized *by the date of the beginning of treatment*, or sometimes by the date of randomization. Then, the timing of the events and related outcomes (duration of treatment, duration of follow up, survival time, etc.) is computed relative to this index date.

Another time frame is *chronological*. It means that the events of the trial are studied as they have occurred by calendar dates. A chronological approach is suitable primarily for analysis of logistical, organizational aspects and the trial monitoring and management. Also, it can be instrumental for studying of the external events and factors potentially influencing the trial population.

The time frame with synchronization *by the date of the outcome* of the study might suit best for studying the factors preceding the outcome and their potential impact on the treatment response and especially on the timing of its realization.

Also, the synchronization *by the date of the event of interest* or *by the date of the outcome* can be used for analysis of incomplete longitudinal records, and limited subsets of individuals e.g., in safety monitoring of the clinical trial, in analysis of observational data, keeping in mind the natural limitations imposed by the nature of these datasets.

Asynchronous data sets can be used for various analytical purposes, where timing is not the issue. For instance, in subgroup analysis the sets of the patient records are defined by a clinically significant variable not necessarily with a reference to duration, chronology, etc.

Aggregation in a longitudinal context

Formally, a population of a longitudinal trial can be described as

$$P = P\begin{vmatrix} I_{1,2,\dots,i} \\ V_{1,2,\dots,j} \\ t_{1,2,\dots,k} \end{vmatrix};$$

where at each moment of time t_i the patient's status can be described by a vector

$$V_{t_i} = Tx_{t_i}, Y_{t_i}, X_{t_i}, E_{t_i}, U_{t_i}\ .$$

Each patient record is described longitudinally as a series of sets of conditions defined at each consecutive moment of time t_k as a set of values of the vectors V_{t_i}.

Aggregation is a fundamental concept in analysis of the treatment response. In a longitudinal context the *aggregation* refers to a *subset of variables,* which is identical in a *subset of individuals* during a *subset of sequential cycles (equally spaced increments of time).*

If, for instance, the aggregation Ag_1 (from the study[*]) was observed during four cycles $(t_1, t_2, t_3\ t_4)$ it could be written as $Ag_1 = I_{2,7,14,15,18}\ V_{(Tx\_Y\_),C_+,F\_,H_+,K_+,O_+,S_+,V_+}\ t_{1,2,3,\dots,4}$ or

$$Ag_1 = Ag\begin{vmatrix} I_{2,7,14,15,18} \\ V_{(Tx\_Y\_),C_+,F\_,H_+,K_+,O_+,S_+,V_+} \\ t_{1,2,3,\dots,4} \end{vmatrix};$$

i.e., the aggregation presents as a multidimensional set of variables inscribed in the 3D space created by the axes of I , V, and t.

The indices defining a size of the aggregation (Sz_{Ag_i}), the probability of random gathering of the elements of the aggregation (Pr_{Ag_i}), and "density" (Dn_{Ag_i}) of the aggregation are described in the cited article [*].

Population as a set of aggregations

In this context, the population can be described as a set of aggregations. In the data matrix, which is a framework for our model, the aggregations can be observed apart, next to each other, overlap entirely or partially, can include one another, etc. Various programmatic and computational approaches to identify the aggregations can be considered. In some cases, primarily in small data sets, the aggregation can be observed directly, by reordering the observations to bring relevant rows together.

Since the order of the members in the sets I and V does not matter (unless there are special considerations), the elements (lines, columns) of axes I and V can be sorted in any order. However, if reordering of aggregations by time is performed for some analytical reason, it should be kept in mind that the natural space and time structure of the events and their relationships can be violated. This can have a substantial impact on the interpretation of the results of analysis.

Explorations and inferences

Single case, one cycle

In medicine, drug safety, litigation, and many other disciplines, *causal assessment* in *single cases* and *small groups* rather than populations or statistically significant samples, is one of the major areas of operations. Almost exclusively, such inferences employ the methods of qualitative analysis. The attempts of quantification of the causal assessment in these cases pose specific and difficult-to-solve methodological problems. The available literature does not appear to contain methods that would fit the practical needs for causal assessment in individual cases in the areas listed above. [xxv, xxvi, xxvii, xxviii, xxix, xxx]

Yu Xie described 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." A common notion among statisticians is that making causal valid inferences regarding the effect of treatment in a single case is impossible. [xxxi] Some of them even state that "identifying individual causal effects is generally not possible, or even does not make sense."[xxxii]

D. Rubin defined a 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." [xxxiii]

One can see that this intuitive definition has 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 "my headache is now gone" because aspirin was taken, which is not necessarily true either. Therefore, the definition by D. Rubin still contains substantial uncertainty, which, however, is acceptable for statistical analysis of the treatment effect.

We will not continue the attempts to define an individual case form a statistical or probabilistic position. Rather we will consider the inferences on a basis of exploring the conditions (properties) *necessary* for the development of this outcome, i.e., consider the situation from the position of the treatment response.

Logically, the outcome of *the* treatment applied to *the* individual to is determined by 1) the ability of *the* treatment to exert its effect on *the* individual and 2) the ability of *the* individual react to *the* treatment with the development of the outcome. Therefore, we consider each of four possible combinations of *the* treatment and *the* outcome (Tx_+, Tx_+; Tx_+, Tx_-; $Tx_- Tx_+$; $Tx_- Tx_-$) as the "treatment–outcome complex," which provides limited but still substantial opportunity for causal inferences regarding the properties determining the individual's response to the treatment. These categories were thoroughly considered from historical, clinical, epidemiological, and experimental positions.[xxxiv]

The infinitely numerous factors forming the individual treatment response can be generalized into two not mutually exclusive categories: the capacity (or a lack of capacity) of a subject to respond positively to the treatment, i.e., be sensitive or not sensitive to the

treatment [*"Sensitive"* (St_+)], and the capacity (propensity, predisposition, readiness) to recover spontaneously, regardless of treatment [*"Spontaneous"* (Sp_+)]. [*]

Table 2 schematizes the relationships between the treatment and outcome using the categories of factors forming a response of the individual to the treatment (St_+, St_-, Sp_+ and Sp_-). In our longitudinal model, the status on the categories St_+, St_-, Sp_+ and Sp_- is considered for each cycle separately.

Table 2. Relationships between the Categories of Treatment, Outcome, "Sensitive" and "Spontaneous"

	Y-	
	St-	St+
Sp-	OK	No
Sp+	OK	No

	Y+	
	St-	St+
Sp-	No	OK
Sp+	No	OK

	St-	
	Y-	Y+
T-	OK	No
T+	OK	No

	St+	
	Y-	Y+
T-	No	OK
T+	No	OK

	Y-	
	St-	St+
Sp-	OK	No
Sp+	No	No

	Y+	
	St-	St+
Sp-	No	OK
Sp+	OK	OK

	St-	
	Y-	Y+
T-	OK	No
T+	No	OK

	St+	
	Y-	Y+
T-	No	OK
T+	No	OK

The left side of the table indicates a presence of the categories of "Spontaneous" and "Sensitive" in all possible combinations of the treatment and outcome. The right side of the table demonstrates the outcome prospectively expected in individuals having all possible combinations of the categories "Spontaneous" and "Sensitive."

Thus, the logical inferences regarding the categories of the treatment response are possible, although in some cases they are limited. In all cases, the negative options are indicted with certainty. The positive options in some cases are the only choice that can be logically made. In other cases, it can be a choice between two options (with or without a possibility that both are true), and in some special cases the inference regarding one of the categories cannot be made.

Sequences in individual subject

In the cited article, the approach towards exploring the treatment response was considered cross-sectionally, i.e., during a single aggregate period of time, within the entire trial population. [*] In the longitudinal context, the treatment process is being considered during a sequence of the equally spaced time segments (cycles). The relationship between treatment, outcome, and the categories of the treatment response, can be an instrument for logical analysis of sequences of cycles.

Above, it was demonstrated that typically during a clinical trial, both treatment and conditions change over time. For simplicity, assume the treatment to be constant until the occurrence of the outcome or the end of the observation. Then, we can hypothesize that the time of the occurrence of the outcome depends on the change of the conditions. Unlike survival analysis, where the timing of the occurrence of the outcome is a random variable conditioned on treatment and on a set of the pre-treatment measurements and baseline characteristics as they are described above, we are interested in the association between the change of the conditions and the timing of the occurrence of the outcome of the named individual.

A naïve question is: why did the outcome, which did not occur during any previous cycle, occur specifically at a given cycle, e.g., t_4.

Let us consider the simple sequence of the events in an individual patient.

$$t_1;\quad t_2;\quad t_3;\quad t_4;\quad t_5;\quad \ldots.;\ t;$$

$$\frac{Cn_-}{Y_-};\frac{Cn_-}{Y_-};\frac{Cn_-}{Y_-};\frac{Cn_+}{Y_+};\frac{Cn_-}{Y_-};\ldots;\frac{Cn_-}{Y_-};$$

where Cn is the condition. In this example, the patient was treated with Tx_+ during the entire period of observation.

The condition Cn (variable Cn) assumed independent from treatment. During the cycles preceding Cn_+ the outcome was negative (Y_-). The change of Cn_- to Cn_+ at the cycle t_4 was followed by the change in the outcome from Y_- to Y_+. There are several possible variants of interpretation. The simplest (and most natural, but not the only possible) interpretation is that the patient was not sensitive to the treatment Tx_+ (did possess neither St_+ nor Sp_+); the condition Cn_+ represents the property St_+, Sp_+, or both.

More complex sequences with analogous interpretations are

$$\begin{matrix} t_1; & t_2; & t_3; & t_4; & t_5; & t_6; & t_7; & t_8; & t_9; & t_{10} \\ \frac{Cn_-}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_+}{Y_+}; & \frac{Cn_-}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_+}{Y_+}; & \frac{Cn_-}{Y_-}; \end{matrix}$$

or

$$\begin{matrix} t_1; & t_2; & t_3; & t_4; & t_5; & t_6; & t_7; & t_8; & t_9; & t_{10} \\ \frac{Cn_-}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_+}{Y_+}; & \frac{Cn_-}{Y_-}; & \frac{Cn_+}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_+}{Y_+}; & \frac{Cn_-}{Y_-}; \end{matrix}$$

$$\begin{matrix} t_1; & t_2; & t_3; & t_4; & t_5; & t_6; & t_7; & t_8; & t_9; & t_{10} \\ \frac{Cn_-}{Y_-}; & \frac{Cn_-}{Y_+}; & \frac{Cn_-}{Y_-}; & \frac{Cn_+}{Y_+}; & \frac{Cn_-}{Y_-}; & \frac{Cn_+}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_-}{Y_-}; & \frac{Cn_+}{Y_+}; & \frac{Cn_-}{Y_-}; \end{matrix}$$

In accordance with the principle of maximum achievable certainty, the content and certainty of the conclusion from the explorations depend on the content, number and structure of the sequence available for analysis. The possibility of a causal assessment in quantitative terms depends substantially on the availability of temporal characteristics of the treatment process, such as the duration of the cycle and the incidence rate of the event. The appropriate duration of the cycle can be selected using theoretical, observational, and experimental data, or, for exploratory purposes, from available data.

Challenge, de-challenge, re-challenge (CDR)

The protocol “challenge, de-challenge, re-challenge” is one of the instruments used for causal assessment in individual cases in many areas. The sequences subject to the CDR protocol can be observed in analysis of the segments of the uncompleted trial (e.g., safety monitoring), or in post marketing surveillance, or in clinical practice.

Schematically, the sequence subject to CRD looks like

$$\begin{matrix} t_1; & t_2; & t_3; & t_4; & t_5; & t_6; & t_7; & t_8; & t_9; & t_{10} \\ \frac{Tx_-}{Y_-}; & \frac{Tx_-}{Y_-}; & \frac{Tx_-}{Y_-}; & \frac{Tx_+}{Y_+}; & \frac{Tx_-}{Y_-}; & \frac{Tx_-}{Y_-}; & \frac{Tx_-}{Y_-}; & \frac{Tx_-}{Y_-}; & \frac{Tx_+}{Y_+}; & \frac{Tx_-}{Y_-} \end{matrix}$$

In the frame of the CDR protocol, the emergence of the outcome of interest or adverse event after the initiation of treatment, its disappearance after discontinuation of treatment, and reappearance after resuming the treatment is an intuitive argument in favor of the association between the treatment and the adverse event, but this intuition can rarely be supported with quantitative arguments. The influence of conditions (and moreover, changing conditions) is either not accounted for or considered in the same intuitive fashion.

Sequences in the context of co-variates

The occurrence of the outcome of interest can be accompanied with the change in the set of conditions during one or several cycles in a single patient or in several patients. The simplest sequence of events can look as follows

$$Ag_i \begin{vmatrix} I_{7,19,34} \\ V_{(Tx_+Y_-),B_-,F_+,R_+} \\ t_{k-4,k-4,k-4} \end{vmatrix}; Ag_{i+1} \begin{vmatrix} I_{2,15,18} \\ V_{(Tx_+Y_+),B_+,F_-,R_-} \\ t_{k+1} \end{vmatrix};$$

where the occurrence of the outcome Y_+ is accompanied with the change in the conditions B_-, F_+, R_+ to B_+, F_-, R_-. In fact, we are considering here a sequence of aggregations. The possible sequences are as numerous as the sequences considered above without the context of the changing and unchanged conditions.

Among the various sequences, the protocol CDR also can be considered in the context of changing multiple conditions. For instance, the CDR sequences also can be expected in trials designed for analysis of the disorders with a potentially repeated outcome, for instance, myocardial infarction, or seizure, etc. In such a case, with treatment assumed constant, conditions can change, which potentially can be related to the timing of the occurrence of the outcome.

The series

$$Ag_1 \begin{vmatrix} I_{2,15,18} \\ V_{(T_Y_-),F_-,O_+,V_+} \\ t_{1,2,3} \end{vmatrix}; Ag_2 \begin{vmatrix} I_{2,15,18} \\ V_{(T_Y_-),F_+,O_-,V_-} \\ t_{4,5} \end{vmatrix}; Ag_3 \begin{vmatrix} I_{2,15,18} \\ V_{(T_Y_-),F_-,O_+,V_+} \\ t_{6,7} \end{vmatrix}; Ag_4 \begin{vmatrix} I_{2,15,18} \\ V_{(T_Y_-),F_+,O_-,V_-} \\ t_{8,9,10} \end{vmatrix};$$

can then be considered in this context from a logical perspective, and the inference can be supported quantitatively in the context of available data.

The changes in treatment and outcome during this period, and co-occurrence of the events of treatment and outcome become the object of logical analysis, which can then be complemented with quantitative arguments.

Steps of exploration

The first step of the exploration is an intuitive and logical assessment of the association between the change in the condition and in the outcome. The next step is the estimation of the probability of the random association between the change of the condition (in the case of a steady treatment) and the occurrence of the outcome. Deriving from the assumption that the probability of the occurrence of the outcome of interest at the cycle t_i is satisfactory described, for instance, by the exponential distribution

$$Pr(t) = ae^{bt};$$

one can assess risk (chance) for the outcome of interest during a segment of time (or a combination of separate cycles) during the known sequence or during the entire trial.

The third step is the identification of the properties determining the treatment response (St_+, Sp_+, St_-, Sp_-) in sequential cycles. The complexity of the logical analysis depends on the structure of the analyzed sequence and on number and the character of the accepted assumptions. Algorithmic assistance might be required including big data approaches.

The quantitative assessment of the sequence of aggregations derives from the assessment of a single aggregation. The latter, i.e., computing the probability of random occurrence of the set of elements comprising the aggregation, is described in [*].

Summary

A randomized controlled trial is a statistical experiment to examine a hypothesis, typically about the higher efficacy of one treatment over placebo. The process of generating the hypothesis involves analysis of numerous theoretical, observational, and experimental data.

The exploration of the treatment response is the process of generation of the data driven hypotheses. It stems from the assumptions that

I. *Each subject is a unique individual.*

II. *Two or more co-occurring events are related unless the contrary is proven.*

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

These are the assumptions (not empirical assertions) designating a starting position of the exploratory analysis – the same way as the assumption about the by chance co-occurrence of the events designates a starting position (null hypothesis) of examining a statistical hypothesis.

Treatment is a process. Even when our knowledge is limited to information that "treatment was effective (or not effective) in the patient A," it implies that 1) there was a baseline status of the patient, 2) the treatment was applied; 3) the outcome has developed. These events unfolded during some time required for the development of the outcome, or for ascertaining that the outcome has not developed.

Treatment is an experiment. In our example, there was a baseline status; there was a hypothesis that the treatment Tx can change the status in a desirable way. The treatment has been applied. The status after the treatment is compared with the status before the treatment. There is an outcome – positive or negative. The outcome depends on the ability of the treatment to exercise a desirable effect – not on everyone (a panacea does not exist), but rather on the patients capable to positively react to the treatment. To determine whether these two factors were or were not in agreement is a matter of inference.

Changes in conditions and outcome can be basis for inferences. At the next cycle, the experiment can be repeated, and it can be re-iterated a number of times. The object of the experiment remains generally the same, but some changes in the internal and/or external properties (conditions) can change over time. A reaction of the patient to the treatment under changing conditions can be a basis for logical and quantitative assessment of the properties determining a reaction of the individual to the treatment.

Time limitation for inferences on treatment effect. Some changes in conditions are either not known to the investigator or cannot be accounted for. Also, sometime several conditions change simultaneously, and is difficult (if possible) to assess an individual contribution of each of them. An amount of the unaccounted and unknown, potentially influential factors entering to the play is growing over time. There is no reason to believe that these factors are distributed evenly across the cohorts and by the trial time. Thus, the inferences regarding *the treatment effect* (as it is defined in clinical trials) can be reasonably considered valid during the time period in which it is acceptable to assume – based on experience, intuition and common sense - that the influence of the unaccounted changes is not substantial. There is no formal way to determine the cut-off point, because of a fundamental difficulty in assessing the temporal changes. Substantial attrition, which is a common factor for virtually any long-term trial, makes comparing sequential segments of time by risk for the negative outcome practically impossible.

Monitoring treatment response. In contrast, a step by step, cycle by cycle exploration the factors (conditions) affecting the timing of the occurrence of the outcome allows for monitoring the treatment response. The timing of these events and changes in the individuals does not depend on the assessment of the entire population and on entire duration of the trial (as it is in the case of the treatment effect). Therefore, these data can be analyzed individually and be integrated at any moment/segment of time. Analysis should be focused on the dynamics of the categories of the treatment response.

Apparently, the treatment process should be monitored for both positions. The dynamic of risk should be monitored from the position of the treatment effect, and it should be complemented with monitoring on the individual treatment response under changing conditions. With all limitations of each approach, these explorations can provide an important information potentially helpful in efforts to influence the survival time.

***

Apparently, analysis of the treatment response requires the development of the conceptual and mathematical apparatus not less voluminous and thorough than analysis of the treatment effect. Our task here was limited only to designating the area and indicating the direction in which, in our opinion, the further development should go.

Literature

[i] 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

[ii] Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Velentgas P, Dreyer NA, Nourjah P, et al., editors. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Jan.

[iii] Tukey, John W. (1977). Exploratory Data Analysis. Pearson. ISBN 978-0201076165.

[iv] Behrens-Principles and Procedures of Exploratory Data Analysis-American Psychological Association-1997

[v] Peduzzi P., Henderson W., Hartigan P., Lavori P. Analysis of Randomized Controlled Trials. Epidemiol. Rev., Vol. 24, No. 1, 2002, pp. 26-38.

[vi] Verbeke, G. and Molenberghs, G. (2000) Linear Mixed Models for Longitudinal Data, Springer.

[vii] Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004) Applied Longitudinal Analysis, Wiley.

[viii] Weiss, R.E. (2005) Modeling Longitudinal Data, Springer

[ix] Diggle, P.J., Heagerty, P., Liang, K.-Y., and Zeger, S.L. (2002) Analysis of Longitudinal Data, 2nd Edition, Oxford University Press.

[x] Davidian, M. and Giltinan, D.M. (1995) Nonlinear Models for Repeated Measurement Data, Chapman and Hall/CRC Press.

[xi] Vonesh, E.F. and Chinchilli, V.M. (1997) Linear and Nonlinear Models for the Analysis of Repeated Measurements, Marcel Dekker.

[xii] Tsay, RS. Multivariate Time Series Analysis: with R and Financial Applications. Wiley, 2014

[xiii] Peterson, J.C., Pirraglia, P.A., Wells, M.T. et al. Attrition in longitudinal randomized controlled trials: home visits make a difference. BMC Med Res Methodol 12, 178 (2012). https://doi.org/10.1186/1471-2288-12-178

[xiv] Chung-Chou H. Chang, Hsiao-Ching Yang, Gong Tang, Ganguli M. Minimizing attrition bias: a longitudinal study of depressive symptoms in an elderly cohort. Int Psychogeriatr. 2009 Oct; 21(5): 869–878.

[xv] Gustavson K, von Soest , Karevold E, Røysamb E. Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. BMC Public Health. 2012; 12: 918.

[xvi] West SG, Duan N, Pequegnat W, Gaist P, Jarlais Don CD, Holtgrave D, Szapocznik J, Fishbein M, Rapkin B, Clatts M, Mullen PD. Alternatives to the randomized controlled trial. Am J Public Health. 2008 Aug; 98(8):1359-66.

[xvii] Fisher, RA. (1971) [1935]. The Design of Experiments. Hafner Publishing Copm[any. New York.

[xviii] Streptomycin in Tuberculosis Trials Committee (1948). Streptomycin treatment of pulmonary tuberculosis. A Medical Research Council investigation. Br Med J. 2 (4582): 769–82
[xix] Montori, VM, Guyatt GH. Intention-to-treat principle. CMAJ. 2001;165(10):1339–41.
[xx] Gupta SK. Intention-to-treat concept: A review. Perspect Clin Res. 2011;2(3):109–112.]
[xxi] C. Eric McCoy. Understanding the Intention-to-treat Principle in Randomized Controlled Trials. West J Emerg Med. 2017 Oct; 18(6): 1075–1078.
[xxii] Plesser, HE. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. Front Neuroinform. 2018 Jan 18;11:76.
[xxiii] Beck JR, Pauker S. The Markov process in medical prognosis. *Med Decis Making.* 1983;3:419-458.
[xxiv]Sonnenberg FA, Beck Jr. Markov models in medical decision making: a practical guide. *Health Econ*. 1983;13:322-338.
[xxv] Ganz JB1, Ayres KM. Methodological standards in single-case experimental design: Raising the bar. Res Dev Disabil. 2018 Aug;79:3-9.
[xxvi] Smith JD. Single-case experimental designs: a systematic review of published research and current standards. Psychol Methods. 2012 Dec;17(4):510-50
[xxvii] Tate RL, McDonald S, Perdices M, Togher L, Schultz R, Savage S. Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the Single-Case Experimental Design (SCED) Scale. Neuropsychol Rehabil. 2008 Aug;18(4):385-401
[xxviii] Manolov R, Gast DL, Perdices M, Evans JJ. Single-case experimental designs: reflections on conduct and analysis. Neuropsychol Rehabil. 2014;24(3-4):634-60.
[xxix] De Young KP, Bottera AR. A summary of reporting guidelines and evaluation domains for using single-case experimental designs and recommendations for the study of eating disorders. Int J Eat Disord. 2018 May 30.
[xxx] Jamshidi L, Heyvaert M, Declercq L, Fernández-Castilla B, Ferron JM, Moeyaert M, Beretvas SN, Onghena P, Van den Noortgate W. Methodological quality of meta-analyses of single-case experimental studies. Res Dev Disabil. 2018 Aug;79:97-115.
[xxxi] Mises, Richard von. Probability, Statistics and Truth. 2nd Revised ed. New York: Dover, 1981 [1957] p. 11
[xxxii]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)
[xxxiii] Rubin, D. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". J. Educ. Psychol. 66 (5): 688–701 [p. 689]. doi:10.1037/h0037350.
[xxxiv] 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.