

Statistical methods to support diagnosing rare diseases

Guenter F. Pilz (✉ Guenter.Pilz@jku.at)

Johannes Kepler Universitat Linz <https://orcid.org/0000-0001-7099-2963>

Frank Weber

German Air Force Center of Aerospace medicine

Werner G. Mueller

Johannes Kepler Universitat Linz

Juergen R. Schaefer

Philipps-Universitat Marburg

Research

Keywords: Diagnosing designs, rare diseases, statistics, regression, block designs

Posted Date: February 13th, 2020

DOI: <https://doi.org/10.21203/rs.2.23479/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Far too often, one meets patients who went for years or even decades from doctor to doctor, without getting a valid diagnosis. This brings pain to millions of patients and their families, not to speak of the enormous costs. Often patients do not know well enough which factors (or combinations thereof) trigger their problems. **Results** If conventional methods fail, we propose the use of statistics and algebra to give doctors much more precise inputs from patients. We propose statistical regression for independent triggering factors for medical problems, and “balanced incomplete block designs” for non-independent factors. These methods might change a useless statement like “I feel very tired after meals” to a much more valuable “After a meal with many carbohydrates, but few vegetables, a moderate physical activity will usually force me into a wheel-chair”. In order to show that these methods do work, we briefly describe a real case in which these methods helped to solve a 60 year old problem in a patient, and give some more examples where these methods might be very useful. **Discussion** In this paper, we present a way of getting medical diagnoses when the methods in medicine are insufficient, too time consuming, or very expensive. By asking the patient to conduct tests (often at home) according to a very well-prepared schedule, statistics can use regression analysis to identify the factors (or combinations thereof) which trigger or worsen the patient’s problems. This (very inexpensive) analysis can often give the patient’s doctor(s) a very good and precise input for their diagnosis. **Conclusions** While regression is used in clinical medicine, it seems to be widely unknown among diagnosing doctors. In finding the reason(s) of rare diseases, doctors face very tough problems. So they deserve to know all tools which could offer some help. This can save the health systems much money, and the patients also a lot of pain.

Introduction

In medicine, a diagnosis of a problem of a patient is usually generated by medical knowledge and experience, often using results of labs and other tests. The success rate for correct diagnoses is high if the inputs tell a clear message, like in case of a broken bone. In other cases, however, like for heavy headache, extreme weakness, etc., the situation is not so simple, and might require a much deeper search. Often enough, a satisfactory diagnosis is not found.

In fact, the number of patients without a valid and correct diagnosis is frighteningly high in areas, where a diagnosis is non-trivial, e.g., in cases of rare diseases, or if decisive parameters are hardly measurable (like stress). A center for rare diseases in Germany presently has a backlog of more than 9,500 desperate requests; a quick and informal search among an organized group of patients for a special rare disease revealed that more than 85% of them did not have a valid diagnosis.

In Europe, a rare disease is defined by a prevalence of ≤ 1 to 2.000 inhabitants. However, due to the fact that there are an estimated number of more than 8.000 different rare diseases, the total number of patients with rare diseases is rather high (at about 5% of the European population). So one might estimate that more than 300 million people on earth suffer from a rare disease. Even much more patients

are afflicted with “incomplete” diagnoses due to hardly measurable or subjective (but wrong) inputs of patients.

This dramatic situation might be improved by an increasingly expensive medical machinery, but also by the use of statistical regression, which tells patients (and their doctors) much more about their triggering factors than they are aware of. Surprisingly little was done so far in this direction, except in clinical research. A rather new book gives a first systematic account on regression in medicine, but with no emphasis on diagnosing, and block designs for dependent factors are not covered there at all.

The situation is intensified by the fact that a small change in the input might result in a big change of the output (= diagnosis), no matter whether the search for the diagnosis is computer-aided or not. In mathematical language, the output does not depend continuously on the input. Hence, in crucial situations, it might be highly desirable to improve the quality of the inputs. The statistical approach usually does need the assistance of a statistician (in the near future maybe simplified by an App) and the cooperation of the patient, but nevertheless it is far less expensive than a complicated medical machinery. Or a wrong diagnosis.

Results And Methods

STATISTICAL METHODS, I: REGRESSION ANALYSIS

The role of statistics in life sciences is ubiquitous, simply think of the millions of statistical tests for the efficiency of medications or medical treatments, or trials on (sometimes many thousands of) patients (see, e.g., Cleophas et al. [1]). Less common is the use of statistics to identify one or more of a large number of factors which might trigger pain or discomfort in a single patient („Precision Medicine“); an account was only given recently by Cleophas and colleagues [2]). And very rarely, a search is done to find positive or negative synergy effects (interactions) between these factors which go far beyond a mere addition of these factors. The reason for that is, of course, the huge number of possible combinations of two or more factors. For the sake of the patients, however, the number of tests should be as small as possible. We present a solution to this dilemma. The identification of these „suspicious“ factors can be very valuable in getting a diagnosis when this turns out to be difficult.

One of the best tools to discover the reasons for medical problems in a patient is the well-known method of (statistical) regression. Regression analysis is a very powerful tool. In medical statistics, regression analysis is usually used to analyze large samples, e.g., stroke risk as a function of age, hypertension, smoking habits etc. Concerning diagnosis, medical statistics considers terms of sensitivity, specificity, pre- and posttest probabilities etc. Here we show that statistical regression can be very useful in the diagnosis of an individual case by detecting unknown connections between a number of „suspicious“ factors. For that, the combination doctor and patient must identify these factors which might trigger a certain event. The patient has to undergo several tests in which (s)he is always exposed to some of the suspicious factors. The number n of factors determines roughly the number of required tests; a common rule of thumb says that n factors require about n^2 tests. So, if the number of possible factors is large, a

very large number of tests are required. Hence a careful selection of possible factors is essential. The tests should follow a suitable test plan (experimental design) so that all factors are tested an approximately equal number of times. In statistics, this is usually called a “screening experiment”.

Of course, this method is much more conspicuous than a usual diagnosis, and so it will only be used in cases where conventional methods have failed. But doctor’s waiting rooms often contain patients who have run through an unsuccessful series of many tests generating numerous diagnoses. This can be very frustrating and sometimes also dangerous for them and usually takes much longer than the method we are demonstrating here. Especially the diagnostic path in patients with rare diseases may be troublesome. It is important to know that the number of rare diseases is much higher than people usually think. Sometimes, the patient can undertake the tests and measure the results by himself.

As an example, take a patient with unknown factors which trigger an allergy, where the usual diagnostic measures did not yield a satisfactory result. Suppose that the patient and the doctor suspect that n more factors x_1, x_2, \dots, x_n might explain the allergy, e.g.,

- x_1 = exhaust air of the vacuum cleaner (measured in minutes of exposure)
- x_2 = intake of certain candies (measured in pieces), ...

and so on. Then a test plan might ask the patient for an exposure to x_1, x_4 and x_9 , and to rank the degree y_1 of allergy on – say – a scale of ten degrees of severeness. The second test might involve x_2, x_4 and x_6 , with a result of y_2 , and so on.

Then well-known statistical algorithms (see, e.g., Morris [3]) will yield a „formula“ of the type

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (*)$$

where β_0 is a constant (the „intercept“) and β_i estimates the influence of x_i to the overall allergy level. Usually, one also determines confidence intervals $[\beta_i - c_i, \beta_i + c_i]$ so that they cover the β_i with a confidence level of – say - 95%. If this interval covers 0, like for example in $[-0.4, 0.7]$, one usually reacts in the way that the influence of the corresponding factor x_i is to be doubted (not statistically significant) and so x_i is eliminated from the list of interesting factors. This usually happens for many factors, such that eventually a small list of suspicious factors remains, and the doctors will pay their attention to these few factors. This reduction is often essential, because it makes a huge difference whether 100 or 3 factors have to be medically investigated. The patient might already be dead when the doctors come to explore factor # 50...

STATISTICAL METHODS; II: EXPERIMENTAL DESIGNS

But much more care must then be taken to the design of the experiments. It might be that x_i and x_j are never (or only once) tested together, and so no clarification of a synergy is possible.

The fairest way would be to test every x_i the same number of times, and also to test every pair x_i and x_j the same number of times. A new problem now comes from the quick rise of binomial coefficients. With 5 factors, we have 10 possible pairs, but with 20 factors, we already have 190 pairs. So a clever trick is needed: we utilize some particular experimental designs:

A *Balanced Incomplete Block Design (BIB-design)*, see, e.g., Lidl & Pilz [4], consists of a set $P = \{p_1, p_2, \dots, p_v\}$ of v „points“ and a collection B of b subsets B_1, B_2, \dots, B_b of P (called „blocks“), such that

- Each point in P belongs to the same number r of blocks
- Each B_j has the same number k of elements
- Each pair p_i, p_j of points belongs to the same number λ of blocks.

The pair (P, B) is then called a (v, b, r, k, λ) -design. The design is complete if B is just the collection of all k -element subsets, otherwise incomplete.

For an experiment like the one above (concerning allergies), a BIB-design can be turned into an experimental design as follows.

- The points are the factors (e.g., possible triggers for an allergy);
- Every block lists the factors which will be tested simultaneously in a test.

So a (v, b, r, k, λ) -design will test v suspected triggering factors; each test requires k suspected factor (at the same time), and one will need b tests. Number (i) above assures that every possible triggering factor will be tested the same number (namely r) of times, and every pair of possible factors will be tested together in exactly λ tests. So a BIB-design gives an experiment which is “fair” both to the factors and the tests.

It is not trivial at all to get such a design. Constructions usually come from areas „far away“, like from finite geometries or abstract algebra (structures like groups or near-rings).

Example: From mathematical considerations (see Ke-Pilz [5]) we might get the following design (which comes “out of the blue” now, but we will not give the long mathematical derivations):

$P = \{1, 2, 3, 4, 5, 6, 7\}$ and B consists of the 14 collections

$B_1 = \{2, 4, 5\}, B_2 = \{1, 3, 7\}, B_3 = \{1, 2, 6\}, B_4 = \{1, 5, 7\}, B_5 = \{1, 3, 4\}, B_6 = \{2, 3, 7\}, B_7 = \{4, 5, 7\},$

$B_8 = \{1, 2, 4\}, B_9 = \{2, 6, 7\}, B_{10} = \{2, 3, 5\}, B_{11} = \{3, 4, 6\}, B_{12} = \{3, 5, 6\}, B_{13} = \{1, 5, 6\}, B_{14} = \{4, 6, 7\}$

This gives a $(7, 14, 6, 3, 2)$ -design. Suppose we have 7 factors x_1, x_2, \dots, x_7 . For the first test, we try the factors x_2, x_4 and x_5 , since $B_1 = \{2, 4, 5\}$, and so on:

Fact.\Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_1	●	●	●	●	●								●		
x_2	●		●			●		●	●	●					
x_3		●			●	●				●	●	●			
x_4	●				●		●	●			●			●	
x_5				●			●	●		●		●	●		
x_6			●						●		●	●	●	●	
x_7		●		●		●	●		●					●	
Results	49	-2	-28	3	54	-1	51	98	-31	69	18	-35	-25	22	3

Table 1: An experimental design for testing 7 factors, each of them 6 times.

So we have $v=14$ tests, plus a „zero test“ for „technical reasons“ (the information matrix would otherwise not have full rank). One sees:

- Every test (except #15) involves $b=3$ factors (3 dots in every column)
- Each factor is tested in $r=6$ tests (6 dots in each row)
- Each pair of factors is tested together in $\lambda=2$

Observe that we have tested the 21 possible pairs x_i and x_j twice in only 15 rather than $2 \cdot 21 = 42$ tests! This „magic reduction“ to only $\frac{1}{3}$ of the tests might be attributed to the fact that in each test three synergies are considered (in terms of experimental designs: we allow certain interactions to be aliased).

In the last row, we have supplemented some (fictitious) results of the tests. Linear regression gives the best estimates according to (*) in the section „Statistical Methods, I“ as

$$y = 3 + 51x_4 + 19x_5 - 41x_6 \quad \dots \text{Model 1}$$

If one also uses interaction terms („synergies“), one gets instead

$$y = 2 + 47x_4 - 31x_6 + 58x_2x_5 \quad \dots \text{Model 2}$$

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
real	49	-2	-28	3	54	-1	51	98	-31	69	18	-35	-25	22	3
Mod.1	54	3	-38	19	54	3	73	73	-38	22	13	-19	-19	13	3
Mod.2	49	-2	-29	2	49	2	49	107	-29	60	18	-29	-29	18	2

Table 2: Comparing two test results with reality.

One easily sees that Model 2 describes „the reality“ considerably better than Model 1.

Let us remark that estimating possible product terms creates a problem because of the very small number of tests. We first look at the „main effects“ x_i , remove the irrelevant ones, and always add one if the $x_i x_j$ to check which of them seemed to be statistically relevant. These are then added to the relevant main effects (thereby following the so-called hereditary principle). The final result might depend on these choices and their ordering, but doing the calculations several times with different choices and orderings may yield a robust choice. And for patients it does make a big difference if they have to undergo many more tests. Of course, repetitions of these 15 tests would also give much better results.

STATISTIC WORKS...

This describes the search for the reason for periodic paralysis in a patient which started in a patient at the age of 15 years and was ongoing for more than 50 years. Despite the consultation of more than 120 doctors and hospitals over decades, no reason was found. Since the patient reported an association of the paralysis attacks with a low intake of carbohydrates, the doctors first thought that a low blood sugar level might be the reason. This did not turn out to be true in this form, so doctors specialized for rare diseases considered the possibility of a periodic paralysis due to a defect in ion channels. But the ion-levels within the blood were normal during the attacks, which is typical in patients suffering on normokalemic periodic paralysis. Thus it was tricky to identify the intracellular mechanism, namely too high or too low sodium or potassium, respectively. The doctors mentioned that it might take a long time to check all these channels.

The patient (a mathematician) tried to support the research team by speeding up the search process. He noted each of his meals and how much calcium, sodium, potassium, protein, etc. he had eaten. He checked his body strength by pressing a bathroom scale, right after the meals, and again one hour later. The differences came up to about ± 10 kilograms. So he first used statistics (model 1) above to identify the few most probable intake components which explain the differences, using the software package *Mathematica*[®]. After a period of about 5 weeks of carefully documented eating, he had the result

$$y = y(p,s) = -0.5 - 0.0048p + 0.0085s$$

with the interpretation that 1 hour after the intake of p mg of potassium and s mg of sodium, the patient could press the scale (on average) $y(p,s)$ kg harder. The signs of the coefficients for p and s were significant on the 99% level, which means that potassium hurts the patient, while sodium helps him. The exact values of these coefficients are not so important, except that the patient now knows in advance, that, for examples, a typical burger will strengthen him by approximately $y(420,1000) = 6.0$ kg, while 100 g of banana will weaken him by $y(390,1) = -2.4$ kg.

Then the patient ran another test, this time with a BIB-design similar to the example above to check, if possibly a combination of other substances might overthrow this result. But no relevant combination was found, so the above result was accepted. In this case, however, no test strictly according to the BIB-plan

was possible, since no food contains only potassium, calcium, and sodium, and no other substances. We will return to this point later.

After this finding, the doctors knew that they had to search for a defect in a potassium channel and, more importantly, that lowering potassium should be beneficial in this special patient. The subsequent analysis of various ion channels revealed a so far unreported defect of the promotor of the KCNJ18 channel. Interestingly enough, this channel was considered so far as a less likely candidate for paralysis and was no target candidate in well-established screening panels (see Kuhn and colleague [6]). The study and results were reported recently elsewhere in Soufi and colleagues [7]. Note that without the doctor's hint to consider ion channels, the patient would never have conducted these experiments. And without medical competence, the results of the experiments could not have been properly interpreted. So this case might be considered as a fine and successful interplay between medicine, statistics, and abstract algebra.

MORE EXAMPLES

1. **Food-dependent exercise-induced anaphylaxis:** The contact with some allergens might be harmless, physical exercise can help a lot, while the combination can be disastrous. So one factor is neutral for the patient, the other one positive, but the combination is really negative! See Romano et al. [8].
2. **Phototoxic Dermatitis:** Smaller amounts of UV-doses can usually be tolerated without any reactions; the same applies to substances which enhance photosensitivity, like Phenothiazine or Furosemid. But if they are combined, they can create skin reactions similar to heavy sunburn effects.
3. **Hyperkalemic Periodic Paralysis:** In a mild form, this disease can usually be tolerated, but in combination with a pathogenic gene mutation, it can create severe paralysis.
4. **Stomach Problems:** A patient complains about stomach pains after some meals. His doctor suspects that seafood might be a reason, but this can hardly explain the pains. And he can exclude a large number of food components which do not hurt the patient. But 15 "suspicious" factors remain. The following might be a typical progression of the statistical investigation. A simple regression test like in "Statistical Methods, I" excludes quickly 8 of them. For the remaining 7 components, this test does not give satisfactory results. So we might use the test in "Statistical Methods, II". Suppose that the remaining 7 factors are sugar (=S), apples (=A), lactose (=L), walnut (=W), pepper (=P), crabs (=C), and mustard (=M). So, according to the experimental design in "Statistical Methods, II", the first test would be a meal with S, A, and W. Then a statistician quickly finds out that C does hurt a bit (as single factor), but the combination A & P is the main reason for the pains, while A and P alone do not really hurt.

Discussion

Many statisticians might be unhappy with several parts of the statistics used above. Metric and ranked data were mixed, the number of tests (especially in model 2) can be dangerously low, the BIB-plan above should be filled with 0-1-data and not with real numbers, the ranking of pain by patients is highly subjective, and so on.

Also, almost all statistics on patients lack an important feature: they are not reproducible, like statistics in technical sciences. See, for instance, the brilliant article by Homes [9]. Our approach using artificial intelligence should only assist the diagnosis and not replace doctors.

However, medicine is not pure natural science and it might be better to use a partly „dirty“ statistics than to do nothing. And – most of all – the statistical results are NOT the diagnosis, but just a suggestion (or suggestions) giving process for an appropriate medical investigation. Still the medical part of getting the diagnosis is by far the most important one. But – as can be seen in our case report - statistics can be very helpful.

In fact, the statistician plays an important role: (s)he has to identify, together with the doctors and the patient, which of the thousands of possible factors might be relevant. As mentioned above, the number of required tests grows exponentially with the number of possible factors. So one has to keep the number of “suspicious” factors low. But the lower this number is, the risk of missing the relevant factors grows. Hence the statistician must be good in “model building”.

Let us note that the statistical models mentioned above are also useful (and have been employed) in many other areas, cf. the survey on spatial applications by Mueller [10]). In agriculture, the factors x_i might be fertilizers, irrigation, etc., in paint manufacturing, they might be additives against weather attacks, and so on. The corresponding author of this paper also coaches a hammer thrower in Munich, where the x_i are technique elements and where Model 2 is especially helpful.

Conclusions

We present a seemingly unknown and inexpensive tool for obtaining valid diagnoses in difficult cases, especially for rare diseases. The basic idea is to employ statistics (mainly regression analysis) in order to get much more precise inputs from patients. They often do not know exactly which substances, actions, and circumstances (or combinations thereof) trigger their problems. Statistics can often easily explain which of these factors contribute to the worsening of the patient’s problems. We found that this precise information often leads the way to the correct diagnoses.

Typically, the patients can gather the necessary data by themselves, in measuring parameters like blood pressure, intake of food and drugs, body strength, degrees of the pains, and so on. We plan to develop an app to facilitate the collection for the patients. Since more and more medical information will be gathered by so-called “wearable sensors”, we can expect a rapid increase of data. These data must be well-organized to be useful for the physicians.

Of course, patients must be able to measure an improvement or worsening of their condition (= the output), and at least some of the measured input parameters (patients do not have to know, which of them) must have an influence on the output. If the pain stays at the same level all the time, statistics is of no use.

There is a vast literature on the use of statistics in medicine, where the data are collected by patients (see, e.g., Saunders and colleagues [11]), but they have a completely different approach.

So we believe that the use of statistics can help physicians to resolve cases of rare diseases, when the usual methods seem to fail. This, for sure, will greatly relieve many deeply unhappy people among the large number of undiagnosed patients.

Declarations

Ethical approval and consents: Not applicable

Consent for publication: Not applicable

Availability of data and materials: There is no restriction on data. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests: The authors declare no competing or conflicts of interests.

Funding: Mueller's research was partially supported by project grants LIT-2017-4-SEE-001 funded by the Upper Austrian Government, and Austrian Science Fund (FWF): "Statistical. J.R. Schaefer's has received research support from the Dr. Reinfried Pohl Foundation. Both funding institutions helped in doing basic research.

Authors' contributions: GFP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing the original draft.

FW: Investigation, Methodology, Resources, Writing and reviewing.

WGM: Formal analysis, Investigation, Methodology, Validation, Writing and reviewing.

JRS: Conceptualization, Data curation, Investigation, Methodology, Resources, Writing and reviewing.

Acknowledgements: Not applicable

References

1. Cleophas, T.J., Zwinderman, A.H., Cleophas, T.F., Cleophas, E.P. Statistics applied to clinical trials. Springer Netherlands; 2009.
2. Cleophas, T.J., Zwinderman, S.H. Regression analysis in medical research. Springer International; 2018.
3. Morris, M.D. Design of experiments. CRC Press, Chapman & Hall, Boca Raton; 2011.
4. Lidl R, Pilz, G.F. Applied abstract algebra, 2nd, Springer New York (Undergraduate Texts in Mathematics), 1997.

5. Ke, WF, Pilz, GF. Abstract algebra in statistics. *J. of Algebraic Statistics* 2010;1: 6-12
6. Kuhn M, Jurkatt-Rott K, Lehmann-Horn F. Rare KCNJ18 variants do not explain hypokalaemic periodic paralysis in 263 unrelated patients. *J. Neurol. Neurosurg. Psychiatry*, 87(1); 49-52; 2016. doi: [10.1136/jnnp-2014-309293](https://doi.org/10.1136/jnnp-2014-309293).
7. Soufi M, Ruppert V, Rinné S, Mueller T, Kurt B, Pilz GF, Maieron A, Dodel R, Decher N, Schaefer JR. Increased KCNJ18 promoter activity as a mechanism in atypical normokalemic periodic paralysis. *Neurology: Genetics*, 2018; 4: e274.
8. Romano A, Fonso M.D, Giufredda F, Pappa G, Artesiani MC, Viola M, Venuti A, Palmieri V, Zeppili P. Food-dependent exercise-induced anaphylaxis: clinical and laboratory findings in 54 subjects. *Int. Arch. Allergy Immunol.* 2001; 125(3): 264-272.
9. Holmes S. Statistical proof? The problem of irreproducibility. *Bulletin of the American Mathematical Society* 2018; 55: 31-55.
10. Mueller WG. *Collecting spatial data – optimum design of experiments for random fields*, Springer Verlag, Berlin-Heidelberg; 2007.
11. Saunders MJ, Wingfield T, Datta S, Montova R, Ramos E, Baldwin MR, Tovar MA, Evans BE., Gilman RH, Evans CA. A household-level score to predict the risk of tuberculosis among contacts of patients with tuberculosis: a derivation and external validation prospective cohort study. *Lancet Desinfect Dis.* 2019; doi: [1016/S1473-3099\(19\)30423-2](https://doi.org/10.1016/S1473-3099(19)30423-2)