Article

Statistical methods to support difficult diagnoses

Guenter F. Pilz[1], Frank Weber[2], Werner G. Mueller[3], Juergen R. Schaefer[4]

[1] Institute of Algebra, Johannes Kepler University, Linz, Austria (corresponding author); guenter.pilz@jku.at
[2] German Air Force Center of Aerospace Medicine, Fuerstenfeldbruck, Germany; Dr.Frank.Weber@t-online.de
[3] Institute for Applied Statistics, Johannes Kepler University, Linz, Austria; Werner.Mueller@jku.at
[4] Center for Undiagnosed and Rare Diseases (ZusE), Philipps University, Marburg, Germany; juergen.schaefer@mailer.uni-marburg.de

Abstract: 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 cannot tell precisely enough which factors (or combinations thereof) trigger their problems. If conventional methods fail, we propose the use of statistics and algebra to give doctors much more useful inputs from patients. We use statistical regression for triggering factors of medical problems, and in particular "balanced incomplete block designs" for factors detection. These methods can supply doctors with much more valuable inputs, and can also find combinations of multiple factors by incredibly few tests. In order to show that these methods do work, we briefly describe a 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. As a conclusion, while regression is used in clinical medicine, it seems to be widely unknown in diagnosing. Statistics and algebra can save the health systems much money, and the patients also a lot of pain.

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

1. 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, if there is a huge number of possible triggers, 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 had no valid diagnosis.

A rare disease is defined by a prevalence of ≤ 1 to 2.000 inhabitants (see, e.g., https://www.orpha.net/consor/cgi-bin/Education_AboutRareDiseases.php?lng=EN). However, due to the fact that there is 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 (see [2]) 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.

Here are some examples where a traditional medical search might be too slow, too complicated, or too expensive, but where mathematics and statistics can give reliable results is a very short time:

- Some (especially elderly) people often take a large number of drugs. Often enough, some of these drugs (or combinations of them) can be the reason for further severe problems. Far not all of these interactions are known well enough. Simply think that for the 1,000 most frequent drugs, there are half a million possible interactions, which also differ from patient to patient. Statistics provides an almost incredible tool to test many of these interactions at the same time, using sophisticated mathematical methods, such as block designs and matrix calculus. Please observe that it is useless to find drug interactions for a large number of patients; they have an individual character. For instance, "dizziness" is on almost all package inserts, hence useless. We describe this method in more detail in Section 3.1 below. Note that this application does not concern rare diseases, but remarkably frequent cases in treating patients!
- Reactions to the intake of food (components) can give valuable hints for the diagnosis. But one cannot expect patients to know that, for instance, they react to an imbalance of magnesium intake. In the section "Statistics Works" below, we describe a case where problems through the intake of too much potassium, but too few sodium caused severe and frequents attacks of paralysis for more than 50 years. This lead doctors to investigate a gene which was before not considered as a cause of paralysis (see [6]), and to find the defect. The solution of this case and the fact that that this

defect seems to be unique world-wide shows the power of the statistical approach. This case was described in detail in [7].

- Combinations of allergens can be tricky. [8] describes cases in which one allergen is neutral for the patient, another one positive, but the combination is a real disaster! Our methods can detect cases like this without problems.

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.

2. Materials 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.

Our statistical method, as already briefly described, is that of (statistical) regression. First, the patient and the doctor together try to find out, which parameters (“factors”, say x_1, x_2, …) might improve or worsen the patient’s situation. These must be in some way “measurable”, and they also must find a numerical indicator y which describes the patient’s situation. The “degree of stress” or “pain on a scale from 0 to 10” are OK, but parameters like “blood pressure” would be better, of course. The number of parameters should not be “too low” (danger of missing the most useful parameters), and not “too large” (resulting in a huge number of tests); usually, a size between 5 and 15 might do

the job. Of course, if the situation of the patient stays at the same level all the time, statistics is of no use.

Then a statistician selects an efficient experimental design, telling which factors (like pills or food components) the patient should try on – say – day 1, which factors should be tested on day 2, etc. The patient notes the resulting state y of his situation after each test. The "protocol" might then look like

- Day 1: I tried x_1, x_3, x_4, x_7, the result was 23
- Day 2: I tried x_2, x_3, x_5, x_9, the result was 19

And so on. Usually one wants a design in which all factors are tested an approximately equal number of times. In statistics, this is usually called a "screening experiment".

The patient sends this protocol to the statistician (or doctor). The statistician, using regression, finds those factors (or combinations thereof) which are very likely to improve the patient's situation, and the "bad" factors which worsen it. Irrelevant factors are detected automatically.

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. Up to now, the use to find a diagnosis, however, is very rare. 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.

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. Often, the patient can undertake the tests and measure the results by himself.

The method used can perhaps be seen best via 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 $x_1, x_2, \ldots, x_n$ 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

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \qquad (*)$$

where β_0 is a constant (the „intercept“) and β_i estimates the "true" influence b_i 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 b_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…

If the tests conducted do not give results which are significant enough, one should continue the tests (for instance, by repetition). Then the software will "learn" more and more about the case of investigation, which is the principal of Artificial Intelligence (AI).

STATISTICAL METHODS; II: EXPERIMENTAL DESIGNS

Up to now, we applied a known algorithm to a rather new situation. Things become less simple if the important factors cannot be accounted for additively, but interactions („synergy effects“) are possible. Then one usually adds terms like $x_i x_j$ to $x_1, x_2, \ldots, x_n$ in the analyzed model. In medicine and biology, often two substances (substance and antidote,…) work together to produce an effect. For more examples, see below.

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:

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

(i) Each point in P belongs to the same number r of blocks

(ii) Each B_i has the same number k of elements

(iii) Each pair p_i, p_j of points belongs to the same number λ of blocks.

The pair $(P, \mathcal{B})$ is then called a *(v,b,r,k,λ)-design*. The design is *complete* if $\mathcal{B}$ 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 $\mathcal{B}$ 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 $x_1, x_2, \ldots, x_7$. 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 λ=2 tests.

Observe that we have tested the 21 possible pairs x_i and x_j twice in only 15 rather than 2*21=42 tests! This „magic reduction" to only ⅓ of the tests can be attributed to the fact that in each test three synergies are considered simultaneously (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

$$y = 3 + 51x_4 + 19x_5 - 41x_6 \qquad \text{…. Model 1}$$

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

$$y = 2 + 47x_4 - 31x_6 + 58x_2x_5 \qquad \text{…. Model 2}$$

Now we can compare the actual results with the predicted ones using these two models:

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	22	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_ix_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 sharper results.

3. Results

3.1. Statistic works …

We will now use the described method in "real" situations. First, as already addressed in the introduction, we describe the search for a diagnosis in a patient. His problems started at the age of 15 years with unusual tiredness attacks, always in the late afternoon. In the following 50 years the tiredness worsened to complete paralysis attacks, in which the patient was fully conscious, but could not move any part of his body. There was no way to communicate with the world around him. 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 levels of 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

$$y = y(p,s) = -0{\cdot}5 - 0{\cdot}0048p + 0{\cdot}0085s$$

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 bananas 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 gain of function by increased expression of the inward rectifier potassium channel Kir2.6, due to highly increased promotor activity of the gene KCNJ18. Interestingly enough, this gene was considered so far as a less likely candidate for paralysis and was no target candidate in well-established screening panels (see Kuhn and colleagues [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.

3.1. More Examples

(1) **Side effects of combinations of drugs:** This was also mentioned in the introduction. We had a case of a person (age 75) who developed a strong and permanent dizziness which did not allow him to drive a car any more. He took 10 types of drugs per day, and we added the consumption of a standardized amount of alcohol as "drug # 11". So we used the following (11,22,5,10,4)- block design like in Table 1, and added another "test", this time the usual drug consumption of the patient:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
●	●			●	●					●	●					●	●	●	●			●
			●		●	●		●		●		●			●		●	●		●		●
			●	●			●	●		●			●	●				●	●		●	●
●		●	●					●	●		●	●	●				●		●			●
	●	●								●		●	●		●	●			●	●	●	●
	●	●	●	●	●	●	●				●	●									●	●
	●					●			●		●		●	●			●	●		●	●	●
●		●				●	●	●		●	●			●		●				●		●
●	●		●		●		●		●					●	●				●	●		●
●				●		●	●		●			●	●		●	●		●				●

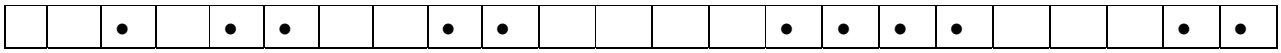

Table 3: *A* (11,22,5,10,4)- block design

So in the 1st test we gave drugs no. 1, 4, 7, 8, 9, and so on. If we exclude "test" 23, every drug was tested 10 times, and each pair of drugs was tested 4 times together. If one used single tests, it would have needed 10*11+4*55=330 tests. If the physicians (and not the statistician!) say that each test needs 1, 2, or 3 days, this program needs 22, 33, or 44 weeks. With single tests, this would be almost 1, 2, or 3 years! No patient would agree.

Let us mention that such a design greatly reduces placebo- and nocebo-effects (which are typical and critical for tests with single medications), since the patients will be "confused" by the relatively large numbers of drugs prescribed / not prescribed per day.

Here we found, by the way, that we could exclude (on the 95% level) any side effect coming from the drugs. This is much more than to say "We did not find any side effect".

(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 Dermatosis:** We usually tolerate sunlight at a usual dose very well since it is essential for our survival. Frequently used medications such as certain antibiotics, non-steroidal anti-inflammatory drugs (NSAIDs), diuretic and anti-arrhythmic drugs have usually no direct side effects at the skin. However, these drugs are known to enhance photosensitivity. In combination with usually well-tolerated sunlight, these drugs can create severe sunburn like skin reactions (see [9] and [10]).

(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". Sup-

pose 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.

4. 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.

An important point concerns the kinds of "dependences" involved. Dependent factors for a regression can be treated using BIB-designs, as we have seen. The tests suggested are by no means independent of the particular patient (see the next paragraph), as we are making no patient-to-patient comparisons. We do not aim at general results when they do not exist. Also, there might be a time-dependence between the tests. For mastering that, physicians (and not statisticians) have to decide on the optimal time distance between the tests, to guarantee independent observations as we assumed. Should one want to relax those assumptions one might have to employ more complex design strategies, such as, e.g., given in Kiefer & Wynn, [11]. Also, almost all statistical investigations on patients lack an important feature: they are not reproducible, like statistics in technical sciences. See, for instance, the brilliant article by Homes [12].

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 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 hundreds of possible factors might be relevant. A careful selection is necessary. 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 [13]). In agriculture, the factors x_i might be fertilizers, irrigation, etc., in paint manufacturing, they might be additives against weather attacks, and so on.

5. Conclusion

We presented a seemingly unknown and inexpensive tool for obtaining valid diagnoses in difficult cases, especially for rare diseases, but also for every-day-problems. The basic idea is to employ statistical regression 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 diagnosis.

Typically, the patients can gather the necessary data by themselves, in measuring parameters like blood pressure, intake of food and drugs, physical 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 be so-called "wearable sensors", we can expect a rapid increase of data. These data must be well-organized to be useful for the physicians.

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 [14]), but they have a completely different approach.

So we believe that the use of statistics can help physicians to resolve difficult cases, for instance in 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.

6. List of Tables

Table 1: *An experimental design for testing 7 factors, each of them 6 times*
Table 2: *Comparing two test results with reality*
Table 3: *A (11,22,5,10,4)- block design*

Author's 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.

All authors have read and agreed to the submitted version of the manuscript.

Funding: W. Mueller's research was partially supported by project grant LIT-2017-4-SEE-001 funded by the Upper Austrian Government, and the Austrian Science Fund (FWF), grant I 3903-N32. J.R. Schaefer's has received research support from

the Dr. Reinfried Pohl Foundation. Both funding institutions helped in doing basic research.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: 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.

Acknowledgments: >The authors want to thank the reviewers for their most useful comments which substantially improved this paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Cleophas, TJ., Zwinderman, AH, Cleophas, TF, Cleophas, EP. *Statistics applied to clinical trials*; Springer Netherlands; 2009.
2. Cleophas, TJ, Zwinderman, SH. *Regression analysis in medical research*; Springer International; 2018.
3. Morris, MD. *Design of experiments;* CRC Press, Chapman & Hall, Boca Raton; 2011.
4. Lidl R, Pilz, GF. *Applied abstract algebra*, 2nd ed.; Springer New York (Undergraduate Texts in Mathematics), 1997.
5. Ke, WF, Pilz, GF. Abstract algebra in statistics. *J. of Algebraic Statistics* **201,** *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*, **2016**, *87(1)*, 49-52;. doi: 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. Gould JW, Mercurio MG, Elmets CA: Cutaneous photosensitivity diseases induced by exogenous agents. J Am Acad Dermatol. 1995; 33(4):551
10. Hofmann GA, Weber B. Drug-induced photosensitivity: culprit drugs, potential mechanisms and clinical consequences. J Dtsch Dermatol Ges. 2021 Jan;19(1):19-29. doi: 10.1111/ddg.14314. PMID: 33491908; PMCID: PMC7898394.

11. Kiefer J, Wynn HP. Optimum Balanced Block and Latin Square Designs for Correlated Observations. *Ann. Statist.* 9 (4) 737 - 757, **1981**. doi: 10.1214/aos/1176345515

12. Holmes S. Statistical proof? The problem of irreproducibility**.** *Bulletin of the American Mathematical Society*; **2018**; *5*, 31-55.
13. Mueller WG. *Collecting spatial data – optimum design of experiments for random fields*, Springer Verlag, Berlin-Heidelberg; 2007.
14. 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: 10.1016/S1473-3099(19)30423-2

Affiliations:

(1) Dr. phil. Dr.h.c. Guenter F. Pilz, Professor emer. at the Institute of Algebra (Applied Algebra), Johannes Kepler University, Linz, Austria (corresponding author); guenter.pilz@jku.at

(2) Dr. Med. Frank Weber, Priv. Dozent at the German Air Force Center of Aerospace Medicine, Fuerstenfeldbruck, Germany; Dr.Frank.Weber@t-online.de

(3) Mag. Dr. Werner G. Mueller, Professor at the Institute for Applied Statistics, Johannes Kepler University, Linz, Austria; Werner.Mueller@jku.at

(4) Dr. med. Juergen R. Schaefer, Professor and Director of the Center for Undiagnosed and Rare Diseases (ZusE), Philipps University, Marburg, Germany; juergen.schaefer@mailer.uni-marburg.de

Biographies:

Günter F. Pilz (born 1945) is Professor emeritus at the Johannes Kepler University Linz (JKU, Institute of Algebra). He got his PhD (1967) from the University of Vienna. After being Assistant Professor (Technical University of Vienna, Dept. of Statistics), he was Researcher at the University of Arizona (1969/70). Then he changed to the University of Linz, where he became Professor in 1974. He was Head of Department for 25 years. From 2000-2007, he was Vice Rector for Research at JKU, and Austria's delegate to the OECD-group for Scientific Integrity. Research interests: algebra and applications. G. Pilz has about 70 publications in refereed journals and wrote 10 books, e.g., "Applied Abstract Algebra" (Springer, New York, 1998, translated into Russian). Pilz is member of the American Mathematical

Society. Many research stays abroad. He holds a Honorary Doctorate (Ural State University, Russia) and is Honorary Professor at Shandong University of Technology (China).

Frank Weber is a clinical neurologist and, as a military physician, is a certified flight surgeon. After many years in hospital based neurology, he is currently head of the department of research of the German Air Force Center of Aerospace Medicine. He has been trained in clinical neurology at the University of Ulm and at the Technical University in Munich, where he met Frank Lehmann-Horn. He has a special interest in neuromuscular disorders and in clinical neurophysiology. His scientific interests lie on disorders of nerve and muscle and in channelopathies of the peripheral nervous system and of the muscle system, particularly on sodium channels. He is a member of several scientific societies, e.g., the German Neurological Society and the European and the American Academy of Neurology.

Werner G. Müller is currently full professor at and head of the Department of Applied Statistics at the Johannes Kepler University Linz, Austria (JKU). He gained his PhD in Statistics from the University of Vienna in 1991. His research interests are in the areas of experimental design and spatial statistics. Author of the book *Collecting Spatial Data* Müller has co-edited six other books, has published more than 50 research papers in refereed journals with currently over 3000 citations on Google Scholar. He serves as (coordinating) Editor-in-Chief of *Statistical Papers* (Springer) since 2010. He has lead several research projects from various funding bodies (FWF, ANR, LIT, ÖNB, ÖAD) with total acquired funding to date of approximately 900000€ and has successfully supervised six phD-students (two further ongoing). Müller also served as President of the Austrian Statistical Society (2014-2017) and is an elected member of the International Statistical Institute.

Prof. Dr. Jürgen Rolf Schaefer (borne 1956) is director of the "Center for undiagnosed and rare diseases" (ZusE) of the University Clinic in Marburg, a well known high volume center for patients with rare diseases in Germany. He studied biochemistry in Tübingen, (1977 – 1978), medicine in Marburg (1978 – 1983) and Frankfurt (1983 – 1984) and was a research fellow at the National Institutes of Health (NIH), NHLBI, Molecular Disease Branch (Head: B. Brewer Jr) (1986 – 1989). Since 1990 he is physician in different positions at the Philipps University in Marburg, Germany. He is editor for the rare disease section of the Journal "Der Internist" of Springer Medizin in Germany. He has published more than 100 scientific, peer-reviewed papers and several books. In addition, hestarted a Podcast on Rare Diseases (https://diagnose-selten.podigee.io/).