Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Olan A. A Unique Set of Physiological Changes Which Causes Breast Cancer: A discovery with a novel method of finding non-infectious disease causes which combines multiple research. British Journal of Cancer Research. 2025; 8(2): 778- 77. doi: 10.31488/bjcr.205

(PDF updated March 28, 2025.)

1. Analogy to clarify a method

A **simplified analogy** which explains why we looking for a match between physiological parameters could be this. Imagine a large town. We observe it from the top. A person arrives to downtown for some *personal business* on a regular day. How likely this person meet *a friend* or colleague in a totally random place of the downtown?..Very unlikely. We observe this person is going to downtown for few days to different places and he never meets a friend or a colleague. Now, we observe from the top that some rare folks meet someone in downtown often and sometime in the same location. We know **there is a pattern explaining these meetings**, they are not random in most cases. There might be someone they have agreed to meet with before (a colleague they travel together with, a friend, etc). If we find these folks meeting we know we found very likely some pattern.

In this analogy, we can treat a downtown as a human body, a person arriving to downtown as a physiological parameter change caused by some disease causing factor. People which meet each other in the downtown are an analogy of physiological parameters which "meet up" as they cause a disease and not just a random meeting. If we find those folks who meet up we know there is some cause there. In the method presented, these "meetings" between physiological parameters are represented by "intersections".

2. Mathematical foundation of the method

Here we will go into a mathematical foundation of the method in details.Let's look at the simple case of a disease where it was prior determined via experiments that multiple factors are causing changes in **only 2** physiological parameters of human body (parameters further).

Let the factors be **F1, F2, ... Fn** and the parameters be **C1, C2.** Now, let's look at the case where *F1, F2,..., Fn factors separately causing **only 1** change* in physiological parameters either C1 or C2 beyond 1-sigma (this actually often takes place in practice). That means that only C1 *or* C2 changes by some factor Fj (j E {1,2,..n }). Let's **P1, ..., Pn**, where n > 2 be *the sets of all physiological parameters which are related to factors F1, F2, .. Fn* accordingly. For example, P1: { r12, r15, r29, 43 }, P2: { r15, r28, r34, r89, r34, r12, r98 }, etc.

Let's look at standalone factor **Fj.** As we know, a factor **Fj** (where *j is some integer from 1 to n*) impacts the specific physiological parameters either **C1 or C2** then we know that *this params C1 or C2 should be part of its set of* **Pj** as **it contains ALL the related to Fj parameters (a complete set) .** *Let's take a factor F1 such that its set P1 contains **C1**, and choose some F2 such that its set P2 contains **C2*** (it is possible as we know factors impact either C1 or C2*), then if we choose any other factor as F3 then its set of P3 must contain either C1 or C2 (as F3 also impacts these physiological parameters - either C1 or C2 and P3 is a complete set). If P3 contains C1 then it intersect with P1. If P3 contains a C2 it intersects with P2. So **P3 must intersect with either P1 or P2 (either in C1 or C2).** In similar way we can apply this to P4, P5,... Pn. So this brings us to conclusion that **a set of physiological parameters Pn, where n > 2 must intersect with either P1 or P2 either in C1 or C2.** This means parameters Pn intersect with each other either in C1 or C2.* We can see representation of this set's behavior on the Pic 2. ***All sets Pj, Pk,... matching to Fj, Fk,.. on the Pic 1 are crossing and only in C1 or C2 but not both.***

We don't know the values of C1 and C2 but if *we can find where parameters Pn intersects with each other we can determine* ***a subset** of physiological parameters **Pm**:* { Ry, Rx, Rz,.. , Rt } which contains values of C1 and C2. This subset of **Pm** will be much smaller then set of all possible params included in P1, P2, .. Pn (as it is a subset and similarities in params of P1,.. Pn are not very probable and that is addressed below) but may contain more then 2 parameters and *only 2 parameters of this subset **Pm*** can be real physiological parameters causing a disease as they are C1 and C2.

In order to eliminate the incorrect parameters from subset **Pm** we need to notice:

1) that the params C1 or C2 should ***be such so all P1, P2, ..., Pn intersect in them*** and if some parameters of ***Pm**:* { Ry, Rx, Rz,.. Rt } don't fit this rule *their need to be eliminated.* Practically it means this. We take random (or using a common sense) a combination of some 2 parameters **Rk** and **Rm** from a set of **Pm** *and*

   *check if the P1, ... P2 all intersect in them if not then the Rk and Rm combination is not a valid set of C1, C2 and we may need to check another set of 2 parameters **Rk** and **Rg***

2) if some parameter of set ***Pm****:* { Ry, Rx, Rz,.. Rt } ***is known as not changed beyond 1-sigma it should be eliminated*** as disease is caused by change in param beyond 1-sigma (as per our model).

3) if some parameter of ***Pm****:* { Ry, Rx, Rz,.. Rt } ***is causing some set Pn intersect 2 times with some other set Pk then it should be eliminated*** as factors F1, F2, .. Fn can *only* impact 1 physiological parameter in this case and cannot impact / intersect 2 or more due to this.

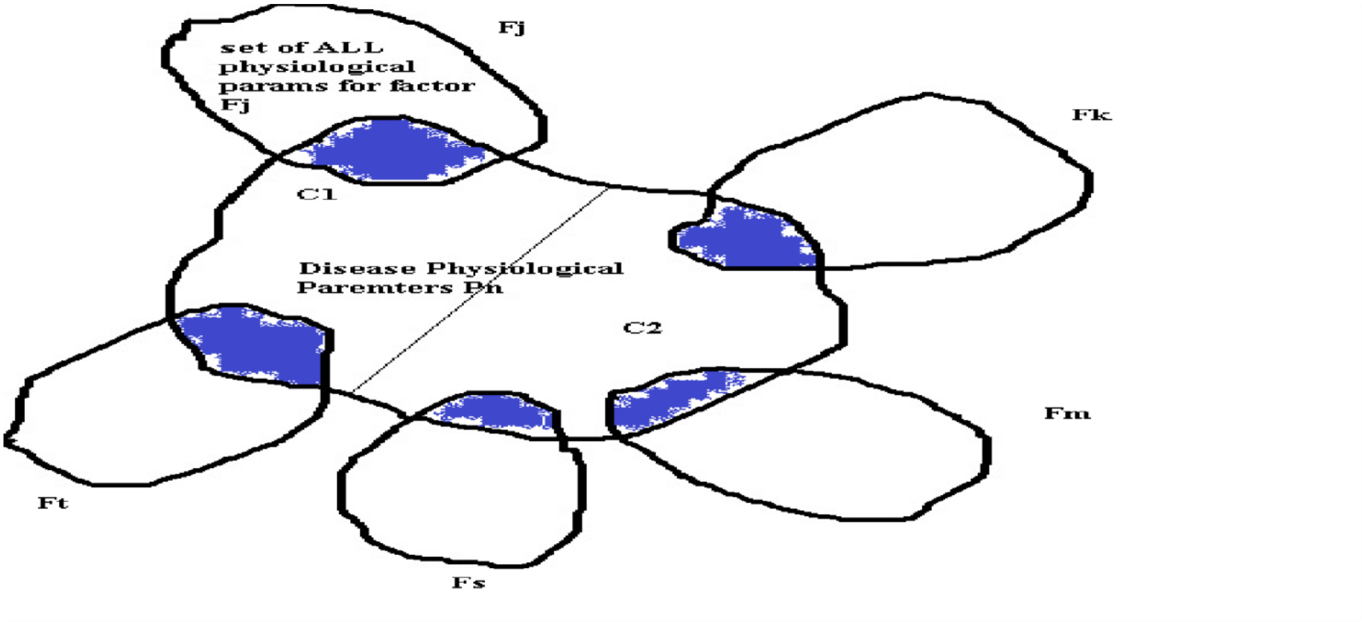
Fj
set of ALL physiological params for factor Fj
Fk
C1
Disease Physiological Paremters Pn
C2
Fm
Ft
Fs

Figure2. (Blue areas are area where Pj, Pk, etc for factors Fj, Fk, etc. intersect with a set of physiological parametersC1, C2 which are part of set Pn and which are causing diease)

The method above was described for a case of factors F1, ..., Fn impacting ***only 2*** parameters but it can be extended to 3 and more parameters.

How likely are random matches between physiological parameters?

As we discussed above the set of physiological parameters ***Pm****:* { Ry, Rx, Rz,.. , Rt } where we observe intersections may contain more parameters then needed (more than 2 in our case and due to other reasons).

We need to be concerned with a question such as if we find one intersection of sets P1 and P2 in a physiological parameter belonging to 2 different external factors how likely it can be a random intersection? To answer this question let’s formulate the problem mathematically.

Let’s have a set A of integers from k = 1 to very large N. Let’s randomly select **n** numbers in set of P1 = {Ak, Ag, .., At } and then randomly select **n** numbers into set P2 = { Af, As, ... Al } from our original set A (k= 1 to N) such that each element repeats only once in set P1 and only once in P2 (it is a unique element to sets P1, P2). For example, if we chose a number 3 as part of the set P1 then it only exist one time in the set P1. What is a probability that we find element **Ai** in set P1 and P2 ?

To answer this question let’s do next steps. Let’s limit set A by some top element enumerated by **t** (so set is not infinite).

1. We can take **n** elements from **t** elements of set A with number ways **tCn**
2. Number of ways to take **n** elements with an element **Ai** equals the number of ways to select **n-1** elements (we exclude Ai) from **t -1** (set of A elements) and is **t-1Cn-1**
3. Then probability to take **n** elements which include element **Ai** in set P1 (or P2) is P(Ai E Psel) = **t-1Cn-1 / tCn,** where Psel is P1 or P2 sets
4. The probablity that element Ai will be in P1 and P2 is P (Ai E P1 and Ai E P2) = P (Ai E P1) * P (Ai E P2) as events independent.
5. So probability P (Ai E P1 and Ai E P2) = P (Ai E P1) * P (Ai E P2) = (**t-1Cn-1 / tCn) ^ 2**

6. Or finally**,** the probablity that element Ai wil be in P1 and P2 is P (Ai E P1 and Ai E P2) = (**t-1Cn-1 / tCn**

) ^ 2

Using a formula above let's calculate a probability of match in element Ai if we take randomly elements from a sequence of numbers from 1 to 1000 (t = 1000, assuming so many physiological parameters exist) and take only n = 10 element into sets P1 and P2 accordinly. P(Ai E P1) = 999C9 / 1000C10 = (2.63 * 10^21) /(2.63 * 10^23) = 1 / 10^2 = 0.01 the same is fare for P(Ai E P2) = 0.01 and so the probability of getting element Ai in sets P1 and P2 is (t-1Cn-1 / tCn) ^ 2 = 0.01 ^ 2 = 0.0001

This is a probability of radom match. **The probability of non-random match** is **1 - P (Ai E P1 and Ai E P2)** = 1 - 0.0001 = 0.9999 ~= 1 so very close to 1. It means ***if we see a match between set P1 and set P2 in some element Ai it exremely likely it is not random.*** This is an important conclusion. *The only matches we find practically are not random but are caused by some reason* and in our case it is due to same physiological parameter impacted by 2 different factors. We need to notice that number of paramaters actaully much more as most agree that there are around 20,000 different proteins in our body and each is a potential physiological parameter. So the probability of the match selecting 10 of 20,000 will be much smaller! In practice there are about over **t=150** physiological parameters known to medicine and for a single factor we usually find about **n=30** related paramters. Doing calculations for this case we ge**t** P(Ai E P1) = **149C29** / **150C30** = (6.43 * 10^30) /(3.2 * 10^31) = 0,2 and so the probability of getting element Ai in sets P1 and P2 is

(t-1Cn-1 / tCn) ^ 2 = (0.2)^2 = 0.04

We see the probability of random match is higher in practice (4%) and so in practice we can see more random matches. The probability of at least one random match will also increases as we find intersections for dozens of different causation factors.

This random matches still can be eliminated with methods described in this article by applying other restrictive conditions, including using our criteria for disease causes.