Research Article

A Connection between Factors Causing Diseases and Diseases Frequencies: Its Application in Finding Disease Causes

Alan Olan*

Independent Researcher

ABSTRACT

In this work an author is building a theoretical model of a non-infectious disease, which shows that there is a connection between diseases frequencies and their causes. This connection allows to determine how many factors are causing a specific non-infectious disease if we know the disease rate in a population. The model shows that for a majority of non-infectious diseases there are at least two simultaneously acting factors which cause a disease and in many cases there are more factors simultaneously involved. This helps researchers to improve an understanding of specific disease causation and physiological mechanisms behind it and will lead to a research for additional, still missing causes to complete these mechanisms. This work determines a number of simultaneously acting factors causing diseases such as a breast cancer, Coronary Heart Disease (CHD), multiple sclerosis, etc. and explains so called French Paradox for CHD. The work also deduces a formula and *a method of determining that a specific risk factor is the one which really causes a disease* or it is not. Applying a method developed in this work the author shows how three different simultaneously acting causes of atrial fibrillation are determined using an existing research data. This method should allow medical researchers to determine if a found risk factor for a disease is really a cause of the disease or not and covers a significant gap in current understanding of risk factors nature and its connection to the physiological parameters of the human body.

Keywords: Causes; Disease; Probability; Frequency; Factor; Prevention, Root cause; Diseases rate; Physiological parameter

INTRODUCTION

Using statistical and experimental data a mathematical model of non-infectious disease is created which is applicable to any non-infectious disease (and to some degree to infectious diseases as well).

The model is considering that **a cause of a disease is a change in a physiological parameter** of the body approximately beyond 1-sigma interval of its measurements, slightly less.

The model predicts and the experimental data confirm its predictions that rate of non-infectious diseases is closely connected to the number of changes to physiological parameters of the human body which are causing the disease. Based on this connection **the model allows to determine number of disease causes (as number of physiological parameters' changed) for any non-infectious disease by only knowing the disease rate.** For example, if the disease rate is 72 per 1000 people then the model determines that disease has 2 causes. Despite the differences in the rates of a specific disease in different countries or populations the model determines the same number of causes. The model introduces a formula to calculate a number of disease causes as below:

$$\text{Nubmer of Causes} = \log_{0.22}(\text{Rate of Disease})$$

where a rate of disease can be for example as 45 per 10000 people and number of causes(as physiological parameters' changes) is an integer number like 1,2,3, etc.

According to this non-infectious disease model in order to cause a disease all of specific for each disease physiological parameters should be changed by actions of the external environment so their measurements will be beyond 1-sigma interval. Actually, slightly less than this interval.

The model shows the more the rate of non-infectious disease the less number of simultaneous changes to the physiological parameters is required to cause it. The changes can be acquired over the lifetime of individual and have to be such so each physiological parameter's measurement goes beyond 1-sigma interval*.

The work also allows to use vast volume of existing medical research's data about diseases' risk factors in order to determine real causes of non-infectious diseases using a simple criteria which mathematically derived from the model.

The work mathematically **derives the criteria which allows to determine which risk factor is a cause of disease** and which is not. It mathematically proves that in order to be a cause of the specific disease the found risk factors (calculated as (risk_cases-normal_cases)/normal_cases) should rich the value of 3.5 (or 350%) or 19.7 (1970%), etc. plus/minus a practical margin (recommendation is ± 50%)

The calculations based on the model has shown **that a majority of non-infectious diseases are caused by at least 2 simultaneously occurring changes to physiological parameters of human body or more.** For example it shows that Coronary Heart Disease (CHD) is caused by 4 physiological parameters changes taking place at the same time.

As the model derived in this work predicts that non-infectious diseases are caused at least by 2 changes (beyond 1-sigma) to physiological parameters of human body then it shows there is no reason and actually, often it is incorrect to search for a single cause of the non-infectious disease. The single cause of non-infectious disease does not exist for a majority of them, according to this mathematical model.

Based on the smoking risk factor value, the developed in this work criteria determines that smoking is one of the causes of lung cancer and that is matching to already a well recognized medical fact, and supports the validity of the criteria used to make this prediction.

The model shows that Multiple Sclerosis in men is caused by changes to 5 physiological parameters (beyond 1 sigma interval*) while the same disease is caused by changes to 4 physiological parameters in women. This explains why a rate of the Multiple Sclerosis is higher in women than in men as the more physiological parameters needs to be changed to cause the disease the less the rate of disease.

The work shows why breast cancers and leukemias are caused by a change to 4 and 6 physiological parameters of the human body accordingly.

The model leads to a conclusion that **the majority of non-infectious diseases cannot occur if only one physiological parameter of human body changes beyond 1-sigma interval.** The multiple and simultaneously taking place changes to physiological parameters needs to be there for a non-infectious disease to occur but the changes can be acquired over the time.

The work leads to a conclusion that by controlling few physiological parameters of the human body so they are located within the 1-sigma of its measurements it is possible to prevent a non-infectious disease such cancers or strokes from occurring at all in the individual [1-3].

The work shows that the fact that few of physiological parameters changes needed to simultaneously occur in order to cause the non-infectious disease, allows to select some risks factors as real causes of the disease and that we have some room for an error in selecting these risk factors as disease causes because if few risk factors are correctly chosen they will compensate for an incorrectly chosen one. This way we can prevent many diseases from occurrence by keeping the right risk factors (physiological parameters) under control (meaning within 1-sigma interval).

The work mathematically **derives a formula which connecting the numbers of disease causes** (physiological parameters' changes) determined in populations with a risk factor and without, **to a risk factor value determined for the population** impacted by this risk factor. The Risk Factor value here is 0 and more and defined as (cases_with_risk-cases_without_risk)/cases_without_risk. Here Risk Causes / No Risk Causes are numbers of causes (physiological parameters changes) determined for a disease in a population with a risk factor and without it using a formula presented above and are integer numbers (0,1,etc.). The formula is:

$$\text{Risk factor} = 0.22 \wedge (\text{Risk Causes - No Risk Cause}) - 1$$

METHODOLOGY

Development of non-infectious diseases model

In this work, we will look at the causes of diseases in humans and will try to find a generic way to determine which possible reasons are causing a specific disease using mathematical statistics. The same rules derived in this study will likely apply to multi cell animals as well. We will build a model of non-infectious disease which, as every model, will be missing some details not considered important for the disease analysis.

Let us consider the human body as a physiological system placed inside an external environment that has the capability to impact the physiological parameters of the human body, called parameters further. This external environment, by acting on parameters of the human body, can cause a disease. If this environment does not negatively impact the human body, then the disease should not occur as there is no reason. In this model, we ignored the fact that even a healthy human body working in a non-harmful environment may have aging effects which cause diseases.

Let us describe the human body as a physiological system that has a set of certain parameters An as {A1, A2, A3,..., An}. Suppose the external environment causes a change in one and only one of these parameters, Aj where j E (1, n) and so others remain unchanged; then, this change should cause the same disease. In reality, we observe many different diseases, so different diseases should be caused by changes at least in different parameters of An {A1, A2, A3,..., An} or by a group of some of parameters included in a set of {A1, A2, A3,..., An}.

Suppose that each disease is caused by only one distinct parameter of the group {A1, A2, A3,..., An}. From experience, we know that the values of the parameters of the human body are usually random variables with a normal (Gaussian) distribution from Amin<Aj<Amax (for any j E (1, n)), such as weight and blood pressure variations, etc. Let's recall that about 95.4% of all values of this parameter will be located within 2σ from a mathematical expectation μ of this parameter, 68.3% of values of this parameter will be located within 1σ interval from its a mathematical expectation μ. For a purpose of estimate, let's accept that value of physiological parameter negatively affected by some external factor causing a disease is located beyond the 1σ interval meaning outside {-1σ, 1σ}.

We take 1σ from a mathematical expectation μ as it is pretty far from average value of the parameter but is not exceeding the normal "healthy" range at the same time, also only a small portion (about 32%) of all values located in this interval. We use this range as also we consider a fact that actions of external factors on human

body is compensated by its capabilities to sustain physiological parameters the same *via* homeostasis, for example a temperature of the body is around 98.6° F and can decrease or increase slightly but in a healthy person it tries to stay around this "normal" 98.6° F value. Homeostasis provides dynamic equilibrium for physiological parameters. The significant difference in the physiological parameters from their average "healthy" range under the actions of external factors is not likely due homeostasis.

Let us look at the case where the change in the parameter's value Ak occurs in the positive direction and its value grows above 1σ and so will be located inside the interval where Ak E ($\mu+\sigma$, ∞). The probability that parameters Ak will be in this zone is $P(1<z)=\Phi(z=3.99)-\Phi(z=1)=0.5-0.34=0.16$ based on the normal distribution probability tables. The probability 0.16 that parameter Ak will have a value above ($\mu+\sigma$) zone means that on average one or two persons out of 10 should be sick with a disease if it is caused by a change in one and only one distinct parameter. The experience contradicts our original assumption that disease can be caused by just one parameter change, as for many non-infectious diseases, their probabilities in a society is about 5 per 100000 people (0.00005 for Leukemias), 303 per 100000 people (0.00303 for Multiple Sclerosis), or similar ranges and far not as high as 0.16. This contradiction implies that changes in one and only one distinctive physiological parameter is not enough to cause the majority of non-infectious diseases (where specific infections are not the main known cause) if the human body parameter's change is just above the 1 σ (sigma) from math expectation μ in a positive(negative) direction. If we consider the possibility that the parameter can be lowered in the same fashion (to $\mu-\sigma$) to the low border of its range, then the probability of these two events ($\mu+\sigma$ top and $\mu-\sigma$ bottom) would double and be 0.16 × 2=0.32, which is still very high compared to practical observations of non-infectious disease occurrences.

Now, let's us estimate what would happen if an physiological parameter after some negative action on it by external environment would be located beyond 2σ in a positive or negative direction? This would be practically outside of interval where majority of all possible values located. In this case $P(2<z)=\Phi(z=3.99)-\Phi(z=2)=0.5-0.477=0.023$. That would mean that the non-infectious disease would be happening as often as 2 in one 100 people. This also contradicts to frequency of occurrence of most non-infectious diseases, for example it is about 3 per 1000 for Multiple Sclerosis (0.00303), etc. We can restate our conclusion that **a change in only one physiological parameter of human body by 1 σ (sigma) above math expectation cannot cause a majority non-infectious diseases as per observation of frequency of these diseases in a real life.**

We can state then that a disease generally is a result of change in multiple parameters of the group of An {A1, A2, A3,..., An}. The probability of disease which is caused by multiple independent factors would be $P(D)=P(Ak1) \times P(Ak2) \times P(Ak3) ...\times P(Akn)$ where kn E (k1, n) is number of parameters of human body which were changed by more then 1 σ (sigma) in positive (or negative) direction. Based on the above calculations, we already know the probability of one parameter change above/below 1 σ P(Awn)=0.16 and let's take it as an estimate that the same probability applies to each of the parameters. In this case, the probability of disease $P(D)=P(Awn)^n$ (power of n), where n is the number of physiological parameters which change causes the disease.

From the above equation, we can obtain number n of physiological parameters of the human body (number of disease causes) which change during a disease as

$n=\log_{P(Awn)}P(D)$ (1.1)

or

$$n=\log_{0.16}P(D)$$

Where P(D) is the frequency(rate) of disease

The frequency(rate) of disease P(D) can be equal to 0.001 if it occurs in 1 in 1000 people. Number n of physiological parameters can be a real number and its needs to be approximated by closest integer numbers (for example 2.4 is approximated by 2, 2.7 is approximated by 3 as per usual math practices).

Now, P(Awn)=0.16 was our theoretical estimate. Let's try to determine the P(Awn) value more precisely from a real data observations. We know that occurrence of chronic diseases increases as people age and according to our model this means environment has time to change physiological parameters of the body during an individual's live. Let's do a practical estimate for P(Awn) based on children where less impact to physiological parameters is expected.

According to National Institute of Mental Health (NIH) occurrence of autism in boys is 36.5 per 1000 as year 2018 data, and in girls it is 8.8 per 1000 on the same year.

For boys $Nb=\log_{0.16}(0.0365)=1.8=2$ (**2 physiological parameters causing autism in boys**)

For girls $Ng=\log_{0.16}(0.0088)=2.58=3$ (3 parameters changes causing austim in girls)

If Pb(D) and Pg(D) is disease frequency for boys and girls accordingly then we know

$Pb(D)=P(Awn)^2$ and $Pg(D)=P(Awn)^3$

$Pg(D)/Pb(D)=P(Awn)^3/P(Awn)^2=P(Awn)$

Or

$P(Awn)=Pg(D)/Pb(D)$

and substituting values for Pb(D) and Pg(D) we get

P(Awn)=0.0088/0.0365=0.24 (a value based on real data)

So based on the data for a real disease rate in humans (for autism) we get P(Awn)=0.24. From practical usage perspective, to improve n value precision as an integer we reduce the value found to P(Awn)=0.22 which does not impact significantly the results. The final suggested formula for practical usage is following:

$$n=\log_{0.22}P(D), \text{ where P(D) is a rate of the disease (1.2)}$$

As already mentioned above, the frequency(rate) of disease P(D) can be equal to 0.001 if it occurs in 1 per 1000 people. Parameter n can be a real number and its needs to be approximated by closest integer numbers (for example if n=2.4 is approximated by 2, n=2.7 is approximated by 3 as per usual math practices).

We deduced a formula 1.2 based on our model. Based on this model the formula is expected to show at least 2 physiological parameters changes or more causing non-infectious diseases. Also, as we currently understand, the same number of causes is expected for each disease despite the fact that rate of the same disease vary significantly. Let's see now how the formula 1.2 is matching the real

world data (Tables 1-3).

Table 1: The rates and causes of CHD in 5 regions of different countries.

Disease	Country / Region	Disease Rate (number of sick per 100 000 individuals)	Number of disease causes ($\log_{0.22}P(D)$)	Comments
Coronary Heart Disease(CHD)	Glasgow (UK)	777	3.20=3	CHD is caused by 3 physiological parameters changes
	Belfast (UK)	695	3.28=3	CHD is caused by 3 physiological parameters changes
	Barcelona (north eastern Spain)	210	4.07=4	CHD is caused by 4 physiological parameters changes
	Lille (northern France)	298	3.84=4	CHD is caused by 4 physiological parameters changes (French Paradox)
	Toulouse (south western France)	233	4.00=4	CHD is caused by 4 physiological parameters changes (French Paradox)

Table 2: MS diseases in 6 countries.

Disease	Country/ Region	Disease Rate (number of sick per 100 000 individuals)	Number of disease causes ($\log_{0.22}P(D)$)	Comments
Multiple Sclerosis	Germany	303	3.83=4	4 physiological parameters changes in a person causes MS
	USA	288	3.86=4	
	Canada	250	3.95=4	
	Switzerland	180	4.17=4	
	France	155	4.27=4	
	Netherlands	150	4.29=4	

Table 3: Application of formula to cancers.

Disease	Country/ Region	Disease Rate (number of sick per 100 000 individuals)	Number of disease causes ($\log_{0.22}P(D)$)	Comments
Lung and Bronchus	New York, 2019	55	4.95 = 5	Lung and Bronchus Cancer is caused by 5 physiological parameters changes
	Pennsylvania, 2019	58	4.92 = 5	
Female Breast	New York, 2019	141	4.3 = 4	Female Breast is caused by 4 physiological parameters changes
	Pennsylvania, 2019	131	4.33 = 4	
Leukemias	New York, 2019	5.5	6.48=6	Leukemias is caused by 6 physiological parameters changes
	Pennsylvania, 2019	6.3	6.38=6	

We can see from the calculated values in the tables that Coronary Heart Disease(CHD) is caused by same number of 3 physiological parameters simultaneous changes in UK but in France, and Spain the number of causes is actually higher as it increased to 4 which explains so called French Paradox which is a fact that despite of high consumption of saturated fats, etc. the CHD occurrence in France (and as you can see in some places in Spain) is low comparing to other countries with high consumption of saturated fats as with UK in the above example. We can see that in France to get CHD people needs to experience a change of 4 parameters outside 1σ (sigma) where in UK only 3 parameters need to be affected and this means in UK some parameter is outside of 1σ (sigma) for most of people (level of "good" cholesterol possibly which in France is increased by high wine consumption?).**The more physiological parameters changes are causing the disease the less the disease rate.**

We can see our deduced formula 1.2 and so our model gives results which are matching the predicted ones in the real world: 1) the diseases are caused by multiple factors (in the examples from 3 to 6) and 2) Despite the differences in diseases occurrences the number of physiological parameters changes (causes) is the same for the same diseases for the same areas or located close (like New York and Pennsylvania States where Lung and Bronchus cancer is caused by 5 physiological parameters changes simultaneously occurring). 3) It can explains why the occurrence of diseases significantly different (2 to 3 times more) in different countries, for example like in UK and France for CHD, by differences in number of factors causing it. 4) It does not give the extremely high number of physiological parameters changes as it would be improbable to have them all happening at the same time (like 10 or 20 for example). So practical data observations of non-infectious diseases look like matching the model we created and confirming it.

Experimental prove of formula for disease causes

The derived formula (1.2) shows that the number of disease causes (physiological parameters changes) is related to the disease rate.

Let's prove that experimental data support the validity of the formula and a model it is based upon.

From a common sense, we can expect from this formula that for different disease rates it would give same number of disease causes -Event A1. If we take a look at Table 1 showing Multiple Sclerosis Rates and Causes we can see 6 countries having a different rate but all having 4 disease cause for MS.

Let's suppose the event M is that "a disease Rate is matching to the number of disease cause" is a coincidence. Let introduce the probability of this event Pms E {0,1}.

Let's determine a probability Pall of the event that "for all countries disease rates for MS matches to the same number of disease causes N" which consist of independent events for each country.

Pall(M1 and M2 and M3 and...and Mn)=P(M1) × P(M2) × P(M3) ×...× P(Mn)

where M1, M2, M3, .. Mn are events that MS is matching to the same disease causes in different countries. Let's take some value for P(Mn) from an interval {0,1} which is equal to Pms as 0.1(for a purpose of easy calculations) which is a pretty high probability the event is a coincidence and for an estimate let's apply it to all P(Mn) where n={1, n}. Then we get Pall(M1 and M2 and M3 and ... and Mn)=Pms^n=0.1^n. From the table 2 for MS diseases we can find already n=6 at least so Pms=0.1^6 and it means Pall=0.000001 which can be reduced as we add more countries.

It means the probability that events of type "MS disease Rate matching to the number of disease cause" happening together by coincidence is small 0.000001 and the complimentary event which is "event happens not by a coincidence" is 1-Pall=1-0.000001~= 1 (it is very high so the fact is true) and this probability can be increased as we like with an addition of other countries causes. We can decrease P(Mn) from 0.1 to a smaller value and this will make Pall is even smaller. We proved that a match between the Disease Rates and a same number (which we hope is a disease cause) is not a coincidence but a fact. The probability of this fact P1=1.

The formula calculating the number of disease causes should give a positive number of diseases-Event A2.

We can see from Table for MS Disease Rates and Causes that it is the fact the number is positive. The probability of this fact P2=1 (means this fact is true, takes place).

The formula should give the number of disease for a majority of disease as 2 and more (as predicted by our model)-Event A3.

We can see from Table for MS Disease Rates and Causes that it is the fact the number is more than 2. The probability of this fact P3=1.

We don't know how many causes is a valid number for a non-infectious disease (based on experience) but we can expect that the difference in disease rates for same disease in women and men is caused by a small number of causes (physiological parameter changes) within a range {1,3} or so-Event A5 as biochemically men and women are very similar but difference is in amount variations of small number of hormones (testosterone, estrogen, etc).

We can see from autism causes above and calculations provided below for a Multiple Sclerosis disease's causes that for a case of autism the number of causes (physiological parameters) differences between boys and girls is 1 cause, same for MS disease in men and women - the difference is in 1 cause only. The probability of this fact above P5=about 1 (let's say P2~=0.9) as we might need to see other similar cases.

What is the probability that our formula is valid? It will be valid if probability of all events A1, A2,,,,, An above together is 1 and this probability is:

P(A1 and A2 and A3 amd A4 and A5)=P(A1) × P(A2) × P(A3) × P(A4) × P(A5)=1 × 1 × 1 × 1 × P5=P5

but P5 is about 1 (or we take it as 0.9 or 0.95) so we can conclude that experimental data support that our formula is valid with about 0.9-0.95 probability or the chance that the derived formula matching the real world is about 90%-95% which makes it predictions pretty precise for a practical usage (and the precision can be made higher with an increased number of data).

Practical conclusions from disease rate to disease cause connection

We can make next conclusions out of analysis of the formula (1.2) above:

A specific disease only occurs if all of n found physiological parameters are changed above 1 sigma (are located outside of interval (μ-σ,μ+σ), slightly less than this). **The disease cannot occur if only one 1 of n physiological parameters changed.** It resembles a lock with few wheels for unlocking where each wheel is a separate physiological parameter. Only unlocking ALL wheels of such a lock will cause a specific disease. Based on this, disease cure (at least in the early stages) can occur if at least one of the parameters is within the normal range (μ-σ,μ+σ),. It also states that the human body has multiple levels of defense mechanisms, and the human body's functions are supported at multiple levels which keeps a human body disease free mostly.

The more a rate of disease (occurrence per 1 thousand people, for example), the less the number of physiological parameters of the human body are changed to cause the disease. **The more frequently a disease occurs the less a number of physiological parameters' changes is causing it**. This can be seen in Figure. 1 below, where y is the number of physiological parameters of the body and x is a rate disease.

The parameters found from the formula (1.2) are physiological parameters of the human body and not parameters of the external environment (it is not environmental factors, etc).

Physiological parameters needed to cause a disease are generally independent of each other (as they treated like independent events above).

In some cases, we can observe that in one area of some country, the population has a disease due to one number of physiological parameters changes and in another part of the country due to another number of physiological parameters changes. For example, in area 1 of the country, the disease occurs due to 4 physiological parameters changed and in area 2, the disease happen due 3 physiological parameters changed or the same can happen in the same area but at different times (e.g., 10 years ago or now). In this case, the additional physiological parameters (4th here) may be completely changed for the entire a group of population and only three remain unchanged. The simultaneous change in those 3 physiological parameters causes the disease.

Usually, diseases in practice occur due to changes in two to six physiological parameters changes in the human body depending on the disease rate. Infections are usually caused by a single factor and may occur very frequently (Figure 1).

Let's see a practical application of the above formula (1.2).

In the USA, a rate of stroke in some of the states is 1.6-8 per 1000 people.

Let us determine the number of disease causes (physiological parameters changed) in the impacted people.

Let us determine Nmin=$\log_{0.22}(0.0016)$=4.25~ = 4(parameter)

Let us determine Nmin=$\log_{0.22}(0.0083)$=3.16~ = 3(parameter)

It means in some US states a stroke is caused by 3 causes and there are some where a stroke caused by 4 causes. The reason for this

may be that in some states one physiological parameter out of 4 is already beyond 1-one sigma interval for a majority of population and does not protect it any more from stroke.

We need to notice here that physiological parameters' changes going beyond 1-sigma interval do not always mean that they are outside of usual "healthy" range presented for a reference on different medical tests (blood work, urine test, etc). So in many cases doctors will not notice these kind of changes.
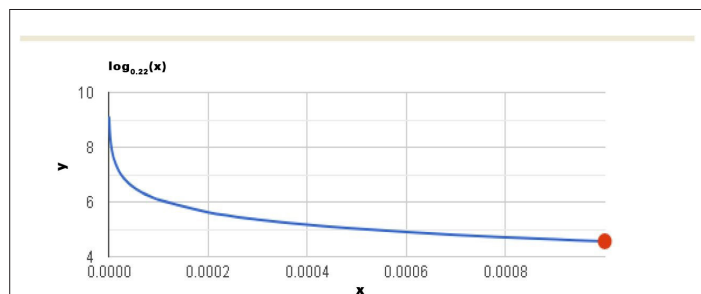

Figure 1: Number of disease causes (physiological parameters changed) related to a frequency of disease (disease rate). Here y (vertical)-number of causes, x (horizontal)-frequency of the disease (2 per 10000 as 0.0002).

RESULTS AND DISCUSSION

Some benefits of the disease causes formula for practical applications

Based on the disease rate a derived formula (1.2) allows to determine a number of disease causes (as physiological parameters changes) in a specific population. It states that **a rate of the disease is related to the number of physiological parameters changes which causing** it and allows to find number of these physiological parameters changes and it means allows to find the number of disease causes.

As there is usually some research on the risk factors of certain diseases, the equation (1.2) above allows to use of these factors as if they are disease causes (physiological parameters changes) and prevent the disease from occurring in the population by keeping these physiological parameters in a healthy range {μ-σ, μ+σ}, slightly less. This reduces the significance of the specific risk factor in the analysis of disease causes, as there are multiple physiological parameters involved and causing the disease, but at the same time, **it allows to take preventive measures for a disease by controlling these factors of risk to the point that disease never occurs [4-6].** For example, if we know that stress hormones and blood pressure can increase the risk of stroke, then by controlling all these factors, we can increase the chance that the disease never occurs for a patient. Currently, we know that blood pressure is a risk factor for a stroke, but we cannot say that it is a reason for a stroke, as not all people with high blood pressure have strokes; if we decide based on significance of research that blood pressure is an important factor and can be one of the real physiological parameters the change of which causes the strokes, then controlling a blood pressure should prevent strokes from occurring. In addition, as a blood pressure is part of the set of other factors, the error in controlling it is compensated for by controlling other factors as well. Yes, we may have chosen a blood pressure as a causing disease factor wrong but we likely chose other factors correctly and controlling those factors as physiological parameters changes within 1-sigma interval will still prevent the disease [7].

The equation (1.2) above allows to use some of already found risk

factors for a disease as disease causes as there are usually few (2-6) physiological parameters changes which are causing the disease and an error in choosing a risk factor as a disease cause gets reduced as if we chose some factors as physiological parameters correctly and some not but if we can control all of them or most so they are within healthy range{μ-σ, μ+σ} then we still can prevent the disease even when one of physiological parameters chosen as a disease cause happens to be a wrong reason. It is usually sufficient to control only one right physiological parameter to prevent the disease, as mentioned above, as all needed to cause a disease [8-11].

This shows that majority of non-infectious diseases are caused by multiple causes (physiological parameters changes) which should take place at the same time. It is not correct to search for just one specific cause of the disease for a majority non-infectious diseases, which is currently done by multiple researchers.

This shows that many risk factors found for a specific disease do not work along and definitely, other factors exist. Controlling some of those risk factors can prevent disease from occurring though.

It shows that human body functions are provided by multiple mechanisms which back up each other.

As there are multiple physiological parameters' changes needed to cause a disease it becomes easier to pick and choose them as they should be working as parts of the physiological mechanisms, and based on these physiological mechanisms, they can be selected more correctly as a set of multiple disease causes which should take place at the same time [12]. Currently, we usually select a factor as a potential cause for a disease based on studying the factor as a standalone.

As a disease is caused by physiological parameters changes beyond 1-sigma interval the reference ranges on medical tests (blood work, urine,etc.) may need to be updated so doctors can see when a specific parameter goes beyond this range and are able to notify patients that they are moving to some disease and need to take corrective actions.

Formula for determining if a specific risk factor is causing a disease

Let's deduce the formula connecting a risk value with an associated change in few out of n physiological parameters of the human body. Let's take 100 people(or 100%) without a risk factor and let the number m (or m%) be a number of sick people having a certain disease (this is usually found through a medical research). Let's take another 100 people(or 100%) with a risk factor and let the number l (or l%) be a number of sick people having the same disease. Let's look at the cases where where l>m (which is usually practically occurring as we estimate how the specific risk increases number of sick people).Then the risk K is:

K=(l-m)/m   (2.0)

For the those not having a risk factor a probability of disease then Pnr(D)=m/100 and with a risk factor Pr(D)=l/100.

From our disease factor formula we know that

Nnr=$\log_{0.22}$Pnr(D)=$\log_{0.22}$(m/100)

From here we get m= (0.22 ^ Nnr) - 100

Same way we get for Nr(number of factors causing disease with a risk factor present)

Nr=$\log_{0.22}$Pr(D)

from here we get I=0.22^Nr.100

substituting the I and m values into formula 2.0 above we get:

K=(1-m)/M=(0.22^Nr*100-0.22^Nnr*100)/(0.22^Nnr*100)=(0.22^Nr-0.22^Nnr)/(0.22^Nnr)=(0.22^Nr/0.22^Nnr)-1=0.22^(Nr-Nnr)-1

We get a formula (2.1) connecting a number of physiological parameters causing a disease with a risk factor K creating these physiological parameters changes where Nr is number of physiological parameters causing disease with a risk factor present and Nnr is a number of physiological parameters causing disease without a risk factor present.

$$K=0.22^{(Nr-Nnr)}-1 \quad (2.1)$$

Where Nr>Nnr, Nr and Nnr E {1,2,3,4,5,..}

The formula 2.1 shows that with an increase of disease risk factor K, the number of physiological parameters Nr necessary to cause the disease gets decremented by only the integer numbers, starting with 1. The disease risk factor's K value associated with each a decrease in number of physiological parameters Nr, takes up the one and only possible theoretical value which uniquely matching to every integer number by which the physiological parameters amount gets reduced. For example, as it will be shown below, for a risk factor caused by 1 physiological parameter decrease (in other words its disappearance) the theoretical value of disease risk K=3.55 (or 355%).

Criteria to determine if a risk factor is a disease cause

If we are doing an experiment and found from it a number Nr of physiological parameters causing a disease in a population with risk factor and we found a number Nnr of physiological parameters causing the same disease without a risk factor and now we want to check if the risk factor we working with impacts only one physiological parameter of the body, then Nr - Nnr=-1 (−1 as the Nr should be less than Nnr because with an increased risk number of physiological parameters changes gets reduced as per disease frequency formula (1.2), also a number of parameters are integers only 1, 2, etc). Then for n=1 we can write down.

K=0.22 ^ (-1) -1=3.55

This means **if a risk factor is causing only one physiological parameter change in the human body and is really causing a disease then the risk factor theoretical value should be 3.55 (or 355%).** We know that a risk factor is causing a disease if the factor is an external one, as external factors are causing a change of physiological parameters and this way are causing a disease as per our model [13]. For example, a solar radiation is an external factor which may cause a skin cancer and the criteria above allows to determine if this radiation is a cause of the disease or not.

Then for n=2 we can write down

K=0.22^(-2)-1=19.67=19.67 or 1967%

This means if a risk factor changing 2 physiological parameters is causing a disease then its value should be theoretically equal 19.67 (or 1967%).

The author suggests to use a rule that the practical risk factor can be located within a theoretical +/-50% range as it is clear that practical data will be somewhat different than theoretical. For example, for a risk factor responsible for just one physiological parameter change its practical value should be within range {355-178, 355+178} or

between 177% and 533%.

Smoking is one cause of lung cancer

Let's look at one example of estimation if a risk factor is really causing a disease.

For people who smoke the risk of lung cancer is at least 15 times bigger than for non-smoking people. So if number of non-smoking sick people is m then l=15 × m and risk K=(15m-m)/m=14 or 1400% it is clear it does not match to 1 physiological parameter change which is 355% but pretty close to 1967% for 2 physiological parameter changes. So let's estimate the deviation as (1967-1400)/1967=0.29 or 29% which is less than 50%.

So the estimate shows that the smoking is really one of the causes of lung cancer and impacting 2 disease causing physiological parameters.

Low beta-carotene is one cause of breast cancer

Let's look at another example of risk factor estimation. The research shows that women having beta-carotine in lower quartile of the range are more than 220% likely to get a breast cancer than those having beta-carotene in the top quartile. It means the value is pretty close to 355% for 1 physiological parameter change and is really within 50% of the practical range as it is above 177% and less then 533% for 1 physiological parameter change as shown above [14]. Based on this risk factor estimate we conclude that low beta-carotene is one of multiple causes of breast cancer.

Accordingly to early made conclusions if women control beta-carotene level so it is pretty high (within one sigma) then breast cancer should not happen unless there is a damaging factor (like a carcinogen, etc) impacting them and in this situation it is necessary to control 2 or more physiological parameters causing the disease in order not to get sick and remove action of carcinogen. This will be discussed further.

Examples of practical applications for the formula of number of disease causes

Let's us see some examples of practical applications for the number of disease causes formula (1.2) derived in this work.

Multiple sclerosis: In USA this disease happen in 277 people per 100000, in women the disease happens 3 times more often than in men.

Let's calculate the number of physiological parameters which changes beyond 1-sigma interval are causing the disease in men and women:

If we consider as x-number of men sick, then women are 3x and then out of equation x+3x=277 we get 69.25=69 men per 100000 people sick and for women 3x=3 × 69.25=207.75 women per 100000 people sick

Number of factors causing Mutliple Sclerosis in men then 5:

$Nm=\log_{0.22}69 \times 10^{-5}=4.8=5$

Number of factors causing Mutliple Sclerosis in women then 4:

$Nm=\log_{0.22}207 \times 10^{-5}=4.08=4$

In women Multiple Sclerosis happens more often because to cause it the one less physiological parameter change outside of 1 sigma is required. **It requires 4 changes to physiological parameters for women to get sick vs. 5 physiological parameters changes for men**

to get sick with the Multiple Sclerosis [15].

It mean the 5th physiological parameter for every women either always outside of 1 sigma during the disease vulnerability times or possibly does not exist at all.

Alzheimer

In USA this disease happens to women in 108 people per 10000, in men 56 per 10000 people *(Qing Cao 1, Chen-Chen Tan, et al, The Prevalence of Dementia: A Systematic Review and Meta-Analysis)*. Doing calculation we get:

For women Alzheimer happens because of 3 internal parameters changes

$Nm=\log_{0.22}108 \times 10^{-4}=2.99=3$

For men Alzheimer happens because of 3 internal parameters changes

$Nm=\log_{0.22}56 \times 10^{-4}=3.42=3$

Alzheimer disease happens to men and women due to 3 physiological parameters changes beyond 1-sigma interval. Alzheimer is not happening just because one protein accumulation in the brain according to this.

Atrial fibrillation: In USA this disease happen to people under age 65 as 1 in 50 people. Over age of 65 this disease happen to 1 in 10 people.

Doing calculation we get:

For age under 65 Artrial fibrilation happens because of 3 parameters changes

$NIs=\log_{0.22}1/50=2.58=3$

For people over 65 Artrial fibrilation happens because of 2 parameters changes

$Nover=\log_{0.22}0.1=1.52=2$

With aging the environmental factors(as per our model) reduce a number of physiological parameters of the body which needs to be changed outside of 1-sigma of their measurements in order to cause the disease from 3 physiological parameters changes to 2 physiological parameters changes and **this explains why atrial fibrillation happens more frequently for people over 65 years old** as the less number of physiological parameters changes required to cause a disease then more a rate of the disease.

Three causes of atrial fibrillation: There are multiple risk factors found in medical research which increase a disease rate of Atrial Fibrillation(AF) in person but not all them matching criteria set by formula 2.1 regarding a risk factor value which required for the risk factor to be a cause of the disease. Let's see some risk factors of atrial fibrillation which actually do pass this criteria. There should be 3 physiological parameters changes they impact as we found above, to cause the AF disease.

People with Obstruction Sleep Apney (OSA) found 5 times more likely to develop AF than without OSA. *(Evan L. O'Keefe, MD, et al., Prevention and Treatment of Atrial Fibrillation via Risk Factor Modification)*. If x-number of people without OSA then $K=(5x/100-x/100)/(x/100)=4$ or $K\%=400\%$ which is in range between 177% and 533% and per discovered above criteria for 1 physiological parameter risk this means that Obstruction Sleep Apney is one of reasons of atrial fibrillation. We can count it as reason 1 of disease.

It was found that epicardial fat increase over 1-standard deviation was associated with 5.4 fold high odds of persistent AF. Calculating in same way as above $K=((5.4x-x)/100:x/100)=4.4$ or $K\%=440\%$ which is in range between 177% and 533%.

It means an epicardial fat increase over 1-standard deviation is causing atrial fibrillation. We can count it as reason 2 of the AF disease. The fact that the fat should be increased above 1 sigma to cause the disease also is according to our model and formula 1.2

Aldosterone found as risk factor for AF as it has been reported to create 12 fold higher risk of AF in hypertension patients with primary hyperaldosteronism when compared with patients with essential hypertension. Calculating $K=(12x-x)/100/x/100=11$ or $K\%=1100$. As we can see the risk factor is too high for 1 physiological parameter change, but looks like within a range for 2 physiological parameters changes. Let's estimate a deviation from a theoretical risk value of 1967%. $dK=(1967-1100)/1967=0.44$ or 44%, as it is below 50% we can consider the risk in the range for 2 physiological parameters changes as it was shown above. It means Aldosterone is a one of the causes for atrial fibrillation. We can count as reason number 3.

According to presented in this article model the 3 risk factors (or here 3 matching them physiological parameters changes) should exist at the same time to cause the disease and this explains why they are called just risk factors as if one of them is present does not mean other 2 exist at the same time and so person should not be sick with AF. **It means the atrial fibrillation develops in people if they have Obstruction Sleep Apney(OSA) AND high level of aldosterone AND and an epicardial fat increase over 1-standard deviation.** Or they can have high level of aldosterone and one additional factor from above as aldosterone covers 2 physiological parameters changes of parameters as to cause the AF disease we need 3 physiological parameter changes for people under 65.

Now, we can see above that for people over 65 the AF requires 2 physiological parameters changes only and this means that these people should have one of the discovered causes applicable to this entire group. Which causation factor(physiological parameter) is changed for entire group needs to be found. It looks like a good hypothesis to research is that an epicardial fat increase over 1-standard deviation is applicable to majority of people over 65 and causing an increase in frequency of atrial fibrillation in this group of population as OSA probably does not apply to all people after 65, and it is doubtful that most people over 65 have high level of aldosterone levels as well.

Coronary Artery Disease (CAD): In USA this disease happen to 20.1 Million of adults over age 20. It is approximately in 7.2% of adult population in USA.

Doing calculation we get that Coronary Artery Disease happens due to 2 causes (2 physiological parameters changes beyond 1-sigma interval).

$Ncad=\log_{0.22}0.072=1.7=2$

Only 2 physiological parameters changes outside of 1 sigma needed at the same time to cause the CAD disease and it explains why CAD disease so frequently occurs in population as the less the number of causes for a disease the higher rate of the disease in population.

As we can see in few example above some physiological parameters are changed for the entire group of particular population-men/women, older 65/under 65. etc. and this causes different disease rates in these groups of population.

CONCLUSION

This work showed that non-infectious diseases are caused by multiple simultaneously taking place factors and this factors are the physiological parameters changes approximately beyond 1 sigma interval. It explains why some disease occurring in men or women, different world regions, etc. at different rates by showing that in many cases it is due to different number of factors which cause a disease and the work allows to determine these specific differences in numbers. The model introduced here allows to improve efficiency of medical research by allowing to convert disease frequency information into a useful information about a physiological mechanism structure of a specific disease first and then allows to find disease causes using provided criterias by applying them to the results of existing or new medical researches. This work is covering some gap in understanding of disease causation which is a significant roadblock of modern medical research. By introducing a concept of multiple factors causation it allows to make an easier selection of the risk factors as diseases causes as it shows that there is a room for an error in this selection and this determination of causes should allow more effective disease prevention and potentially their cure at early stages by controlling appropriate factors.

FUNDING STATEMENT

ETHICAL COMPLIANCE

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

CONFLICT OF INTEREST DECLARATION

The authors declare that they have NO affiliations with or involvement in any organization or entity with any financial interest in the subject matter or materials discussed in this manuscript.

AUTHOR CONTRIBUTIONS

Alan Olan is the sole contributor to this study.

DISCLAIMER

This work is not intended to be used in diagnosing and treatment of any disease in any patient, those actions should be done by qualified professionals only. The work is intended as a reference for the researchers only. The author is not responsible for any issues created by unintentional usage or consequences of this work.

REFERENCES

1. O'Keefe EL, Sturgess JE, O'Keefe JH, Gupta S, Lavie CJ. Prevention and treatment of atrial fibrillation *via* risk factor modification. Am J Cardiol. 2021;160:46-52.

2. Wong CX, Sun MT, Odutayo A, Emdin CA, Mahajan R, Lau DH, et al. Associations of epicardial, abdominal, and overall adiposity with atrial fibrillation. Circ Arrhythm Electrophysiol. 2016;9(12): 4378.

3. Ferrières J. The French paradox: lessons for other countries. Heart. 2004;90(1):107-111.

4. Matveychuk D, Dursun SM, Wood PL, Baker GB. Reactive aldehydes and neurodegenerative disorders. Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology. 2011;21(4):277-288.

5. Taveggia C, Feltri ML, Wrabetz L. Signals to promote myelin formation and repair. Nat Rev Neurol. 2010;6(5):276-287.

6. Zhang Y, Yang Y, He X, Yang P, Zong T, Sun P, et al. The cellular function and molecular mechanism of formaldehyde in cardiovascular disease and heart development. J Cell Mol Med. 2021;25(12):5358-5371.

7. Lebwohl B, Ludvigsson JF, Green PH. The unfolding story of celiac disease risk factors. Clin Gastroenterol Hepatol. 2014;12(4):632-635.

8. Sereda MW, Meyer zu Hörste G, Suter U, Uzma N, Nave KA. Therapeutic administration of progesterone antagonist in a model of Charcot-Marie-Tooth disease (CMT-1A). Nat Med. 2003;9(12):1533-1537.

9. Franco PG, Silvestroff L, Soto EF, Pasquini JM. Thyroid hormones promote differentiation of oligodendrocyte progenitor cells and improve remyelination after cuprizone-induced demyelination. Exp Neurol. 2008;212(2):458-467.

10. Bodar V, Chen J, Gaziano JM, Albert C, Djoussé L. Coffee consumption and risk of atrial fibrillation in the physicians' health study. J Am Heart Assoc. 2019;8(15):11346.

11. Neefs J, Van den Berg NW, Limpens J, Berger WR, Boekholdt SM, Sanders P, et al. Aldosterone pathway blockade to prevent atrial fibrillation: a systematic review and meta-analysis. Int J Cardiol. 2017;231:155-161.

12. Alzheimer's Association. Alzheimer's Disease Facts and Figures 2023. 2023.

13. Jin MN, Yang PS, Song C, Yu HT, Kim TH, Uhm JS, et al. Physical activity and risk of atrial fibrillation: a nationwide cohort study in general population. Sci Rep. 2019;9(1):1-9.

14. World Health Rankings. France: Coronary Heart Disease. 2020.

15. Arrhythmia Alliance. Arrhythmia Facts and Statistics: What You Need to Know. 2022.