Parvovirus B19 - The cause of systemic sclerosis.

Ilija Barukčić
Internist: Horandstrase, DE-26441, Jever, Germany.
Email: Barukcic@t-online.de

How to cite this paper: Ilija Barukčić (2018) Parvovirus B19 – The cause of systemic sclerosis, **viXra**, **1**, 1-28.
http://vixra.org/author/ilija_barukcic

Received: 2018 01, 31
Accepted: 2018 01, 31
Published: 2018 01, 31

Abstract

Objective: Parvovirus B19 appears to be associated with several diseases, one among those appears to be systemic sclerosis. Still, there is no evidence of a causal link between parvovirus B19 and systemic sclerosis.
Methods: To explore the cause-effect relationship between Parvovirus B19 and systemic sclerosis, a systematic review and re-analysis of studies available and suitable was performed. The method of the conditio sine qua non relationship was used to proof the hypothesis *without* Parvovirus B19 infection *no* systemic sclerosis. The mathematical formula of the causal relationship k was used to proof the hypothesis, whether there is a cause effect relationship between Parvovirus B19 and systemic sclerosis. Significance was indicated by a p-value of less than 0.05
Result: The data analyzed support the Null-hypothesis that without Parvovirus B19 infection no systemic sclerosis. In the same respect, the studies analyzed provide evidence of a (highly) significant cause effect relationship between Parvovirus B19 and systemic sclerosis.
Conclusion: This study supports the conclusion that Parvovirus B19 is the cause of systemic sclerosis.

Keywords

Parvovirus B19, systemic sclerosis, causal relationship

1. Introduction

Historically, the first documented description of a scleroderma-like disease is credited to Curzio [1] [2] in Naples in 1753. About 100 years later, in 1847 Gintrac [3] coined the term scleroderma, as the skin was the most obvious organ involved. Systemic sclerosis (SSc) is a rare and clinically heterogeneous generalized autoimmune disorder (AID). In point of fact, the number of people in North America, Australia and Europe which suffer from systemic sclerosis (SSc) right now *(prevalence)* [4]-[7] is between (13 - 105) and (13 - 140) per million. The number of new systemic sclerosis cases (incidence) [8]-[13] has been reported between 2.6 and 20 to 28 per million per year.

Systemic sclerosis (SSc) may affects the skin, the internal organs such as lungs, heart and gastrointestinal tract and kidneys and is characterized among other features by a massive deposition of collagen and an excessive extracellular matrix deposition and fibrosis of the connective tissues of organs. Several survival studies [14] have indicated that systemic sclerosis is a life threatening [15] disease. Most patients die of renal or cardiopulmonary disease. Today, systemic sclerosis is treatable, but not curable. Thus far, an early diagnosis [16] is very important and can help to individually tailor a therapy and to manage this disease. Today, there is no ideal drug available for the treatment of SSc. The therapy [14] of this multifactorial disease involves vasoactive substances (Nifedipin, Captopril, Enalapril, Iloprost), antiinflammatory and immunesuppressive substances (Methylprednisolone, Azathioprine, Cyclophosphamide), antifibrotic substances (D-Penicillamin, Penicillin G, PUVA), gastroenterologics (Omeprazol, Ranitidin, Metoclopramid) et cetera. In view cases, physical therapy and psychotherapy are also important adjunctive therapies.

2. Material and methods

2.1.0. Search strategy

A systematic literature search and review according to a predefined protocol in PubMed, Google scholar and other sites was conducted to identify relevant studies published while reporting follows the PRISMA statements as much as possible [17]. A combination of different keywords like: review, bacterium, colorectal cancer, virus et cetera has been used in the search filed to search for eligible articles. In addition, the reference lists of the relevant articles including review articles was additionally used as a possible source for identifying studies related to the topic. Titles and abstracts of all identified articles were checked. Studies with potential relevance for the study topic underwent a review only if detailed data information could be extracted without any data access barriers.

2.1.1. Study of Ferri et al. 1999 (Italy)

Ferri et al. [18] investigated the prevalence of human parvovirus B19 (B19) infection in the bone marrow of twenty-one consecutive systemic sclerosis (SSc) patients and 15 sex- and age-matched subjects. The presence of human parvovirus B19 DNA was demonstrated in the bone marrow biopsies from SSc patients (12/21; 57%) and was never detected in the control group. The data as obtained by Ferri et al are presented by the 2 by 2-table (**Table 1**).

Table 1. Human parvovirus B19 and systemic sclerosis due to Ferri et al. (1999)

		Systemic sclerosis		
		yes	**no**	**Total**
Parvovirus B19 DNA	**yes**	12	0	12
	no	9	15	24
	Total	21	15	36

2.1.2. Study of Ohtsuka et al. 2004 (Japan)

Ohtsuka et al. [19] investigated patients with SSc (n = 48) and normal subjects (n = 97) using nested polymerase chain reaction (PCR) to determine whether human parvovirus B19 DNA can be detected in SSc skin tissue specimens. The occurrence rate of parvovirus B19 DNA in normal controls (50 of 97, 52%) in comparison with in SSc skin tissues (36 of 48, 75%) was significantly differtent. The data as obtained by Ohtsuka et al. are presented by the 2 by 2-table (**Table 2**).

Table 2. Human parvovirus B19 and systemic sclerosis due to Ohtsuka et al. (2004)

		Systemic sclerosis		
		yes	**no**	**Total**
Parvovirus B19 DNA	**yes**	36	50	86
	no	12	47	59
	Total	48	97	145

2.1.3. Study of Zakrzewska et al. 2009 (Italy)

Zakrzewska et al. [20] re-investigated a possible association between parvovirus B19 (B19V) infection and systemic sclerosis (SSc). B19V DNA was detected in 17/29 SSc patients **bone marrow** compared to 0/10 healthy controls.The data as obtained by Zakrzewska et al. are presented by the 2 by 2-table (**Table 3**).

Table 3. Human parvovirus B19 and systemic sclerosis due to Zakrzewska et al. (2009)

		Systemic sclerosis		
		yes	**no**	**Total**
Parvovirus B19 DNA	**yes**	17	0	17
	no	12	10	22
	Total	29	10	39

2.1.4. Study of Bilgin et al. 2015 (Turkey)

Bilgin et al. investigated [21] the presence of different antibodies against Helicobacter pylori, cytomegalovirus (CMV), Epstein-Barr virus (EBV), and parvovirus B19 in in sera samples obtained from 30 SSc patients and 30 healthy controls. Bilgin et al. found antibodies against parvovirus B19 elevated in 26/30 of the cases compared to 19/30 of controls. The data as obtained by Bilgin et al. are presented by the 2 by 2-table (**Table 4**).

Table 4. Human parvovirus B19 and systemic sclerosis due to Bilgin et al. (2015)

		Systemic sclerosis		
		yes	**no**	**Total**
Parvovirus B19 DNA	**yes**	26	19	45
	no	4	11	15
	Total	30	30	60

2.3. Statistical Analysis

All statistical analyses were performed with Microsoft Excel version 14.0.7166.5000 (32-Bit) software (Microsoft GmbH, Munich, Germany).

2.3.1. Bernoulli trials

Among some discrete distributions like the hypergeometric distribution, the Poisson distribution et cetera the binomial distribution is of special interest. Sometimes, the binomial distribution is called the Bernoulli distribution in honor of the Swiss mathematician Jakob Bernoulli (1654-1705), who derived the same. Bernoulli trials are an essential part of the Bernoulli distribution. Thus far, let us assume two fair coins named as ${}_0W_t$ and as ${}_RU_t$. In our model, *heads* of such a coin are considered as success T (i.e. true) and labeled as +1 while *tails* may be considered as failure F (i. e. false) and are labeled as +0. Such a coin is called a *Bernoulli-Boole coin*. The probability of success of ${}_RU_t$ at one single Bernoulli trial t is denoted as

$$p({}_RU_t = +1) \equiv p({}_RU_t) \tag{1}$$

The probability of failure of ${}_R\underline{U}_t$ at one single Bernoulli trial t is denoted as

$$p({}_RU_t = +0) \equiv p({}_R\underline{U}_t) \equiv 1 - p({}_RU_t) \tag{2}$$

Furthermore, no matter how many times an experiment is repeated, let the probability of a head or the tail remain the same. The trials are independent which implies that no matter

how many times an experiment is repeated, the probability of a single event at a single trial remain the same. Repeated independent trials which are determined by the characteristic that there are always only two possible outcomes, either +1 or +0 and that the probability of an event (outcome) remain the same at each single trial for all trials are called *Bernoulli trials*. The definition of Bernoulli trials provides a theoretical model which is of further use. However, in many practical applications, we may by confronted by circumstances which may be considered as approximately satisfying Bernoulli trials. Thus far, let us perform an experiment of tossing two fair coins simultaneously. Suppose two fair coins are tossed twice. Then there are 2^2=4 possible outcomes (the sample space), which may be shown as

$([{}_R U_t = +1], [{}_0 W_t = +1]), ([{}_R U_t = +1], [{}_0 \underline{W}_t = +0]), ([{}_R \underline{U}_t = +0], [{}_0 W_t = +1]), ([{}_R \underline{U}_t = +0], [{}_0 \underline{W}_t = +0])$

This may also be shown as a 2-dimensional sample space in the form of a contingency table (**Table 5**).

Table 5. The sample space of a contingency table

		Conditioned		
		Yes = +1	**No = +0**	**Total**
Condition	**Yes =+1**	**$([{}_R U_t = +1], [{}_0 W_t = +1])$**	$([{}_R U_t = +1], [{}_0 \underline{W}_t = +0])$	**${}_R U_t$**
	No = +0	$([{}_R \underline{U}_t = +0], [{}_0 W_t = +1])$	$([{}_R \underline{U}_t = +0], [{}_0 \underline{W}_t = +0])$	${}_R \underline{U}_t$
	Total	**${}_0 W_t$**	${}_0 \underline{W}_t$	**${}_R W_t$**

In the following, the contingency table is defined more precisely (**Table 6**).

Table 6. The sample space of a contingency table

		Conditioned		
		Yes = +1	**No = +0**	**Total**
Condition	**Yes =+1**	a	b	**${}_R U_t$**
	No = +0	c	d	${}_R \underline{U}_t$
	Total	**${}_0 W_t$**	${}_0 \underline{W}_t$	**N = ${}_R W_t$**

In general it is $(a+c) = {}_0 W_t$, $(a+b) = {}_R U_t$, $(c+d) = {}_0 \underline{W}_t$, $(b+d) = {}_R \underline{U}_t$ and $a+b+c+d=N={}_R W_t$. Equally, it is ${}_0 W_t + {}_0 \underline{W}_t = {}_R U_t + {}_R \underline{U}_t = {}_R W_t = N$. Thus far, if one fair coin is tossed *n* times, we have *n* repeated Bernoulli trials and an *n* dimensional sample space with 2^n sample points is generated. In general, when given *n* Bernoulli trials with *k* successes, the probability to obtain exactly *k* successes in *n* Bernoulli trials is given by

$$p(k) = \binom{n}{k} \times p({}_R U_t = +1)^k \times \left(1 - p({}_R U_t = +1)\right)^{n-k} \qquad (3)$$

The random variable k is sometimes called a *binomial variable*. The probability to obtain

k events or more (*at least k events*) in *n* trials is calculated as

$$p(k \geq X) = p(k = X) + p(k > X) = \sum_{k=X}^{k=n} \left(\binom{n}{k} \times p({}_{R}U_{t} = +1)^{k} \times (1 - p({}_{R}U_{t} = +1))^{n-k} \right) \quad (4)$$

The probability to obtain less than *k* events in *n* Bernoulli trials is calculated as

$$p(k < X) = 1 - p(k \geq X) = 1 - \sum_{k=X}^{k=n} \left(\binom{n}{k} \times p({}_{R}U_{t} = +1)^{k} \times (1 - p({}_{R}U_{t} = +1))^{n-k} \right) \quad (5)$$

2.3.2. Sufficient condition (conditio per quam)

The formula of the conditio per quam [22]-[36] relationship was derived as

$$p(\text{Parvovirus B19 DNA} \rightarrow \text{Systemic sclerosis}) \equiv \frac{a + c + d}{N} \quad (6)$$

and used to proof the hypothesis: *if* presence of Parvovirus B19 infection (EBV DNA) *then* presence of systemic sclerosis.

2.3.3. Necessary condition (conditio sine qua non)

The formula of the conditio per quam [22]-[36] relationship was derived as

$$p(\text{Parvovirus B19 DNA} \leftarrow \text{Systemic sclerosis}) \equiv \frac{a + b + d}{N} \quad (7)$$

and used to proof the hypothesis: *without* presence of Parvovirus B19 infection *no* presence of systemic sclerosis.

2.3.4. Necessary and sufficient condition

The necessary and sufficient condition relationship was defined [22]-[36] as

$$p(\text{Parvovirus B19 DNA} \leftrightarrow \text{Systemic sclerosis}) \equiv \frac{a + d}{N} \quad (8)$$

Scholium.

Historically, the notion sufficient condition is known since thousands of years. Many authors testified original contributions of the notion material implication only for Diodorus Cronus. Still, Philo the Logician (~ 300 BC), a member of a group of early Hellenistic philosophers (the Dialectical school), is the main forerunner of the notion material impli-

cation and has made some groundbreaking contributions [36] to the basics of this relationship. As it turns out, it is very hard to think of the "conditio per quam" relationship without considering the historical background of this concept. Remarkable as it is, Philo's concept of the material implications came very close to that of modern concept material implication. In propositional logic, a conditional is generally symbolized as "$\mathbf{p \rightarrow q}$" or in spoken language "if p then q". Both q and p are statements, with q the consequent and p the antecedent. Many times, the logical relation between the consequent and the antecedent is called a **material implication**. In general, a conditional "if p then q" is false only if p is true and q is false otherwise, in the three other possible combinations, the conditional is always true. In other words, to say that p is a sufficient condition for q is to say that the presence of p guarantees the presence of q. In particular, it is impossible to have p without q. *If p is present, then q must be present too*. To show that p is not sufficient for q, we come up with cases where p is present but q is not. It is well-known that the notion of a necessary condition can be used in defining what a sufficient condition is (and vice versa). In general, p is a necessary condition for q if it is impossible to have q without p. In fact, the absence of p guarantees the absence of q. **Example (Condition: Our earth).** Without oxygen no fire. The following table (**Table 7**) may demonstrate this relationship.

Table 7. Without Oxygen no fire (on our planet earth).

		Fire		
		Yes = +1	**No = +0**	**Total**
Oxygen	**Yes =+1**	a	b	$\mathbf{{}_{R}U_{t}}$
	No = +0	**0**	d	${}_{R}\underline{U}_{t}$
	Total	$\mathbf{{}_{0}W_{t}}$	${}_{0}\underline{W}_{t}$	$\mathbf{N = {}_{R}W_{t}}$

In contrast to such a point of view, the opposite point of view is correct too. Thus far, there is a straightforward way to give a precise and comprehensive account of the meaning of the term necessary or sufficient condition itself. In other words, **if** fire is present **then** oxygen is present too. The following table (**Table 8**) may demonstrate this relationship.

Table 8. If fire is present **then** oxygen is present too (on our planet earth).

		Oxygen		
		Yes = +1	**No = +0**	**Total**
Fire	**Yes =+1**	a	**0**	$\mathbf{{}_{R}U_{t}}$
	No = +0	c	d	${}_{R}\underline{U}_{t}$
	Total	$\mathbf{{}_{0}W_{t}}$	${}_{0}\underline{W}_{t}$	$\mathbf{N = {}_{R}W_{t}}$

Especially, necessary and sufficient conditions are converses of each other. Still, *the fire is not the cause of oxygen* and vice versa. *Oxygen is note the cause of fire*. In this example before, oxygen is a necessary condition, *a conditio sine qua non*, of fire. A necessary condition is sometimes also called "an essential condition" or a conditio sine qua non. In propositional logic, a necessary condition, a condition sine qua non, is generally symbolized as

"**p ← q**" or in spoken language "**without** p **no** q". Both q and p are statements, with p the antecedent and q the consequent. To show that p is not a necessary condition for q, it is necessary to find an event or circumstances where q is present (i. e. an illness) but p (i. e. a risk factor) is not. On any view, (classical) logic has as one of its goals to characterize the most basic, the most simple and the most general laws of objective reality. Especially, in classical logic, the notions of necessary conditions, of sufficient conditions of necessary and sufficient conditions et cetera are defined very precisely for a single event, for a single Bernoulli trial t. In point of fact, no matter how many times an experiment is repeated, the relationship of the conditio sine qua or of the conditio per quam which is defined for every single event will remain the same. Under conditions of independent trials this implies that no matter how many times an experiment is repeated, the probability of the conditio sine qua or of the conditio per quam of a single event at a single trial *t* remain the same which transfers the relationship of the conditio sine qua or of the conditio per quam et cetera into the sphere of (Bio-) statistics. Consequently, (Bio) statistics generalizes the notions of a sufficient or of a necessary condition from one single Bernoulli trial to N Bernoulli trials. However, in many practical applications, we may by confronted by circumstances which may be considered as approximately satisfying the notions of a sufficient or of a necessary condition. Thus far, under these circumstances, we will need to perform some tests to investigate, can we rely on our investigation.

2.3.5. The central limit theorem

Many times, for some reason or other it is not possible to study exhaustively a whole population. Still, sometimes it is possible to draw a sample from such a population which itself can be studied in detail and used to convince us about the properties of the population. Roughly speaking, statistical inference derived from a randomly selected subset of a population (a sample) can lead to erroneous results. The question raised is how to deal with the uncertainty inherent in such results? The concept of confidence intervals, closely related to statistical significance testing, was formulated to provide an answer to this problem. Confidence intervals, introduced to statistics by Jerzy Neyman in a paper published in 1937 [38], specifies a range within a parameter, i. e. the population proportion π, with a certain probability, contain the desired parameter value. Most commonly, the 95% *confidence interval* is used. Interpreting a confidence interval involves a couple of important but subtle issues. In general, a 95% confidence interval for the value of a random number means that there is a 95% probability that the "true" value of the value of a random number is within the interval. Confidence intervals for proportions or a population mean of random variables which are not normally distributed in the population can be constructed while relying on the central limit theorem as long as the sample sizes and counts are big enough (i. e. a sample size of *n=30* and more). A formula, justified by the central limit theorem, is known as

$$p_{Crit} = p_{Calc} \pm \left(z_{Alpha/2} \times \left(\sqrt[2]{\frac{1}{N} \times p_{Calc} \times (1 - p_{Calc})} \right) \right)$$

(9)

where p_{Calc} is the sample proportion of successes in a Bernoulli trial process with N trials yielding X successes and N-X failures and z is i.e the 1 – (Alpha/2) quantile of a standard normal distribution corresponding to the significance level alpha. For example, for a 95% confidence level alpha = 0.05 and z is z = 1.96. A very common technique for calculating binomial confidence intervals was published by Clopper-Pearson [39]. Agresti-Coull proposed another different method [40] for calculating binomial confidence intervals. A faster and an alternative way to determine the lower and upper "exact" confidence interval is justified by the F distribution [41].

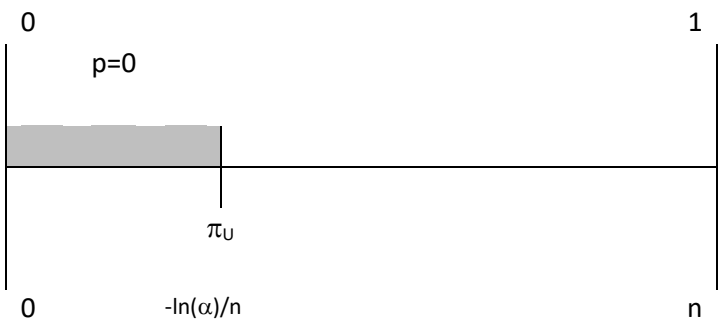
2.3.6. The rule of three

Furthermore, an approximate and conservative (one sided) confidence interval was developed by Louis [42], Hanley et al. [43] and Jovanovic [44] known as the rule of three. Briefly sketched, the rule of three can be derived from the binomial model. Let π_U denote the upper limit of the one-sided $100 \times (1 - \alpha)\%$ confidence interval for the unknown proportion when in N independent trials **no events occur** [44]. Then π_U is the value such that

$$\pi_U = \left(\frac{-\ln(\alpha)}{n} \right) \approx \left(\frac{3}{n} \right) \tag{10}$$

assuming that $\alpha = 0{,}05$. In other words, an one-sided approximate upper 95% confidence bound for the true binomial population proportion π, the rate of occurrences in the population, based on a sample of size n where no successes are observed (p=0) is 3/n [44] or given approximately by $[0 < \pi < (3/n)]$. The rule of three is a useful tool especially in the analysis of medical studies. The following table (**Table 9**) will illustrate this relationship.

Table 9. The one-sided approximate upper $100 \times (1 - \alpha)\%$ confidence bound where **no successes** (p=0) are observed.



Under conditions where **a certain event did not occur** [42] in a sample with *n* subjects (i. e. **p=0**) the interval from 0 to $(-\ln(\alpha)/n)$ is called a $100 \times (1 - \alpha)\%$ confidence interval for

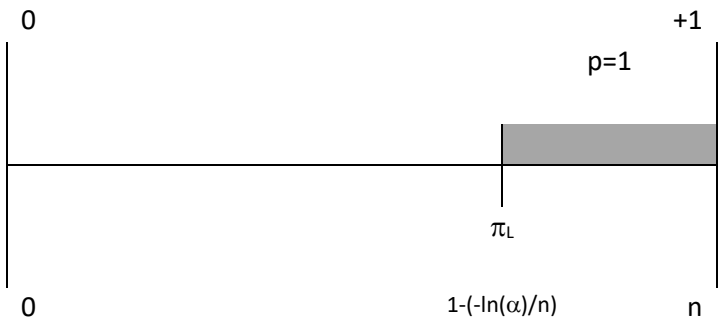
the binomial parameter for the rate of occurrences in the population.

Another special case of the binomial distribution is based on a sample of size *n* where **only successes are observed (p=1)**. Accordingly, the lower limit of a one-sided 100 × (1- α)% confidence interval for a binomial probability π_L, the rate of occurrences in the population, based on a sample of size *n* where only successes are observed is given approximately by $[(1-(-\ln(\alpha)/n)) < \pi < +1]$ or (assuming $\alpha=0{,}05$)

$$\pi_L = 1-\left(\frac{-\ln(\alpha)}{n}\right) \approx 1-\left(\frac{3}{n}\right) \tag{11}$$

The following table (**Table 10**) may illustrate this relationship.

Table 10. The one-sided approximate upper 100 × (1 - α)% confidence bound where **only successes** are observed.



To construct a two-sided 100 × (1 – (α))% interval according to the rule of three, it is necessary to take a one-sided 100 × (1 - (α/2))% confidence interval. In this study, we will use the rule of three [45] too, to calculate the confidence interval for the value of a random number.

2.3.7. Fisher's exact test

A test statistics of independent and more or less normally distributed data which follow a chi-squared distribution is valid as with many statistical tests due to the central limit theorem. Especially, with large samples, a chi-squared distribution can be used. A sample is considered as large when the sample size n is n = 30 or more. With a small sample (n < 30), the central limit theorem does not apply and erroneous results could potentially be obtained from the few observations if the same is applied. Thus far, when the number of observations obtained from a population is too small, a more appropriate test for of analysis of categorical data i. e. contingency tables is R. A. Fisher's exact test [46]. Fisher's exact test is valid for all sample sizes and calculates the significance of the p-value (i. e. the deviation from a null hypothesis) exactly even if in practice it is employed when sample

size is small. Fisher's exact test is called exact because the same uses the exact hypergeometric distribution to compute the p-value rather than the approximate chi-square distribution. Still, computations involved in Fisher's exact test can be time consuming to calculate by hand.

2.3.8. Hypergeometric distribution

The hypergeometric distribution, illustrated in a table (**Table 11**), is a discrete probability distribution which describes the probability of ***a*** events/successes in a sample with the size ${}_0W_t$, without replacement, from a finite population of the size N which contains exactly ${}_RU_t$ objects with a certain feature while each event is either a success or a failure. The formula for the hypergeometric distribution, a discrete probability distribution, is

$$p(a) = \frac{\binom{{}_RU_t}{a} \times \binom{N - {}_RU_t}{{}_0W_t - a}}{\binom{N}{{}_0W_t}} \tag{12}$$

Table 11. The hypergeometric distribution

		Conditioned		
		Yes = +1	**No = +0**	**Total**
Condition	**Yes =+1**	**a**	$b=({}_RU_t - a)$	**${}_RU_t$**
	No = +0	$c=({}_0W_t - a)$	$N-{}_RU_t - {}_0W_t + a$	$N-{}_RU_t$
	Total	**${}_0W_t$**	$N - {}_0W_t$	**N**

The hypergeometric distribution has a wide range of applications. The Hypergeometric distribution can be approximated by a Binomial distribution. The elements of the population being sampled are classified into one of two mutually exclusive categories: **either** conditio sine qua non **or** no conditio sine qua non relationship. We are sampling without replacement from a finite population. How probable is it to draw specific ***c*** events/successes out of ${}_0W_t$ total draws from an aforementioned population of the size N? The hypergeometric distribution, as shown in a table (**Table 12**) is of use to calculate how probable is it to obtain $c=({}_0W_t - a)$ events out of N events.

Table 12. The hypergeometric distribution and conditio sine qua non

		Conditioned		
		Yes = +1	**No = +0**	**Total**
No Condition	**Yes =+1**	$c=({}_0W_t - a)$	$N-{}_RU_t - {}_0W_t + a$	$N-{}_RU_t$
	No = +0	a	$b=({}_RU_t - a)$	${}_RU_t$
	Total	**${}_0W_t$**	$N - {}_0W_t$	**N**

2.3.9. Statistical hypothesis testing

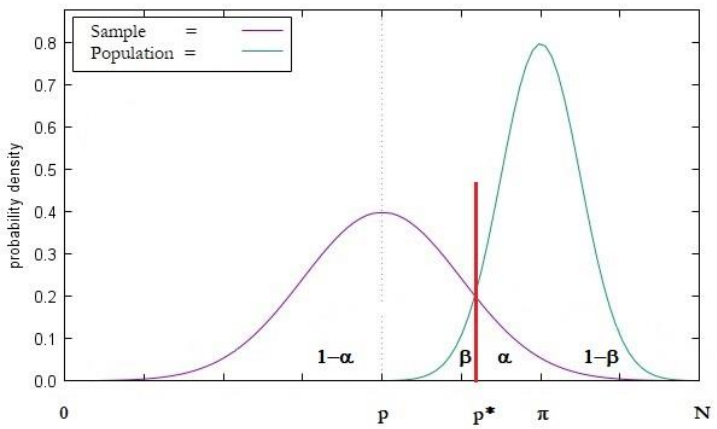
A statistical hypothesis test is a method to extract some inferences from data. A hypothesis is compared as an alternative hypothesis. Under which conditions does the outcomes of a study lead to a rejection of the null hypothesis for a pre-specified level of significance. According to the rules of a proof by contradiction, a null hypothesis (H_0) is a statement which one seeks to disproof. The related specific alternative hypothesis (H_A) is opposed to the null hypothesis such that if null hypothesis (H_0) is true, the alternative hypothesis (H_A) is false and vice versa. If the alternative hypothesis (HA) is true then the null hypothesis (H0) is false. In principle, a null hypothesis that is true can be rejected (type I error) which lead us to falsely infer the existence of something which is not given. The significance level, also denoted as α (alpha) is the probability of rejecting a null hypothesis when the same is true. A type II error is given, if we falsely infer the absence of something which in reality is given. A null hypothesis can be false but a statistical test may fail to reject such a false null hypothesis. The probability of accepting a null hypothesis when the same is false (type II error), is denoted by the Greek letter ß (beta) and related to the power of a test (which equals 1-ß). The power of a test indicates *the probability by which the test correctly rejects the null hypothesis (H_0) when a specific alternative hypothesis (H_A) is true*. Most investigator assess the power of a tests using **1-ß** = 0.80 as a standard for adequacy. A tabularized relation between truth/falseness of the null hypothesis and outcomes of the test are shown precisely within a table (**Table 13**).

Table 13. Table of error types

		Null Hypothesis (H_0) is		
		True	**False**	**Total**
Null Hypothesis (H_0)	**Accepted**	1-α	β	1-α+ß
	Rejected	α	**1-ß**	1+α-ß
	Total	1	1	**2**

In general, it is 1- α + α = 1 or (1- α -β) + α = **1- β**. The following figure may illustrate these relationships (**Figure 1**).

Figure 1 . The relationship between error types.



2.3.10. The mathematical formula of the causal relationship k

The mathematical formula of the causal relationship k [22]-[36] defined as

$$k\left({}_{R}U_{t}, {}_{0}W_{t}\right) \equiv \frac{\left(\left(N \times a\right) - \left({}_{R}U_{t} \times {}_{0}W_{t}\right)\right)}{\sqrt[2]{\left({}_{R}U_{t} \times {}_{R}\underline{U}_{t}\right) \times \left({}_{0}W \times {}_{0}\underline{W}_{t}\right)}} \tag{13}$$

and the chi-square distribution [47] were applied to determine the significance of causal relationship between a EBV and HL. A one-tailed test makes it much more easier to reject a null hypothesis (no causal relationship) while a two-tailed test makes it more difficult to reject a null hypothesis and is more conservative on this account. For this reason, in causal relationship testing, a two-tailed test is preferred. In general, a p value of less than 0.05 is considered as significant. In this context, what is the necessary connection between a cause and effect? What ties a cause and its own effect together? Is there a necessary connection between a cause and effect at all? Theoretically, it is neither justified nor necessary to reduce causation as such to an act of observation or measurement. Sill, case-control studies, experiments, observations et cetera can help us to recognize cause effect relationships. In this context it is necessary to stress out that **every single event (effect) has its own cause**, which is the logical foundation of the mathematical formula of the causal relationship k. It is therefore entirely clear that this is the fundamental difference to Pearson's methodological approach. Obviously, although under some certain specified circumstances Pearson's product-moment correlation coefficient [48] or Pearson's Phi [49] coefficient can yield the same numerical result as the mathematical formula of the causal relationship k, there is nothing truly exciting about such a coincidence. Nevertheless, when conducting experiments and analyzing data, views in which correlation and causation are brought very close together are incorrect and worthless. The mathematical formula of the causal relationship k is neither identical nor can the same mathematical formula be reduced to Pearson's product-moment correlation coefficient [48] or to Pearson's Phi [49] Coefficient (Mean Square Contingency Coefficient). In contrast to Pearson's product-moment correlation coefficient and to Pearson's Phi Coefficient (Mean Square Contingency Coefficient) the mathematical formula of the causal relationship k is defined and valid at every single Bernoulli trial *t* or at every single event. Sir Austin Bradford Hill (1897 - 1991), an English epidemiologist, proposed 1965 a set of nine criteria (Strength, Consistency, Specificity, Temporality, Biological gradient, Plausibility, Coherence, Experiment, Analogy) [50] to establish epidemiologic evidence of a causal relationship (Bradford Hill criteria). In point of fact, Bredford's "fourth characteristic is the temporal relationship of the association" [50] and in last consequence the "*post hoc ergo propter hoc*" logical fallacy.47 Causation cannot be derived from the "*post hoc ergo propter hoc*" [35] logical fallacy. Consequently, the Mathematical Formula of the causal relationship k can neither be reduced to the Bradford Hill criteria nor is the same just a mathematization of Bradford Hill criteria.

2.3.11. The chi square distribution

The chi-squared distribution [47] is a widely known distribution and used in hypothesis testing, in inferential statistics or in construction of confidence intervals. The critical values of the chi square distribution are visualized by **Table 14**.

Table 14. The critical values of the chi square distribution (degrees of freedom: 1).

	p-Value	One sided X^2	Two sided X^2
The chi square distribution	0,1000000000	1,642374415	2,705543454
	0,0500000000	**2,705543454**	**3,841458821**
	0,0400000000	3,06490172	4,217884588
	0,0300000000	3,537384596	4,709292247
	0,0200000000	4,217884588	5,411894431
	0,0100000000	5,411894431	6,634896601
	0,0010000000	9,549535706	10,82756617
	0,0001000000	13,83108362	15,13670523
	0,0000100000	18,18929348	19,51142096
	0,0000010000	22,59504266	23,92812698
	0,0000001000	27,03311129	28,37398736
	0,0000000100	31,49455797	32,84125335
	0,0000000010	35,97368894	37,32489311
	0,0000000001	40,46665791	41,82145620

2.3.12. The X^2 goodness of fit test

A chi-square goodness of fit test can be applied to determine whether sample data are consistent with a hypothesized distribution. The chi-square goodness of fit test is appropriate when some conditions are met. A view of these conditions are simple random sampling, categorical variables and an expected value of the number of sample observations which is at least 5. The null hypothesis (H_0) and its own alternative hypothesis (H_A) are stated in such a way that they are mutually exclusive. In point of fact, if the null hypothesis (H_0) is true, the other, alternative hypothesis (H_A), must be false; and vice versa. For a chi-square goodness of fit test, the hypotheses can take the following form.

H_0: The sample distribution agrees with the hypothetical (theoretical) distribution.
H_A: The sample distribution does not agree with the hypothetical (theoretical) distribution.

The X^2 Goodness-of-Fit Test can be shown schematically as

$$\chi^2 \equiv \sum_{t=+1}^{t=+N} \left(\frac{\left(\text{Observed}_t - \text{Expected}_t\right)^2}{\text{Expected}_t} \right) \tag{14}$$

The degrees of freedom are calculated as N-1. If there is no discrepancy between an observed and a theoretical distribution, then $X^2=0$. As the discrepancy between an observed

and a theoretical distribution becomes larger, the X^2 becomes larger. This X^2 values are evaluated by the known X^2 distribution.

The original X^2 values are calculated from an original theoretical distribution, which is continuous, whereas the approximation by the X^2 Goodness of fit test we are using is discrete. Thus far, there is a tendency to underestimate the probability, which means that the number of rejections of the null hypothesis can increase too much and must be corrected downward. Such an adjustment (*Yate's correction for continuity*) is used only when there is one degree of freedom. When there is more than one degree of freedom, the same adjustment is not used. Applying this to the formula above, we find the X^2 Goodness-of-Fit Test *with continuity correction* shown schematically as

$$\chi^2 \equiv \sum_{t=+1}^{t=+N} \left(\frac{\left(\left| \text{Observed}_t - \text{Expected}_t \right| - \left(\frac{1}{2} \right) \right)^2}{\text{Expected}_t} \right) \quad (15)$$

When the term ($|\text{Observed}_t - \text{Expected}_t|$) is less than ½, the continuity correction should be omitted.

2.3.12.1. The X^2 goodness of fit test of a sufficient condition

The theoretical (hypothetical) distribution of a sufficient condition is shown schematically by the 2x2 table (**Table 15**).

Table 15. The theoretical distribution of a sufficient condition (conditio pre quam).

		Conditioned		
		Yes = +1	**No = +0**	**Total**
Condition	**Yes =+1**	a	**b=0**	(a+b)
	No = +0	c	d	(c+d)
	Total	(a+c)	(b+d)	(a+b+c+d)

The theoretical distribution of a sufficient condition (conditio pre quam) is determined by the fact that **b=0**. The X^2 Goodness-of-Fit Test *with continuity correction* of a sufficient condition (conditio per quam) is calculated as

$$\chi^2(\text{IMP}) \equiv \left(\frac{\left(\left| a - (a+b) \right| - \left(\frac{1}{2} \right) \right)^2}{(a+b)} \right) + \left(\frac{\left(\left| (c+d) - (c+d) \right| - \left(\frac{1}{2} \right) \right)^2}{(c+d)} \right) = \left(\frac{\left(\left| a - (a+b) \right| - \left(\frac{1}{2} \right) \right)^2}{(a+b)} \right) + 0 \quad (16)$$

or more simplified as

$$\chi^2(\mathrm{IMP}) \equiv \left(\frac{\left(|-b| - \left(\frac{1}{2}\right) \right)^2}{(a+b)} \right) + 0 \tag{17}$$

Under these circumstances, the degree of freedom is d.f. = N-1=2-1=1.

2.3.12.2. The X^2 goodness of fit test of a necessary condition

The theoretical (hypothetical) distribution of a necessary condition is shown schematically by the 2x2 table (**Table 16**).

Table 16. The theoretical distribution of a necessary condition (conditio sine qua non).

		Conditioned		
		Yes = +1	**No = +0**	**Total**
Condition	**Yes =+1**	a	b	(a+b)
	No = +0	**c=0**	d	(c+d)
	Total	(a+c)	(b+d)	(a+b+c+d)

The theoretical distribution of a necessary condition (conditio sine qua non) is determined by the fact that **c=0**. The X^2 Goodness-of-Fit Test *with continuity correction* of a necessary condition (conditio sine qua non) is calculated as

$$\chi^2(\mathrm{SINE}) \equiv \left(\frac{\left(\left|(a+b)-(a+b)\right| - \left(\frac{1}{2}\right) \right)^2}{(a+b)} \right) + \left(\frac{\left(\left|(d)-(c+d)\right| - \left(\frac{1}{2}\right) \right)^2}{(c+d)} \right) = 0 + \left(\frac{\left(\left|d-(c+d)\right| - \left(\frac{1}{2}\right) \right)^2}{(c+d)} \right) \tag{18}$$

or more simplified as

$$\chi^2(\mathrm{SINE}) \equiv \left(\frac{\left(|-c| - \left(\frac{1}{2}\right) \right)^2}{(c+d)} \right) + 0 \tag{19}$$

Under these circumstances, the degree of freedom is d.f. = N-1=2-1=1.

2.3.12.2. The X^2 goodness of fit test of a necessary and sufficient condition

The theoretical (hypothetical) distribution of a necessary and sufficient condition is shown schematically by the 2x2 table (**Table 17**).

Table 17. The theoretical distribution of a necessary and sufficient condition.

		Conditioned		
		Yes = +1	**No = +0**	**Total**
Condition	**Yes =+1**	a	**b=0**	(a+b)
	No = +0	**c=0**	d	(c+d)
	Total	(a+c)	(b+d)	(a+b+c+d)

The theoretical distribution of a necessary and sufficient condition is determined by the fact that **b=0** and that **c=0**. The X^2 Goodness-of-Fit Test *with continuity correction* of a

necessary and sufficient condition is calculated as

$$\chi^2\left(\text{Necessary AND Sufficient}\right) \equiv \left(\frac{\left(\left| (a) - (a+b) \right| - \left(\frac{1}{2} \right) \right)^2}{(a+b)} \right) + \left(\frac{\left(\left| (d) - (c+d) \right| - \left(\frac{1}{2} \right) \right)^2}{(c+d)} \right) \tag{20}$$

or more simplified as

$$\chi^2\left(\text{Necessary AND Sufficient}\right) \equiv \left(\frac{\left(\left| -b \right| - \left(\frac{1}{2} \right) \right)^2}{(a+b)} \right) + \left(\frac{\left(\left| -c \right| - \left(\frac{1}{2} \right) \right)^2}{(c+d)} \right) \tag{21}$$

Under these circumstances, the degree of freedom is d.f. = N-1=2-1=1.

3. Results

3.1. *Without* a human parvovirus B19 *no* systemic sclerosis

Claims.

Null hypothesis:

An infection by parvovirus B19 is a conditio sine qua non of systemic sclerosis.

(The sample distribution does agree with the theoretical distribution of a conditio sine qua non relationship.)

Alternative hypothesis:

An infection by parvovirus B19 is not a conditio sine qua non of systemic sclerosis.

(The sample distribution does not agree with the theoretical distribution of a conditio sine qua non relationship.)

Significance level (Alpha) below which the null hypothesis will be rejected: 0.05.

Proof.

The data of **Ferri et al. (1999)** [18] of an infection by parvovirus B19 and systemic sclerosis are viewed in the 2 × 2 table (**Table 1**). The X^2 Goodness-of-Fit Test *with continuity correction* of a necessary condition (conditio sine qua non) known to be defined as p(Parvovirus B19 ← Systemic sclerosis) is calculated as

$$\chi^2(\mathrm{SINE}) \equiv \left(\frac{\left(|-c| - \left(\frac{1}{2}\right) \right)^2}{(c+d)} \right) + 0 = \left(\frac{\left(|-9| - \left(\frac{1}{2}\right) \right)^2}{(9+15)} \right) = 3,010416667$$

Under these circumstances, the degree of freedom is d. f. = N-1=2-1=1. The *two sided* critical X^2 (significance level alpha = 0.05) is known to be 3,841458821 (**Table 14**). The calculated X^2 value = 3,010416667 and less than the critical X^2 = 3,841458821. Hence, our calculated X^2 value = 3,010416667 is not significant and we accept our null hypothesis. Due to this evidence, we do not reject the null hypothesis in favor of the alternative hypotheses. In other words, the sample distribution does agree with the theoretical distribution of a necessary condition (conditio sine qua non) relationship. Thus far, the data as published by **Ferri et al. (1999)** [18] do support our null hypothesis: an infection by parvovirus B19 is a conditio sine qua non of systemic sclerosis. In other words, *without* an infection by human parvovirus B19 *no* systemic sclerosis.

Q. e. d.

3.2. *Without* a human parvovirus B19 *no* systemic sclerosis

Claims.

Null hypothesis:

An infection by parvovirus B19 is a conditio sine qua non of systemic sclerosis.

(The sample distribution does agree with the theoretical distribution of a conditio sine qua non relationship.)

Alternative hypothesis:

An infection by parvovirus B19 is not a conditio sine qua non of systemic sclerosis.

(The sample distribution does not agree with the theoretical distribution of a conditio sine qua non relationship.)

Significance level (Alpha) below which the null hypothesis will be rejected: 0.05.

Proof.

The data of **Ohtsuka et al.** [19] of an infection by parvovirus B19 and systemic sclerosis are viewed in the 2 × 2 table (**Table 2**). The X^2 Goodness-of-Fit Test *with continuity correction* of a necessary condition (conditio sine qua non) known to be defined as p(Parvovirus B19 ← Systemic sclerosis) is calculated as

$$\chi^2(\mathrm{SINE}) \equiv \left(\frac{\left(|-c| - \left(\frac{1}{2}\right) \right)^2}{(c+d)} \right) + 0 = \left(\frac{\left(|-12| - \left(\frac{1}{2}\right) \right)^2}{(12+47)} \right) = 2,241525424$$

Under these circumstances, the degree of freedom is d. f. = N-1=2-1=1. The *one sided* critical X^2 (significance level alpha = 0.05) is known to be 2,705543454 (**Table 14**). The calculated X^2 value = 2,241525424 and less than the critical X^2 = 2,705543454. Hence, our calculated X^2 value = 2,241525424 is not significant and we accept our null hypothesis. Due to this evidence, we do not reject the null hypothesis in favor of the alternative hypotheses. In other words, the sample distribution does agree with the theoretical distribution of a necessary condition (conditio sine qua non) relationship. Thus far, the data as published by **Ohtsuka et al.** [19] do support our null hypothesis: an infection by parvovirus B19 is a conditio sine qua non of systemic sclerosis. In other words, *without* an infection by human parvovirus B19 *no* systemic sclerosis.

Q. e. d.

3.3. *Without* a human parvovirus B19 *no* systemic sclerosis

Claims.

Null hypothesis:

An infection by parvovirus B19 is a conditio sine qua non of systemic sclerosis.

(The sample distribution does agree with the theoretical distribution of a conditio sine qua non relationship.)

Alternative hypothesis:

An infection by parvovirus B19 is not a conditio sine qua non of systemic sclerosis.

(The sample distribution does not agree with the theoretical distribution of a conditio sine qua non relationship.)

Significance level (Alpha) below which the null hypothesis will be rejected: 0.05.

Proof.

The data of Bilgin et al. [21] of an infection by parvovirus B19 and systemic sclerosis are viewed in the 2 × 2 table (**Table 4**). The X^2 Goodness-of-Fit Test *with continuity correction* of a necessary condition (conditio sine qua non) known to be defined as p(Parvovirus B19 ← Systemic sclerosis) is calculated as

$$\chi^2(\mathrm{SINE}) \equiv \left(\frac{\left(|-c| - \left(\frac{1}{2}\right) \right)^2}{(c+d)} \right) + 0 = \left(\frac{\left(|-4| - \left(\frac{1}{2}\right) \right)^2}{4+11} \right) = 0{,}816666667$$

Under these circumstances, the degree of freedom is d. f. = N-1=2-1=1. The *one sided* critical X^2 (significance level alpha = 0.05) is known to be 2,705543454 (**Table 14**). The calculated X^2 value = 0,816666667 and less than the critical X^2 = 2,705543454 while the number of sample observations was 4 and not at least 5. Hence, our calculated X^2 value = 0,816666667 is not significant and we accept our null hypothesis. Due to this evidence, we do not reject the null hypothesis in favor of the alternative hypotheses. In other words, the sample distribution does agree with the theoretical distribution of a necessary condition (conditio sine qua non) relationship. Thus far, the data as published by **Bilgin et al.** [21] do support our null hypothesis: an infection by parvovirus B19 is a conditio sine qua non of systemic sclerosis. In other words, *without* an infection by human parvovirus B19 *no* systemic sclerosis.

Q. e. d.

3.4. Human parvovirus B19 is the cause of systemic sclerosis

Claims.

Null hypothesis: (**no causal relationship**)

There is no significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis.

(k=0).

Alternative hypothesis: (**causal relationship**)

There is an significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis.

(k<>0).

Conditions.

Alpha level = 5%.

The two tailed critical Chi square value (degrees of freedom = 1) for alpha level 5% is 3.841458821.

Proof.

The data for this hypothesis test were provided by Ferri et al. [18] and are illustrated in the 2 × 2 table (**Table 1**). The causal relationship k(human parvovirus B19, systemic sclerosis) is calculated [22]-[36] as

$$k\left(\text{Parvovirus B19, Systemic sclerosis}\right)=\frac{\left((36\times12)-(12\times21)\right)}{\sqrt[2]{(21\times15)\times(12\times24)}}=+0{,}5976143046671970$$

The value of the test statistic k=+0,59761430466719700000 is equivalent to a calculated [22]-[36] chi-square value of

$$\chi^2_{\text{Calculated}}=36\times\left(\frac{\left((36\times12)-(12\times21)\right)}{\sqrt[2]{(21\times15)\times(12\times24)}}\right)\times\left(\frac{\left((36\times12)-(12\times21)\right)}{\sqrt[2]{(21\times15)\times(12\times24)}}\right)$$

$$\chi^2_{\text{Calculated}}=36\times0{,}59761430466719700000\times0{,}59761430466719700000$$

$$\chi^2_{\text{Calculated}}=12{,}85714286$$

The chi-square statistic, uncorrected for continuity, is calculated as X^2 = 12,85714286 and thus far equivalent to a P value of 12,85714286. The calculated chi-square statistic exceeds the critical chi-square value of 3.841458821 (**Table 14**). Consequently, we reject the null hypothesis and accept the alternative hypotheses. There is a significant causal relationship between an by infection human parvovirus B19 and systemic sclerosis (k=+0,59761430466719700000, p Value=0,00033619350474802500). The result is significant at $p < 0.01$.

Q. e. d.

3.5. Human parvovirus B19 is the cause of systemic sclerosis

Claims.

Null hypothesis: (**no causal relationship**)

There is no significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis.

(k=0).

Alternative hypothesis: (**causal relationship**)

There is an significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis.

(k<>0).

Conditions.

Alpha level = 5%.

The two tailed critical Chi square value (degrees of freedom = 1) for alpha level 5% is 3.841458821.

Proof.

The data for this hypothesis test were provided by **Ohtsuka et al.** [19] and are illustrated in the 2 × 2 table (**Table 2**). The causal relationship k(human parvovirus B19, systemic sclerosis) is calculated [22]-[36] as

$$k\left(\text{Parvovirus B19, Systemic sclerosis}\right)=\frac{\left((145\times 36)-(86\times 48)\right)}{\sqrt[2]{(48\times 97)\times(86\times 59)}}=+0{,}2246678457782880$$

The value of the test statistic k=+0,2246678457782880 is equivalent to a calculated [22]-[36] chi-square value of

$$\chi^2_{\text{Calculated}}=145\times\left(\frac{\left((145\times 36)-(86\times 48)\right)}{\sqrt[2]{(48\times 97)\times(86\times 59)}}\right)\times\left(\frac{\left((145\times 36)-(86\times 48)\right)}{\sqrt[2]{(48\times 97)\times(86\times 59)}}\right)$$

$$\chi^2_{\text{Calculated}}=145\times 0{,}22466784577828800000\times 0{,}22466784577828800000$$

$$\chi^2_{\text{Calculated}}=7{,}318967934$$

The chi-square statistic, uncorrected for continuity, is calculated as X^2 = 7,318967934 and thus far equivalent to a P value of 0,00682305838508812000. The calculated chi-square statistic exceeds the critical chi-square value of 3.841458821 (**Table 14**). Consequently, we reject the null hypothesis and accept the alternative hypotheses. There is a significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis (k=+0,2246678457782880, p Value = 0,00682305838508812000). The result is significant at $p < 0.05$.

Q. e. d.

3.6. Human parvovirus B19 is the cause of systemic sclerosis

Claims.

Null hypothesis: (**no causal relationship**)

There is no significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis.

(k=0).

Alternative hypothesis: (**causal relationship**)

There is an significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis.

(k<>0).

Conditions.

Alpha level = 5%.

The two tailed critical Chi square value (degrees of freedom = 1) for alpha level 5% is 3.841458821.

Proof.

The data for this hypothesis test were provided by **Zakrzewska et al.** [20] and are illustrated in the 2 × 2 table (**Table 3**). The causal relationship k(human parvovirus B19, systemic sclerosis) is calculated [22]-[36] as

$$k\left(\text{Parvovirus B19, Systemic sclerosis}\right)=\frac{\left((39\times17)-(17\times29)\right)}{\sqrt[2]{(29\times10)\times(17\times22)}}=+0{,}51619538960628400000$$

The value of the test statistic k=+0,51619538960628400000 is equivalent to a calculated [22]-[36] chi-square value of

$$\chi^2_{\text{Calculated}}=39\times\left(\frac{\left((39\times17)-(17\times29)\right)}{\sqrt[2]{(29\times10)\times(17\times22)}}\right)\times\left(\frac{\left((39\times17)-(17\times29)\right)}{\sqrt[2]{(29\times10)\times(17\times22)}}\right)$$

$$\chi^2_{\text{Calculated}}=39\times0{,}51619538960628400000\times0{,}51619538960628400000$$

$$\chi^2_{\text{Calculated}}=10{,}39184953$$

The chi-square statistic, uncorrected for continuity, is calculated as X^2 = 10,39184953 and thus far equivalent to a P value of 0,00126572775613513000. The calculated chi-square statistic exceeds the critical chi-square value of 3.841458821 (**Table 14**). Consequently, we reject the null hypothesis and accept the alternative hypotheses. There is a significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis (k=+0,5161953896062840, p Value =0,00126572775613513000). The result is significant at $p < 0.05$.

Q. e. d.

3.7. Human parvovirus B19 is the cause of systemic sclerosis

Claims.

Null hypothesis: (**no causal relationship**)

There is no significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis.

(k=0).

Alternative hypothesis: (**causal relationship**)

There is an significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis.

(k<>0).

Conditions.

Alpha level = 5%.

The two tailed critical Chi square value (degrees of freedom = 1) for alpha level 5% is 3.841458821.

Proof.

The data for this hypothesis test were provided by **Bilgin et al.** [21] and are illustrated in the 2 × 2 table (**Table 4**). The causal relationship k(human parvovirus B19, systemic sclerosis) is calculated [22]-[36] as

$$k\left(\text{Parvovirus B19, Systemic sclerosis}\right)=\frac{\left(\left(60\times 26\right)-\left(45\times 30\right)\right)}{\sqrt[2]{\left(30\times 30\right)\times\left(45\times 15\right)}}=+0{,}26943012562182500$$

The value of the test statistic k=+0,26943012562182500000 is equivalent to a calculated [22]-[36] chi-square value of

$$\chi^2_{\text{Calculated}}=60\times\left(\frac{\left(\left(60\times 26\right)-\left(45\times 30\right)\right)}{\sqrt[2]{\left(30\times 30\right)\times\left(45\times 15\right)}}\right)\times\left(\frac{\left(\left(60\times 26\right)-\left(45\times 30\right)\right)}{\sqrt[2]{\left(30\times 30\right)\times\left(45\times 15\right)}}\right)$$

$$\chi^2_{\text{Calculated}}=60\times 0{,}26943012562182500000\times 0{,}26943012562182500000$$

$$\chi^2_{\text{Calculated}}=4{,}355555556$$

The chi-square statistic, uncorrected for continuity, is calculated as X^2 = 4,355555556 and thus far equivalent to a P value of 0,03688842570704990. The calculated chi-square statistic exceeds the critical chi-square value of 3.841458821 (**Table 14**). Consequently, we reject the null hypothesis and accept the alternative hypotheses. There is a significant causal relationship between an infection by human parvovirus B19 and systemic sclerosis (k=+0,26943012562182500000, p Value = 0,03688842570704990). The result is significant at $p < 0.05$.

Q. e. d.

4. Discussion

The spectrum of diseases associated with human parvovirus, a small, single-stranded DNA virus, is constantly growing and very wide. Many times a parvovirus B19 infection is more or less asymptomatic or present with flu like clinical presentation. Human parvovirus infection has been found even in a previously healthy pediatric population [51]. About half of 15 year old adolescents have been identifies as parvovirus specific IgG antibody positive [52]. Several bacterial and viral infectious agents have been suspected to be contributing the pathology of systemic sclerosis. The etiology of systemic sclerosis is elusive, although systemic sclerosis is not an inherited disease.

In the current study, we investigated the relationship between and parvovirus B19 infection and systemic sclerosis. Besides of all, this study had several limitations. First, systemic sclerosis is a rare disease. Still, the study groups were very small. Second, the design of the studies was very different.

Besides of the severe limitations, in the present study, the studies presented support the hypothesis that parvovirus B19 is a necessary condition of systemic sclerosis. In other words, without a parvovirus B19 infection no systemic sclerosis. In particular, in the same context, the studies presented support the conclusion that there is a significant relationship between cause and effect. Since without a parvovirus B19 no systemic sclerosis will develop we can deduce that a parvovirus B19 infection is not only a cause but the cause of systemic sclerosis. The clear this point exactly and definitely more and systematic studies are necessary. Until this can be clarified we should accept for preliminary purposes the following unescapable conclusion.

5. Conclusion

A parvovirus B19 infection is the cause of systemic sclerosis.

Acknowledgements

To Michael Aurek Germany.

The public domain software GnuPlot was use to draw the figure.

References

[1] Curzio C. (1753) Discussionil anatomico-pratiche di un raro, e stravagante morbo cutaneo in una giovane donna felicemente curato in questo grande ospedale degl' incurabili. Napoli, Presso Giovanni di Simone. https://catalog.hathitrust.org/Record/009295544 https://babel.hathitrust.org/cgi/pt?id=ucm.5320215940;view=1up;seq=7

[2] Nollet A, Watson R. (1753) An account of an extraordinary disease of the skin and its cure. Extracted from the Italian of Carlo Crusio. *Philosophical Transactions of the Royal Society of London*, **48**, 579-587. https://doi.org/10.1098/rstl.1753.0077

[3] Gintrac M. (1847) Note sur la sclerodermie. *Rev Med Chir. Paris*, **2**, 263-81.

[4] Maricq HR, Weinrich MC, Keil JE, Smith EA, Harper FE, Nussbaum AI, LeRoy EC, McGregor AR, Diat F, Rosal EJ. (1989) Prevalence of scleroderma spectrum disorders in the general population of South Carolina. *Arthritis Rheum*, **32**, 998-1006.

[5] Black CM. (1990) Systemic sclerosis and pregnancy. *Clin Rheumatol*, **4**, 105-124.

[6] Silman AJ. (1997) Scleroderma - Demographics and survival. *J Rheumatol*, **24**, 58-61.

[7] Englert H, O'Connor H, Small-McMahon J, Chambers P, Davis K, Brooks P. (1999) Systemic sclerosis prevalence and mortality in Sydney 1974-88. *Aust NZ J Med*, **29**, 42-50.

[8] Medsger TA, Masi AT, Rodnan GP, Benedek TG, Robinson H. (1971) Survival with systemic sclerosis (scleroderma): a life-table analysis of clinical and demographic factors in 309 patients. *Ann Intern Med*, **75**, 369-76.

[9] Silman AJ, Jannini S, Symmons D, Bacon P. (1988) An epidemiological study of scleroderma in the West Midlands. *Br J Rheumatol*, **27**, 286-290.

[10] Maricq HR, Weinrich MC, Keil JE. (1989) Prevalence of scleroderma spectrum disorders in the general population of South Carolina. *Arthritis Rheum*, **32**, 998-1006.

[11] Altman RD, Medsger TA Jr, Bloch DA, Beat AM. (1991) Predictors of survival in systemic sclerosis (scleroderma). *Arthritis Rheum*, **34**, 403-13.

[12] Silman AJ. (1997) Scleroderma - Demographics and survival. *J Rheumatol*, **24**, 58-61.

[13] Hesselstrand R, Scheja A, Akesson A. (1998) Mortality and causes of death in a Swedish series of systemic sclerosis patients. *Ann Rheum Dis*, **57**, 682-686.

[14] Haustein UF. (2002). Systemic sclerosis-scleroderma. *Dermatol Online J.*, **8**, 3. https://escholarship.org/uc/item/0vd8p0xw

[15] Hesselstrand R, Scheja A, Akesson A. (1998) Mortality and causes of death in a Swedish series of systemic sclerosis patients. *Ann Rheum Dis*, **57**, 682-686.

[16] Subcommittee for Scleroderma Criteria of the American Rheumatism Association. (1980) Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). *Arthritis Rheum*, **23**, 581-90.

[17] Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. (2013) Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS One*, **8**, e83138. https://doi.org/10.1371/journal.pone.0083138

[18] Ferri C, Zakrzewska K, Longombardo G, Giuggioli D, Storino FA, Pasero G, Azzi A. (1999) Parvovirus B19 infection of bone marrow in systemic sclerosis patients. *Clin Exp Rheumatol.*, **17**, 718-20.

[19] Ohtsuka T, Yamazaki S. (2004) Increased prevalence of human parvovirus B19 DNA in systemic sclerosis skin. *Br J Dermatol.*, **150**, 1091-5.

[20] Zakrzewska K, Corcioli F, Carlsen KM, Giuggioli D, Fanci R, Rinieri A, Ferri C, Azzi A. (2009) Human parvovirus B19 (B19V) infection in systemic sclerosis patients. *Intervirology*, **52**, 279-82. https://doi.org/10.1159/000232945

[21] Bilgin H, Kocabas H, Kesli R. (2015) The prevalence of infectious agents in patients with systemic sclerosis. *Turk J Med Sci.*, **45**, 1192-7.

[22] Barukčić, I. (1989) Die Kausalität. Hamburg: Wissenschaftsverlag, pp. 218.

[23] Barukčić, I. (1997) Die Kausalität. Wilhelmshaven: Scientia, 1997. pp. 374.

[24] Barukčić, I. (2005) Causality. New Statistical Methods. Hamburg - Norderstedt: Books on Demand, pp. 488.

[25] Barukčić, I. (2006) Causality. New Statistical Methods, Second English Edition. Hamburg-Norderstedt: Books on Demand, pp. 488.

[26] Barukčić, I. (2006) New method for calculating causal relationships. Proceeding of XXIII[rd] International Biometric Conference. July 16-21; McGill University, Montréal, Québec, Canada. p. 49.

[27] Barukčić, I. (2011) Causality I. A Theory of Energy, Time and Space. Morrisville: Lulu, pp. 648.

[28] Barukčić, I. (2011) Causality II. A Theory of Energy, Time and Space. Morrisville: Lulu, pp. 376.

[29] Barukčić, I. (2012) The deterministic relationship between cause and effect. *International International Biometric Conference*, Kobe, JAPAN, 26 - 31 August 2012. https://www.biometricsociety.org/conference-abstracts/2012/programme/p1-5/P-1/249-P-1-30.pdf

[30] Barukčić, I. (2016) The Mathematical Formula of the Causal Relationship k. *International Journal of Applied Physics and Mathematics*, **6**, 45-65. https://doi.org/10.17706/ijapm.2016.6.2.45-65

[31] Barukčić K, Barukčić, I. (2016) Epstein Barr Virus - The Cause of Multiple Sclerosis. *Journal of Applied Mathematics and Physics*. **4**, 1042-53. https://doi.org/10.4236/jamp.2016.46109

[32] Barukčić, I. (2016) Unified Field Theory. *Journal of Applied Mathematics and Physics*. **4**, 1379-1438. https://doi.org/10.4236/jamp.2016.48147

[33] Barukčić, I. (2017) Helicobacter pylori-The Cause of Human Gastric Cancer. *Journal of Biosciences and Medicines.* **5**, 1-19. https://doi.org/10.4236/jbm.2017.52001

[34] Barukčić, I. (2017) Anti Bohr - Quantum Theory and Causality. *International Journal of Applied Physics and Mathematics*. **7**, 93-111. https://doi.org/10.17706/ijapm.2017.7.2.93-111

[35] Barukčić, I. (2017) Theoriae causalitatis principia mathematica. Hamburg - Norderstedt: Books on Demand, pp. 244. https://www.bod.de/buchshop/theoriae-causalitatis-principia-mathematica-ilija-barukcic-9783744815932

[36] Barukčić, I. (2018) Epstein Bar Virus-The Cause of Hodgkin's Lymphoma. *Journal of Biosciences and Medicines*, **6**, 75-100. https://doi.org/10.4236/jbm.2018.61008

[37] Astorga ML,. (2015) Diodorus Cronus and Philo of Megara: Two Accounts of the Conditional. *Rupkatha Journal on Interdisciplinary Studies in Humanities*. **7**, 9-16.

[38] Neyman, J,. (1937) Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Philosophical Transactions of the Royal Society A*. **236**, 333-380. https://doi.org/10.1098/rsta.1937.0005

[39] Clopper, C. and Pearson, E. S. (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, **26**, 404–413. http://dx.doi.org/10.1093/biomet/26.4.404

[40] Agresti, A. and Coull, B. A. (1998) Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, **52**, 119–126. http://dx.doi.org/10.2307/2685469

[41] Leemis, L. M. and Trivedi, K. S. (1996) A Comparison of Approximate Interval Estimators for the Bernoulli Parameter. *The American Statistician*, **50**, 63-68. http://dx.doi.org/10.2307/2685046

[42] Louis, T. A. (1981) Confidence Intervals for a Binomial Parameter After Observing No Successes. *The American Statistician*, **35**, 154. http://dx.doi.org/10.1080/00031305.1981.10479337

[43] Hanley, J. A. and Lippman-Hand, A. (1983) If Nothing Goes Wrong, Is Everything All Right? *The Journal of the American Medical Assn.*, **249**, 1743-1745. http://dx.doi.org/10.1001/jama.1983.03330370053031

[44] Jovanovic, B. D. and Levy, P.S. (1997) A Look at the Rule of Three. *The American Statistician*,

51, 137-139. http://dx.doi.org/10.1080/00031305.1997.10473947

[45] Rumke C. L. (1975) Implications of the Statement: No Side Effects Were Observed. *N Engl J Med*, **292**, 372-373. http://dx.doi.org/10.1056/NEJM197502132920723

[46] Fisher RA,. (1922) On the interpretation of X^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*. **85**: 87 - 94. https://doi.org/10.2307/2340521

[47] Pearson, K. (1900) On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling. *Philosophical Magazine Series*, **5**, 157-175. http://dx.doi.org/10.1080/14786440009463897

[48] Pearson K,. (1896) VII. Mathematical contributions to the theory of evolution.-III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Ser. A.* **187**, 253-18. https://doi.org/10.1098/rsta.1896.0007

[49] Pearson K. (1904) Mathematical contributions to the theory of evolution. -XIII. On the Theory of Contingency and Its Relation to Association and Normal Correlation. London, Dulau and Co., pp. 1-35. https://archive.org/details/cu31924003064833

[50] Hill, AB,. (1965) The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*. **58**, 295-300. https://doi.org/10.1177/0141076814562718

[51] Barash J, Dushnitzky D, Sthoeger D, Bardenstein R, Barak Y. (2002) Human Parvovirus B19 Infection in Children: Uncommon Clinical Presentations. *Israel Medical Assosiation Journal*, **4**,763-5. https://www.ima.org.il/FilesUpload/IMAJ/0/56/28472.pdf

[52] Centers for Disease Control (CDC). (1989) Review Risks associated with human parvovirus B19 infection. *MMWR Morb Mortal Wkly Rep*., **38**, 81-8, 93-7.

