BIO Web of Conferences **194**, 01091 (2025) https://doi.org/10.1051/bioconf/202519401091
BFT-2025

Without cytomegalovirus infection, no lung cancer

Ilija Barukčić[11]

[1]GIM Wittmund, Horandstrasse, DE-26441 Jever, Germany

Abstract. Cytomegalovirus (CMV) is a ubiquitous herpesvirus with recognized roles in immunomodulation and oncogenesis in specific contexts. Recent evidence suggests a potential association between CMV infection and lung cancer, though existing data are limited and at times contradictory. It was applied novel statistical approaches based on the concept of a "necessary condition" to evaluate the relationship between CMV infection and lung cancer, re-analysing data from multiple studies. Analysis indicates that CMV infection may act as a necessary condition for the development of lung cancer. Lung cancer incidence was consistently higher among CMV-seropositive individuals, supporting a non-random association. While these findings do not establish definitive causality, they underscore CMV infection as a potentially necessary condition in lung cancer development. Further studies are warranted to clarify this relationship and its clinical implications.

1 Introduction

Lung cancer remains the primary cause of cancer-related deaths worldwide, affecting both men and women in the US and beyond. Non–small [1] cell lung cancer (NSCLC) accounts for the majority, around 84%, while small-cell lung cancer (SCLC) makes up approximately 13% of cases. Historically, NSCLC has carried a bleak prognosis, especially in advanced stages. However, significant advancements in treatment over the last decade, including widespread lung cancer screening and improved radiation techniques, have resulted in notable [2] changes. These improvements are believed to have contributed to the observed decline in NSCLC mortality rates while the significance of the globally increasing intake of statins in this context is neglected. In terms of the cause of lung cancer, unfortunately, we have not made much progress yet. Several behavioural (tobacco smoking, second-hand smoke, alcohol use) and environmental (carcinogenic chemicals, heavy metals, such as radon gas, asbestos, arsenic, chromium, beryllium, nickel et cetera) risk factors [3], including viruses, have been linked with the development of lung cancer. Despite the significant public health impact of lung cancer, there is limited evidence about a CMV infection and lung cancer.

Human cytomegalovirus [4], CMV, initially documented by Ribbert in 1904 through inclusion-bearing cells in infant kidneys, was later proposed as a viral agent by Goodpasture

1 Corresponding author: Barukcic@t-online.de

and Talbert in 1921. The viral etiology was confirmed through independent isolations by Smith, Rowe, and Weller between 1956-1957, with the term cytomegalovirus formally established by Weller et al. in 1960. Human cytomegalovirus or human herpesvirus-5 (HHV-5), a double-stranded DNA beta-herpesvirus, is distinguished from other Herpesviridae family members by its substantial genomic size and complexity. CMV contains approximately 200 genes and demonstrates a seroprevalence exceeding 60% in adults, increasing with age to over 90% in individuals aged ≥80 years. Primary infection establishes lifelong latency within myeloid cells, with reactivation contributing to disease pathogenesis. Current CMV antiviral therapies remain limited, consisting primarily of ganciclovir, valganciclovir, foscarnet, and cidofovir. Emerging alternatives include leflunomide, which inhibits virion assembly and demonstrates viral clearance within approximately one month, and CMV hyperimmune globulin (Cytotect), showing efficacy in transplant recipients. Notably, CMV has been implicated in various malignancies, with detectable viral components in tumor tissues. Recent evidence also suggests associations with chronic obstructive [5] pulmonary disease (COPD) progression and cancer mortality. It has been found that CMV seropositivity is strongly associated with increased mortality risk in chronic obstructive pulmonary disease (COPD). Emerging evidence suggests a potential oncogenic role for cytomegalovirus (CMV) in lung carcinogenesis. This study aims to systematically investigate this theoretical possibility.

2 Material and methods

2.1 Material

2.1.1 The PRISMA guidelines

A systematic review following PRISMA guidelines was not feasible for this topic. A comprehensive search of the PubMed database revealed no suitable case-control or other relevant studies on the association between CMV and lung cancer. Consequently, there is no empirical evidence base to synthesize according to PRISMA standards. Our discussion is therefore limited to theoretical considerations and inferences drawn from indirectly relevant studies and well-established methods. Future empirical research is needed to generate data suitable for a formal systematic review.

2.1.2 The study of Kuo et al., 2008

Kuo et al. found that "Of the 15 participants ... 11 had head and neck cancers, two had lung cancer, one had lymphoma and one had rectal cancer ... **All participants** were seropositive for CMV-specific IgG before chemotherapy." [6] We added **2** young children free of lung cancer and CMV negative, to complete the control group. These measures have no statistical effect on the calculations performed.

Table 1. CMV and Lung Cancer.

		Lung cancer		
		Yes	No	
CMV IgG	**Yes**	2	13	15
	No	**0**	**2**	2
		2	15	17

This is allowed, because all lung cancer patient where CMV positive, including CMV DNA.

Table 2. CMV and Lung Cancer.

		Lung cancer		
		Yes	No	
CMV DNA	**Yes**	2	12	14
	No	**0**	1	1
		2	13	15

Kuo et al. found in their small Pilot study that the maximum recorded cytomegalovirus load of 281 (copies per mL) in serum or more represents a critical threshold at which even relatively small increases in viral load correspond to rapid increases in the probability of dying.

Table 3. CMV and Lung Cancer.

		Death (Chemoth.)		
		Yes	No	
CMV > 281 copies per mL	**Yes**	6	5	11
	No	**0**	4	4
		6	9	15

The CMV viral load increased in 14 of 15 patients after chemotherapy commenced, while **without** a cytomegalovirus load of 281 (copies per mL) in serum or more **no** death under chemotherapy.

2.1.3 The study of Li et al., 2017

Li et al. [7] reported a seroprevalence of cytomegalovirus (CMV) immunoglobulin G (IgG) in the Han Chinese population of 547/563, corresponding to 97.16%. They also provided additional data on the relationship between CMV infection and nicotine consumption. According to their findings, in the absence of CMV infection, no nicotine consumption was observed.

Table 4. CMV and smoking.

		Smoking		
		Yes	No	
CMV IgG	**Yes**	192	355	547
	No	**5**	11	16
		197	366	563

The association between exposure (CMV IgG positive) and outcome (smoking) was quantified using the odds ratio (OR) with a 95% confidence interval (CI). The calculations were performed as follows. The odds ratio was calculated using the formula:

$$OR = \frac{a \times d}{b \times c} = \frac{192 \times 11}{355 \times 5} = 1.19$$

where $a = 192$ (exposed with outcome), $b = 355$ (exposed without outcome), $c = 5$ (unexposed with outcome), and $d = 11$ (unexposed without outcome). The standard error of the natural logarithm of the OR was calculated as:

$$SE(\ln(OR)) = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} = \sqrt{\frac{1}{192}+\frac{1}{355}+\frac{1}{5}+\frac{1}{11}} = 0.547$$

The upper 95% confidence interval was derived using:

$$\text{Upper } 95\%\text{CI} = \exp\left(\ln(OR) + 1.96 \times SE(\ln(OR))\right) = 3.47$$

The lower 95% confidence interval was derived using:

$$\text{Lower } 95\%\text{CI} = \exp\left(\ln(OR) - 1.96 \times SE(\ln(OR))\right) = 0.41$$

The analysis revealed a non-statistically significant association (OR = 1.19, 95% CI [0.41, 3.47]). The wide confidence interval that includes the null value of 1.0 suggests considerable uncertainty in the point estimate, and the observed association may be due to chance. The necessary condition relationship between CMV infection and smoking is given as p(CMV IgG <- Nicotine consumption) = ((192+366)/563) = 0.9911190053 (P Value = 0.011), a significant result. Without CMV infection, no smoking (P Value = 0.011).

2.1.4 The study of Khurana et al., 2007

Over time, numerous risk factors for lung cancer have been identified, with cigarette smoking remaining the most prominent modifiable risk factor. Nevertheless, it is important to emphasize that 10–15% of **lung cancer cases occur in lifelong never-smokers**. Moreover, no study to date has demonstrated that smoking is a sufficient cause of lung cancer in itself. Vikas Khurana et al. [8] reported data on the relationship between smoking and lung cancer.

Table 5. Tobacco use and lung cancer.

		Lung cancer		
		Yes	No	
Tobacco use	**Yes**	3410	176247	179657
	No	**1262**	121422	122684
		4672	297669	302341

The association between exposure and outcome was quantified using the odds ratio (OR) with a 95% confidence interval (CI). The calculations were performed as follows. The odds ratio was calculated using the formula:

$$OR = \frac{a \times d}{b \times c} = \frac{3410 \times 121422}{176247 \times 1262} = 1.86$$

where $a = 3410$ (exposed with outcome), $b = 176247$ (exposed without outcome), $c = 1262$ (unexposed with outcome), and $d = 121422$ (unexposed without outcome). The standard error of the natural logarithm of the OR was calculated as:

$$SE(\ln(OR)) = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} = \sqrt{\frac{1}{3410}+\frac{1}{176247}+\frac{1}{1262}+\frac{1}{121422}} = 0.033$$

The upper 95% confidence interval was derived using:

$$\text{Upper } 95\%\text{CI} = \exp\left(\ln(OR) + 1.96 \times SE(\ln(OR))\right) = 1.987$$

The lower 95% confidence interval was derived using:

$$\text{Lower } 95\%\text{CI} = \exp\left(\ln(OR) - 1.96 \times SE(\ln(OR))\right) = 1.745$$

The analysis revealed a statistically significant positive association (OR = 1.86, 95% CI [1.74, 1.99]) between nicotine consumption and lung cancer.

The necessary condition relationship between nicotine consumption and lung cancer is given as p(Tobacco use <- Lung cancer) = ((3410+297669)/ 302341) = 0.9958259052 (P Value = 0.0042), a significant result. **Without** tobacco use, **no** lung cancer. These results are corroborated by findings from other studies.

2.1.5 Modus ponens disjunctivus

We consider at this point a potential approach to the relationship between CMV and lung cancer, particularly from the perspective of classical logic. Our analysis indicates that in the absence of CMV infection, no smoking occurs (p(CMV IgG <- Nicotine consumption) = 0.9911190053; P Value = 0.011). Moreover, robust data support the conclusion that without smoking, lung cancer does not develop (p(Nicotine consumption <- Lung cancer) = 0.9949328738; P Value = 0.0042). Taken together, and following the rules of classical logic, one may theoretically infer that in the absence of CMV infection, lung cancer would not occur. However, this conclusion is purely theoretical; empirical, data-driven evidence would, of course, be preferable. In the meantime, we fortunately also have robust empirical data on the relationship between CMV and lung cancer.

2.1.6 The study of Geris et al., 2022

Geris et al. [9] assessed cancer incidence across CMV risk groups based on donor (D) and recipient (R) IgG serostatus. Data were drawn from the U.S. Solid Organ Transplantation (SOT) registry linked with 32 cancer registries, enabling analysis of cancer outcomes according to CMV risk stratification.

Table 6. CMV and lung cancer.

		Lung cancer		
		Yes	No	
CMV positive	**Yes**	1658	204142	205800
	No	274	41244	41518
		1932	245386	247318

In the absence of a CMV infection (CMV R-negative / D-negative group, n = 41518), 274 cases of lung cancer were observed.

The association between exposure (CMV positive) and outcome (Lung cancer) was quantified using the odds ratio (OR) with a 95% confidence interval (CI). The calculations were performed as follows. The odds ratio was calculated using the formula:

$$OR = \frac{a \times d}{b \times c} = \frac{1658 \times 41244}{204142 \times 274} = 1.223$$

where $a = 1658$ (exposed with outcome), $b = 204142$ (exposed without outcome), $c = 274$ (unexposed with outcome), and $d = 41244$ (unexposed without outcome). The standard error of the natural logarithm of the OR was calculated as:

$$SE(\ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} = \sqrt{\frac{1}{1658} + \frac{1}{204142} + \frac{1}{274} + \frac{1}{41244}} = 0.0654$$

The upper 95% confidence interval was derived using:

$$\text{Upper 95\%CI} = \exp\left(\ln(OR) + 1.96 \times SE(\ln(OR))\right) = 1.3898$$

The lower 95% confidence interval was derived using:

$$\text{Lower 95\%CI} = \exp\left(\ln(OR) - 1.96 \times SE(\ln(OR))\right) = 1.0754$$

The analysis revealed a statistically significant positive association (OR = 1.22, 95% CI [1.08, 1.39]). Since the entire confidence interval lies above 1.0, this indicates that the exposure is associated with approximately 22% higher odds of the outcome compared to non-exposure. The necessary condition relationship between CMV infection and lung cancer is given as p(CMV IgG <- lung cancer) = ((1658+245386)/ 247318) = 0.9988921146 (P Value = ((274/247318) + (1/247318)) = 0.0011), a significant result. In other words, **without** CMV infection, **no** lung cancer (P Value = 0.0011).

2.2 Methods

2.2.1 Definition. Odds ratio (OR)

Given the case-control design of this retrospective study, odds ratios with corresponding 95% confidence intervals were calculated as the appropriate measure of association, as the sampling was based on outcome status precluding direct estimation of risk ratios. The odds ratio (OR) measures association between exposure and outcome, calculated as:

$$OR = \frac{a \times d}{b \times c}$$

from a 2×2 contingency table where a = exposed with outcome, b = exposed without outcome, c = unexposed with outcome, and d = unexposed without outcome. Statistical methods like Risk Ratio (RR) and Odds Ratio (OR) Conventional measures such as risk ratios and odds ratios are designed to quantify the strength and direction of the absence of statistical independence. They do not, on their own, demonstrate any kind of causation. A value significantly different from 1.0 indicates that the exposure and the outcome are not independent. While a value deviating from 1.0 signifies a departure from independence, it does not establish a causal relationship, nor does it elucidate the nature of that relationship in terms of necessary or sufficient condition—a distinction that is critical for our research question regarding CMV and lung cancer pathogenesis.

2.2.2 Definition. Necessary condition (SINE))

Conventional association measures such as risk ratios or odds ratios demonstrate limited utility when modeling necessary conditions between events, as they primarily quantify associations rather than necessary condition relationships. To address this methodological gap, we employed the conditio sine qua non approach, a framework specifically developed to identify necessary conditions in data. This methodology is specifically designed to evaluate necessary condition relationships, has undergone formal validation [10], and provides analytical capabilities that transcend conventional statistical measures. Mathematically, the necessary condition (SINE) relationship, denoted by $p(A_t \leftarrow B_t)$ in terms of statistics and probability theory, is defined as

$$\begin{aligned} p\left(A_t \leftarrow B_t\right) &\equiv p\left(a_t\right) + p\left(b_t\right) + p\left(d_t\right) \\ &\equiv \frac{N \times \left(p\left(a_t\right) + p\left(b_t\right) + p\left(d_t\right)\right)}{N} \\ &\equiv \frac{a + b + d}{N} \\ &\equiv \frac{A + d}{N} \\ &\equiv \frac{a + (Not\ B)}{N} \\ &\equiv +1 \end{aligned}$$

In general, it is $p\left(A_t \rightarrow B_t\right) \equiv 1 - p\left(A_t \leftarrow B_t\right)$. A necessary condition A_t is characterized itself by the property that another event B_t will not occur if A_t is not given, if A_t did not occur . The following table gives an overview.

Table 7. Necessary condition relationship.

		Conditioned B		
		Yes	No	
Condition A	**Yes**	a	b	A
	No	c = 0	d	Not A
		B	Not B	N

Taking into account Kolmogorov's definition of an n-dimensional probability density of random variables A_t, B_t et cetera at the (period of) time t, we obtain

$$\begin{aligned} p\left(A_t \leftarrow B_t\right) &\equiv +1 \\ &\equiv +1-p\left(c_t\right) \\ &\equiv +1-p\left((\text{Not A})_t \cap B_t\right) \\ &\equiv \left(\int_{-\infty}^{A_t}\int_{-\infty}^{B_t} f\left(A_t,B_t\right)dA_t dB_t\right)+\left(1-\int_{-\infty}^{B_t} f\left(B_t\right)dB_t\right) \end{aligned}$$

while $p\left(A_t \leftarrow B_t\right)$ would denote the cumulative distribution function of random variables of a necessary condition. Another adequate formulation of a necessary condition is possible too. If certain conditions are met, then necessary conditions and sufficient conditions are one way or another converses of each other, too. It is

$$p\left(A_t \leftarrow B_t\right) \equiv \left(A_t \vee (\text{Not } B_t)\right) \equiv \left((\text{Not } B_t)_t \vee A_t\right) \equiv p\left(B_t \rightarrow A_t\right)$$

In general, there are circumstances under which

$$p\left(A_t \leftarrow B_t\right) \equiv \left(A_t \vee (\text{Not } B_t)\right) \equiv \left((\text{Not } A_t) \vee B_t\right) \equiv p\left(A_t \rightarrow B_t\right)$$

However, this relationship does not apply under any circumstances.

Example I.

A wax candle is characterized by various properties, but is also subject to certain conditions. **Without** sufficient amounts of gaseous oxygen **no** burning wax candle, gaseous oxygen is a necessary condition of a burning candle. However, the converse relationship **if** burning wax candle, **then** sufficient amounts of gaseous oxygen are given is at the same (period of) time t / Bernoulli trial t true. The following tables are illustrating these relationships.

Example II.

Once again, a human being cannot live without water. A human being cannot live without gaseous oxygen, et cetera. Water itself is a necessary condition for human life. However, gaseous oxygen is a necessary condition for human life too. Thus far, even if water is given and even if water is a necessary condition for human life, without gaseous oxygen there will be no human life. In general, if a conditioned or an outcome B_t depends on the necessary condition A_t and equally on numerous other necessary conditions, an event B_t will not occur if A_t itself is not given independently of the occurrence of other necessary conditions.

Example III.

Another different aspect of a necessary condition relationship is appropriate to be focused upon here. As a direct consequence of a necessary condition **without** sufficient amounts of gaseous oxygen **no** burning wax candle is a special case of an exclusion relationship. The absence of sufficient amounts of gaseous oxygen A_t excludes a burning wax candle B_t. Thus

far, if we want to stop the burning of a wax candle, we would have to significantly reduce the amounts of gaseous oxygen A_t. Under these conditions, a wax candle will stop burning. The necessary condition relationship follows approximately as

$$p\left(A_t \leftarrow B_t\right) \geq 1 - \frac{p\left(c_t\right)}{p\left(B_t\right)}$$

and equally as

$$p\left(A_t \leftarrow B_t\right) \geq 1 - \frac{p\left(c_t\right)}{p\left((\text{Not A})_t\right)}$$

The Chi-square goodness of fit test of a necessary condition relationship

Under some circumstances, hypothesis about the conditio sine qua non relationship $p(A_t \leftarrow B_t)$ can be tested by the chi-square distribution (also chi-squared or χ^2-distribution), first described by the German statistician Friedrich Robert Helmert and later rediscovered by Karl Pearson in the context of a goodness of fit test. The $\tilde{\chi}^2$ goodness of fit test of a conditio sine qua non relationship with degree of freedom (d. f.) of d. f. = 1 is calculated as

$$\begin{aligned}\tilde{\chi}^2_{\text{Calculated}}\left(A_t \leftarrow B_t \mid B\right) &\equiv \frac{(a-(a+c))^2}{B} \\ &\quad + \frac{((b+d)-\text{B})^2}{\text{B}} \\ &\equiv \frac{c^2}{B} + 0 \\ &\equiv \frac{c^2}{B}\end{aligned}$$

or equally as

$$\begin{aligned}\tilde{\chi}^2_{\text{Calculated}}\left(A_t \leftarrow B_t \mid (\text{Not A})\right) &\equiv \frac{(d-(c+d))^2}{(\text{Not A})} \\ &\quad + \frac{((a+b)-A)^2}{A} \\ &\equiv \frac{c^2}{(\text{Not A})} + 0 \\ &\equiv \frac{c^2}{(\text{Not A})}\end{aligned}$$

and can be compared with a theoretical chi-square value at a certain level of significance α. It has not yet been finally clarified whether the use of Yate's continuity correction is necessary at all.

2.2.3 One-Sided Right-Tailed Binomial Hypothesis Test

The **one–sided right–tailed binomial hypothesis test** (with the rejection region in the right tail — i.e., higher values of p) is generally formulated as:

$$H_0{:}p \leq p_0 \qquad \text{vs.} \qquad H_1{:}p > p_0$$

We propose a one-sided right-tailed hypothesis test to assess whether the true probability p of a relationship resembling a (necessary) condition exceeds the benchmark of $1 - \alpha$ (e.g., 95% if $\alpha = 0.05$). Specifically, the null hypothesis

$$H_0{:}p \leq 1 - \alpha$$

posits that the success rate is at most this threshold, while the alternative hypothesis

$$H_1: p > 1 - \alpha$$

tests whether the true probability p of a relationship like a (necessary) condition is significantly higher. This testing framework controls the Type I error rate at the α level, ensuring rigorous statistical inference. Rejecting H_0 provides strong evidence that the true probability p of a necessary condition relationship is significantly higher in clinical practice. Thus far, let $X \sim \text{Binomial}(N,p)$, where $p = 1 - \alpha$. Then the left-tailed cumulative distribution function, denoted as $P(X \leq k)$, up to k is given by:

$$P(X \leq k) = \sum_{t=0}^{k} \binom{N}{t} p^t (1-p)^{N-t} = \sum_{t=0}^{k} \binom{N}{t} (1-\alpha)^t \alpha^{N-t}$$

Symbol	Meaning
N	Total number of trials
t	Summation index (number of successes, single Bernoulli trial)
k	Observed number of successes
$p = 1 - \alpha$	Hypothetical success probability

This expression gives the **left-tailed cumulative distribution function (CDF)** of the binomial distribution — the total probability of getting k or fewer successes in N trials. Often, $1 - \alpha$ represents the confidence level, and the confidence interval provides a range of values within which the true population parameter (such as the probability of a (necessary) condition relationship) is likely to fall, with a specified level of confidence.

The p-value indicates the probability of obtaining the observed data (or something more extreme) assuming that the null hypothesis is true. A low p-value suggests that the observed data is highly unlikely under the null hypothesis.

If the probability (or estimate) of a necessary condition lies outside the confidence interval, this means the parameter falls outside the range of likely values based on the data. This suggests that the observed data is inconsistent with the null hypothesis.

A very small p-value indicates strong evidence against the null hypothesis. Specifically, a p-value of 0.00021 means there's only a 0.021% chance of observing data as extreme (or more extreme) as what was observed, assuming the null hypothesis is true. This indicates that the result is statistically significant and provides strong support for rejecting the null hypothesis.

The binomial probability mass function based one–sided right tailed binomial P–value is given as:

$$p(X > (N-1)) = p(X = N) = p^N$$

In general it is:

$$p(X \leq (N-1)) + p(X > (N-1)) = 1 - p(X = N) + p(X = N) = 1$$

The cumulative probability up to N - 1 successes is given as:

$$p(X \leq (N-1) = 1 - p(X = N)$$

and right tailed one sided P–value is given as:

$$p(X > (N-1)) = p(X = N) = 1 - p(X \leq (N-1)$$

In general, based on the binomial probability mass function, the one–sided right tailed binomial P–value is given by:

$$p(X > (N-1)) = p(X = N) = \binom{N}{N} p^N (1-p)^{N-N} = 1 \cdot p^N (1-p)^0 = p^N$$

where: N is the total number of trials, $k = N$ is the number of successful trials, p is the probability of success in a single trial, $1 - p$ is the probability of failure.

Theoretically, it is possible to **reject the null hypothesis even when it corresponds to reality**. In other words, a statistically significant effect may be accepted when, in truth, no actual effect exists—this constitutes a *Type I error*. Technically speaking, the *significance level* α of a statistical test, commonly set at $\alpha = 0.05$ or $\alpha = 0.01$, represents the maximum allowable probability of committing such an error; that is, falsely **rejecting the null hypothesis when it is in fact true**.

The inequality $p(X_t) > 1 - \alpha$ indicates that the probability of the event X_t is sufficiently high to exceed the confidence threshold $1 - \alpha$ (e.g., 95% when $\alpha = 0.05$). This outcome is interpreted as *statistically significant* evidence in support of the assertion that X_t is not merely a product of random variation. Under these circumstances, it follows that:

$$p(X > N - 1) = p(X = N) = (1 - \alpha)^N$$

Tests based on $1 - \alpha$ are typically referred to as *confidence-based tests*, or more formally, as tests that operate at a specified *confidence level*. The value $1 - \alpha$ is known as the *confidence level*, representing the probability that the test will not commit a Type I error. A statistical test that uses $1 - \alpha$ (e.g., 95% or 99%) as a threshold for rejecting the null hypothesis is said to operate at the corresponding confidence level.

Example.

Assume a necessary condition relationship between A_t and B_t represented by the binary random variable

$$X_t = (A_t \vee \neg B_t) \in \{0,1\}.$$

In an investigation, we examined $N = 100$ individuals and observed $k = N - 1 = 99$ occurrences of the event of interest, X_t. To assess the statistical significance of this observation, we calculate the probability of observing at most $k = 99$ successes in $N = 100$ independent trials, under the assumption that the true success probability is $p = 1 - \alpha = 0.95$. This probability is given by:

$$P(X \leq (N-1))) = 1 - p(X = N) = 1 - 0.95^{100} = 1 - 0.00592052922 = 0.9940794708$$

The one–sided right–tailed binomial P–value is given as:

$$p - \text{value}_{\text{One–sided right–tailed binomial}}: \quad P(X > N - 1) = P(X = N) = 0.95^{100} \approx 0.005920$$

Decision and Interpretation

We compare the p-value to the significance level $\alpha = 0.05$:

- If $\text{p} - \text{value} \leq \alpha$, reject H_0
- If $\text{p} - \text{value} > \alpha$, do not reject H_0

In our case, the p-value is:

$$\text{p} - \text{value} = 0.00592052922 \ll (\alpha = 0.05)$$

In other words, we have strong statistical evidence that the true success probability p is greater than 0.95. In other words, observing all 100 successes is so unlikely under the null hypothesis $p \leq 0.95$, that we can confidently conclude $p > 0.95$.

Conclusion: The observed perfect success rate (100/100) is statistically significantly higher than the benchmark threshold of 95%, with a p-value of 0.00592052922. Hence, **we do reject the null hypothesis** H_0 that

$$H_0{:}p \leq 0.95$$

and conclude that the true probability p of the necessary condition being satisfied is greater than 95% with high confidence ($p < 0.006$).

There are conditions of a one–sided right–tailed binomial test, where the p-value is given as $\mathrm{p-value} = (1-\alpha)^N$. Under these circumstances, the confidence interval $1-\alpha$ is based on the sample size N and the p–value as

$$1-\alpha = e^{\frac{\ln(p-value)}{N}}$$

We consider conditions of a one–sided right–tailed binomial test, where the p-value is given as:

$$\mathrm{p-value} = (1-\alpha)^N$$

Take the natural logarithm of both sides:

$$\ln(\mathrm{p-value}) = \ln\left((1-\alpha)^N\right)$$

$$\ln(\mathrm{p-value}) = N\ln(1-\alpha)$$

$$\ln(1-\alpha) = \frac{\ln(\mathrm{p-value})}{N}$$

$$1-\alpha = e^{\frac{\ln(\mathrm{p-value})}{N}}$$

For the special case: $\mathrm{p-value} = \frac{1}{N} = (1-\alpha)^N$ the expression for $(1-\alpha)$ under the these conditions becomes:

$$1-\alpha = e^{-\frac{\ln(N)}{N}} = N^{-\frac{1}{N}}$$

For N = 100, it is

$$1-\alpha = e^{-\frac{\ln(100)}{100}} = 100^{-\frac{1}{100}} = 0.954992586$$

3 Results

3.1 CMV infection and lung cancer

From a theoretical standpoint, it is conceivable that CMV may play a causal role in lung cancer. However, without rigorous empirical verification, this remains speculative. It is therefore essential to conduct well-designed studies to clarify the relationship, rather than relying on assumptions or circumstantial data. Only through systematic investigation can we determine whether CMV is genuinely implicated in lung carcinogenesis or if the observed associations are coincidental. Until such evidence is available, any assertion of causality must be considered provisional.

Null Hypothesis (H_0): Cytomegalovirus infection (CMV) **is not** a necessary condition for lung cancer (LC).

Alternative Hypothesis (H_1): Cytomegalovirus infection (CMV) is a necessary condition for lung cancer (LC).

Proof By Hypothesis Test.

Hypothesis testing is a fundamental tool in the of science, allowing researchers to evaluate the plausibility of a proposed relationship within a framework of empirical uncertainty. By formally testing hypotheses, we can distinguish patterns likely to reflect genuine associations from those arising by chance, thereby providing epistemic rigor to scientific inference.

We reanalyzed the data published by Geris et al. [9] to investigate the potential necessary condition relationship between CMV infection and lung cancer. Using a necessary-condition approach, we calculated the probability that lung cancer occurs in the absence of CMV infection. The necessary-condition relationship was quantified as follows:

$$p\left(A_t \leftarrow B_t\right) \equiv \frac{a+b+d}{N} = \frac{1658+204142+41244}{247318} = 0.9988921146$$
$$\equiv \frac{A+d}{N} = \frac{205800+41244}{247318} = 0.9988921146$$
$$\equiv \frac{a+(Not\ B)}{N} = \frac{1658+(245386)}{247318} = 0.9988921146$$

The resulting P value was

$$P\ Value_{Right\ Tailed} \equiv \frac{c}{N} + \frac{1}{N} = \frac{274}{247318} + \frac{1}{247318} = 0.0011,$$

indicating a highly significant association.

Taken together, these results allow us to reject the Null Hypothesis (H_0) and accept the Alternative Hypothesis (H_1). Specifically, in this dataset, **without** CMV infection, **no** lung cancer occurred (P value = 0.0011).

Quod erat demonstrandum.

4 Discussion

Jennifer M. Geris et al. [9] provided compelling data on the association between CMV seropositivity and lung cancer, highlighting a potential role of CMV infection in the pathogenesis of this malignancy. The CMV status of recipients (R) and donors (D) was as follows: 41,518 patients were CMV-negative (R) *and* CMV-negative (D). Among the total cohort of 247,318 patients, this implies that 205,800 patients were CMV-positive. Within the CMV-negative (R) / CMV-negative (D) group, 274 cases of lung cancer were documented, whereas a total of 1,932 lung cancer cases were observed in the entire study CMV positive population. Based on these data, the following statistical measures can be derived. The necessary condition relationship between CMV infection and lung cancer is given as p(CMV IgG <- lung cancer) = ((1658+245386)/ 247318) = 0.9988921146 (P Value = ((274/247318) + (1/247318)) = 0.0011), a significant result. The confidence interval is given as

$$1-\alpha = e^{\frac{\ln(p-value)}{N}} = e^{\frac{\ln(0.001111928772)}{247318}} = 0.9999724987.$$

In other words, based on the data reported by Jennifer M. Geris et al., it follows that **without** CMV infection, **no** lung cancer (P Value = 0.0011).

However, in contrast to the findings of the present study, Selena Rashid et al. [11] reported that human cytomegalovirus (CMV) infection was associated with a decreased risk of bronchogenic carcinoma. In their analysis, the incidence of bronchogenic carcinoma was 1.69% (243/14,319 patients) in the CMV group compared with 6.08% (871/14,319 patients) in the control group. This retrospective analysis exhibits significant methodological limitations that challenge the validity of its findings.

The temporal relationship between CMV infection and lung cancer diagnosis remains ambiguous, precluding any causal inferences. Therefore, the study's methodology for establishing a causal relationship between cytomegalovirus (CMV) infection and bronchogenic carcinoma is fundamentally flawed. The authors assigned patients to a CMV-exposed group based solely on a documented history of *prior CMV infection* and subsequently assessed for the development of lung cancer. This approach implicitly assumes that any cancer diagnosed after the infection is a consequence of it—a post hoc ergo propter hoc logical fallacy. This logical error is analogous to concluding that day causes night simply because night follows the day. Crucially, the analysis did not account for the patients' CMV status at the time of the lung cancer diagnosis. This likely results in a significant number of patients with active CMV infection at the time of their cancer diagnosis being erroneously included in the non-exposed control group, thereby invalidating the core comparison. Another major limitation of the study by Rashid et al. concerns both the ascertainment of bronchogenic carcinoma and the determination of CMV seropositivity. The methodology used to establish the diagnosis of lung cancer is not specified—whether histopathological confirmation (e.g., biopsy or bronchoalveolar lavage) was required, or whether cases were identified solely on the basis of imaging such as computed tomography of the chest (CT). In the absence of histological verification, conditions including lymphomas, sarcoidosis, and other pulmonary diseases are at risk of being misclassified as bronchogenic carcinoma. Such diagnostic uncertainty could result in a substantial number of patients being incorrectly classified as lung cancer cases, disproportionately among those without CMV infection, thereby introducing serious misclassification bias. Relying on administrative codes for "bronchogenic carcinoma" can result in misclassification (e.g., miscoding of other lung cancers, benign lesions) bias. The study relies only on ICD-9/ICD-10 billing codes to define prior CMV infection. These codes are notoriously incomplete and inaccurate for latent viral infections. The vast majority of CMV seropositive individuals never receive a corresponding diagnosis code. Consequently, a substantial proportion of truly exposed patients who develop lung cancer are misclassified as unexposed, leading to profound under-ascertainment in the CMV cohort and a systematic dilution of the observed association. Equally concerning, the authors provide no information regarding the sensitivity of the assay used to determine CMV IgG seropositivity. This is critical when evaluating the relationship between CMV infection and bronchogenic carcinoma. The kinetics of cytomegalovirus (CMV) IgG and the reliability of serological CMV diagnosis warrant further detailed evaluation. For CMV-IgG assays, reported sensitivity and specificity range from 90% to 100%, depending on the specific assay employed. For instance, if the sensitivity of the CMV assay were 95%, then among 100 patients who are truly CMV positive, only 95 would be correctly identified as such, while 5 would be misclassified as CMV negative. Misclassification of exposure status, combined with uncertainty in outcome classification, could substantially distort the reported association between CMV infection and reduced lung cancer risk, raising serious questions regarding the validity of the study's conclusions. The data published by Rashid et al. are markedly inconsistent and, as such, cannot be considered a reliable basis for advancing our understanding of the relationship between CMV infection and bronchogenic carcinoma.

A CMV infection appears to be linked to various diseases, including lung cancer. Therefore, the development of a vaccine against CMV is imperative for substantially and enduringly decreasing the significant disease burden posed by CMV infections. Presently, there is no approved vaccine [12] offering protection against CMV, even if numerous vaccine candidates, are undergoing evaluation in clinical trials. Possibly, the breakthrough of mRNA-based vaccinations against the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) or COVID-19 disease with a fatality rate of about 2.19% could accelerate efforts in this regard.

Based on the emerging evidence provided by this publication, it strongly suggests that cytomegalovirus (CMV) infection is a necessary condition for the development of lung cancer, indicating a crucial and causal role of CMV in lung carcinogenesis (P Value = 0.0011). However, the available data are somewhat inconsistent. Consequently, caution is warranted when interpreting the relationship between CMV infection and lung cancer as found in this publication until further, and more systematic research can provide more definitive conclusions. However, this publication underscores the need for additional well-designed and highly systematic studies with robust methodologies to elucidate the role of CMV in lung carcinogenesis and to clarify any discrepancies in the existing data. Such research efforts will be crucial for informing clinical practice and guiding future therapeutic strategies targeting CMV-associated lung cancer.

5 Conclusion

The available evidence is somewhat contradictory; nevertheless, the data presented cannot be disregarded. Provisionally, and until further evidence emerges to the contrary, it is reasonable to consider CMV infection as a potential causal factor in the development of lung cancer.

References

1. L.E.L. Hendriks, J. Remon, C. Faivre-Finn, et.al., Non-small-cell lung cancer. Nat Rev Dis Primers. **10**, 71 (2024). https://doi.org/10.1038/s41572-024-00551-9
2. G.S. Jones, D.R. Baldwin, Recent advances in the management of lung cancer. Clin Med (Lond). **18**, s41–s46 (2018). https://doi.org/10.7861/clinmedicine.18-2-s41
3. Y. Alduais, H. Zhang, F. Fan, J. Chen, B. Chen, Non-small cell lung cancer (NSCLC): A review of risk factors, diagnosis, and treatment. Medicine (Baltimore). **102**, e32899 (2023). https://doi.org/10.1097/MD.0000000000032899
4. H.D. Riley, History of the cytomegalovirus. South Med J. **90**, 184–190 (1997). https://doi.org/10.1097/00007611-199702000-00004
5. D.B. Tan, F.S. Amran, T.H. Teo, P. Price, Y.P. Moodley, Levels of CMV-reactive antibodies correlate with the induction of CD28(null) T cells and systemic inflammation in chronic obstructive pulmonary disease (COPD). Cell Mol Immunol. **13**, 551–553 (2016). https://doi.org/10.1038/cmi.2015.4
6. C.-P. Kuo, C.-L. Wu, H.-T. Ho, et.al., Detection of cytomegalovirus reactivation in cancer patients receiving chemotherapy. Clin Microbiol Infect. **14**, 221–227 (2008). https://doi.org/10.1111/j.1469-0691.2007.01895.x
7. Z. Li, Y. Tang, N. Tang, et.al., High anti-human cytomegalovirus antibody levels are associated with the progression of essential hypertension and target organ damage in Han Chinese population. PLoS ONE. **12**, e0181440 (2017). https://doi.org/10.1371/journal.pone.0181440
8. V. Khurana, H.R. Bejjanki, G. Caldito, M.W. Owens, Statins reduce the risk of lung cancer in humans: a large case-control study of US veterans. Chest. **131**, 1282–1288 (2007). https://doi.org/10.1378/chest.06-0931
9. J.M. Geris, L.G. Spector, R.M. Pfeiffer, et.al., Cancer risk associated with cytomegalovirus infection among solid organ transplant recipients in the United States. Cancer. **128**, 3985–3994 (2022). https://doi.org/10.1002/cncr.34462

10. I. Barukčić, The causal relationship k. MATEC Web Conf. **336**, 09032 (2021). https://doi.org/10.1051/matecconf/202133609032
11. S. Rashid, A. Ardeljan, L.R. Frankel, et.al., Human Cytomegalovirus (CMV) Infection Associated With Decreased Risk of Bronchogenic Carcinoma: Understanding How a Previous CMV Infection Leads to an Enhanced Immune Response Against Malignancy. Cureus. **15**, e37265 (2023). https://doi.org/10.7759/cureus.37265
12. L.R. Smith, M.K. Wloch, J.A. Chaplin, M. Gerber, A.P. Rolland, Clinical Development of a Cytomegalovirus DNA Vaccine: From Product Concept to Pivotal Phase 3 Trial. Vaccines (Basel). **1**, 398–414 (2013). https://doi.org/10.3390/vaccines1040398