Are There More UAP Reports on Nuclear Test Days?

A daily-count comparison with a placebo test for the years 1945 to 2017

Ralf van Thuyl

In 2025, Stephen Bruehl and Beatriz Villarroel reported that photographs of the sky from the 1950s showed more short-lived point-like light sources around above-ground nuclear tests. Reportedly, they also found more of these light points on days with more UAP reports.[1] This study asks whether a link with nuclear tests also shows up in the reported sightings themselves, that is, in eyewitness reports from the ground. It neither confirms nor refutes that study, because the two do not measure the same thing (Section 7).

Result: In the report data examined, no more UAP sightings are reported on nuclear test days than on the days before and after. On the test day, reports were on average about 4% higher (11.4 versus 10.9 per day; Fig. 2). This is not a conspicuous difference: even if there were no association at all, a deviation of this size would occur by chance about one time in eight. Under the pre-specified rule, only a value of z = 2 or above would have counted as conspicuous; here z = 1.70. Large effects are unlikely in these daily counts; small or regional effects are not ruled out. In a subanalysis using US sightings and exclusively US tests (849 test days), the value is z = 2.20. Because several analyses were examined, this is a lead that needs replication, not evidence.

Restricting the analysis to days with above-ground tests, as the photographic-plate study does, gives an estimated difference of +6% to +7% (417 to 464 test days; z = 0.83 to 1.03). The direction matches the primary analysis, but the value lies within normal variation.

The basis is 1,762 nuclear test days (1945 to 2017) and the database of the *Gesellschaft zur Erforschung des UFO-Phänomens (GEP)* with about 376,000 reports. Each test day was compared with the seven days before and after.

As a caveat, on 671 of the 1,762 test days (38.1%), at least one of the countries Russia (Soviet Union), China, India, Pakistan or North Korea was involved. Sightings from Russia, China and North Korea are barely represented in the catalogs, and no national databases exist for India and Pakistan. A possible effect of these countries' tests would therefore not be detectable in these figures.



***Fig. 1**: Locations of nuclear tests between 1945 and 2017.*
Map material: © OpenStreetMap contributors/esri

Key points

- **Question:** Are more UAP sightings reported on days with nuclear tests than on the days before and after?
- **Result:** About 4% more sightings on the test day, within normal variation (z = 1.70, threshold 2). No established association.
- **Lead:** For US tests and US sightings, z is 2.20. Because several analyses were examined, this is a lead, not evidence.
- **Above-ground tests:** Days with above-ground tests only: +6% to +7% (z = 0.83 to 1.03). Same direction, no established association.
- **Limitation:** On about 38% of test days, a country was involved from which the catalogs contain few sightings. Small or regional effects are not ruled out.
- **Context:** The photographic-plate study is neither confirmed nor refuted.

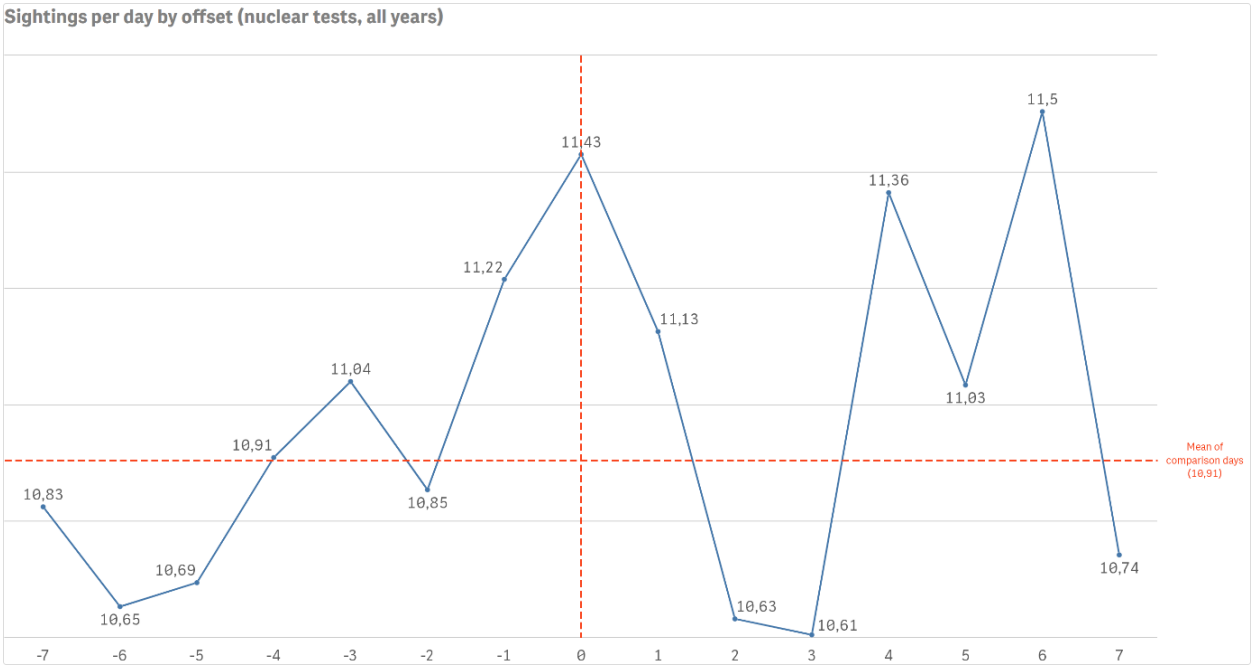


Fig. 2 *Sightings per day on nuclear test days by offset (y-axis truncated).*

1 Introduction

1.1 Background

In 2025, Stephen Bruehl and Beatriz Villarroel analyzed the photographic plates of the first Palomar Observatory Sky Survey (1949 to 1957).[1] They report that short, star-like points of light ("transients") appeared more often when an above-ground nuclear test was nearby. On days within one day of a test they were 45% more likely, and on the day after a test 68% more likely. Reportedly, the authors additionally found a small correlation with UAP reports from the UFOCAT catalog.[2] A critical examination of this study exists (Watters et al., 2026)[3] as well as a response by the study authors.[4]

The author of this study took this as an occasion to pose the related question with ground-based data: Is there also a link between eyewitness reports and nuclear tests up to 2017? A direct comparison with the photographic plates is not possible. Eyewitness reports are subjective and hard to verify, and the test sites were remote and closed off, so direct on-site observation is practically ruled out. The locations of the nuclear tests are shown in Fig. 1.

1.2 Research question

Question: Are more UAP sightings reported on days with nuclear tests than on other days?

- **H_0 (null hypothesis):** There is no difference beyond normal variation.
- **H_1 (alternative hypothesis):** On nuclear test days, the number of reported sightings is higher than on comparable days without a test.

In plain terms: The null hypothesis is the "normal case" that is assumed until the data show otherwise. The question is whether anything appears on the test day that cannot be explained by chance. Results are evaluated two-sided, that is, for deviations upward and downward.

2 Data

Sightings. The basis is the GEP database, which merges ten national and international sighting catalogs. Of the currently 376,182 records, 348,638 have a readable date. This corresponds to about 347,600 distinct case numbers.[5]

Records outside the study period (for example years before 1940) fall into none of the windows. In the period 1961 to 1968, about 92% of reports come from a single catalog, UFOCAT.[2] At least one of the ±7-day windows around a nuclear test day contains 120,829 distinct reports (Tab. 1).

Source data by Data base / Decade

Source DB	Decade							
	Totals	1940's	1950's	1960's	1970's	1980's	1990's	2000's	2010's
Totals	**120.829**	**282**	**21.983**	**44.560**	**37.650**	**9.820**	**4.907**	**503**	**1.125**
Canada (CA)	**772**	-	-	-	-	120	381	82	189
DEGUFO (DE)	**57**	-	2	4	6	24	11	-	10
GEP (DE)	**528**	-	2	20	94	207	135	19	51
Greece (GR)	**142**	-	25	31	43	33	8	2	-
HUMCAT	**1.201**	2	215	542	442	-	-	-	-
Larry Hatch (US)	**10.688**	26	2.248	2.835	3.790	1.247	542	-	-
MUFON-CES (DE)	**91**	-	1	3	29	36	22	-	-
UFOCAT (US)	**106.851**	254	19.444	41.003	33.053	8.064	3.772	398	864
UPDB (DE)	**221**	-	24	61	79	27	21	1	8
USO (US)	**278**	-	22	61	114	62	15	1	3

Tab. 1 *Sightings in the ±7-day windows around nuclear test days, by source database and decade (nuclear tests only)*

Nuclear events. The data on the nuclear tests and on the 24 nuclear incidents come from Johnston's Archive.[6] They contain 2,081 entries dated from 1945 to 2017. This yields 1,762 nuclear test days and 24 days with nuclear incidents. Several tests on the same day count as one test day. The list was checked for data-entry and formatting errors and cleaned (Section 5). Individual entries were not checked against the original source. For the analysis in Section 6.6, each of the 2,057 tests was classified as above-ground (525), underground (1,473) or unknown (59), in sequence by design type, depth entry and, last, by test site and date (Appendix F).

Which countries tested? On 71 of the 1,762 test days (4.0%), tests by several countries took place, mostly the Soviet Union and the USA. On days when only one country tested, the figures were: USA 849 days, Soviet Union 561, France 192, United Kingdom 39, China 39, others 11 (Fig. 3).

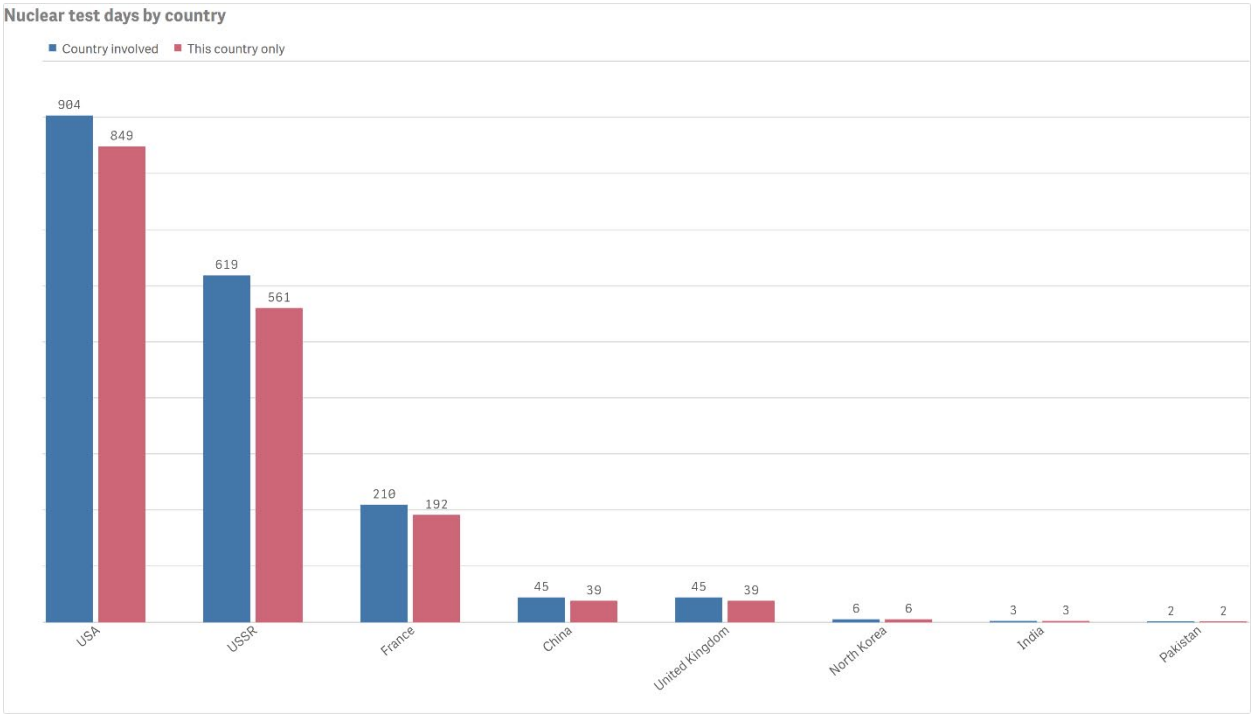


Fig. 3 *Test days by country, once with all participating countries, once only days on which exclusively this country tested.*

Important gap. On 671 of the 1,762 test days (38.1%), at least one of the countries Russia (Soviet Union), China, India, Pakistan or North Korea was involved. The sum of country participations is 675 because on four days the Soviet Union and China tested together.

Russia, China and North Korea (666 test days, 37.8%): Sightings from these countries are barely represented in the catalogs of the GEP database. This is not a shortcoming of this study alone but a problem of Western-accessible sighting catalogs in general. Whether separate collections exist in these countries was not investigated, nor was the cause. Three explanations suggest themselves:

- The catalogs come predominantly from the USA and Central Europe.
- In these countries, during the study period, there was no free press and there were no independent reporting bodies.
- The tests were initially kept secret there.

India and Pakistan (five test days, 0.3%): No national databases exist for these two countries. An effect there therefore cannot be validly examined. For the overall result this carries little weight.
For the tests of these countries, a local association cannot be examined with the available data. The result therefore applies mainly to tests in which the USA, the United Kingdom or France were involved.

3 Method

Basic idea. One cannot say that a day has "many" sightings without knowing what is normal. Each test day is therefore compared with comparison days from its surroundings.

1. **Time window.** Around each test day lie 15 days: seven before, the test day itself ("offset 0") and seven after.
2. **Roles of the days.** Offset 0 is the test day. The days immediately before and after (±1) are treated separately because delayed reports or time-zone effects could lie there. The days ±2 to ±7 are the comparison days, provided they are not themselves test days.
3. **Difference D.** For each test day: sightings on the test day minus the average of its comparison days. Without an association, the average of all D values should be close to zero.
4. **Measured variability.** Daily counts vary more than simple chance models expect (waves, media coverage, catalog peculiarities).[7] The standard deviation was therefore measured from the data rather than assumed.
5. **Placebo test.** The same difference is also computed for the twelve days with no connection to a test (offsets ±2 to ±7, the placebo positions). These are also called sham dates: they are treated like test days although no test took place there. These twelve values show how far the difference deviates by chance alone. The distance of the real test day from this range yields the z value.

Pre-specified rules (set before the results were viewed): The main comparison is offset 0 against the comparison days ±2 to ±7. As the threshold, the customary value z = 2 was chosen, at which a deviation would occur by chance alone in about 5% of cases. This value is a convention, not a natural boundary. What matters is that it was fixed before the analysis. The 1,762 nuclear tests are the primary analysis; the 24 incidents are exploratory only.

Not pre-specified: After the primary analysis, an additional analysis was run using only days on which above-ground tests took place, because the photographic-plate study considers above-ground tests. It is regarded as a sensitivity analysis (Section 6.6).

In plain terms: The placebo test answers the question "Does the real test day fall outside the range of normal variation?" If perfectly ordinary days deviate just as strongly as the test day, there is nothing special to explain.

4 Mathematical derivation and verification

This chapter describes the calculation so that it can be reproduced. Each section is followed by a paragraph headed "In plain terms".

4.1 Notation

- t_i is the calendar day of the i-th test day, $i = 1, \ldots, n$, with $n = 1{,}762$.
- $s(t)$ is the number of distinct case numbers reported on day t.
- k is the offset in days, $k \in \{-7, \ldots, +7\}$. The days $t_i + k$ form the window of test day i.

- K_i is the set of offsets k ∈ {±2, …, ±7} for which the day t_i + k is not itself a test day or incident day. n_i = |K_i| is the number of these comparison days. On average n_i ≈ 9.9 (17,407 comparison days for 1,762 test days).

4.2 The difference D

For each test day, the mean of its comparison days is formed:
$c_i = (1/n_i) \cdot \Sigma_{k \in K_i} s(t_i + k)$

The difference is the number of reports on the test day minus this mean:
$D_i = s(t_i) - c_i$

The reference value R is the mean over all comparison days of all windows:
$R = \Sigma_i \Sigma_{k \in K_i} s(t_i + k) / \Sigma_i n_i$ (here R = 10.91)

The relative effect is $\bar{D} / R$, where $\bar{D}$ is the mean of all D_i.

In plain terms: Each test day is compared with its own surroundings. This removes differences in reporting level between years: in the 1960s it is about 15 reports per day, in the other years about 9. Without an association one expects $\bar{D} \approx 0$.

4.3 Mean, standard deviation and t value

$\bar{D} = (1/n) \cdot \Sigma D_i = +0.4847$
s_D = standard deviation of the D_i = 11.72
SE_day = s_D / √n = 11.72 / 41.98 = 0.279
t_day = $\bar{D}$ / SE_day = 0.4847 / 0.279 = 1.74
The relative effect is 0.4847 / 10.91 = +4.4%.

Because tests in the same year are related, there is a more cautious variant. For each test year y the mean $\bar{D}_y$ is formed, here with Y = 55 years, and the years are treated as observations:
SE_year = (standard deviation of the $\bar{D}_y$) / √Y
t_year = (mean of the $\bar{D}_y$) / SE_year = 1.28

In plain terms: The t value states how many "standard errors" the mean is away from zero. The standard error indicates how much such a mean would vary under random repetition. Computed per test year, the t value is smaller (1.28) because only 55 observations count instead of 1,762. The per-year calculation is the more conservative one.

4.4 The placebo test

For each of the twelve placebo positions k′ ∈ {±2, …, ±7}, the same calculation as for the test day is carried out. The day t_i + k′ is treated like a test day and compared against the remaining comparison days (leave-one-out):
$D_i(k') = s(t_i + k') - (1/(n_i - 1)) \cdot \Sigma_{j \in K_i, j \neq k'} s(t_i + j)$

This applies only if t_i + k′ is not itself a test day. For each position, the average over the test days is taken ($\bar{D}(k')$). From the twelve means follow:
m = mean of the twelve $\bar{D}(k')$ = −0.0004
s_P = standard deviation of the twelve $\bar{D}(k')$ = 0.2847
The test day is measured against this range:
$z = (\bar{D}(0) - m) / s_P = (0.4846 + 0.0004) / 0.2847 = 1.70$

For the probability, the t distribution with 12 − 1 = 11 degrees of freedom is used (two-sided). For 1.70 this gives about p = 0.12. Strictly speaking, the standard deviation of the placebo D must be multiplied by $\sqrt{(1 + 1/12)} = 1.04$, because their mean is also only an estimate. This gives z = 1.64 instead of 1.70. The interpretation does not change.

In plain terms: Twelve days with no connection to a test are calculated exactly like the test day. This shows how strongly such values fluctuate by chance alone. The real test day deviates by 1.7 times this usual variation. That is not conspicuous: even if there were no association with nuclear tests, a deviation of this size would occur by chance about one time in eight.

4.5 Why a simple chance model is not enough

If the daily count s(t) were purely random (Poisson distribution), its variance would equal its mean (λ). For D_i this would give:
$\mathrm{Var}(D_i) \approx \lambda \cdot (1 + 1/n_i) \approx 10.9 \cdot 1.10 = 12.0$, that is, a standard deviation of about 3.5

The measured value was 11.72. The daily values thus fluctuate about three times as strongly as in the simple model. The standard error would be 3.5 / 41.98 = 0.08 instead of the measured 0.28, and the t value would be 5.9, about three times as large as the measured 1.74. This contributes to the earlier, seemingly high values (Section 5). The figures are approximations ($n_i \approx 9.9$ and $\lambda \approx 10.9$).

In plain terms: Reports come in waves. Media coverage and catalog peculiarities make them cluster. Anyone who treats them as independent random events considers small deviations significant. The variability was therefore measured from the data.

4.6 Dependent test days and the informative value of the placebo test

Test days often lie close together (Fig. 4). Their windows overlap, and overlapping windows share days. As a result, the D_i are not independent, and the mean $\bar{D}$ fluctuates more than $s_D/\sqrt{n}$ suggests.

Two figures show how large this effect is. First, the standard deviation of the twelve placebo D ($s_P = 0.2847$) differs by only about 2% from the standard error per test day (0.279). The placebo positions arise from the same overlapping windows and therefore already contain the dependence. Second, the standard error per test year (0.379) is about 36% larger than that per test day. The dependence is therefore not negligible, which is why the per-year measure is reported as the more cautious variant. The standard deviation from only twelve placebo positions is itself accurate to only about ±20%.

In plain terms: When nuclear tests follow each other closely, their comparison periods share the same days. Each individual test then provides less new information than if it stood alone. The placebo test captures this approximately, because the sham dates also come from the same overlapping periods and therefore fluctuate equally strongly. Anyone wanting to be very cautious computes per test year instead of per test day: each year counts only once there, and the t value is smaller.

4.7 How large could an effect be?

As an approximation, the interval is $\bar{D} \pm 1.96 \cdot SE$, divided by R.[8]

Measure	SE	Interval for D	Relative to R = 10.91
Per test year	0.379	−0.26 to +1.23	−2.4% to +11.3%
Per test day	0.279	−0.06 to +1.03	−0.6% to +9.5%

The detection limit (80% power, two-sided 5% significance level) is $(1.96 + 0.84) \cdot SE$: per test day $2.8 \cdot 0.279 = 0.78$, which is 7.2% of R; per test year $2.8 \cdot 0.379 = 1.06$, which is 9.7% of R. An effect of +45% would correspond to $0.45 \cdot 10.91 = 4.9$ reports per day, almost eighteen times the standard error per test day.

In plain terms: A real effect of about 7% to 10% of the normal daily volume would have been detected in these data with a probability of about 80%. Smaller effects disappear in the usual variation and cannot be distinguished from chance. For comparison: the photographic-plate study reports for light points an approximately 45% higher probability around a test.[1] If this showed up in the reports in the same way, there would have had to be about 5 more reports on the test day on average. About half a report was found. An effect of this size would thus not have been missed. Because light points and reports do not measure the same thing, however, this is no contradiction of the study (Section 7).

Limitation: The standard error per test year (0.379) was back-calculated from the mean $\bar{D}$ and the t value (1.28). This holds only if the mean of the yearly means roughly equals $\bar{D}$. The values for this variant are therefore approximations.

4.8 Multiple testing and rank probability

In the placebo test, 13 values stand side by side (the test day and 12 placebo positions). If all were exchangeable, the test day would rank first with probability 1/13 = 7.7% and rank first or second with probability 2/13 = 15.4%.

Anyone who examines m subanalyses will, by chance alone, find at least one conspicuous value with probability 1 − 0.95^m. For m = 5 this is about 23%, for m = 10 about 40%.

In plain terms: The more subsets one examines, the more likely a chance hit becomes. Conspicuous results in subanalyses therefore count only as leads.

4.9 Consistency checks

Check	Calculation	Result
All test days included in D	Test days in D = test days	1,762 = 1,762
Bridge rows	1,786 · 15	26,790
Roles (all events)	1,786 + 3,572 + 17,651 + 3,781	26,790
Nuclear tests only	1,762 · 15	26,430
t per test day	0.4847 / (11.72 / √1,762)	1.74
Relative effect	0.4847 / 10.91	+4.4%
Weighted mean of the blocks	(524 · 1.133 + 1,238 · 0.210) / 1,762	0.4846
Mean of the placebo D	m	−0.0004 (near zero)
SD of placebo D versus standard error	0.2847 versus 0.279	Deviation about 2%
z	(0.4846 + 0.0004) / 0.2847	1.70
Cluster pairs	2.4852 · 1,762	4,379
Pairs at ±1	4,379 − 3,737	642
Rank probability	1/13	7.7%

In plain terms: Every figure can be calculated from others. That all calculations add up shows that the model and the figures are internally consistent.

5 Data checking

A large part of the work consisted of finding errors and repeatedly checking results for plausibility. Initial intermediate values (up to about z = ±6) were artifacts: counted test entries, an unequal daily basis, a comparison that was too narrow (only the following day) and an overly simple variability assumption. With broad comparison days and measured variability, nothing of them remained. These values are not included in this study.

- **Data loss.** 27,544 records (7.3%) have no readable date and are not included in the analysis. 94% of them come from UFOCAT. Whether the loss is related to the test days was not checked. No mechanism for this is apparent.
- **Duplicate rows.** About 0.3% of records carry the same case number on the same day, mostly from the source USO. Distinct case numbers per day are counted. Duplicate rows are not conspicuously distributed around the test day (0.021 per day on the test day, 0.033 on average for the comparison days).
- **Connected test days.** 85.3% of test days have at least one other test day or incident day within ±7 days, 31.0% within ±1 (Fig. 4). Such days do not

count as comparison days. The test days are thus not independent of each other (Section 4.6).

- **Wartime uses.** Two of the 1,762 test days (5 and 9 August 1945) are the attacks on Hiroshima and Nagasaki. They are listed as nuclear tests in the data and were not removed.
- **Nuclear test list.** Corrections included the signs of the latitudes of the three Argus tests in the South Atlantic, an incorrect longitude entry (N_80055), a typo in the design type (SHAFT) and corrupted test names. One case number occurred twice (N_74044, USA and France) and was made unique. Two tests without coordinates do not appear on the map. The number of test days (1,762) and the main results remained unchanged.

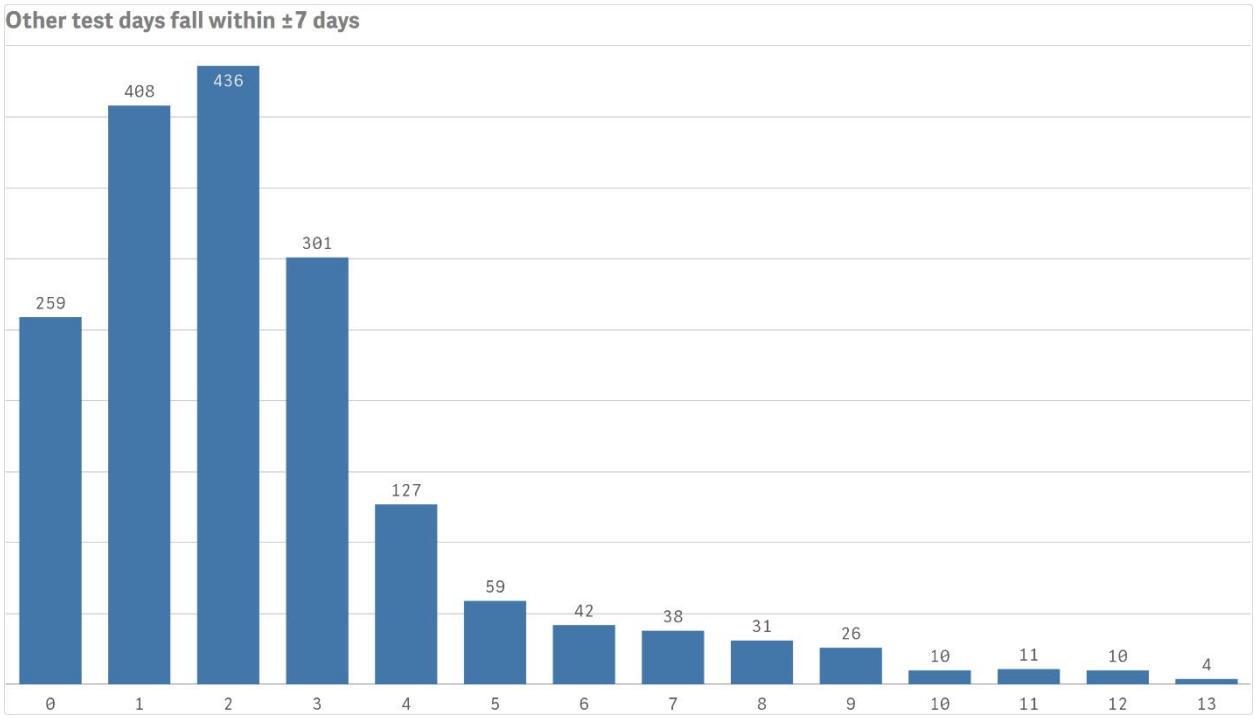


Fig. 4 *Distribution of test days by number of other test days within ±7 days.*

6 Results

6.1 Sightings per day around the test day

The average of the comparison days is 10.9 sightings per day; on the test day it is 11.4 (+4.8% in the raw profile, +4.4% after the difference calculation). The difference is of the same order as the variation among the comparison days themselves (about ±0.3; Fig. 2).

6.2 Main statistics (nuclear tests, all years)

Measure	Value
Test days / test years	1,762 / 55
Mean D (excess on the test day)	+0.48 sightings per day
Reference (normal day)	10.91
Relative effect	+4.4%
t per test day	1.74
t per test year	1.28
Placebo test: z	1.70
SD of placebo D	0.28

Tab. 2 *Main statistics (nuclear tests, type N). Selection: Type = N.*

In plain terms: On the test day there was on average almost half a report more than on comparison days. A z value of 1.70 means: the excess is 1.7 times the usual variation. Under the pre-specified rule, a value below 2 counts as no effect. The t per test year (1.28) is the more cautious variant and is also clearly below 2 (Tab. 2).

How large could a real effect be? The possible excess lies roughly between −2.4% and +11.3% of the normal daily volume (per test year) and between −0.6% and +9.5% (per test day). Both ranges include zero (derivation in Section 4.7). Magnitudes such as +45% or +68% are not seen in these data. This is no contradiction of the photographic plates, because the two studies do not measure the same thing (Section 7).

6.3 The period 1961 to 1968

Broken down by time block, the block 1961 to 1968 shows the largest deviation (Fig. 5): 524 test days, mean D +1.13, z = 2.02. Without this block, 1,238 test days remain with D = +0.21 and z = 0.79.

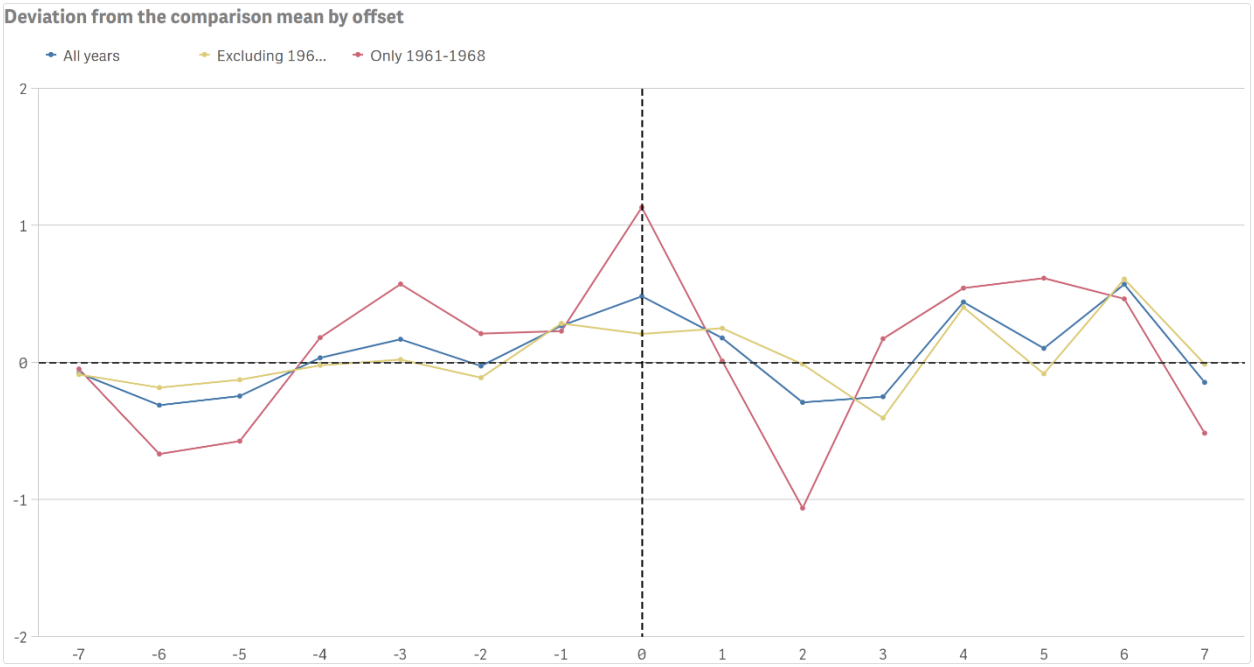


Fig. 5 *Difference by offset for all years, for 1961 to 1968 only, and without this block.*

In plain terms: This period comprises only eight years, contains particularly dense series of tests and the largest waves of reports (1966/67). Among the 13 compared positions (test day and twelve sham dates), the real test day ranks first here. By chance alone this would happen in about one case in twelve to thirteen (about 8%, Section 4.8). In addition, this period was examined more closely only after the primary analysis. It is therefore a lead that would have to prove itself in a new, pre-specified test, but not a finding.

6.4 Subanalysis: sightings from the USA

Because about 97% of the reports in the windows come from catalogs of US origin (Tab. 1), the placebo test was additionally computed using only sightings from the USA, for the 849 test days on which exclusively the USA tested. The reference here is 5.84 sightings per comparison day. On the test day the excess is +0.90 sightings per day (+15.3%). The twelve placebo D average +0.00 with a standard deviation of 0.41; the z value is 2.20 (Tab. 3; p ≈ 0.05). With the correction factor from Section 4.4 it is about 2.12 (p ≈ 0.06, Tab. 3). As an approximation (excess ± 2.2 standard deviations), the excess ranges from about 0% to +31% of the US reference.

The variant with all days on which the USA participated (904 test days) was not computed (Section 9.4).

Placebo test (N)

D at offset 0	Mean of placebo D	Stand. deviation of the placebo D	z
0,4846	-0,0004	0,2847	1,7036

Placebo test (USA only) (N)

D at offset 0	Mean of Placebo-D	Stand. deviation of the placebo D	z
0,8953	0,0022	0,4055	2,2026

Tab. 3 *Placebo test. Left: all sightings, 1,762 test days. Right: US sightings only, 849 test days (selection: Type = N and test countries = USA).*

In plain terms: In the US sightings, the test day lies more clearly above its surroundings than in the worldwide count. The value (z = 2.20) narrowly exceeds the pre-specified threshold of 2. However, several analyses were examined (Section 8, item 5), and anyone who examines many subsets finds a chance hit more easily. A stricter standard is therefore needed: with two analyses examined, the threshold would be about 2.6; with five, about 3.1. This standard was applied only afterwards and is not part of the pre-specified rules. The value therefore remains a lead that needs replication with independent data, not evidence.

6.5 Nuclear incidents

The 24 nuclear incidents are too few for a statistical statement and were not analyzed with the procedure described here. An earlier, simpler calculation suggested that a few individual days drive the result. This is not reliable.

6.6 Sensitivity analysis: above-ground tests only

The photographic-plate study considers above-ground tests. The calculation was therefore additionally run using only test days on which an above-ground test took place. Variant 1 counts only days classified with certainty as above-ground; Variant 2 additionally counts days with unclear classification (Appendix F). Entries flagged as non-explosion, SAM test or wartime use are excluded in both variants.

Measure	All tests	Variant 1	Variant 2
Test days	1,762	417	464
Reference	10.91	12.91	12.43
Mean D	+0.48	+0.78	+0.87
Relative effect	+4.4%	+6.0%	+7.0%
t per test day	1.74	1.01	1.23
t per test year	1.28	1.14	1.92
SD of placebo D	0.28	0.94	0.84
Placebo test: z	1.70	0.83	1.03

Tab. 4 *Main statistics and placebo test for all tests and for days with above-ground tests (selection: Type = N and one value in "Analysis").*

In plain terms: Both variants show the same direction as the primary analysis: on the test day, reports lie somewhat above the surroundings. The z value is lower, however, because only 417 to 464 days enter instead of 1,762 and the sham dates therefore vary more. The two variants are close together, so the result does not depend on the 59 tests with unclear classification. The possible excess lies roughly between −6% and +18% of the normal daily volume (mean D ± 1.96 standard errors per test day; Variant 1: −5.7% to +17.8%; Variant 2: −4.2% to +18.2%). An effect of +45% would have been detected here as well. This analysis is an additional check and not a second main finding.

7 Context: the photographic-plate study

This study neither confirms nor refutes the study by Bruehl and Villarroel. It merely tries to pose the same basic question with ground-based data. The differences at a glance:

- **Different measure.** The plates measure light points with an instrument at one location. Here, reports from people worldwide are counted. Reporting frequency depends on attention, press and collection practice.
- **Different period.** The plate study covers 1949 to 1957. Here the most conspicuous period is 1961 to 1968.
- **Same catalog.** Both studies rely on UFOCAT, here in the version in the GEP database. Whether both versions agree in data status was not checked. This study is therefore not an independent confirmation.
- **Methodological parallel**. The criticism of the plate study concerns mainly the definition and validation of the datasets used.[3] In this study, too, a large part of the effort went into data checking (Section 5).
- **Above-ground tests**. The plate study considers above-ground tests, the primary analysis all tests. The analysis in Section 6.6 is closer to the plate study's selection but remains a different measure (reports instead of light points, worldwide instead of on site).

The photographic-plate studies were assessed on the basis of publications and accessible reports. The original data were not fully re-examined.

8 Limits of the findings

1. **Daily counts only.** Regional effects (for example sightings near the test site) were not examined.
2. **Missing countries.** On 38.1% of test days at least one of the countries Russia (Soviet Union), China, India, Pakistan or North Korea was involved. The available catalogs contain few sightings from these countries, which holds for Western-accessible sighting catalogs in general (Section 2). An effect cannot be validly examined for these tests. The result therefore applies mainly to tests in which the USA, the United Kingdom or France were involved.
3. **Catalog quality.** Report date, duplicate reports and rounded dates (for example, the 1st or 15th of a month used as a blanket value) are not completely cleaned. The test dates are presumably given in universal time (5 August 1945 for Hiroshima suggests this), while the sightings are in local time, which can shift individual days by ±1. In addition, sighting catalogs and research methods are not standardized internationally or nationally.
4. **Only twelve placebo positions.** The chance variability was estimated from twelve comparison values that all come from the immediate surroundings of the tests. This makes the z value somewhat imprecise. With only twelve positions, a rank comparison is also coarse: the smallest possible p value is 1/13 = 7.7% (one-sided).
5. **Multiple testing.** Besides the pre-specified primary analysis, at least six subanalyses were examined, including time blocks, sightings from the USA and two variants using only above-ground tests (for Variant 2, t per test year is 1.92, just below 2). With each additional selection, the chance increases that chance alone produces a conspicuous value (as with dice: roll often enough and you eventually get a six). Conspicuous results in subsets are therefore treated only as leads, not as findings (Section 4.8).

6. **Above-ground/underground classification.** This classification is not present in the original data and is therefore based partly on inference from design type, depth, site and date. 59 of 2,057 tests are unclear (Variant 2 includes them), and two entries were assigned by hand (N_57018, N_58022). A comparison with the original source is pending. "Above-ground" is defined broadly (including water, balloon, rocket and space).

9 Conclusion

9.1 What the study shows

No established association. On nuclear test days, no more UAP sightings are reported than on the days before and after: the difference of almost half a report per day lies within normal variation. Even on ordinary days without a test there are deviations of this size; the test day does not stand out from them. The more cautious calculation per test year confirms this.

In the worldwide count, large effects are unlikely and small ones not ruled out. Effects clearly above 10% would probably have been detected. In the subanalysis for US tests and US sightings, the estimate is about +15%; an effect of this size cannot be ruled out there. An effect in only one region would be lost in a worldwide count.

Even when restricted to above-ground tests, the difference remains within normal variation (+6% to +7%, z = 0.83 to 1.03).

9.2 What the study does not show

- The subanalysis for US tests and US sightings (z = 2.20) is a lead, not evidence (Section 6.4). The same holds for the period 1961 to 1968 (Section 6.3). Both would need replication with independent data.
- The photographic-plate findings are neither refuted nor confirmed (Section 7).
- The scope is limited: on about 38% of test days, Russia (Soviet Union), China, India, Pakistan or North Korea was involved. The result therefore applies mainly to tests in which the USA, the United Kingdom or France were involved.
- The classification of tests as above-ground or underground is partly inferred (Appendix F); 59 tests remained unclear.

9.3 What the study shows nonetheless

A null result is a result. The study shows that a carefully computed daily count yields no conspicuous association, and it shows how easily spurious effects creep in. This is a useful warning for anyone examining similar questions with similar catalogs. In addition, the transparent procedure with pre-specified rules allows third parties to reproduce the calculation.

9.4 Open questions

Not examined were regional effects, the variant with all days under US participation, the period 1949 to 1957, and a replication of the US subanalysis with independent data. Whether the classification of tests as above-ground or underground holds up against the original source is open.

Appendix

A Glossary

Term	Meaning
Test day	Date on which at least one nuclear test took place
Offset	Distance of a day from the test day, in days
Comparison days	Days with offset ±2 to ±7 that are not themselves test days
Placebo test	The same calculation as on the test day, applied to days with no connection to a test. Shows how strongly results vary by chance alone (also: sham-date test)
Placebo position	One of the twelve positions (offset ±2 to ±7) that are computed like a test day (also: sham date)
Placebo D	The difference D, computed at a placebo position
D (difference)	Sightings on a day minus the average of the comparison days
z value / t value	Distance of a result from chance variation, in units of variation. Values below 2 are regarded as not established
Standard error	How much a mean would vary under random repetition
p value	If there were no effect: how often a deviation at least this large would occur by chance
Overdispersion	Daily values vary more than simple chance models expect
t per test year	More cautious variant of the t value, because tests in the same year are related
Reference	Mean sightings per comparison day, the “normal” level
Median	The middle value of a sorted series, unaffected by outliers
Above-ground	Test with a design type or depth entry above ground, in or on water, on a balloon, by rocket or in space (broad definition)
Unknown	Test that could be classified neither by design type nor by depth, nor by site and date (59)
Sensitivity analysis	The same calculation with a changed selection or assumption (here: above-ground tests only), to check whether the result depends on it. It counts as an additional check, not a second main result

B Key figures of the main statistics

Field	Meaning
Test days	Number of nuclear test days in the selection
Test days in D	Control figure: must equal "test days", otherwise days have dropped out of the calculation unnoticed
Test years	Number of years with tests (tests in the same year count for less than independent observations)
Mean D	Average excess on the test day, in reports per day
SD of D	How strongly the individual values vary
Relative effect	Mean D divided by reference
Share above control mean	Share of test days above their surroundings (without an effect not exactly 0.5, because a few peak days skew the distribution)
Median D	Typical test day. Comparison with the mean shows whether a few peak days drive the mean (here −0.92 versus +0.48)

C Reproducibility and data availability

The analysis was run in Qlik Cloud with a load script (calendar, event table, bridge table, placebo table). The load script exists and can be requested from the author. The figures were recalculated independently from the exports. The nuclear test data (Johnston's Archive) and the UFOCAT catalog (CUFOS) are publicly accessible. The GEP database and the load script are not freely available and can be requested from the author. The nuclear test list is cleaned and classified in a separate script, which can be requested from the author.

D Footnotes

1. Bruehl, S. & Villarroel, B. (2025): *Transients in the Palomar Observatory Sky Survey (POSS-I) may be associated with nuclear testing and reports of unidentified anomalous phenomena*. *Scientific Reports*, published 20 October 2025. https://www.nature.com/articles/s41598-025-21620-3

2. Center for UFO Studies (CUFOS): UFOCAT. https://cufos.org/cufos-publications-databases/ufocat/

3. Watters, W. A., Dominé, L., Little, S., Pratt, C. & Knuth, K. H. (2026): *Critical Evaluation of Studies Alleging Evidence for Technosignatures in the POSS1-E Photographic Plates*. arXiv:2601.21946 (preprint, not peer-reviewed; several versions). The focus is on the validation of the datasets and the so-called Earth-shadow deficit.

4. Villarroel, B., Streblyanska, A., Bruehl, S. & Geier, S. (2026): *A Response to paper "Critical Evaluation of Studies Alleging Evidence for Technosignatures in the POSS1-E Photographic Plates" by Watters et al. (2026)*. arXiv:2602.15171 (preprint, submitted 16 February 2026, second version of 30 April 2026). A peer-review status was not discernible in the arXiv entry (as of October 2026). https://arxiv.org/abs/2602.15171

5. In the entire catalog, 348,638 rows have a readable date; they belong to about 347,600 distinct case numbers because individual reports have several rows. In the windows around nuclear tests lie 121,197 rows, which are 120,829 distinct reports. The column sum in Tab. 1 is 120,830 because one case number occurs in two decades. In all analyses, distinct case numbers per day are counted (Section 5).

6. Johnston, W. R.: *Johnston's Archive, Nuclear Tests – Databases and Other Material*. https://johnstonsarchive.net/nuclear/tests/index.html (accessed 28 October 2026). The list was cleaned (Section 5), and the case numbers were made unique.

7. The difference D varies with a standard deviation of 11.72, about three times as strongly as would be expected under a simple chance model (Poisson) (about 3.5). Derivation in Section 4.5.

8. Rough interval: mean D ± 1.96 standard errors, divided by the reference 10.91. Standard error per test year about 0.38 (back-calculated from t_year), per test day 0.28. It is an approximation and not an exact confidence-interval calculation.

E References (alphabetical)

- Bruehl, S. & Villarroel, B. (2025): Transients in the Palomar Observatory Sky Survey (POSS-I) may be associated with nuclear testing and reports of unidentified anomalous phenomena. *Scientific Reports.* https://www.nature.com/articles/s41598-025-21620-3

- Center for UFO Studies (CUFOS): UFOCAT. https://cufos.org/cufos-publications-databases/ufocat/

- Johnston, W. R.: Johnston's Archive, Nuclear Tests – Databases and Other Material. https://johnstonsarchive.net/nuclear/tests/index.html (accessed 28 October 2026)

- Villarroel, B., Streblyanska, A., Bruehl, S. & Geier, S. (2026): A Response to paper "Critical Evaluation of Studies Alleging Evidence for Technosignatures in the POSS1-E Photographic Plates" by Watters et al. (2026). arXiv:2602.15171. https://arxiv.org/abs/2602.15171

- Watters, W. A., Dominé, L., Little, S., Pratt, C. & Knuth, K. H. (2026): Critical Evaluation of Studies Alleging Evidence for Technosignatures in the POSS1-E Photographic Plates. arXiv:2601.21946.

F Classification as above-ground/underground

Order of classification: (1) design type, (2) sign of the depth entry, (3) site and date (last test at the site classified with certainty as above-ground and first test classified with certainty as underground, with a 60-day buffer), otherwise unknown.

Classification	Tests
above-ground	525
underground	1,473
unknown	59
Total	**2,057**

Basis of classification	Tests
Design type	1,710
Depth	45
Site and date	243
open	59
Total	**2,057**

Assigned by hand (inferred from date and yield, to be checked against the original source): N_57018 underground (consistent with a tunnel test in 1957), N_58022 above-ground (consistent with an underwater test in 1958). Crater explosions (N_65016, N_62067) count as above-ground. N_58016 has coordinates that fit no test site and remained unknown.

Author: Ralf van Thuyl (member of GEP), contact: homuell2@gmail.com

Thanks to the Gesellschaft zur Erforschung des UFO-Phänomens (GEP). https://www.uap-research.org, https://www.ufo-forschung.de

Notes on AI tools: Text drafting and computational cross-checking of the exports were supported by Claude (Anthropic). The analysis in Qlik and the responsibility for the content rest with the author.

© 2026 Ralf van Thuyl. This work is licensed under CC BY 4.0 (attribution). Map material in Fig. 1: © OpenStreetMap contributors/esri, excluded from the license.