Gibt es mehr UAP-Meldungen an Atomtest-Tagen?
Ein Tageszählungs-Vergleich mit Placebo-Test für die Jahre 1945 bis 2017

Ralf van Thuyl

2025 berichteten Stephen Bruehl und Beatriz Villarroel, dass auf Himmelsfotos der 1950er Jahre rund um oberirdische Atomtests mehr kurze Lichtpunkte zu sehen waren. Nach Berichten fanden sie außerdem an Tagen mit mehr UAP-Meldungen mehr Lichtpunkte.[1] Diese Arbeit fragt, ob sich ein Zusammenhang mit Atomtests auch in den gemeldeten Sichtungen selbst zeigt, also in Zeugenmeldungen von der Erde aus. Sie will diese Studie weder bestätigen noch widerlegen, denn beide messen nicht dasselbe (Abschnitt 7).

Ergebnis: In den untersuchten Meldedaten werden an Atomtest-Tagen nicht mehr UAP-Sichtungen gemeldet als an den Tagen davor und danach. Am Testtag lagen die Meldungen im Schnitt etwa 4 % höher (11,4 gegenüber 10,9 pro Tag; Abb. 2). Das ist kein auffälliger Unterschied: Auch wenn es gar keinen Zusammenhang gäbe, käme ein Ausschlag dieser Größe etwa bei jedem achten Versuch allein durch Zufall vor. Als auffällig hätte nach der vorab festgelegten Regel erst ein Wert ab z = 2 gegolten, hier sind es z = 1,70. Große Effekte sind in diesen Tageszählungen unwahrscheinlich, kleine oder regionale Effekte nicht ausgeschlossen. In einer Teilauswertung mit US-Sichtungen und ausschließlich US-Tests (849 Testtage) liegt der Wert bei z = 2,20. Weil mehrere Auswertungen angesehen wurden, ist das ein Hinweis, der eine Wiederholung braucht, kein Beleg.

Beschränkt man die Auswertung auf Tage mit oberirdischen Tests, wie sie auch die Fotoplatten-Studie betrachtet, liegt der geschätzte Unterschied bei +6 % bis +7 % (417 bis 464 Testtage; z = 0,83 bis 1,03). Die Richtung entspricht der Hauptauswertung, der Wert liegt aber innerhalb der normalen Schwankung.

Grundlage sind 1.762 Atomtest-Tage (1945 bis 2017) und die Datenbank der Gesellschaft zur Erforschung des UFO-Phänomens (GEP) mit rund 376.000 Meldungen. Jeder Testtag wurde mit den sieben Tagen davor und danach verglichen.

Einschränkend ist zu erwähnen, dass an 671 der 1.762 Testtage (38,1 %) mindestens eines der Länder Russland (Sowjetunion), China, Indien, Pakistan oder Nordkorea beteiligt war. Aus Russland, China und Nordkorea sind Sichtungen in den Katalogen kaum vertreten, für Indien und Pakistan liegen keine nationalen Datenbanken vor. Ein möglicher Effekt der Tests dieser Länder wäre in den Zahlen daher nicht auffindbar.



***Abb. 1**: Orte der Atomtests zwischen 1945 und 2017.*
Kartenmaterial: © OpenStreetMap contributors/esri.

Das Wichtigste in Kürze

- **Frage:** Werden an Tagen mit Atomtests mehr UAP-Sichtungen gemeldet als an den Tagen davor und danach?
- **Ergebnis:** Etwa 4 % mehr Sichtungen am Testtag, innerhalb der normalen Schwankung (z = 1,70, Schwelle 2). Kein belegter Zusammenhang.
- **Hinweis:** Für US-Tests und US-Sichtungen liegt z bei 2,20. Wegen mehrerer Auswertungen ein Hinweis, kein Beleg.
- **Oberirdische Tests:** Nur Tage mit oberirdischen Tests: +6 bis +7 % (z = 0,83 bis 1,03). Gleiche Richtung, kein belegter Zusammenhang.
- **Grenze:** An etwa 38 % der Testtage war ein Land beteiligt, aus dem die Kataloge kaum Sichtungen enthalten. Kleine oder regionale Effekte sind nicht ausgeschlossen.
- **Einordnung:** Die Fotoplatten-Studie wird weder bestätigt noch widerlegt.

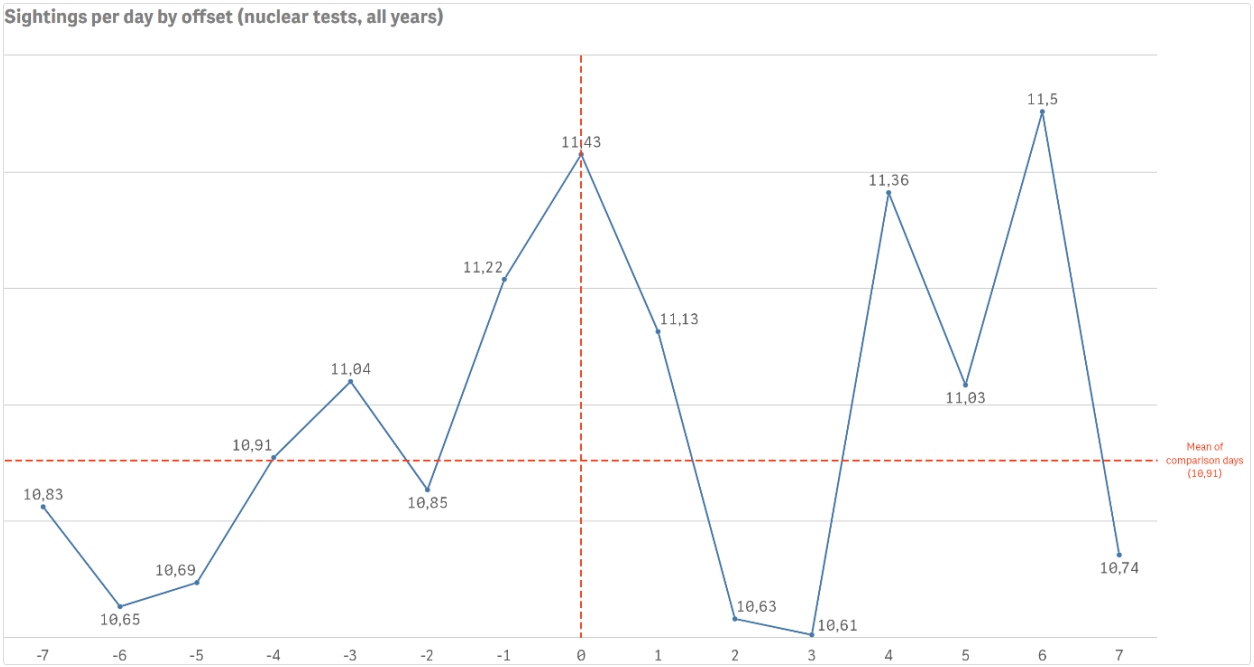


Abb. 2 *Sichtungen Atomtests pro Tag nach Versatz (Y-Achse gekürzt).*

1 Einleitung

1.1 Hintergrund

Im Jahr 2025 werteten Stephen Bruehl und Beatriz Villarroel die Fotoplatten des ersten Palomar-Himmelsatlas (1949 bis 1957) aus.[1] Sie berichten, dass kurze, sternähnliche Lichtpunkte ("Transienten") häufiger auftraten, wenn ein oberirdischer Atomtest in der Nähe lag. An Tagen innerhalb eines Tages um einen Test waren sie 45 % wahrscheinlicher, am Tag nach einem Test 68 % wahrscheinlicher. Nach Berichten fanden die Autoren zusätzlich eine kleine Korrelation mit UAP-Meldungen aus dem Katalog UFOCAT.[2] Zu dieser Studie gibt es eine kritische Auseinandersetzung (Watters u. a., 2026)[3] und eine Erwiderung der Studienautoren.[4]

Der Autor dieser Arbeit nahm das zum Anlass, die verwandte Frage mit erdgebundenen Daten zu stellen: Gibt es auch in den Zeugenmeldungen einen Zusammenhang mit Atomtests bis 2017? Ein direkter Vergleich mit den Fotoplatten ist nicht möglich. Zeugenmeldungen sind subjektiv und schwer überprüfbar, und die Testgebiete waren abgelegen und gesperrt, direkte Beobachtungen vor Ort sind daher so gut wie ausgeschlossen. Die Orte der Atomtests zeigt Abb. 1.

1.2 Fragestellung

Frage: Werden an Tagen mit Atomtests mehr UAP-Sichtungen gemeldet als an anderen Tagen?

- **H_0 (Nullhypothese):** Es gibt keinen Unterschied außerhalb der üblichen Schwankung.
- **H_1 (Gegenhypothese):** An Atomtest-Tagen ist die Zahl der gemeldeten Sichtungen höher als an vergleichbaren Tagen ohne Test.

Im Klartext: Die Nullhypothese ist der "Normalfall", von dem man ausgeht, bis die Daten etwas anderes zeigen. Gesucht wird, ob sich am Testtag etwas findet, das sich nicht mit Zufall erklären lässt. Die Ergebnisse werden zweiseitig bewertet, also bei Abweichungen nach oben und nach unten.

2 Daten

Sichtungen. Grundlage ist die Datenbank der GEP, die zehn nationale und internationale Sichtungskataloge zusammenführt. Von den aktuell 376.182 Datensätzen haben 348.638 ein lesbares Datum. Das entspricht rund 347.600 verschiedenen Fallnummern.[5]
Datensätze außerhalb des Untersuchungszeitraums (zum Beispiel Jahre vor 1940) fallen in keines der Fenster. Im Zeitraum 1961 bis 1968 stammen etwa 92 % der Meldungen aus einem einzigen Katalog, UFOCAT.[2] In mindestens einem der ±7-Tage-Fenster um einen Atomtest-Tag liegen 120.829 verschiedene Meldungen (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 *Sichtungen in den ±7-Tage-Fenstern um Atomtest-Tage, nach Quelldatenbank und Jahrzehnt (nur Atomtests)*

Atomereignisse. Die Daten zu den Atomtests sowie den 24 nuklearen Zwischenfällen stammen aus Johnston's Archive.[6] Sie enthalten 2.081 Einträge mit Datum von 1945 bis 2017. Daraus ergeben sich 1.762 Atomtest-Tage und 24 Tage mit nuklearen Zwischenfällen. Mehrere Tests am selben Tag zählen als ein Testtag. Die Liste wurde auf Eingabe- und Formatfehler geprüft und bereinigt (Abschnitt 5). Ein Abgleich jedes Einzeleintrags mit der Originalquelle erfolgte nicht. Für die Auswertung in Abschnitt 6.6 wurde jeder der 2.057 Tests als oberirdisch (525), unterirdisch (1.473) oder unbekannt (59) eingestuft, und zwar nacheinander nach Bauart, Tiefenangabe und zuletzt nach Testgelände und Datum (Anhang F).

Welche Länder testeten? An 71 der 1.762 Testtage (4,0 %) fanden Tests mehrerer Länder statt, meist Sowjetunion und USA. An Tagen, an denen nur ein Land testete, waren es: USA 849 Tage, Sowjetunion 561, Frankreich 192, Großbritannien 39, China 39, übrige 11 (Abb. 3).

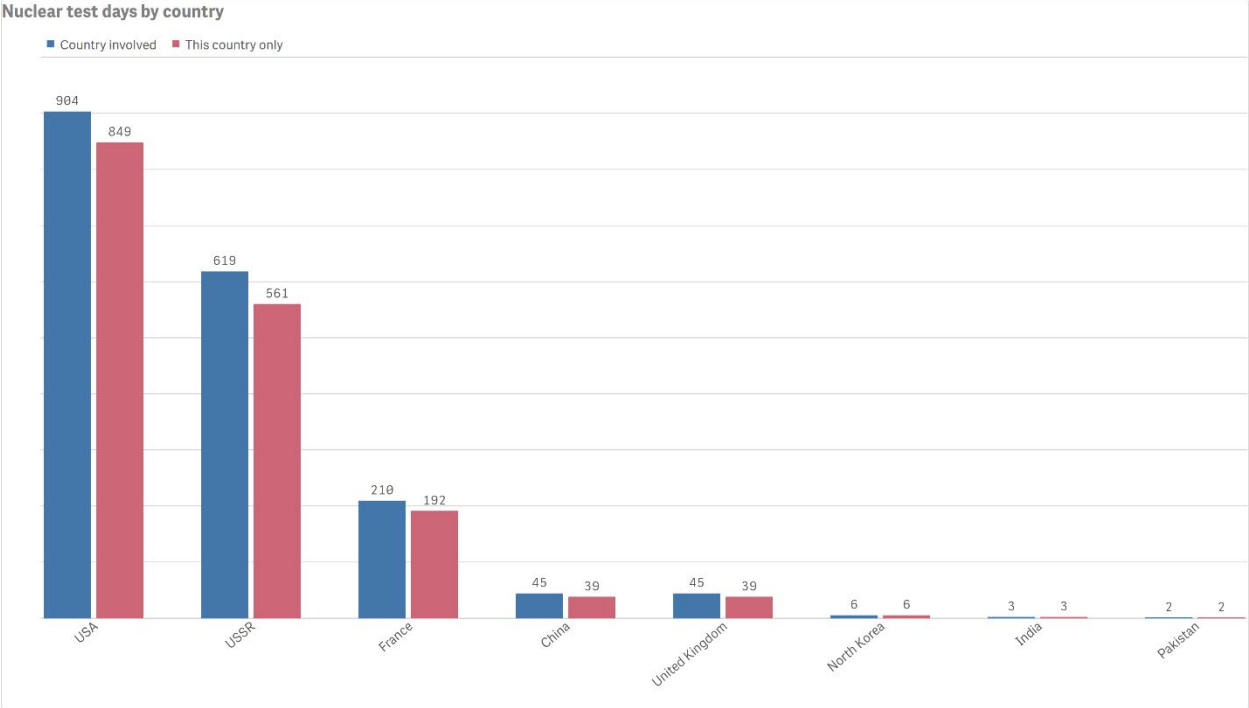


Abb. 3 *Testtage nach Land, einmal mit allen beteiligten Ländern, einmal nur Tage, an denen ausschließlich dieses Land testete.*

Wichtige Lücke. An 671 der 1.762 Testtage (38,1 %) war mindestens eines der Länder Russland (Sowjetunion), China, Indien, Pakistan oder Nordkorea beteiligt. Die Summe der Länderbeteiligungen beträgt 675, weil an vier Tagen Sowjetunion und China gemeinsam testeten.

Russland, China und Nordkorea (666 Testtage, 37,8 %): Sichtungen aus diesen Ländern sind in den Katalogen der GEP-Datenbank kaum vertreten. Das ist kein Mangel dieser Arbeit allein, sondern ein Problem der westlich zugänglichen Sichtungskataloge insgesamt. Ob es in diesen Ländern eigene Sammlungen gibt, wurde nicht untersucht, ebenso wenig die Ursache. Drei Erklärungsansätze liegen nahe:

- Die Kataloge stammen überwiegend aus den USA und Mitteleuropa.
- In diesen Ländern bestanden im Untersuchungszeitraum keine freie Presse und keine unabhängigen Meldestellen.
- Die Tests wurden dort zunächst geheim gehalten.

Indien und Pakistan (fünf Testtage, 0,3 %): Für diese beiden Länder liegen keine nationalen Datenbanken vor. Ein Effekt lässt sich dort daher nicht valide untersuchen. Für das Gesamtergebnis hat das kaum Gewicht.
Für die Tests dieser Länder ist ein örtlicher Zusammenhang mit den vorliegenden Daten nicht prüfbar. Das Ergebnis gilt daher vor allem für Tests, an denen die USA, Großbritannien oder Frankreich beteiligt waren.

3 Methode

Grundidee. Man kann nicht sagen, ein Tag habe "viele" Sichtungen, ohne zu wissen, was normal ist. Deshalb wird jeder Testtag mit Vergleichstagen aus seiner Umgebung verglichen.

1. **Zeitfenster.** Um jeden Testtag liegen 15 Tage: sieben davor, der Testtag selbst ("Versatz 0") und sieben danach.
2. **Rollen der Tage.** Versatz 0 ist der Testtag. Die Tage unmittelbar davor und danach (±1) werden getrennt betrachtet, weil dort verzögerte Meldungen oder Zeitzoneneffekte liegen könnten. Die Tage ±2 bis ±7 sind die Vergleichstage, sofern sie nicht selbst Testtage sind.
3. **Differenz D.** Für jeden Testtag gilt: Sichtungen am Testtag minus Durchschnitt seiner Vergleichstage. Ohne Zusammenhang sollte der Durchschnitt aller D-Werte nahe null liegen.
4. **Gemessene Streuung.** Die Tageszahlen schwanken stärker, als einfache Zufallsmodelle erwarten (Wellen, Medienberichte, Katalogeigenheiten).[7] Deshalb wurde die Streuung aus den Daten gemessen und nicht angenommen.
5. **Placebo-Test.** Dieselbe Differenz wird auch für die zwölf Tage ohne Testbezug (Versatz ±2 bis ±7, die Placebo-Positionen) berechnet. Man nennt sie auch Scheintermine: Sie werden wie Testtage behandelt, obwohl dort kein Test stattfand. Diese zwölf Werte zeigen, wie weit die Differenz allein durch Zufall ausschlägt. Der Abstand des echten Testtags zu dieser Spanne ergibt den z-Wert.

Vorab festgelegte Regeln (bevor die Ergebnisse angesehen wurden): Hauptvergleich ist Versatz 0 gegen die Vergleichstage ±2 bis ±7. Als Schwelle wurde der übliche Wert z = 2 gewählt, bei dem ein Ausschlag bei reinem Zufall in etwa 5 % der Fälle vorkäme. Dieser Wert ist eine Konvention und keine Naturgrenze. Entscheidend ist, dass er vor der Auswertung festgelegt wurde. Die 1.762 Atomtests sind die Hauptanalyse, die 24 Zwischenfälle nur explorativ.
Nicht vorab festgelegt: Nach der Hauptauswertung wurde zusätzlich eine Auswertung nur mit Tagen gerechnet, an denen oberirdische Tests stattfanden, weil die Fotoplatten-Studie oberirdische Tests betrachtet. Sie gilt als Sensitivitätsanalyse (Abschnitt 6.6).

Im Klartext: Der Placebo-Test beantwortet die Frage "Fällt der echte Testtag aus dem Rahmen der normalen Schwankung?". Wenn selbst ganz gewöhnliche Tage ähnlich stark abweichen wie der Testtag, gibt es nichts Besonderes zu erklären.

4 Mathematische Herleitung und Prüfung

Dieses Kapitel beschreibt die Rechnung so, dass sie nachvollzogen werden kann. Nach jedem Abschnitt steht ein Absatz "Im Klartext".

4.1 Bezeichnungen

- t_i ist der Kalendertag des i-ten Testtags, $i = 1, \ldots, n$, mit $n = 1.762$.
- $s(t)$ ist die Zahl der verschiedenen Fallnummern, die am Tag t gemeldet wurden.
- k ist der Versatz in Tagen, $k \in \{-7, \ldots, +7\}$. Die Tage $t_i + k$ bilden das Fenster von Testtag i.

- K_i ist die Menge der Versätze $k \in \{\pm 2, \ldots, \pm 7\}$, bei denen der Tag $t_i + k$ nicht selbst ein Testtag oder Zwischenfalltag ist. $n_i = |K_i|$ ist die Zahl dieser Vergleichstage. Im Mittel ist $n_i \approx 9{,}9$ (17.407 Vergleichstage bei 1.762 Testtagen).

4.2 Die Differenz D

Für jeden Testtag wird das Mittel seiner Vergleichstage gebildet:
$c_i = (1/n_i) \cdot \Sigma_{k \in K_i} s(t_i + k)$

Die Differenz ist die Meldezahl am Testtag minus dieses Mittel:
$D_i = s(t_i) - c_i$

Der Referenzwert R ist das Mittel über alle Vergleichstage aller Fenster:
$R = \Sigma_i \Sigma_{k \in K_i} s(t_i + k) / \Sigma_i n_i$ (hier R = 10,91)

Der relative Effekt ist $\bar{D} / R$, mit $\bar{D}$ als Mittel aller D_i.

Im Klartext: Jeder Testtag wird mit seinem eigenen Umfeld verglichen. Das schaltet Unterschiede im Meldeniveau zwischen den Jahren aus: In den 1960ern liegt es bei etwa 15 Meldungen pro Tag, in den übrigen Jahren bei etwa 9. Ohne Zusammenhang erwartet man $\bar{D} \approx 0$.

4.3 Mittelwert, Streuung und t-Wert

$\bar{D} = (1/n) \cdot \Sigma D_i = +0{,}4847$
s_D = Standardabweichung der D_i = 11,72
$SE_{Tag} = s_D / \sqrt{n} = 11{,}72 / 41{,}98 = 0{,}279$
$t_{Tag} = \bar{D} / SE_{Tag} = 0{,}4847 / 0{,}279 = 1{,}74$

Der relative Effekt beträgt 0,4847 / 10,91 = +4,4 %.

Weil Tests im selben Jahr zusammenhängen, gibt es eine vorsichtigere Variante. Man bildet je Testjahr y das Mittel $\bar{D}_y$, hier mit Y = 55 Jahren, und behandelt die Jahre als Beobachtungen:
SE_{Jahr} = (Standardabweichung der $\bar{D}_y$) / $\sqrt{Y}$
t_{Jahr} = (Mittel der $\bar{D}_y$) / SE_{Jahr} = 1,28

Im Klartext: Der t-Wert gibt an, wie viele "Standardfehler" der Mittelwert von null entfernt ist. Der Standardfehler sagt, wie stark so ein Mittelwert bei zufälliger Wiederholung schwanken würde. Je Testjahr gerechnet ist der t-Wert kleiner (1,28), weil nur 55 statt 1.762 Beobachtungen zählen. Die Jahresrechnung ist die konservativere.

4.4 Der Placebo-Test

Für jede der zwölf Placebo-Positionen $k' \in \{\pm 2, \ldots, \pm 7\}$ wird dieselbe Rechnung wie am Testtag durchgeführt. Der Tag $t_i + k'$ wird dabei wie ein Testtag behandelt und gegen die übrigen Vergleichstage verglichen (Leave-one-out):
$D_i(k') = s(t_i + k') - (1/(n_i - 1)) \cdot \Sigma_{j \in K_i, j \neq k'} s(t_i + j)$

Dies gilt nur, wenn $t_i + k'$ selbst kein Testtag ist. Für jede Position wird über die Testtage gemittelt ($\bar{D}(k')$). Aus den zwölf Mittelwerten folgen:
m = Mittel der zwölf $\bar{D}(k')$ = −0,0004
s_P = Standardabweichung der zwölf $\bar{D}(k')$ = 0,2847

Der Testtag wird an dieser Spanne gemessen:
$z = (\bar{D}(0) - m) / s_P = (0{,}4846 + 0{,}0004) / 0{,}2847 = 1{,}70$
Für die Wahrscheinlichkeit nimmt man die t-Verteilung mit 12 − 1 = 11 Freiheitsgraden (zweiseitig). Für 1,70 ergibt das etwa p = 0,12. Genau gerechnet ist die Streuung der Placebo-D mit $\sqrt{(1 + 1/12)} = 1{,}04$ zu multiplizieren, weil auch ihr Mittelwert nur geschätzt ist. Dann ergibt sich z = 1,64 statt 1,70. Die Einordnung ändert sich dadurch nicht.

Im Klartext: Zwölf Tage ohne Testbezug werden genauso gerechnet wie der Testtag. So sieht man, wie stark solche Werte schon durch Zufall schwanken. Der echte Testtag weicht um das 1,7-Fache dieser üblichen Schwankung ab. Das ist wenig auffällig: Auch wenn es keinen Zusammenhang mit Atomtests gäbe, käme eine Abweichung dieser Größe bei etwa jedem achten Versuch allein durch Zufall vor.

4.5 Warum ein einfaches Zufallsmodell nicht reicht

Wäre die Tageszahl s(t) rein zufällig (Poisson-Verteilung), wäre ihre Varianz gleich dem Mittelwert (λ). Für D_i ergäbe sich:
$\mathrm{Var}(D_i) \approx \lambda \cdot (1 + 1/n_i) \approx 10{,}9 \cdot 1{,}10 = 12{,}0$, also eine Standardabweichung von etwa 3,5

Gemessen wurde 11,72. Die Tageswerte schwanken also etwa dreimal so stark wie im einfachen Modell. Der Standardfehler wäre 3,5 / 41,98 = 0,08 statt der gemessenen 0,28, und der t-Wert fiele mit 5,9 etwa dreimal so groß aus wie die gemessenen 1,74. Das trägt zu den früheren, scheinbar hohen Werten bei (Abschnitt 5). Die Zahlen sind eine Näherung ($n_i \approx 9{,}9$ und $\lambda \approx 10{,}9$).

Im Klartext: Meldungen kommen in Wellen. Medienberichte und Katalogeigenheiten lassen sie klumpen. Wer so tut, als wären sie unabhängige Zufallsereignisse, hält kleine Ausschläge für bedeutsam. Deshalb wurde die Streuung aus den Daten gemessen.

4.6 Abhängige Testtage und Aussagekraft des Placebo-Tests

Testtage liegen oft dicht beieinander (Abb. 4). Ihre Fenster überlappen sich, und überlappende Fenster teilen sich Tage. Dadurch sind die D_i nicht unabhängig, und das Mittel $\bar{D}$ schwankt stärker, als es $s_D/\sqrt{n}$ vermuten lässt.

Zwei Zahlen zeigen, wie groß dieser Effekt ist. Erstens weicht die Streuung der zwölf Placebo-D ($s_P = 0{,}2847$) nur um etwa 2 % vom Standardfehler je Testtag (0,279) ab. Die Placebo-Positionen entstehen aus denselben überlappenden Fenstern und enthalten die Abhängigkeit deshalb bereits. Zweitens ist der Standardfehler je Testjahr (0,379) etwa 36 % größer als der je Testtag. Die Abhängigkeit ist also nicht vernachlässigbar, weshalb das Maß je Testjahr als vorsichtigere Variante mitgeführt wird. Die Streuung aus nur zwölf Placebo-Positionen ist dabei selbst nur auf etwa ±20 % genau.

Im Klartext: Wenn Atomtests dicht aufeinander folgen, teilen sich ihre Vergleichszeiträume dieselben Tage. Dann liefert jeder einzelne Test weniger neue Information, als wenn er allein stünde. Der Placebo-Test fängt das näherungsweise auf, denn auch die Scheintermine stammen aus denselben überlappenden Zeiträumen und schwanken deshalb gleich stark. Wer ganz vorsichtig sein will, rechnet je Testjahr statt je Testtag: Dort zählt jedes Jahr nur einmal, und der t-Wert fällt kleiner aus.

4.7 Wie groß könnte ein Effekt sein?

Näherungsweise gilt als Intervall: $\bar{D} \pm 1{,}96 \cdot SE$, geteilt durch R.[8]

Maß	SE	Intervall für D	relativ zu R = 10,91
je Testjahr	0,379	−0,26 bis +1,23	−2,4 % bis +11,3 %
je Testtag	0,279	−0,06 bis +1,03	−0,6 % bis +9,5 %

Die Nachweisgrenze (Wahrscheinlichkeit 80 %, Schwelle zweiseitig 5 %) liegt bei $(1{,}96 + 0{,}84) \cdot SE$: je Testtag $2{,}8 \cdot 0{,}279 = 0{,}78$, das sind 7,2 % von R, je Testjahr $2{,}8 \cdot 0{,}379 = 1{,}06$, das sind 9,7 % von R. Ein Effekt von +45 % entspräche $0{,}45 \cdot 10{,}91 = 4{,}9$ Meldungen pro Tag, knapp dem Achtzehnfachen des Standardfehlers je Testtag.

Im Klartext: Ein echter Effekt von etwa 7 bis 10 % der normalen Tagesmenge wäre in diesen Daten mit einer Wahrscheinlichkeit von etwa 80 % aufgefallen. Kleinere Effekte gehen in der üblichen Schwankung unter und lassen sich von Zufall nicht unterscheiden. Zum Vergleich: Die Fotoplatten-Studie berichtet für Lichtpunkte eine um etwa 45 % höhere Wahrscheinlichkeit rund um einen Test.[1] Würde sich das in den Meldungen genauso zeigen, hätte es am Testtag im Schnitt rund 5 Meldungen mehr geben müssen. Gefunden wurde knapp eine halbe. Ein Effekt dieser Größe wäre also nicht zu übersehen gewesen. Weil Lichtpunkte und Meldungen aber nicht dasselbe messen, ist das kein Widerspruch zur Studie (Abschnitt 7).

Einschränkung: Der Standardfehler je Testjahr (0,379) wurde aus dem Mittelwert $\bar{D}$ und dem t-Wert (1,28) zurückgerechnet. Das stimmt nur, wenn das Mittel der Jahresmittel ungefähr $\bar{D}$ entspricht. Die Werte für diese Variante sind deshalb Näherungen.

4.8 Mehrfachprüfungen und Rangwahrscheinlichkeit

Beim Placebo-Test stehen 13 Werte nebeneinander (Testtag und 12 Placebo-Positionen). Wenn alle austauschbar wären, läge der Testtag mit der Wahrscheinlichkeit 1/13 = 7,7 % auf Rang 1 und mit 2/13 = 15,4 % auf Rang 1 oder 2.

Wer m Teilauswertungen ansieht, findet bei reinem Zufall mit der Wahrscheinlichkeit 1 − 0,95^m mindestens einen auffälligen Wert. Bei m = 5 sind das etwa 23 %, bei m = 10 etwa 40 %.

Im Klartext: Je mehr Ausschnitte man ansieht, desto wahrscheinlicher ist ein Zufallstreffer. Auffälligkeiten in Teilauswertungen gelten deshalb nur als Hinweise.

4.9 Kontrollrechnungen

Prüfung	Rechnung	Ergebnis
Alle Testtage in D enthalten	Testtage in D = Testtage	1.762 = 1.762
Brückenzeilen	1.786 · 15	26.790
Rollen (alle Ereignisse)	1.786 + 3.572 + 17.651 + 3.781	26.790
Nur Atomtests	1.762 · 15	26.430
t je Testtag	0,4847 / (11,72 / √1.762)	1,74
Relativer Effekt	0,4847 / 10,91	+4,4 %
Gewichtetes Mittel der Blöcke	(524 · 1,133 + 1.238 · 0,210) / 1.762	0,4846
Mittel der Placebo-D	m	−0,0004 (nahe null)
Streuung der Placebo-D gegen Standardfehler	0,2847 gegen 0,279	Abweichung etwa 2 %
z	(0,4846 + 0,0004) / 0,2847	1,70
Cluster-Paare	2,4852 · 1.762	4.379
Paare bei ±1	4.379 − 3.737	642
Rangwahrscheinlichkeit	1/13	7,7 %

Im Klartext: Jede Kennzahl lässt sich aus anderen berechnen. Dass alle Rechnungen aufgehen, zeigt, dass Modell und Kennzahlen in sich stimmig sind.

5 Datenprüfung

Ein großer Teil der Arbeit bestand darin, Fehler zu finden und Ergebnisse immer wieder zu plausibilisieren. Anfängliche Zwischenwerte (bis etwa z = ±6) waren Artefakte: mitgezählte Testeinträge, eine ungleiche Tagesbasis, ein zu schmaler Vergleich (nur der Folgetag) und eine zu einfache Streuungsannahme. Mit breiten Vergleichstagen und gemessener Streuung blieb davon nichts übrig. Diese Werte sind in der Arbeit nicht enthalten.

- **Datenverlust.** 27.544 Datensätze (7,3 %) haben kein lesbares Datum und gehen nicht in die Auswertung ein. 94 % davon stammen aus UFOCAT. Ob der Verlust mit den Testtagen zusammenhängt, wurde nicht geprüft. Ein Mechanismus dafür ist nicht ersichtlich.
- **Doppelzeilen.** Etwa 0,3 % der Datensätze tragen dieselbe Fallnummer am selben Tag, überwiegend aus der Quelle USO. Gezählt werden verschiedene Fallnummern je Tag. Doppelzeilen verteilen sich nicht auffällig um den Testtag (0,021 pro Tag am Testtag, 0,033 im Mittel der Vergleichstage).
- **Zusammenhängende Testtage.** 85,3 % der Testtage haben mindestens einen anderen Testtag oder Zwischenfalltag in ±7 Tagen, 31,0 % bei ±1

(Abb. 4). Solche Tage zählen nicht als Vergleichstage. Die Testtage sind dadurch nicht unabhängig voneinander (Abschnitt 4.6).

- **Kriegseinsätze.** Zwei der 1.762 Testtage (05. und 09.08.1945) sind die Einsätze auf Hiroshima und Nagasaki. Sie sind in den Daten als Atomtest geführt und wurden nicht herausgenommen.
- **Atomtest-Liste.** Korrigiert wurden unter anderem die Vorzeichen der Breitengrade der drei Argus-Tests im Südatlantik, eine falsche Längenangabe (N_80055), ein Tippfehler in der Bauart (SHAFT) und zerstörte Testnamen. Eine Fallnummer kam doppelt vor (N_74044, USA und Frankreich) und wurde eindeutig gemacht. Zwei Tests ohne Koordinaten erscheinen nicht auf der Karte. Die Zahl der Testtage (1.762) und die Hauptergebnisse blieben unverändert.

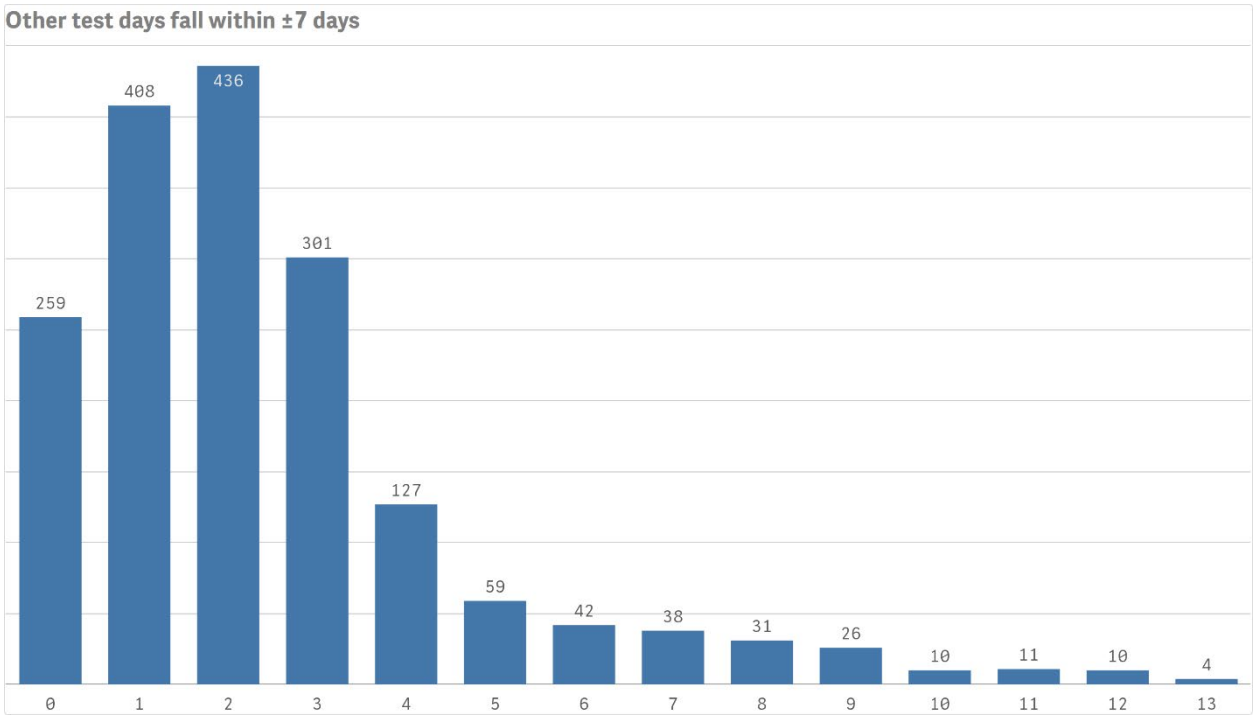


Abb. 4 *Verteilung der Testtage nach Zahl der anderen Testtage in ±7 Tagen.*

6 Ergebnisse

6.1 Sichtungen pro Tag um den Testtag

Der Durchschnitt der Vergleichstage beträgt 10,9 Sichtungen pro Tag, am Testtag sind es 11,4 (+4,8 % im reinen Profil, +4,4 % nach der Differenzrechnung). Der Unterschied liegt in der Größenordnung der Schwankung zwischen den Vergleichstagen selbst (etwa ±0,3; Abb. 2).

6.2 Hauptstatistik (Atomtests, alle Jahre)

Kennzahl	Wert
Testtage / Testjahre	1.762 / 55
Mittel D (Überschuss am Testtag)	+0,48 Sichtungen pro Tag
Referenz (normaler Tag)	10,91
Relativer Effekt	+4,4 %
t je Testtag	1,74
t je Testjahr	1,28
Placebo-Test: z	1,70
Streuung der Placebo-D	0,28

Tab. 2 *Hauptstatistik (Atomtests, Typ N). Auswahl: Typ = N.*

Im Klartext: Am Testtag gab es im Mittel knapp eine halbe Meldung mehr als an Vergleichstagen. Ein z-Wert von 1,70 heißt: Der Überschuss ist 1,7-mal so groß wie die übliche Schwankung. Unter 2 gilt nach der vorab festgelegten Regel als kein Effekt. Das t je Testjahr (1,28) ist die vorsichtigere Variante und liegt ebenfalls klar unter 2 (Tab. 2).

Wie groß könnte ein echter Effekt sein? Der mögliche Überschuss liegt grob zwischen −2,4 % und +11,3 % der normalen Tagesmenge (je Testjahr) beziehungsweise zwischen −0,6 % und +9,5 % (je Testtag). Beide Bereiche schließen null ein (Herleitung in Abschnitt 4.7). Größenordnungen wie +45 % oder +68 % sind in diesen Daten nicht zu sehen. Das ist kein Widerspruch zu den Fotoplatten, weil beide Arbeiten nicht dasselbe messen (Abschnitt 7).

6.3 Der Zeitraum 1961 bis 1968

Aufgeschlüsselt nach Zeitblöcken zeigt der Block 1961 bis 1968 den größten Ausschlag (Abb. 5): 524 Testtage, Mittel D +1,13, z = 2,02. Ohne diesen Block bleiben 1.238 Testtage mit D = +0,21 und z = 0,79.

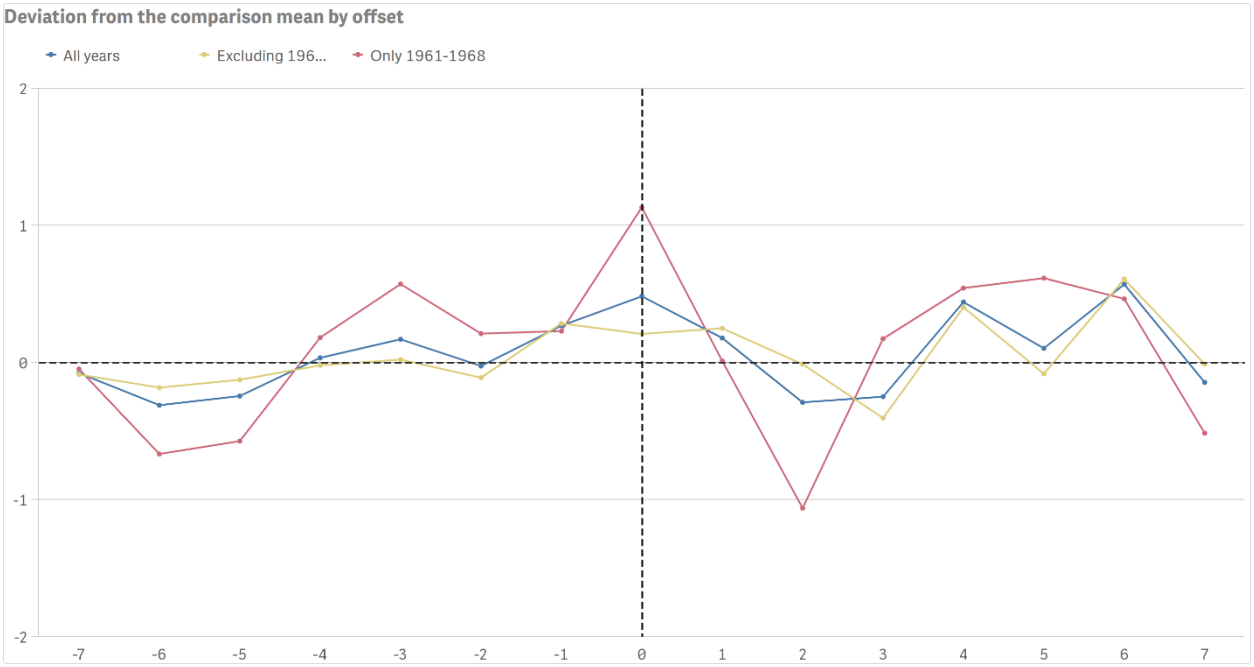


Abb. 5 *Differenz nach Versatz für alle Jahre, nur 1961 bis 1968 und ohne diesen Block.*

Im Klartext: Dieser Zeitraum umfasst nur acht Jahre, enthält besonders dicht aufeinanderfolgende Testserien und die größten Meldewellen (1966/67). Unter den 13 verglichenen Positionen (Testtag und zwölf Scheintermine) liegt der echte Testtag hier auf Rang 1. Das käme bei reinem Zufall in etwa jedem zwölften bis dreizehnten Fall vor (rund 8 %, Abschnitt 4.8). Hinzu kommt: Dieser Zeitraum wurde erst nach der Hauptauswertung genauer angesehen. Er ist deshalb ein Hinweis, der sich in einer neuen, vorab festgelegten Prüfung bewähren müsste, aber kein Befund.

6.4 Teilauswertung: Sichtungen aus den USA

Weil die Meldungen in den Fenstern zu etwa 97 % aus Katalogen US-amerikanischer Herkunft stammen (Tab. 1), wurde der Placebo-Test zusätzlich nur mit Sichtungen aus den USA gerechnet, und zwar für die 849 Testtage, an denen ausschließlich die USA testeten. Die Referenz beträgt hier 5,84 Sichtungen pro Vergleichstag. Am Testtag liegt der Überschuss bei +0,90 Sichtungen pro Tag (+15,3 %). Die zwölf Placebo-D liegen im Mittel bei +0,00 mit einer Streuung von 0,41, der z-Wert beträgt 2,20 (Tab. 3; p ≈ 0,05). Mit dem Korrekturfaktor aus Abschnitt 4.4 ergibt sich etwa 2,12 (p ≈ 0,06, Tab. 3). Als Näherung (Überschuss ± 2,2 Streuungen) reicht der Überschuss von etwa 0 % bis +31 % der US-Referenz.
Die Variante mit allen Tagen, an denen die USA beteiligt waren (904 Testtage), wurde nicht berechnet (Abschnitt 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. Links: alle Sichtungen, 1.762 Testtage. Rechts: nur Sichtungen aus den USA, 849 Testtage (Auswahl: Typ = N und Testländer = USA).*

Im Klartext: In den US-Sichtungen liegt der Testtag deutlicher über seinem Umfeld als in der weltweiten Zählung. Der Wert (z = 2,20) überschreitet die vorab festgelegte Schwelle von 2 knapp. Allerdings wurden mehrere Auswertungen angesehen (Abschnitt 8, Punkt 5), und wer viele Ausschnitte prüft, findet leichter einen Zufallstreffer. Deshalb braucht es einen strengeren Maßstab: Bei zwei angesehenen Auswertungen läge die Schwelle bei etwa 2,6, bei fünf bei etwa 3,1. Dieser Maßstab wurde erst nachträglich angelegt, er gehört nicht zu den vorab festgelegten Regeln. Der Wert bleibt deshalb ein Hinweis, der eine Wiederholung mit unabhängigen Daten braucht, kein Beleg.

6.5 Nukleare Zwischenfälle

Die 24 nuklearen Zwischenfälle sind für eine statistische Aussage zu wenige und wurden mit dem hier beschriebenen Verfahren nicht ausgewertet. Eine frühere, einfachere Rechnung deutete darauf hin, dass wenige Einzeltage das Ergebnis tragen. Belastbar ist das nicht.

6.6 Sensitivität: nur oberirdische Tests

Die Fotoplatten-Studie betrachtet oberirdische Tests. Deshalb wurde die Rechnung zusätzlich nur mit Testtagen gerechnet, an denen ein oberirdischer Test stattfand. Variante 1 zählt nur sicher oberirdisch eingestufte Tage, Variante 2 zusätzlich Tage mit unklarer Einstufung (Anhang F). Einträge, die als Nicht-Explosion, SAM-Test oder Kriegseinsatz markiert sind, sind in beiden Varianten ausgeschlossen.

Kennzahl	Alle Tests	Variante 1	Variante 2
Testtage	1.762	417	464
Referenz	10,91	12,91	12,43
Mittel D	+0,48	+0,78	+0,87
Relativer Effekt	+4,4 %	+6,0 %	+7,0 %
t je Testtag	1,74	1,01	1,23
t je Testjahr	1,28	1,14	1,92
Streuung der Placebo-D	0,28	0,94	0,84
Placebo-Test: z	1,70	0,83	1,03

Tab. 4 *Hauptstatistik und Placebo-Test für alle Tests und für Tage mit oberirdischen Tests (Auswahl: Typ = N und jeweils ein Wert in "Auswertung").*

Im Klartext: Beide Varianten zeigen dieselbe Richtung wie die Hauptauswertung: Am Testtag liegen die Meldungen etwas über dem Umfeld. Der z-Wert ist aber niedriger, weil nur noch 417 bis 464 statt 1.762 Tage eingehen und die Scheintermine dadurch stärker streuen. Die beiden Varianten liegen nah beieinander, das Ergebnis hängt also nicht an den 59 unklar eingestuften Tests. Der mögliche Überschuss liegt grob zwischen −6 % und +18 % der normalen Tagesmenge (Mittel D ± 1,96 Standardfehler je Testtag; Variante 1: −5,7 % bis +17,8 %, Variante 2: −4,2 % bis +18,2 %). Ein Effekt von +45 % wäre auch hier aufgefallen. Diese Auswertung ist eine zusätzliche Prüfung und kein zweiter Hauptbefund.

7 Einordnung zur Fotoplatten-Studie

Diese Arbeit will die Studie von Bruehl und Villarroel weder bestätigen noch widerlegen. Sie versucht lediglich, die gleiche Grundfrage mit erdgebundenen Daten zu stellen. Die Unterschiede im Überblick:

- **Andere Messgröße.** Die Platten messen Lichtpunkte mit einem Instrument an einem Ort. Hier werden Meldungen von Menschen weltweit gezählt. Meldehäufigkeit hängt von Aufmerksamkeit, Presse und Sammelpraxis ab.
- **Anderer Zeitraum.** Die Plattenstudie deckt 1949 bis 1957 ab. Hier ist der auffälligste Zeitraum 1961 bis 1968.
- **Gleicher Katalog.** Beide Arbeiten stützen sich auf UFOCAT, hier in der Fassung der GEP-Datenbank. Ob beide Fassungen im Datenstand übereinstimmen, wurde nicht geprüft. Diese Arbeit ist deshalb keine unabhängige Bestätigung.
- **Methodische Parallele.** Die Kritik an der Plattenstudie betrifft vor allem die Definition und Validierung der verwendeten Datensätze.[3] Auch in dieser Arbeit bestand ein großer Teil des Aufwands in der Datenprüfung (Abschnitt 5).
- **Oberirdische Tests.** Die Plattenstudie betrachtet oberirdische Tests, die Hauptanalyse alle Tests. Die Auswertung in Abschnitt 6.6 liegt der Auswahl der Plattenstudie näher, bleibt aber ein anderes Maß (Meldungen statt Lichtpunkte, weltweit statt am Ort).

Die Fotoplatten-Arbeiten wurden anhand der Veröffentlichungen und zugänglicher Berichte eingeordnet. Eine vollständige Nachprüfung der Originaldaten fand nicht statt.

8 Grenzen der Aussagekraft

1. **Nur Tageszählungen.** Regionale Effekte (zum Beispiel Sichtungen nahe dem Testgelände) wurden nicht geprüft.
2. **Fehlende Länder.** An 38,1 % der Testtage war mindestens eines der Länder Russland (Sowjetunion), China, Indien, Pakistan oder Nordkorea beteiligt. Aus diesen Ländern enthalten die vorliegenden Kataloge kaum Sichtungen, was für westlich zugängliche Sichtungskataloge insgesamt gilt (Abschnitt 2). Für diese Tests lässt sich ein Effekt nicht valide prüfen. Das Ergebnis gilt daher vor allem für Tests, an denen die USA, Großbritannien oder Frankreich beteiligt waren.
3. **Katalogqualität.** Meldedatum, Doppelmeldungen und gerundete Datumsangaben (zum Beispiel pauschal der 1. oder 15. eines Monats) sind nicht vollständig bereinigt. Die Datumsangaben der Tests sind vermutlich in Weltzeit angegeben (der 5.8.1945 für Hiroshima spricht dafür), die der Sichtungen lokal, was einzelne Tage um ±1 verschieben kann. Hinzu kommt, dass Sichtungskataloge und Untersuchungsmethoden auf internationaler und auch auf nationaler Ebene nicht vereinheitlicht sind.
4. **Nur zwölf Placebo-Positionen.** Die Streuung des Zufalls wurde aus zwölf Vergleichswerten geschätzt, die alle aus der unmittelbaren Umgebung der Tests stammen. Das macht den z-Wert etwas ungenau. Mit nur zwölf Positionen ist auch ein Rangvergleich grob: Der kleinste mögliche p-Wert liegt bei 1/13 = 7,7 % (einseitig).
5. **Mehrfachprüfungen.** Neben der vorab festgelegten Hauptauswertung wurden mindestens sechs Teilauswertungen angesehen, darunter Zeitblöcke, Sichtungen aus den USA und zwei Varianten nur mit oberirdischen Tests (bei Variante 2 liegt t je Testjahr mit 1,92 knapp unter 2). Bei jeder zusätzlichen

Auswahl steigt die Chance, dass allein der Zufall einen auffälligen Wert liefert (wie beim Würfeln: wer oft genug würfelt, bekommt irgendwann eine Sechs). Auffälligkeiten in Teilauswahlen werden deshalb nur als Hinweise gewertet, nicht als Befund (Abschnitt 4.8).

6. **Einstufung oberirdisch/unterirdisch.** Diese Einstufung ist in den Originaldaten nicht vorhanden und beruht daher teils auf Ableitungen aus Bauart, Tiefe, Gelände und Datum. 59 von 2.057 Tests sind unklar (Variante 2 rechnet sie mit), zwei Einträge wurden von Hand zugeordnet (N_57018, N_58022). Ein Abgleich mit der Originalquelle steht aus. "Oberirdisch" ist weit gefasst (einschließlich Wasser, Ballon, Rakete und Weltraum).

9 Fazit

9.1 Was die Arbeit zeigt

Kein belegter Zusammenhang. An Atomtest-Tagen werden nicht mehr UAP-Sichtungen gemeldet als an den Tagen davor und danach: Der Unterschied von knapp einer halben Meldung pro Tag liegt innerhalb der normalen Schwankung. Auch an gewöhnlichen Tagen ohne Test gibt es Ausschläge dieser Größe, der Testtag fällt aus dieser Reihe nicht heraus. Die vorsichtigere Rechnung je Testjahr bestätigt das.

In der weltweiten Zählung sind große Effekte unwahrscheinlich, kleine nicht ausgeschlossen. Effekte, die deutlich über 10 % lägen, hätten sich vermutlich gezeigt. In der Teilauswertung für US-Tests und US-Sichtungen liegt der Schätzwert bei etwa +15 %, ein Effekt dieser Größe lässt sich dort nicht ausschließen. Ein Effekt in nur einer Region würde in einer weltweiten Zählung untergehen.

Auch bei Beschränkung auf oberirdische Tests bleibt der Unterschied innerhalb der normalen Schwankung (+6 % bis +7 %, z = 0,83 bis 1,03).

9.2 Was die Arbeit nicht zeigt

- Die Teilauswertung für US-Tests und US-Sichtungen (z = 2,20) ist ein Hinweis, kein Beleg (Abschnitt 6.4). Das Gleiche gilt für den Zeitraum 1961 bis 1968 (Abschnitt 6.3). Beide bräuchten eine Wiederholung mit unabhängigen Daten.
- Die Fotoplatten-Befunde werden weder widerlegt noch bestätigt (Abschnitt 7).
- Die Reichweite ist begrenzt: An etwa 38 % der Testtage war Russland (Sowjetunion), China, Indien, Pakistan oder Nordkorea beteiligt. Das Ergebnis gilt daher vor allem für Tests, an denen die USA, Großbritannien oder Frankreich beteiligt waren.
- Die Einstufung der Tests als oberirdisch oder unterirdisch ist teilweise abgeleitet (Anhang F), 59 Tests blieben unklar.

9.3 Was die Arbeit trotzdem zeigt

Ein Nullergebnis ist ein Ergebnis. Die Arbeit zeigt, dass sich mit einer sauber gerechneten Tageszählung kein auffälliger Zusammenhang ergibt, und sie zeigt, wie leicht sich Scheineffekte einschleichen. Das ist für alle, die ähnliche Fragen mit ähnlichen Katalogen untersuchen, eine nützliche Warnung. Dazu kommt das offengelegte Vorgehen mit vorab festgelegten Regeln, das Dritten ermöglicht, die Rechnung nachzuvollziehen.

9.4 Offene Fragen

Nicht untersucht wurden regionale Effekte, die Variante mit allen Tagen unter US-Beteiligung, der Zeitraum 1949 bis 1957 sowie eine Wiederholung der US-Teilauswertung mit unabhängigen Daten. Ob die Einstufung der Tests als oberirdisch oder unterirdisch einem Abgleich mit der Originalquelle standhält, ist offen.

Anhang

A Glossar

Begriff	Bedeutung
Testtag	Datum, an dem mindestens ein Atomtest stattfand
Versatz	Abstand eines Tages zum Testtag in Tagen
Vergleichstage	Tage mit Versatz ±2 bis ±7, die selbst keine Testtage sind
Placebo-Test	Dieselbe Rechnung wie am Testtag, angewendet auf Tage ohne Testbezug. Zeigt, wie stark die Ergebnisse allein durch Zufall schwanken (auch: Scheintermin-Test)
Placebo-Position	Eine der zwölf Positionen (Versatz ±2 bis ±7), die wie ein Testtag gerechnet werden (auch: Scheintermin)
Placebo-D	Die Differenz D, berechnet an einer Placebo-Position
D (Differenz)	Sichtungen am Tag minus Durchschnitt der Vergleichstage
z-Wert / t-Wert	Abstand eines Ergebnisses zur zufälligen Schwankung, in Schwankungseinheiten. Werte unter 2 gelten als nicht gesichert
Standardfehler	Wie stark ein Mittelwert bei zufälliger Wiederholung schwanken würde
p-Wert	Gäbe es keinen Effekt: Wie oft käme ein mindestens so großer Ausschlag durch Zufall vor
Überdispersion	Die Tageswerte schwanken stärker, als einfache Zufallsmodelle erwarten
t je Testjahr	Vorsichtigere Variante des t-Werts, weil Tests im selben Jahr zusammenhängen
Referenz	Mittlere Sichtungen je Vergleichstag, das “normale” Niveau
Median	Der mittlere Wert einer sortierten Reihe, unbeeinflusst von Ausreißern
Oberirdisch	Test mit Bauart oder Tiefenangabe über der Erde, im oder auf dem Wasser, am Ballon, per Rakete oder im Weltraum (weit gefasst)
Unbekannt	Test, der sich weder nach Bauart noch nach Tiefe oder Gelände und Datum einstufen ließ (59)
Sensitivitätsanalyse	Dieselbe Rechnung mit geänderter Auswahl oder Annahme (hier: nur oberirdische Tests), um zu prüfen, ob das Ergebnis davon abhängt. Sie gilt als zusätzliche Prüfung, nicht als zweites Hauptergebnis.

B Kennzahlen der Hauptstatistik

Feld	Bedeutung
Testtage	Anzahl der Atomtest-Tage in der Auswahl
Testtage in D	Kontrollzahl: muss gleich "Testtage" sein, sonst sind Tage unbemerkt aus der Rechnung gefallen
Testjahre	Zahl der Jahre mit Tests (Tests im selben Jahr zählen weniger als unabhängige Beobachtungen)
Mittel D	Durchschnittlicher Überschuss am Testtag in Meldungen pro Tag
Streuung D	Wie stark die Einzelwerte schwanken
Relativer Effekt	Mittel D geteilt durch Referenz
Anteil über Kontrollmittel	Anteil der Testtage über ihrem Umfeld (Erwartung ohne Effekt nicht genau 0,5, weil wenige Spitzentage die Verteilung verzerren)
Median D	Typischer Testtag. Der Vergleich zum Mittel zeigt, ob wenige Spitzentage den Mittelwert treiben (hier −0,92 gegenüber +0,48)

C Reproduzierbarkeit und Datenverfügbarkeit

Die Auswertung lief in Qlik Cloud mit einem Ladeskript (Kalender, Ereignistabelle, Brückentabelle, Placebo-Tabelle). Das Ladeskript ist vorhanden und kann beim Autor angefordert werden. Die Zahlen wurden anhand der Exporte unabhängig nachgerechnet. Die Daten zu Atomtests (Johnston's Archive) und der Katalog UFOCAT (CUFOS) sind öffentlich zugänglich. Die GEP-Datenbank und das Ladeskript sind nicht frei verfügbar und können beim Autor angefordert werden. Die Atomtest-Liste wird in einem eigenen Skript bereinigt und eingestuft, das beim Autor angefordert werden kann.

D Fußnoten

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, veröffentlicht am 20.10.2025. 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, nicht begutachtet; mehrere Fassungen). Der Schwerpunkt liegt auf der Validierung der Datensätze und dem sogenannten Erdschatten-Defizit.

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, eingereicht am 16.02.2026, zweite Fassung vom 30.04.2026). Ein Begutachtungsstatus war im arXiv-Eintrag nicht erkennbar (Stand: Oktober 2026). https://arxiv.org/abs/2602.15171

5. Im gesamten Katalog haben 348.638 Zeilen ein lesbares Datum, sie gehören zu rund 347.600 verschiedenen Fallnummern, weil einzelne Meldungen mehrere Zeilen haben. In den Fenstern um Atomtests liegen 121.197 Zeilen, das sind 120.829 verschiedene Meldungen. Die Spaltensumme in Tab. 1 beträgt 120.830, weil eine Fallnummer in zwei Jahrzehnten vorkommt. Gezählt werden in allen Auswertungen verschiedene Fallnummern je Tag (Abschnitt 5).

6. Johnston, W. R.: *Johnston's Archive, Nuclear Tests – Databases and Other Material.* https://johnstonsarchive.net/nuclear/tests/index.html (abgerufen am 28.10.2026). Die Liste wurde bereinigt (Abschnitt 5), die Fallnummern wurden eindeutig gemacht.

7. Die Differenz D schwankt mit einer Streuung von 11,72 etwa dreimal so stark wie nach einem einfachen Zufallsmodell (Poisson) zu erwarten wäre (etwa 3,5). Herleitung in Abschnitt 4.5.

8. Intervall grob: Mittel D ± 1,96 Standardfehler, geteilt durch die Referenz 10,91. Standardfehler je Testjahr etwa 0,38 (aus t_Jahr zurückgerechnet), je Testtag 0,28. Es ist eine Näherung und keine exakte Konfidenzrechnung.

E Quellenverzeichnis (alphabetisch):

- 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 (abgerufen am 28.10.2026)

- 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.

Anhang F Einstufung oberirdisch/unterirdisch

Reihenfolge der Einstufung: (1) Bauart, (2) Vorzeichen der Tiefenangabe, (3) Gelände und Datum (letzter sicher oberirdischer und erster sicher unterirdischer Test des Geländes, mit 60 Tagen Puffer), sonst unbekannt.

Einstufung	Tests
oberirdisch	525
unterirdisch	1.473
unbekannt	59
Summe	**2.057**

Herkunft der Einstufung	Tests
Bauart	1.710
Tiefe	45
Gelände und Datum	243
offen	59
Summe	**2.057**

Von Hand zugeordnet (aus Datum und Ertrag erschlossen, gegen die Originalquelle zu prüfen): N_57018 unterirdisch (passt zu einem Tunneltest 1957), N_58022 oberirdisch (passt zu einem Unterwassertest 1958). Kraterexplosionen (N_65016, N_62067) zählen als oberirdisch. N_58016 hat Koordinaten, die zu keinem Testgelände passen, und blieb unbekannt.

Autor: Ralf van Thuyl (Mitglied der GEP), Kontakt: homuell2@gmail.com

Dank an die Gesellschaft zur Erforschung des UFO-Phänomens (GEP). https://www.uap-research.org, https://www.ufo-forschung.de

Hinweis zu KI-Werkzeugen: Textentwurf und rechnerische Gegenprüfung der Exporte wurden mit Claude (Anthropic) unterstützt. Die Auswertung in Qlik und die inhaltliche Verantwortung liegen beim Autor.

© 2026 Ralf van Thuyl. Dieses Werk steht unter der Lizenz CC BY 4.0 (Namensnennung).
Kartenmaterial in Abb.1: © OpenStreetMap contributors/ esri, von der Lizenz ausgenommen.