Can an attribution assessment be made for Yellow Rain?

Systematic reanalysis in a chemical-and-biological-weapons use investigation

Rebecca Katz, Ph.D.*
Department of Health Policy
George Washington University
2021 K Street, NW, Suite 800
Washington, DC 20006
rlkatz@gwu.edu
* Corresponding author.

Burton Singer, Ph.D.
Office of Population Research
245 Wallace Hall
Princeton University
Princeton, NJ 08544
singer@opr.princeton.edu

Abstract. **In intelligence investigations, such as those into reports of chemical- or biological-weapons (CBW) use, evidence may be difficult to assemble and, once assembled, to weigh. We propose a methodology for such investigations and then apply it to a large body of recently declassified evidence to determine the extent to which an attribution can now be made in the Yellow Rain case. Our analysis strongly supports the hypothesis that CBW were used in Southeast Asia and Afghanistan in the late 1970s and early 1980s, although a definitive judgment cannot be made. The proposed methodology, while resource-intensive, allows evidence to be assembled and analyzed in a transparent manner so that assumptions and rationale for decisions can be challenged by external critics. We conclude with a discussion of future research directions, emphasizing the use of evolving information-extraction (IE) technologies, a sub-field of artificial intelligence (AI).**

There is a science to the collection of evidence in an intelligence investigation. Experienced analysts collect information, analyze relationships, draw tentative conclusions, test those conclusions against alternative explanations, and hopefully engage in critical review.[1] The evidence in intelligence investigations, however, often includes only partial information from a variety of sources with variable quality, and not all potentially relevant evidence is available. In many instances, analysts must make causal inferences from single, novel events. Standards of evidence and inference are ill-defined and strengths of conclusions hard to clarify and compare.

Historically, investigations into allegations of chemical- or biological-weapons attack have suffered from an inability to discern essential information.[2, 3] Did an attack happen, and if so, what was the agent employed, and who was the responsible party? Answering these questions has been hindered in part by the innate evidentiary challenges posed by such investigations (*e.g.*,

Disclaimer. R.K. is a consultant to the Office of Biological Weapons Affairs, United States Department of State. The views expressed in this paper are solely those of the authors and do not represent the official views of the Department of State or the United States Government.

distinguishing epidemiologically between naturally occurring and intentionally caused disease, acquiring biomedical and environmental samples, accessing appropriate controls, tracing recovered agents to specific parts of the world and then to individuals and groups plausibly able to obtain them), but also because of the lack of a systematic approach to attribution assessment — particularly the blending of quantitative and qualitative evidence.

Our purpose is to clarify in a systematic way how to assemble and weight evidence in intelligence investigations, specifically investigations into allegations of chemical or biological weapons (CBW) use. We illustrate this methodology using the Yellow Rain case, an historical investigation into CBW allegations in Southeast Asia and Afghanistan, for which a substantial amount of new evidence allows us to reevaluate previous claims. To enable critical review of our conclusions, we offer through Politics and the Life Sciences electronic access to 8,529 pages of recently declassified documents pertaining to the investigation of Yellow Rain.[4]

Background

The late 1970s was a tumultuous time for Southeast Asia and Afghanistan. In Laos, the United States had engaged the Hmong, an ethnic minority, to create a resistance army in the fight against communist Vietnamese and Pathet Lao forces. In 1975, after many years of war, the Pathet Lao took power in the country, the United States pulled out, and the majority of Hmong were left behind, although given their active role fighting the ruling body, many began to flee across the Mekong River into Thailand. In Cambodia, the Khmer Rouge had come to power in 1975; at the end of 1978, Vietnamese forces invaded and ousted the Khmer Rouge. With a Vietnamese-backed government in power, Khmer Rouge forces joined with other Cambodian parties opposed to the government to form a coalition of resistance fighters hidden primarily along the Thai border. Several thousand miles away, Afghanistan was experiencing regime challenges, and, in December 1979, the Soviet Union invaded Afghanistan, beginning what would be a ten-year war of resistance against the Soviets waged by Islamic Mujuhadin.

Starting in the late 1970s, there were reports of chemical- or toxin-weapons use against three peoples — the Hmong in Laos, the Khmer in Cambodia, and the Mujuhadin in Afghanistan. Accounts often described events in which a helicopter or airplane had flown over a village and released a colored gas that would fall in a manner that looked, felt, and sounded like rain. Many colors of gas were reported, but the color most commonly reported was yellow, whence the name "Yellow Rain."[5] If true, these events would have been in direct violation of the Geneva Protocols and, if the agent employed was a toxin, the Biological and Toxin Weapons Convention.[6, 7] Additionally, any intentional use of chemical or toxin weapons against civilians would have been considered a human-rights violation and, in the context of conflict, a war crime.[8, 9]

Hmong in Laos, Khmer Rouge resistance fighters in Cambodia, and Mujuhadin resistance fighters in Afghanistan described similar types of attacks and subsequent symptoms, raising suspicions that the same agent and attack mechanism were being used in all three sites. Common symptoms included nausea, vomiting, diarrhea, dizziness, difficulty breathing, eye irritation, and blistering or other skin rash. In the most severe cases, victims were said to have had bloody vomitus and bloody diarrhea, as well as subconjunctival ("under the lining of the eye") and subungual ("under the nail") bleeding.[10, 11]

Over the course of several years, multiple countries and the United Nations conducted investigations into these allegations of chemical and toxin weapons use.[12, 13, 14, 15] The United States Government announced in 1981 and officially reported in 1982 that the attacks in Southeast Asia and Afghanistan involved trichothecene mycotoxins — specifically T2 — and accused the Soviet Union of sponsoring their use.[16]

Not everyone concurred with the US government reports singling out trichothecene mycotoxin as the Yellow Rain culprit. The primary competing theory came from members of the academic community, led by Dr. Matthew Meselson of Harvard University. Meselson and his team, suspicious of the government's findings and not fully satisfied by the scientific rigor of its published analysis, hypothesized that the events reported by the Hmong might have been due to the cleansing flights of Asian honey bees. These bees periodically defecate *en masse*, creating a shower of pollen appearing as a yellowish brown rain. Charles Darwin was the first scientist to write about this event[17]; a modern account was published in a Chinese journal in 1977.[18] Meselson and his team reviewed laboratory analyses of environmental samples for the presence of pollen and conducted their own field investigation designed to search for evidence of bee cleansing flights in Thailand. They concluded that the evidence examined did not support a claim of chemical or toxin weapons attack, and they determined that Yellow Rain was a natural occurrence attributable to bees.[19, 20] The bee theory applied only to the reports of Yellow Rain in Southeast Asia; it did not address CBW claims in Afghanistan, where bee cleansing flights are not known to occur.

More than twenty-five years after the fact, a vigorous debate persists over whether chemical and toxin weapons were used as charged in any of the three sites described. Limiting this debate have been multiple government decisions to withhold much of the collected evidence. Indeed, the debate around and about the evidence has never been free-ranging, its focus repeatedly being redirected to vulnerabilities in counterfactual arguments. Very recently, the United States government declassified over eight thousand pages of documents pertaining to the evidence collected at the time of the initial Yellow Rain investigation, providing an impetus, along with the current political importance of determining when weapons of mass destruction have been used, to re-evaluate the Yellow Rain investigation and findings in an open, transparent and meaningful manner.

Methods

Three main bodies of evidence were reviewed for this project: 8,529 pages of United States government documents, declassified by the Defense Intelligence Agency and released through a Freedom of Information Act request, including medical records, laboratory reports, diplomatic communications, internal memos, and protocols originating primarily from the Armed Forces Medical Intelligence Center; over 800 documents of previously published material on Yellow Rain, mycotoxins, and chemical weapons; and interviews with 48 individuals with expert knowledge related to Yellow Rain, including 20 who were directly involved in investigating allegations for either the United States, an NGO, or another country.

Information from these sources was combined to create a database of epidemiologic and clinical findings, intelligence information, investigative protocols, and toxicological sampling and analysis. Additionally, we georeferenced all alleged attack sites and created a separate database for results of toxicological analysis on over 1,600 samples. These data sets were used in the overall evaluation of Yellow Rain evidence according to the methodology described herein.

We devised a seven-step strategy for integrating the complex mixture of qualitative and quantitative data and for then establishing in a transparent fashion that one among a range of plausible hypotheses was best supported by available evidence.

The first step was to divide the evidence into blocks or types of information.

The second step was to assign to each evidence block a *veritas* ranking based on a combination of what we refer to as degree of dubiousness and degree of fallacy. The distinction between these notions is that determining degree of dubiousness requires an appraisal of intrinsic ambiguity or likelihood, whereas determining degree of fallacy requires an appraisal of deception — meaning here the purposeful introduction of falsity. These measures draw upon methods previously used to evaluate probability, validity in arguments, and measurement error in a variety of contexts, including the determination of the velocity of light.[21, 22, 23] Because of the difficulty in making fine distinctions in degree of belief when determination of error and level of ambiguity is so fluid, this methodology does not attempt to assign precise weights or probabilities to the blocks of evidence and instead relies upon a rank-order system.[24, 25] Thus, we assigned each block of evidence a "degree of dubiousness" score, to which we added an assigned fallacy score to create an overall *veritas* rank, which we coded as either low, medium, or high. Our objective was to be sufficiently explicit to allow others assessing this same information to identify the rationale behind our ranking and then, as they would deem appropriate, introduce alternative appraisals.

The third step was to develop groups of hypotheses, meaning that multiple plausible possibilities were formally considered and counterfactual explanations explored, so as to build into our method a forced reduction in investigator bias.[24, 26]

The fourth step was to assess each evidence block for the strength of association to each hypothesis, assigning a ranking of strong, medium, or weak.

The fifth step was to organize the evidence blocks by hypothesis into a matrix based on strength of

Table 1. Scoring for strength of evidence block's association with a hypothesis and *veritas* rank.

An evidence block's association with a hypothesis	Veritas *rank*		
	High	*Medium*	*Low*
Strong	10	8	4
Medium	8	6	2
Weak	4	2	0

Darkly shaded cells show relatively strong support.
Lightly shaded cells show minimally strong support.

association and *veritas* rank. Strength of evidence was reexamined by grouping blocks, where appropriate.

The sixth step was to choose the strongest hypothesis based on quality of evidence, quantity of evidence, and strength of explanation based on evidence. While it was often possible to determine the strongest hypothesis visually, comparing competing hypotheses numerically was helpful. To accomplish this comparison, each block was assigned a numerical score in accordance with a coding scheme attached to the strength of association and *veritas* rank for each hypothesis (Table 1).

We then employ six summary statistics: maximum score over all evidence blocks; minimum score over all evidence blocks; average score; average score over evidence blocks in "minimally-strong-support" cells; average score over evidence blocks in "relatively-strong-support" cells; and percent of evidence blocks in "relatively-strong-support" cells. These statistics were applied to the set of evidence blocks relevant to each individual hypothesis, to pairs of hypotheses, and to all hypotheses simultaneously, to produce a numerical and visual representation of hypotheses by strength of support.

Finally, the strongest hypothesis was checked to ensure that it agreed with the current overall state of historical and scientific knowledge, that it satisfied guidelines for causation, and that it was consistent with any definitive proof or admission, not only answering the criteria "beyond a reasonable doubt," but also allowing for conclusions with absolute certainty.

Before utilizing this methodology to evaluate the Yellow Rain evidence, we tested it using cases with known evidence bases and definitive results. The Sverdlovsk anthrax outbreak of 1979 presented a fitting case study to test the described methodology for weighing and interpreting the diverse sets of evidence associated with the investigation of a chemical- or biological-weapons attack. We utilized evidence from an investigation led by Matthew Meselson as assembled in a book by Jeanne Guillemin, *Anthrax: The Investigation of a Deadly Outbreak*; a collection of declassified documents pertaining to U.S. intelligence reports on the Sverdlovsk outbreak; newspaper articles, scientific papers, and Soviet press releases; and the personal notes of individuals involved.[27, 28] We were not expecting to overturn current understanding of this incident, but we were looking for ways our methodology might allow us erroneously to do so. We found no such way and, as anticipated, found the evidence to support firmly the hypothesis of an accident at a biological-weapons facility in Sverslovsk being responsible for the human cases of anthrax in the Spring of 1979.

We also tested our ability to evaluate evidence and successfully choose correct explanations using a series of cases analyzed fictionally by Sherlock Holmes, as described in the short stories of Sir Arthur Conan Doyle; we repeated this exercise using evidence and story lines in multiple episodes of the American television drama series, *Law and Order.*[29] In each case, our methodology proved successful, even when evidence led the reader or viewer down a false inferential path prior to a twist in story line and a surprise ending.

These exercises were useful for testing the methodology, which is then transferable to more complicated investigations. We found, however, that in more complex investigations we could not always complete each step of the methodology. In addition, interpretation of evidence did not always result in selection of preferred hypotheses or the ability to verify demonstrably correct hypotheses. The number of steps that could be completed seemed dependent on the nature and complexity of an investigation and its evidence.

Additionally, we found that the selection of evidence blocks, the *veritas* ranking and the generation and selection of hypotheses were not devoid of researcher bias and that this bias could greatly affect the evaluation and interpretation of evidence. However, the sources of potential bias are transparent. Other analysts can introduce their own judgments to ascertain whether or not they are led to different conclusions, and why. That said, our evaluative framework readily accepted the diversity of evidence expected in a CBW investigation such as the Yellow Rain case, to which we now turn.

Results

We divided all available Yellow Rain information from the investigation conducted by the United States government into 12 blocks separated by type and source, representing a course-grained division of evidence (Table 2). Block 11 (Conduct of investigation) and Block 12 (Sampling methods) by themselves did not provide evidence to support a given hypothesis, but influenced the analysis of evidence when grouped with one of the other blocks.

The raw evidence by itself was not informative, and each evidence block required significant analysis. Applying advances in scientific knowledge that have accrued during the quarter century since the original Yellow Rain investigations, the advantage of hindsight, and a more complete picture of the mission, conduct, analysis, and interpretation of the original investigation, we were able to determine the following:

1. Samples, medical records, and testimony prior to 1983 were more reliable than those from 1983 on, when the investigation was compromised by refugees' knowledge of incentives to claim victim status and by searching for indicators of attacks, rather than coordinating intelligence data and refugee reports to locate attack sites.
2. Between 1979 and 1982, refugee reports of attacks were consistent with other intelligence data, including known battles and flight paths of aircraft, more than 60 percent of the time.
3. Clinical complaints and findings among self-described victims and detailed refugee accounts of attacks were sufficiently similar in Laos, Cambodia, and Afghanistan to suggest a key common factor, most plausibly a Soviet link, in influence and support of direct operational involvement.
4. Clinical complaints and findings of alleged victims as documented by photographs, medical records, autopsy results, and third-hand accounts are consistent with mass simultaneous poisoning and not with any known natural disease endemic to Laos, Cambodia, or Afghanistan or with the potential to affect multiple individuals simultaneously.[11] Signs and symptoms reported by alleged victims and eye witnesses, however, were consistent with trichothecene mycotoxin poisoning but also shared features of exposure to nerve gases, riot-control agents, phosgene, and arsines.

Table 2. Yellow-rain evidence blocks.

E1 Toxicological analysis	**E5** Intelligence reports	**E9** Hmong interviews
E2 Medical records	**E6** Soviet-link evidence	**E10** Investigator interviews
E3 Attack data	**E7** Coincidence analysis	**E11** Conduct of investigation
E4 Autopsy results	**E8** Open-source reports	**E12** Sampling methods

5. We captured detailed information on 766 separate alleged attacks in Laos, Cambodia, Afghanistan, and Thailand from 1975 through 1985 (Figure 1). Attacks were reported to occur in all months of the year, varying more by season in Laos and Cambodia than in Afghanistan (Figure 2). The locations of the reported attacks were consistent with the locations of Hmong, Khmer and Mujuhadin, including the few claims in Thailand, whose border area hosted refugees and guerilla groups. (Figure 3 and 4)
6. Approximately 75 percent of alleged attacks involved seeing or hearing a helicopter or airplane, followed by seeing or smelling a gas or powder fall to the ground. The remainder cited landmines, grenades, pipes, artillery, and contaminated food or water. The most common color reported in association with gas or powder was yellow (57 percent), but other colors were also described. Yellow accounted for almost 70 percent of reports from Laos, but only 48 percent from Cambodia, and 20 percent from Afghanistan.
7. Biomedical samples were collected from 170 alleged victims; samples from only 146 people were analyzed. These samples were of blood, urine, and tissue rendered from autopsies. Twenty-six of 146 people were positive for trichothecene mycotoxin; these 26 were from 11 sites in Laos and 5 in Cambodia. All control samples analyzed as part of the United States Government investigation were negative for trichothecene mycotoxin.
8. Samples were determined to be positive for

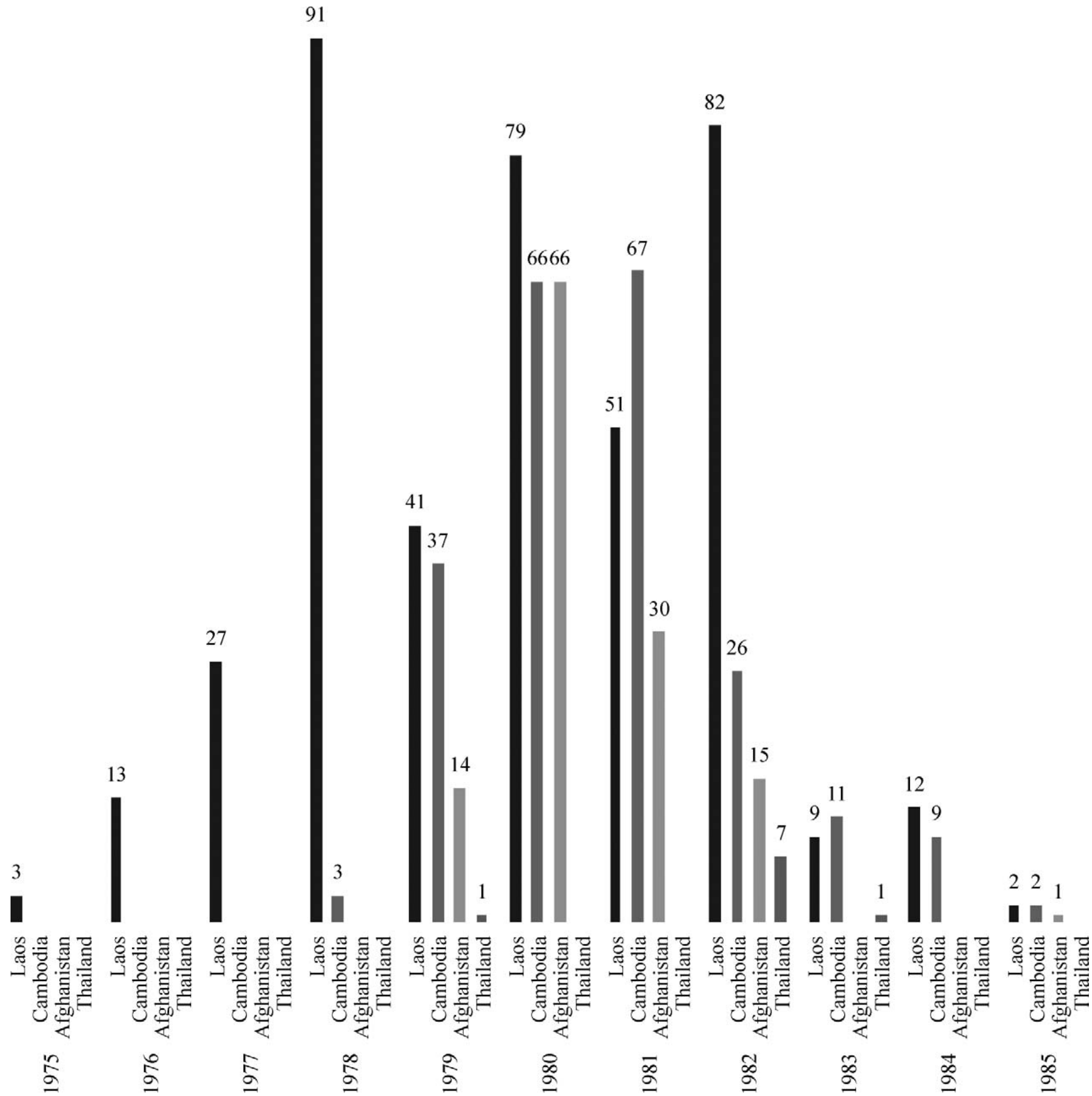

Figure 1. Reported Yellow Rain attacks, by year and location.

trichothecene mycotoxin if they met the following criteria established by the Armed Forces Medical Intelligence Center. Multiple specimens had to contain both T2, a highly toxic stable trichothecene mycotoxin, and HT2, a metabolite of T2.[30] Positive blood samples had to be confirmed by a positive urine sample or highly credible intelligence report of an attack. Data from certain laboratories could not be considered alone; their reports had to be confirmed by another laboratory on a blinded basis. Older samples were given less weight than newer samples, as older samples were at greater risk of degradation. Absent any known clinical complaints and findings, "victim" status, laboratory results notwithstanding, had to be reassessed.[31]

9. Problems associated with the sampling and handling

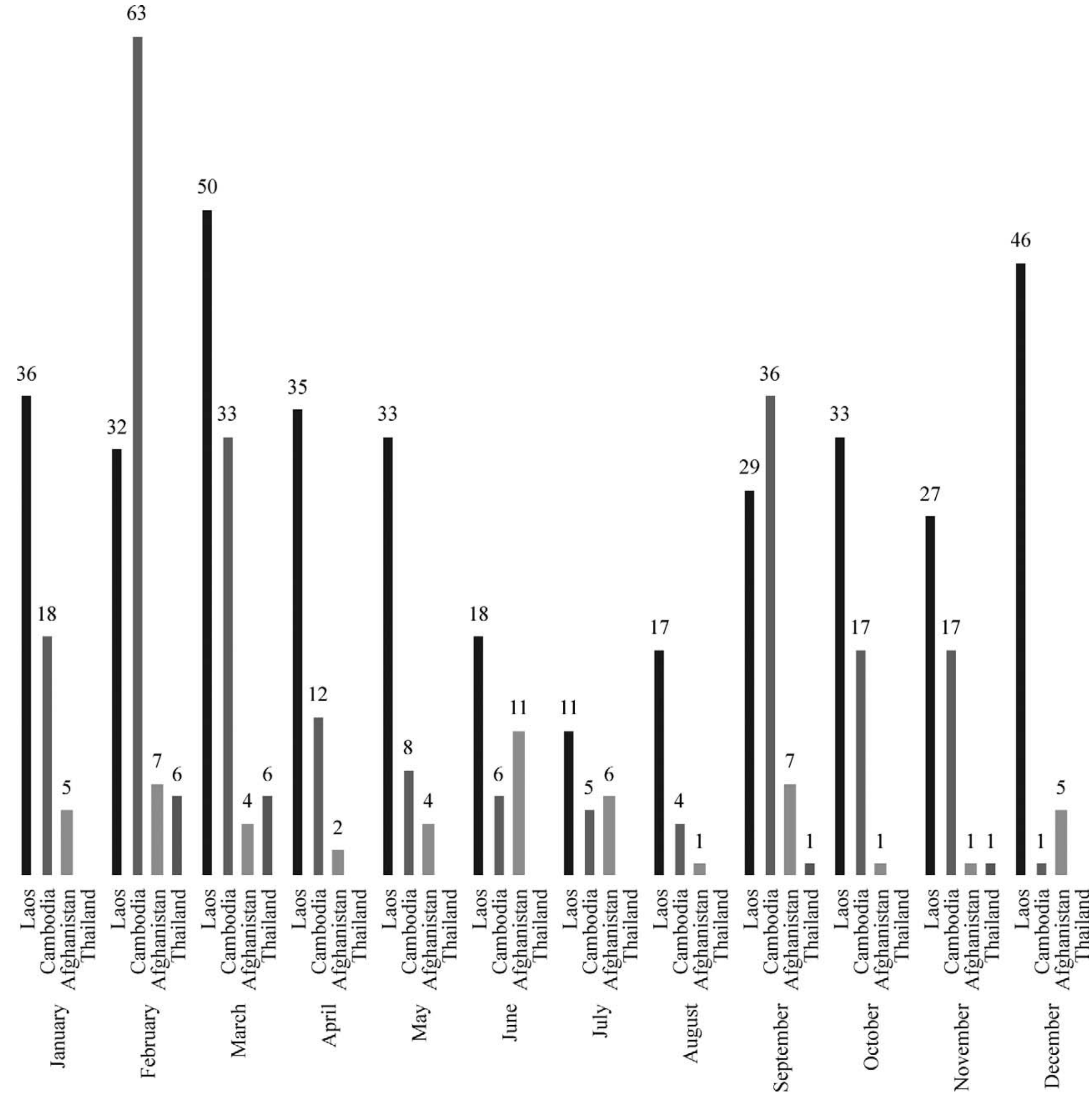

Figure 2. Reported Yellow Rain attacks, 1975–1985, by month and location.

of samples, lack of chain-of-custody documentation, paucity of human controls, difficulty in ascertaining appropriate environmental controls, and lack of baseline toxicological data on populations and the environment in Southeast Asia and Afghanistan combine to mean that we cannot determine with certainty that laboratory findings positive for trichothecene mycotoxin in both biomedical and environmental samples following Yellow Rain claims were consequences of intentional attack.

10. When we examined all 12 evidence blocks for consistency with the bee theory, we found that some, but not all, of the environmental samples contained pollen. Some of the environmental

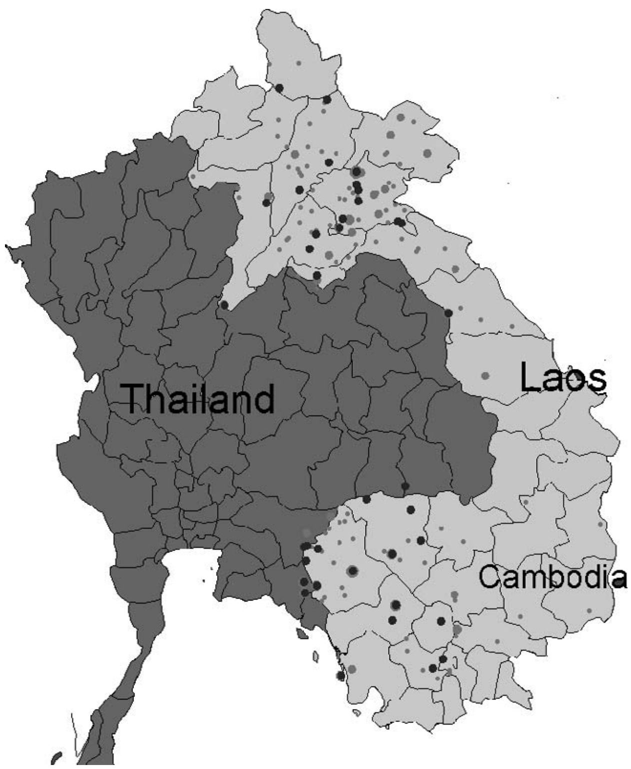

Figure 3. Reported Yellow Rain attacks, 1976–1984, in Laos, Cambodia, and Thailand. Gray dots are proportional in size to number of claims, ranging from 1 to a high of 25 (in a single year). Black dots show claim sites at which multiple sources concurred that an attack took place.

> samples that tested positive for T2 did not contain pollen. The bee theory[19] provides no explanation for the presence of T2 in biomedical samples from alleged victims. The bee theory does not address the 63 percent of reported air attacks that were *not* associated with the color yellow, nor the 25 percent of all reported attacks that did not involve an air assault. The bee theory does not account for the 69 attacks alleged to have occurred in Afghanistan, nor does it explain the morbidity and mortality experienced by the Hmong, Khmer, and Mujuhadin.

The next step in our methodology was to apply to each evidence block a *veritas* rank based on a combination of dubiousness and fallacy. Dubiousness relates to anything that might cause distortion, error, or divergence. Divergence may be due to systematic errors of measurement and sources of bias, *e.g.* selection bias or recall bias. We scored "degree of dubiousness" according to the probability of high (1), moderate (2), or minimal (3) distortion. "Degree of fallacy" referred to the extent to which a piece of evidence was deceptive, misleading, or the result of unreliable reasoning and was scored as follows: 1, if the event's probability was low and evidence for its occurrence doubtful; 2, if supporting information was accurate but event probability low; 3, if we accepted the evidence but doubted a piece of it; and 4, if we accepted all evidence as probably accurate. The overall *veritas* score was (degree of dubiousness) + (degree of fallacy). The *veritas* rank was called high if the *veritas* score was 6 or 7; medium if the score was 4 or 5; and low if the score was 2 or 3. Table 3 includes the dubiousness, fallacy, and overall *veritas* scores for each of the evidence blocks. The details of the evidence contained in each block, the analysis of that evidence, and the rationale for each score can be found in the author's (RK) dissertation.[11]

Once evidence blocks were assigned *veritas* rankings, hypotheses were developed to test explanations for Yellow Rain claims. We determined three rival hypotheses and numerous subsidiary hypotheses to be plausible candidates for consideration (Table 4).

If a chemical or toxin agent had been used intentionally (H1), then what was the composition of the agent (H1A), what were its users' intentions (H1B), and who might its users and their sponsors have been (H1C)? Testing the second hypothesis (H2), that Yellow Rain was a naturally occurring event, would necessitate determining which pieces of evidence in the Yellow Rain investigation seemed plausible and how those events might be explained through natural events. Visual accounts of Yellow Rain could be due to bee cleansing flights (H2A) as described by Meselson and colleagues.[19] Findings of trichothecene mycotoxins in the areas of interest could be due to natural levels of fungi in the region (H2B). A possible explanation evaluated was the elephant grass theory (H2C), as presented to the United Nations by the Soviet Union Academies of Science, asserting that American use of defoliants in Vietnam resulted in region-wide overgrowth of elephant grass. This increase in elephant grass supposedly led to increased amounts of *Fusarium* species and their products, including trichothecene

Figure 4. Reported Yellow Rain attacks, 1979–1984, in Afghanistan. Gray dots are proportional in size to number of claims. Black dots show claim sites at which multiple sources concurred that an attack took place.

mycotoxin.[32, 33] The final possibility evaluated was that morbidity and mortality had increased for reasons — war, malnutrition, infectious disease — unrelated to Yellow Rain (H2D).

The third hypothesis (H3), that events were fabricated, rejects refugee reports of Yellow Rain, as well as morbidity and mortality data and findings of mycotoxins in favor of two possible explanations. The first (H3A) is that the events were fabricated in confusion, ignorance, or mass hysteria. The second (H3B) is that the events were fabricated in order to gain political favor or asylum either by refugees themselves, regional political groups, or by the United States intelligence community.

Each evidence block was assigned a ranking of strong, medium, or weak that reflected the strength of association provided by that block for each hypothesis. In some instances, an evidence block had no association with a given hypothesis. Table 5 contains our rankings for strength of association for each of the Yellow Rain hypotheses, together with a summary of rationale.

With scores and rankings completed, we organized evidence blocks by hypothesis, and arranged them in matrices (Figures 5a, 5b, and 5c). We then re-evaluated the material by grouping evidence blocks where appropriate. For example, information on the conduct of the investigation and sampling methods dilute the strength of the toxicology findings, defensibility of conclusions from medical records, and quality of some of the attack data derived from interviews. Conversely, evidence provided by the attack data is supported by the medical records, coincidence analysis, intelligence reports and interviews.

The strongest hypothesis was chosen based on a visual examination of the matrices and also through scoring and combining the evidence. In particular, each evidence block was assigned a numerical score in accordance with the coding scheme shown in Figures

Table 3. *Veritas* score and *veritas* rank, by evidence block.

Evidence block	*Degree of dubiousness*		*Degree of fallacy*		*Veritas score*		*Veritas rank*
E1 Toxicological analysis	3	+	3	=	6	⇒	High
E2 Medical records	2	+	3	=	5	⇒	Medium
E3 Attack data	2	+	2 to 3	=	4 to 5	⇒	Medium
E4 Autopsy results	2	+	4	=	6	⇒	High
E5 Intelligence reports							
a. Human source	1	+	2	=	3	⇒	Low
b. Intercepts and imagery	3	+	3	=	6	⇒	High
E6 Soviet-link evidence	1 to 2	+	3	=	4 to 5	⇒	Medium
E7 Coincidence analysis	2	+	3	=	5	⇒	Medium
E8 Open-source reports	1	+	3	=	4	⇒	Medium
E9 Hmong interviews	1	+	3	=	4	⇒	Medium
E10 Investigator interviews	1	+	3	=	4	⇒	Medium
E11 Conduct of investigation	3	+	4	=	7	⇒	High
E12 Sampling methods	3	+	4	=	7	⇒	High

5a, 5b, and 5c. Cells in the upper left sections of these tables, where both *veritas* and association rankings are medium or higher, were interpreted to represent relatively strong support for a given hypothesis. Cells on the right-hand and bottom borders, where at least one of the support or *veritas* rankings was low, were interpreted as categories of minimal support, as described in Table 1. All values assigned to the upper-left section of the table were at or above the mid-point on the numerical scale, thereby reflecting "relatively strong" support.

Table 4. Yellow Rain hypotheses.

H1: Chemical or toxin agents used intentionally in Laos, Cambodia, and Afghanistan, 1975–1985
- H1A: Composition
  - Included trichothecene mycotoxin
  - Same composition in all countries
- H1B: Intent
  - Intent to kill and injure
  - Intent to frighten
  - Intent to target animals, crops, foliage
  - Intent to experiment
- H1C: Responsible party
  - Soviet Union
  - Other states

H2: "Yellow Rain" was a naturally occurring event.
- H2A: Visual accounts of Yellow Rain due to natural causes (bee theory)
- H2B: Toxicological findings of mycotoxins due to natural levels of *Fusarium* in region
- H2C: Elephant grass indirectly increased natural levels of trichothecene mycotoxins
- H2D: Morbidity and mortality due to causes unrelated to Yellow Rain

H3: Events fabricated
- H3A: Incorrect explanations invented in confusion, ignorance, or mass hysteria
- H3B: Events fabricated to gain asylum or political favor

Statistical summary of the evidence, based on the numerical scoring of entries in Figures 5a, 5b, and 5c for each of the hypotheses evaluated separately, for pairwise overlapping evidence blocks, and for evidence blocks common to all three hypotheses, were evaluated using six summary statistics (Table 6). The summary statistics consistently indicated that the strongest support was decidedly for H1, with less support for H2 and the least support for H3. Thus, on the basis of all evidence at hand, our conclusion is that lethal chemical or toxin compounds were used in Laos, Cambodia, and Afghanistan in violation of the international conventions operative during the 1970s and 1980s. We cannot, however, identify the specific agents used, the intent, or the root source or sources of the attacks.

The chosen hypothesis, that CBW attacks occurred in Southeast Asia and Afghanistan, was evaluated to ensure it met guidelines for causation, that it agreed with the state of knowledge, and that it was consistent with any definitive proof or admission that might be available. For a check against the state of knowledge, the evidence was evaluated to ensure it was consistent with what is known about intentional and accidental releases of CBW agents, as opposed to natural occurrence of disease. While no official rules exist for determining if disease in a population is due to intentional release of an agent or naturally occurring events, several patterns have emerged that can be used for analysis.[34, 35] We evaluated the evidence for features of a point-source exposure, unusual route of exposure, higher attack rate

Table 5. Strength of association, with rationale, by evidence block.

Evidence block	Strength of association for hypothesis: H1 (Attacks intentional)	H2 (Exposures natural)	H3 (Events fabricated)	Rationale
E1 Toxicological analysis	Medium	Medium	Medium	Mycotoxins were found, but the sample analysis program could not conclusively say they were not from the environment.
E2 Medical records	Strong	Weak	None	Clinical complaints and findings were consistent with exposure to a chemical or toxin attack, but it is possible that symptoms were of natural origin. Illness and death, however, were real and not fabricated.
E3 Attack data	Strong	None	None	Detailed information on each claimed attack gave strong support for attacks having taken place and gave no support for other hypotheses.
E4 Autopsy results	Medium	Medium	None	Autopsies demonstrated large toxin loads at death, but toxin origin could not be clarified. Fabrication not supported.
E5 Intelligence reports **a.** Human source **b.** Intercepts and imagery	Strong	None	Weak	Intelligence reports strongly supported intentional attack, but some reports could have been fabricated. No support for natural exposure.
E6 Soviet-link evidence	Strong	None	None	Evidence consistent with a Soviet link to attacks in Laos, Cambodia, Thai borderlands, and Afghanistan supported intentional attack strongly and exclusively.
E7 Coincidence analysis	Strong	Weak	Weak	Consistencies among claims in Laos, Cambodia, Thai borderlands, and Afghanistan strongly supported intentional attack but also offered support to natural exposure (if plausible in two dissimilar regions) and fabrication (if plausible in two distant regions).
E8 Open-source reports	Medium	Medium	Weak	Open-source reporting offered some support for each hypothesis.
E9 Hmong interviews	Strong	Medium	Weak	Hmong interviews strongly supported intentional attack, but some content was not inconsistent with natural exposure. Fabrication seemed unlikely.
E10 Investigator interviews	Strong	Medium	None	Investigator interviews found evidence strongly suggesting intentional attack but not precluding natural exposure. No content from the interviews suggested fabrication.
E11 Conduct of investigation **E12** Sampling methods	None None	None None	None None	Neither E11 nor E12 supports any hypothesis, but each becomes important when grouped with other evidence.

or larger-than-normal number of cases, atypical or unusually severe disease course, geographically unusual disease or strain, and corroborating intelligence information suggesting a non-natural occurrence. We found that all distinguishing characteristics of a CBW attack can be seen in the Yellow Rain evidence.

The evidence collected in the Yellow Rain investigation is only suggestive of an association between the morbidity and mortality reported by the population and exposure to a chemical or toxin agent. Standard guidelines for causal attribution are used to evaluate whether the association represents a causal relationship.[36, 37, 38, 39] The Yellow Rain evidence supporting a CBW event met the criteria for strength of association, dose-response relationship, temporal relationship, biological plausibility, and cessation. The criteria for consistency of association was difficult to meet, since it was not possible to say definitively what the causative agent was. Specificity of association was also difficult to discern given the evidence available, since it is unclear just how many people might have been exposed and how many of those exposed became ill. Finally, under confounding factors, it is possible that some type of naturally occurring disease affected all of the people who

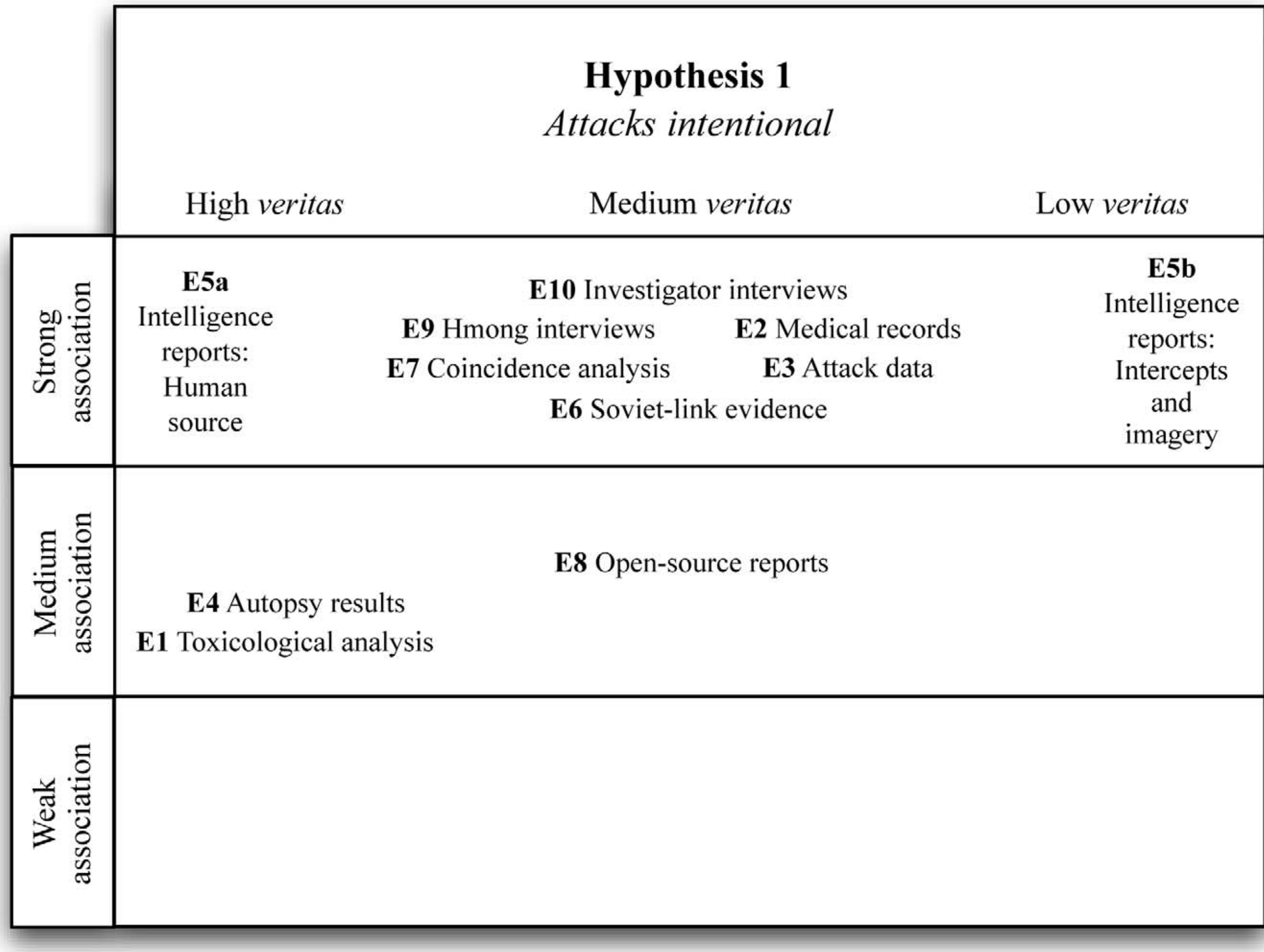

Figure 5a. Evidence blocks by *veritas* rank and strength of association with Hypothesis 1.

reportedly became ill or died. It is also possible that a variety of ailments were responsible. Given the evidence, though, it seems unlikely that all of the reported disease and death were due to anything other than exposure to a chemical or toxic agent.

Lastly, we checked our hypothesis against definitive proof or admission. Regardless of how much evidence there is for or against the use of chemical or toxin agents in Southeast Asia and Afghanistan in the late 1970s and early 1980s, no definitive proof is available. Nothing in the evidence collected during the investigation meets the requirement of "definitive" proof, and, since the biomedical and environmental samples from the investigation were destroyed, there is no way to go back and revisit the evidence for further clues.

Discussion

On the basis of evidence at hand, we conclude that lethal chemical or toxin compounds were used in Laos, Cambodia, in Thai borderlands, and Afghanistan, in violation of the international conventions operative during the 1970s and 1980s. We cannot, however, identify the specific agents used, the intent, or the root source or sources of the attacks.

The evidence analyzed here suggests — but only suggests — an association between reports and exposures. Evidence supporting an intentional attack met criteria for strength of association, dose-response relationship, temporal relationship, biological plausibility, and cessation. The criteria for consistency of

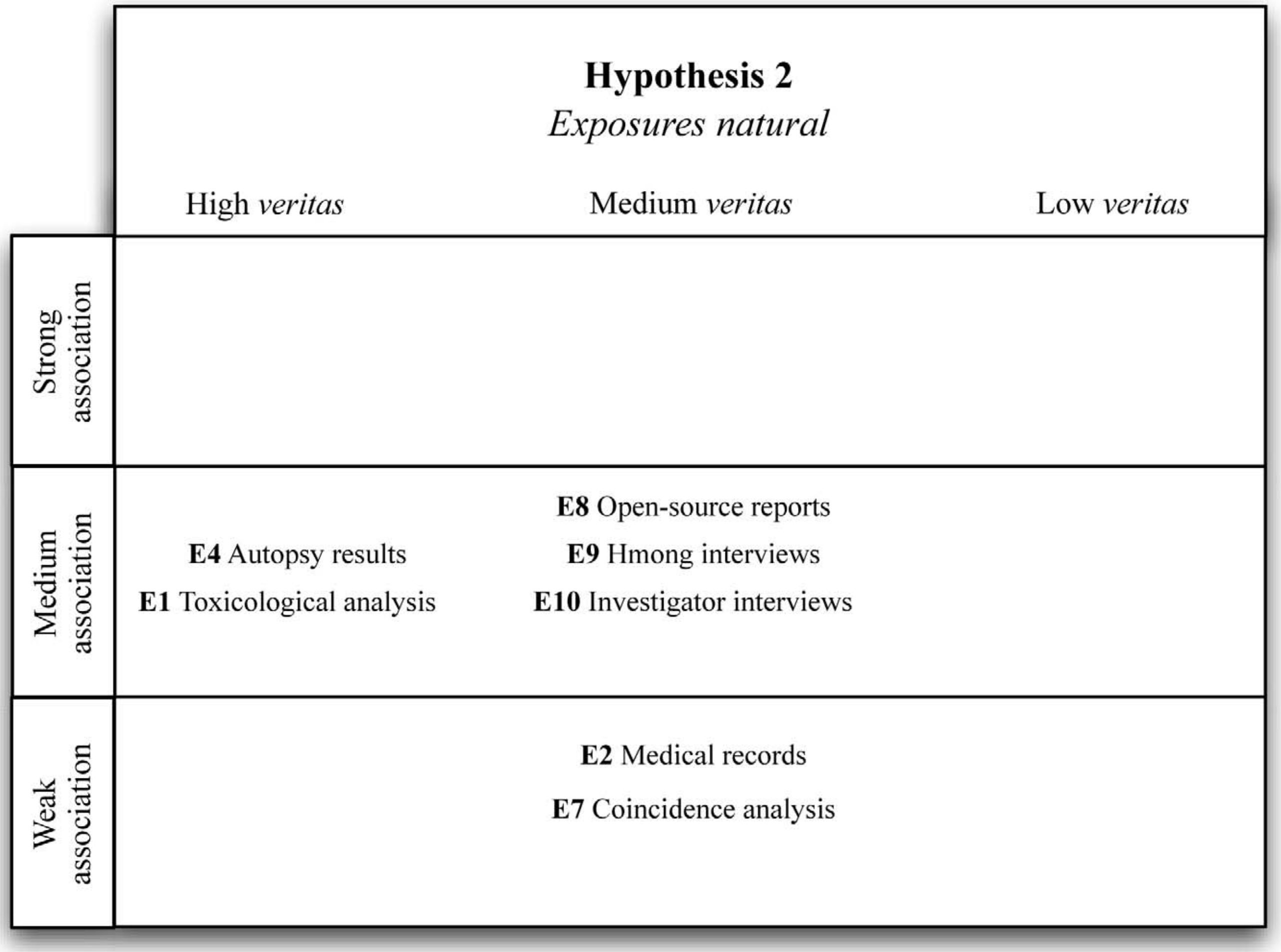

Figure 5b. Evidence blocks by *veritas* rank and strength of association with Hypothesis 2.

association were difficult to meet, since the causative agent or agents could not be identified definitively. Specificity of association was also difficult to discern, since just how many people might have been exposed, and how many of those exposed became ill, could not be well estimated. Finally, we could control for few confounding factors, such as a naturally occurring disease peaking in incidence during the years studied. That said, attributing *no* report of disease and death to a chemical or toxic agent seems the least plausible of all conclusions.

We can offer no assurance that our work has been free of bias, but we have throughout been aware of the risk that it might not be. The value framework of the researcher, as described by Robert Merton, affects not only the definition of the problem, but also the view of the evidence.[40] Awareness of this problem might help alleviate bias either through self-evaluation or through assistance by co-evaluators. A consensus evaluation process, such as the Diagnostic Council as developed by Henry Murray, could help.[41] Bias aside, the methodology we set forth makes all judgments *en route* to conclusions transparent. Anyone can challenge any step and introduce an alternative judgment, which might lead to a different conclusion. Any disparity would then be available for debate among a broader community.

The ambiguities inherent in analyzing evidence in a CBW investigation may in part be lessened through a technology now evolving in the domain of natural language processing systems.

Hypothesis 3
Events fabricated

	High *veracity*	Medium *veracity*	Low *veracity*
Strong association			
Medium association	**E1** Toxicological analysis		
Weak association	**E5a** Intelligence reports: Human source	**E7** Coincidence analysis **E8** Open-source reports **E9** Hmong interviews	**E5b** Intelligence reports: Intercepts and imagery

Figure 5c. Evidence blocks by *veritas* rank and strength of association with Hypothesis 3.

In this regard, if the process we have utilized is to be further mechanized, natural language processing systems that could carry out information extraction from large heterogeneous sets of documents would be needed.

Most natural language processing systems have roots in the work of Zellig Harris,[42, 43] who proposed a theory of sub-languages associated with particular domains of knowledge — *e.g.*, immunology,[44] clinical medicine,[45, 46] law.[47] Harris hypothesized that, in specialized lines of inquiry, information content and structure formed a specialized language that could be delineated as a sub-language grammar that a "language processor" could capture so as to encode salient information and relations found in text. This insight has been operationalized for various technical literatures and even for telegraphic fragments analyzed by naval intelligence officers.[48]

Of particular relevance to CBW investigations based on heterogeneous sets of documents is a sub-field of AI called information extraction (IE). This area was heavily promoted in the late 1980s in the United States under the auspices of the Defense Advanced Research Projects Agency (DARPA),[49] but its origins were in the Linguistic String Project at New York University,[45] where Naomi Sager advanced the work of Zellig Harris. Running roughly in parallel was Roger Schank, who studied story comprehension. Stories followed certain stereotyped patterns, referred to as scripts. Knowing "the script," language analyzers were able to fill in details and make

Figure 6. Laotian child said to be a victim of Yellow Rain. Rash was confined to the right side of the body, reportedly the side exposed to attack.

inferential leaps where information required to make a leap was not present in the text examined.[50]

Gerald de Jong[51] designed and built the first system based on this idea, FRUMP. It was used to extract information from news stories, clearly one of the important and difficult domains that arise in CBW investigations. This work has been complemented by a long series of "Message Understanding Conferences" (MUCs) running from the 1980s to the present and focusing on information extraction in the context of naval intelligence.[49] Although many MUC systems have been implemented, the information extraction (IE) field remains in flux. Hobbs[52] describes the generic IE system as a "cascade of transducers or modules that at each step add structure and often lose information, hopefully irrelevant, by applying rules that are acquired manually and/or automatically." To describe such a system requires identifying modules, identifying each module's input and output, identifying the form of the rules the modules apply, and specifying how the rules are applied and how they are acquired. Our explicit documentation of judgments leading to the assignment of *veritas* rankings and strengths of association is roughly equivalent to these IE steps. However, we are still some distance away from incorporation of such inputs into an operational expert system tuned to the myriad sources of evidence in the Yellow Rain controversy, or any other body of evidence relevant to CBW.

A second broad topic that induces discomfort in readers of conclusions derived from CBW investigations — *e.g.*, the "Results" above — is that they represent the culmination of causal inferences from single, novel events. The method underlying this culmination would seem to fly in the face of a large

Table 6. Summary statistics.

A. Hypotheses evaluated separately

Statistic Statistic	H_1	H_2	H_3
Maximum	10	8	8
Minimum	4	2	0
Average	7.64	5.42	3
Average over minimum-support cells	4	2	2
Average over strong-support cells	8	6.8	8
Percent of evidence blocks in strong-support cells	91	75	16.7

H1: E1, E2, E3, E4, E5, E6, E7, E8, E9, E10
H2: E1, E2, E4, E7, E8, E9, E10
H3: E1, E5, E7, E8, E9

B. Comparative scores based on pairwise overlapping blocks

Evaluated pairs	H_1 AND H_2		H_1 AND H_3		H_2 AND H_3	
Statistic Statistic	H_1	H_2	H_1	H_3	H_2	H_3
Maximum	8	8	10	8	8	8
Minimum	6	2	4	0	2	2
Average	7.7	5.42	7.3	3	5.5	3
Average over minimum-support cells	2	2	2	2	—	—
Average over strong-support cells	8	7.3	7.3	8	7	8
Percent of evidence blocks in strong-support cells	100	60	100	17	67	33

H1 AND H2: E1, E2, E4, E7, E8, E9, E10
H1 AND H3: E1, E5, E7, E8, E9
H2 AND H3: E1, E7, E8, E9

C. Comparative scores based on evidence blocks common to all three hypotheses

Statistic Statistic	H_1	H_2	H_3
Maximum	8	8	8
Minimum	6	2	2
Average	7.3	5.5	3.5
Average over minimum-support cells	2	2	
Average over strong-support cells	7.3	7	8
Percent of evidence blocks in strong-support cells	100	67	33

Blocks common to all three: E1, E7, E8, E9

literature going back to David Hume[53] implying that cause is inferred from the "frequent co-occurrence (spatial and temporal) of two previously experienced events."[54] A growing literature in psychology and neurobiology[54, 55, 56] implies that "although Hume's logic was sound, his psychology was not: cause is often inferred, by human adults and infants from single, novel events. . ."[54] The central point is that this form of causal inference is an unavoidable aspect of the analysis of evidence in CBW investigations. We admittedly know very little at the present time about the evolution of our capacity to generate causal inferences, and this is an active research area.[54] However, the uncertainties associated with this process, and with the evolving conceptual basis for establishing causal inference using natural language texts,[57] are reflected in uncertainties about conclusions derived from the assembly of evidence such as we have carried out or that might be carried out by future information extraction systems tuned to CBW applications. Our, or anyone else's, claims about Yellow Rain will, of necessity, be subject to this form of uncertainty. Whether definitive evidence will become available in the future remains to be seen.

The methodology we present for combining and weighing quantitative and qualitative evidence hopefully will enable analysts systematically to assess large bodies of evidence, particularly those pertaining to CBW investigations, and then transparently to establish a range of hypotheses. Results may be inconsistent, particularly in large complicated cases; inter-observer and inter-analyst comparisons are advisable.

There is no checklist for verifying if chemical or biological weapons have been used, so verification itself is subjective. While we are unable retrospectively to specify agent, origin, responsibility, or intent, we are nonetheless confident that chemical- or toxin-weapons attacks *did* occur in Southeast Asia and Afghanistan in the late 1970s and early 1980s.

Absence of evidence is not evidence of absence. It is entirely possible that some former Soviet or Vietnamese officer has definitive knowledge as to whether or not biological, toxin, or chemical weapons were used against the Hmong, Khmer, and Mujuhadin. And, if these weapons were used, someone knows with certainty the composition of the poison employed. Barring credible testimony, the Yellow Rain question is unlikely ever to rest.

We thank James D. Claxton for his assistance in declassifying the Yellow Rain documents; the many individuals we interviewed who graciously provided their time, personal files, and assistance; Phillip Russell, D. A. Henderson, Donald Kennedy, and James Crow for their assistance and comments on earlier drafts of this paper; Joshua Lederberg for helpful input in the early stages of this work; Princeton University, the Office of Population Research, and the Fellowship of Woodrow Wilson Scholars for financial support (R.K.); anonymous PLS reviewers for constructive comments; and our PLS editor for time, suggestions, and commitment.

Rebecca Katz is an Assistant Research Professor at The George Washington University School of Public Health and Health Services in the Department of Health Policy. Her research is focused on public health preparedness, biological warfare, and the intersection of infectious diseases and national security. Current research projects are focused on evaluating bioterrorism training for clinicians and the implementation of the International Health Regulations. She previously was an Intelligence Research Fellow at the Center for Strategic Intelligence Research in the Joint Military Intelligence College and is a consultant to the Office of Biologic Weapons Affairs in the U.S. Department of State, Bureau of Verification, Compliance and Implementation. Katz received a Ph.D. in Public Affairs from Princeton University and an M.P.H. from Yale University.

Burton Singer is the Charles and Marie Robertson Professor of Public and International Affairs and a Professor of Demography and Public Affairs at Princeton University, Woodrow Wilson School. Prior to joining the Princeton faculty, he was chair of the Department of Epidemiology and Public Health and professor of economics and statistics at Yale University. He has served as chairman of the National Research Council Committee on National Statistics and of the Steering Committee for Social and Economic Research in the World Health Organization. He is a member of the National Academy of Sciences and the Institute of Medicine and was a Guggenheim Fellow in 1981–82. His research interests include the epidemiology of tropical disease, the demography and economics of aging, the health and social consequences of economic development, and the interrelationships between genetics and historical demography. Professor Singer received his Ph.D. in statistics from Stanford University.

References

1. D. Kennedy, “Intelligence Science: Reverse Peer Review?” *Science,* 2004, 303:1945.

2. R. McCreight and S. Weigert, “Up in Smoke: Political Realities and Chemical Weapons Use Allegations during Mozambique’s Civil War,” *International Politics,* June 2001, 38:253–272.

3. L. Cole, *The Anthrax Letters: A Medical Detective Story* (Joseph Henry Press, 2003).

4. http://www.politicsandthelifesciences.org/Contents/Subscribers/Contents-2007-3/poli-26-01-05_24-42_supplemental_documents/.

5. S. Seagrave, *Yellow Rain: A Journey Through the Terror of Chemical Warfare* (New York: M. Evans and Company, 1981).

6. Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of War (Geneva, 1925); available at http://www.state.gov/t/ac/trt/4784.htm.

7. Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction (1972); available at http://www.state.gov/t/ac/trt/4718.htm#treaty.

8. Geneva Convention, Common Article 3 (1949); available at http://www.unhchr.ch/html/menu3/b/91.htm.

9. Principles of International Law Recognized in the Charter of the Nuremberg Tribunal and in the Judgment of the Tribunal (1950); available at http://www.icrc.ch/IHL.nsf/FULL/390?OpenDocument.

10. U.S. Department of State, *Chemical Warfare in Southeast Asia and Afghanistan: An Update. Report to the Congress from Secretary of State George P. Shultz.* (Special Report No. 104, November 1982).

11. R. Katz, *Yellow Rain Revisited: lessons learned for the investigation of chemical and biological weapons allegations.* dissertation (Princeton, NJ: Princeton University, 2005).

12. H. Crone, J. Gee, and R. Barton, "Lessons Learnt from the Yellow Rain Investigation Conducted by Australia," paper presented at the Workshop on Investigations of Alleged Use of Chemical Weapons, Geneva, 1984.

13. J. Dow and G. R. Humphreys, "An Epidemiological Investigation of Alleged CW/BW Incidents in SE Asia," Prepared by Directorate of Preventive Medicine, Surgeon General Branch, National Defence Headquarters, Ottawa, 1982.

14. T. D. Inch, "The Decision to go to War in Iraq: Oral Evidence Taken before the Foreign Affairs Committee, UK Parliament," United Kingdom House of Commons, Foreign Affairs Committee, 18 June 2003; http://www.globalsecurity.org/intell/library/reports/2003/6182003_3_iraq_uk.htm).

15. A. Heyndrickx, N. Sookvanichsilp, and M. Van Den Heede, *Arch Belg*, 1984, Supplement, p. 110.

16. U. S. Department of State, *Chemical Warfare in Southeast Asia and Afghanistan. Report to the Congress from Secretary of State Alexander M. Haig, Jr.,* Special Report No. 98, 22 March 1982.

17. M. J. B., quoting C. Darwin on Yellow Rain, *Gardener's Chronicle,* 18 July 1863, 29:675.

18. Zhang Zhongying, Chen Yu-Ming, Chow Shu, and Li Min, "A Study of the Origin and the Pollen Analysis of "Yellow Rain" in Northern Jiangsu," *Kexue Tongbao,* 1977, 22:409–412.

19. T. D. Seeley, J. W. Nowicke, M. Meselson, J. Guillemin, and P. Akratanakul, "Yellow Rain," *Scientific American,* 1985, 253:128–137.

20. J. Nowicke and M. Meselson, "Yellow Rain — a palynological analysis," *Nature,* 17 May 1984, 309:205–206.

21. N. E. Dorsey, "The Velocity of Light," *Transactions of the American Philosophical Society*, 1944, 34, Part 1, p. 1.

22. M. T. Kane, "An argument-based approach to validity," *Psychological Bulletin,* 1992, 112:527.

23. J. Venn, *The Logic of Chance* (New York: Chelsea Publishing Company, 1888), Ch. 14.

24. I. J. Good, *Probability and the Weighing of Evidence* (London: Charles Griffin and Company Limited, 1950).

25. J. Bunker, W. Forrest, F. Mosteller, and L. Vandam, *The National Halothane Study: A Study of the Possible Association between Halothane Anesthesia and Postoperative Hepatic Necrosis* (Bethesda, MD: The National Institutes of Health, 1969).

26. P. E. Tetlock and A. Belkin, *Counterfactual thought experiments in world politics: logical, methodological, and psychological perspectives.* (Princeton: Princeton University Press, 1996).

27. J. Guillemin, *Anthrax: the investigation of a deadly outbreak* (Berkeley: University of California Press, 1999).

28. National Security Archives, *Volume V: Anthrax at Sverdlovsk, 1979: U.S. Intelligence on the Deadliest Modern Outbreak.* Robert A. Wampler and Thoas S. Blanton, eds., National Security Archive Electronic Briefing Book No. 61. 15 November 2001; https://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB61.

29. Arthur Conan Doyle, *The Complete Sherlock Holmes: All 4 Novels and 56 Short Stories* (Bantam Classic and Loveswept, 1998).

30. R. Wannemacher and S. Weiner, "Trichothecene Mycotoxins," in *Textbook of Military Medicine: Medical Aspects of Chemical and Biological Warfare*, F. R. Sidell, E. T. Takafuji, and D. R. Franz (Washington, DC: Office of the Surgeon General at TMM Publications, 1997), pp. 655–676.

31. Armed Forces Medical Intelligence Center. (1985) "Compilation of CW Use Data, 5 July 1985," (DIA Declassification number 152534).

32. United Nations Group of Experts, Report to the Secretary General (1982). UN Document number A/37/259.

33. E. Marshall, "The Soviet Elephant Grass Theory," *Science,* 1982, 217:32.

34. J. Pavlin, "Epidemiology of Bioterrorism," *Emerging Infectious Diseases,* 1999, 5(4):528.

35. T. Treadwell, D. Koo, K. Kuker, and A. Khan, "Epidemiologic clues to bioterrorism," *Public Health Reports,* 2003, 118:92.

36. A. Evans, "Causation and Disease: A chronological journey," *American Journal of Epidemiology,* 1978, 108(4):249.

37. A. B. Hill, "Heberden Oration 1965: Reflections on the Controlled Trials," *Annals of the rheumatic diseases,* 1966, 25(2):107.

38. A. Evans, "Causation and Disease: the Henle-Koch postulates revisited," *Yale Journal of Biology and Medicine,* 1976, 49:175.

39. S. Falkow, "Molecular Koch's postulates applied to bacterial pathogenicity — a personal recollection 15 years later," *Nature Reviews Microbiology,* 2004, 2(1):67–72.

40. R. K. Merton, "The Role of Applied Social Science in the Formation of Policy: A Research Memorandum," *Philosophy of Science,* 1949, 16(3):161–181.

41. H. A. Murray, *Exploration in Personality* (New York: Oxford Univesrtiy Press, 1938).

42. Z. Harris, *Language and Information* (New York: Columbia University Press, 1988).

43. *The Legacy of Zellig Harris: Language and Information into the 21st Century,* 2 vols., B. E. Nevin, S. B. Johnson, eds. (Amsterdam: John Benjamins Publishers, 2002).

44. Z. Harris, M. Gottfried, T. Ryckman, P. Mattick, A. Daladier, T. N. Harris, and S. Harris S, *The Form of Information in Science: Analysis of an Immunology Sub-language* (Dordrecht: Reidel, 1987).

45. N. Sager and N. T. Nhan, “The computability of strings, transformations, and sublanguage,” in *The Legacy of Zellig Harris: Language and Information into the 21st Century,* 2 vols., B. E. Nevin, S. B. Johnson, eds. (Amsterdam: John Benjamins Publishers, 2002), vol. 2, pp. 79–120.

46. N. Sager, M. Lyman, C. Bucknall, N. Nhan, and L. J. Tick, “Natural language processing and the representation of clinical data, *J Am Med Inform Assoc,* 1994, 1(2):142–60.

47. M. Sergot, “The representation of law in computer programs: a survey and comparison of past and current projects,” in T. Bench-Capon, ed., *Knowledge-Based Systems and Legal Applications* (London: Academic Press, 1991), pp. 3–67.

48. M. C. Linebarger, D. A. Dahl, L. Hirschman, and R. J. Passoneau, “Sentence fragments regular structures,” *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics* (Buffalo, NY: 1988).

49. R. Gaizauskas and R. Wilks, “Information Extraction: Beyond Document Retrieval,” *Computational Linguistics and Chinese Language Processing,* 1998, 3(2):17–60.

50. R. C. Schank and R. P. Abelson, *Scripts, Plans, Goals, and Understanding* (Hillsdale, NJ: Lawrence Erlbaum, 1977).

51. G. Dejong, “An overview of the FRUMP system,” in W. Lehnert and M. H. Ringle, eds., *Strategies for Natural Language Processing* (Hillsdale, NJ: Lawrence Erlbaum, 1982), pp. 149–176.

52. J. R. Hobbs, “The generic information extraction system,” in *Proceedings of the 5th Conference on Message Understanding.* (Morristgown, NJ: Association for Computational Linguistics, 1993), pp. 87–91.

53. D. Hume, *A Treatise of Human Nature* (Oxford, UK: Oxford University Press, 1739/1998).

54. M. Hauser and B. Spaulding, “Wild rhesus monkeys generate causal inferences about possible and impossible physical transformations in the absence of experience,” *Proc National Acad. Sciences,* 2006, 103(18):7181–7185.

55. D. Premack and A. Premack, *Original Intelligence* (New York: McGraw-Hill, 2002).

56. A. Michotte, *The Perception of Causality* (New York: Basic Books, 1962).

57. J. R. Hobbs, “Toward a Useful Concept of Causality for Lexical Semantics,” *J. Semantics,* 2005, 22(2):181–209.