LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-646429

Conditional Dependence in Microbial Forensic Assays - A Primer

S. P. Velsko

November 19, 2013

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Conditional dependence in microbial forensic assays – A primer

Stephan P. Velsko
Lawrence Livermore National Laboratory
November 8, 2013

Summary

This report provides an introduction to the topic of conditional dependence in the context of microbial forensic assays. Conditional dependence between two items of evidence E_1 and E_2 occurs when they are both used to support a hypothesis, but E_1 affects the probability of E_2 and vice versa. Ignoring this dependence can lead to very large errors in estimating the diagnosticity of the combined evidence. To introduce readers to this concept, a number of definitions of conditional dependence that have been used by authors in the past have been collected together and compared. Formal mathematical relationships that constrain conditional dependence are summarized. There are several specific scenarios in which unrecognized conditional dependence can arise in microbial forensic contexts. This report provides some notional examples that illustrate dramatic effects of conditional dependence on the weight of microbial forensic evidence, and discusses the relevance of these observations for the validation of microbial forensic assays. A two-parameter model that describes the transition between various limiting forms of conditional dependence relations is provided in an appendix.

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

1. Introduction

Scientists who present forensic evidence to non-experts may sometimes unknowingly create an erroneous perception of the probative weight of the technical findings they communicate. This is true even for scientists who utilize rather sophisticated statistical concepts in their work, or invoke such concepts when they communicate evidence to others. One of the more subtle statistical concepts that can be problematic in this regard is conditional dependence, which occurs when the results of one test are correlated with those of a second test for ascertaining the same fact in question. Error can arise because the source of correlation may not be obvious, and the intuitive heuristic is to assume that the two tests represent two independent determinations of the same fact.

There is an extensive literature on the problem of conditional dependence in medical tests, where this problem has been recognized for a long time.[1-4] In addition, some aspects of the fallacies that can arise from incorrectly assuming independence have been discussed in relation to general types of forensic evidence.[5,6] A primary concern is that investigators or jurors will overestimate the combined weight of two pieces of evidence offered in support of a common hypothesis because their mutual dependencies will go unrecognized. Understanding this statistical phenomenon takes on additional significance in theories of forensic evidence because recent treatments of context bias have been based on modeling such bias as an induced form of conditional dependence.[7]

The purpose of this report is to assemble a primer that will help scientists who use or develop microbial forensic assays improve their intuitions about the phenomenon of conditional dependence. This will presumably help them to recognize when conditional dependence may be important in their specific domains of knowledge, provide a basis for estimating the magnitudes of conditional dependence on evidence weight, and guide the construction of validation experiments. Toward this end, this report:

- collects together and compares a number of definitions of conditional dependence that have been used by authors in the past,
- summarizes some formal mathematical relationships that constrain conditional dependence,
- describes some specific scenarios in microbial forensic contexts in which unrecognized conditional dependence can arise,
- provides some concrete examples that illustrate dramatic effects of conditional dependence on the weight of microbial forensic evidence, and
- discusses the implications for the validation of microbial forensic assays.

The mathematical treatment assumes that the reader is familiar with some elementary probability theory - particularly the relationships between joint and conditional probability, Bayes theorem, and the notion of marginalization.[8] The approach, however, is not Bayesian *per se*. The notion of the diagnosticity of tests as measured by the likelihood ratio is borrowed directly from the medical literature. The notion of prior probability enters into consideration only briefly, because it is necessary for connecting conditional dependence to direct measures of correlation between two different types of measurements or two different test results.

2. When are two evidence items conditionally *dependent*?

Consider a case where there are two items of evidence, for example the results of two tests for the presence of some substance. Assume that these items of evidence are proffered to support some hypothesis H about the nature or origin of some crime or terrorist event. The likelihood that those test results would be found if H were true is denoted $P(E_1,E_2|H)$, where E_1 and E_2 are variables whose values represent the various possible results for each test. Normally, evidence is used to distinguish between H and some alternative hypothesis A. The degree to which E_1 and E_2 can indicate whether H or A is more likely to be true is called the *diagnosticity* of the evidence.[9] The magnitude of the diagnosticity is given by the likelihood ratio:

$$L = \frac{P(E_1E_2|H)}{P(E_1E_2|A)} \tag{1}$$

When the value of L is greater than 1 the evidence favors H over A, and vice versa when L is less than 1. When L is very close to 1 we say that the evidence is not very diagnostic – i.e. it does not help us determine whether H or A is the more likely hypothesis.

We say that E_1 and E_2 are conditionally independent *given* H if

$$P(E_1E_2|H) = P(E_1|H)P(E_2|H). \tag{2}$$

It is important to note that the independence of E_1 and E_2 with respect to H does not, generally, guarantee that they are independent with respect to any other hypothesis, such as A.

Using the relationship between joint and conditional probabilities we can also write

$$P(E_1E_2|H) = P(E_1|E_2H)P(E_2|H) = P(E_2|E_1H)P(E_1|H) \tag{3}$$

where $P(E_1|E_2H)$ is the likelihood of E_1, given both E_2 and H. Using this equation we can write L in the alternative forms:

$$L = \frac{P(E_1|E_2H)P(E_2|H)}{P(E_1|E_2A)P(E_2|A)} = \frac{P(E_2|E_1H)P(E_1|H)}{P(E_2|E_1A)P(E_1|A)} \tag{4}$$

If E_1 and E_2 are independent with respect to *both* H and A this equation becomes

$$L = \frac{P(E_1|H)P(E_2|H)}{P(E_1|A)P(E_2|A)} = L_1L_2 \text{ where } L_j = \frac{P(E_j|H)}{P(E_j|A)}. \tag{5}$$

Examining equations (2) and (3) it is clear that the conditional independence of E_1 and E_2 with respect to H and A implies

$$P(E_j|E_kX) = P(E_j|X) \text{ where } X = H \text{ or } A,\ j = 1 \text{ or } 2, \text{ and } k \neq j. \tag{6}$$

Equations (2) and (6) are equivalent definitions of conditional independence, and whenever they do <u>not</u> hold we say that E_1 is conditionally dependent on E_2 (or vice versa) given H (or A).

When can intuitive arguments be made either for or against the proposition that two items of evidence are independent with respect to some hypothesis? Some authors have attempted to address this question by identifying conditions under which independence is true, while others have considered reasons for evidence items to be conditionally dependent (i.e. *not* independent.)

For example, in his book on Bayesian Networks Judea Pearl interprets the meaning of the conditional independence expression of $P(E_1|E_2H)) = P(E_1|H)$ in the following way (using our own variables): "Given H, E_2 tells us nothing new about E_1," and "E_2 is irrelevant to E_1, once we learn H." Similarly, "Once H is given, the probability of E_1 will not be affected by the discovery of E_2"(reference 10, page 80.) From Pearl's point of view, the (in)dependence question depends on whether one item of evidence can generate new information that is relevant to the other item – and is not available from H.

As an example of this kind of reasoning, consider two items of evidence offered to support the hypothesis that a certain suspect was present at the scene of the crime: DNA found at the scene that matches the suspect's, and eyewitness testimony. If we assume that he was present (H is true), then the eyewitness testimony does not affect the probability of finding his DNA at the crime scene.

As another example, suppose that E_1 is the finding that traces of *B. anthracis* DNA were found in a defendant's apartment. E_2 is the fact that the defendant works as a microbiology technician at a veterinary lab that has handled anthrax cases. H is the hypothesis that the defendant stole and processed the pathogen for use in a letter. Clearly, if H were true the additional fact regarding the suspect's employment adds no new information relevant to the probability of finding trace amounts of pathogen in his apartment. In contrast, consider the alternative hypothesis that the suspect is

innocent, and someone else sent the pathogen-containing letters. Now the fact that he worked in the lab is highly relevant to the probability of finding traces of *B. anthracis* – the finding could simply represent inadvertent transfer of contamination from work to home, making $P(E_1|E_2A) >> P(E_1|A)$.

A subtly different take on the informational relevance concept has been expressed by Pepe in a discussion of conditional dependence in medical testing: "[I]f a subject's true disease status is known, then knowledge of the result of one test is not informative about the result of [the other test or] any of the other tests." (Reference 11, page 197.) It is clear from the context that by "known" Pepe means "if the subject has been drawn from the class of persons with a particular disease status." Pepe's formulation originates from an explicit concern with testing, while Pearl is concerned with more general categories of knowledge and types of inference:

> "[C]onditional independence is not a grace of nature for which we wait passively, but rather a psychological necessity which we satisfy actively by organizing our knowledge in a specific way." (Reference 10, page 44)

How does one recognize conditional dependence in data? First, clearly, it is necessary to have data in which the two tests or measurements are performed on exemplars taken from the classes that both tests are meant to distinguish. Then "[a] lack of conditional independence for any two variables is signaled by a significant correlation between them within the [H true] cases or within the [A true] cases, or both." (Reference 2, page 425)

What considerations may lead one to suspect that two tests are conditionally dependent? A variety of factors have been identified:

> "[I]f both tests are based on a particular antibody reaction, something which inhibits the reaction or causes a false reaction for one of the tests may have a similar effect on the other." (Reference 3, page 959.)

> "If both tests more easily detect disease when it is advanced or severe, then the conditional independence assumption will fail." (Reference 11, Page 196)

An analogy to this idea would be: *if both tests more easily detect a microbe when it is present in high amounts, then they are conditionally dependent*. Another observation that has relevance to microbial forensics is:

> "If contamination of a specimen causes both [tests] to be positive when no disease is present, then [the tests] will be conditionally dependent." (Reference 11, page 196).

Conceptually, the idea of conditional dependence is linked to the idea of causality. There must be some objective physical or biological reason for two properties to be conditionally dependent. Thus, if expert judgment is involved with deciding whether a test is positive or negative it is important to ask: "might an expert decide the state of E_2 differently if E_1 were positive or negative, *given that he does not know which state E_1 is in*?" Generally, intuition is a poor guide to this issue, and empirical blind testing of the expert is the only way to ascertain the answer.

Suppose one wishes to improve one's ability to discriminate between positive and negative samples by incorporating a second test. In choosing a second test, it is important to know that its performance is not linked to the performance of the first test by the same factors. Lempert explicitly points to this potential source of evidentiary weakness:

> "In attempting to prove a disputable point, an attorney should seek items of evidence that do not share the same sources of possible unreliability. In attempting to destroy an opponent's case, counsel should strive to show that the evidence of the opponent is infected from a common source." (Reference 5, page 1055.)

Lempert calls an item of evidence E_2 that is strongly conditionally dependent on another, E_1, "cumulative" if $\frac{P(E_2|E_1H)}{P(E_2|E_1A)} \approx 1$, in other words the second item of evidence is strongly implied by the first, and adds nothing additional to the support of H. In general, he raises the concern that juries might mistakenly think that the probative value of the two items together is much stronger than the value of E_1 alone. However, he acknowledges that sometimes E_2 has value beyond its nominal diagnosticity because non-experts simply expect it to be part of the case and would be suspicious of E_1 if it weren't presented:

> "A second situation in which cumulative evidence should be admitted is where the jury expects that the evidence will be produced if it exists. The absence of evidence conveys information to the jury, and it is possible for the proven availability of evidence to be cumulative while its proven unavailability has considerable probative value. In these circumstances cumulative evidence should be admissible, despite slight probative value, in order to dispel the implication that it is unavailable." (Reference 5, page 1056)

An example of this situation is one where a "classic" test is well known (perhaps through previous cases) but less diagnostic than a newer test that has replaced it in practice. Convincing non-experts that a positive result from the new test is trustworthy might require showing them that the older test is also positive. Of course, if the older test is negative, considerable explanation might be required to establish the probative value of the newer test.

3. Some general mathematical relationships governing conditional dependence

This section collects together some general mathematical relationships among the conditional probabilities involving two items of evidence E_1 and E_2, and two complementary hypotheses: H and it's alternative A (Not-H). For simplicity we will assume that E_1 and E_2 are binary (Yes/No) variables (for example, the results of dichotomous tests), and the compliment of E_j is denoted $\not{E}_j$. Finally, generalization of the formulas is facilitated by introducing the general symbols X and Y to stand for H, A or E_j, with $\not{X}$ or $\not{Y}$ being the corresponding complements.

Conditional probabilities involving binary variables obey the so-called "attraction" and "repulsion" relationship: if $P(X|Y) > P(X)$ then $P(X|\not{Y}) < P(X)$.[12] Thus either

$$P(E_j|X) \leq P(E_j) \leq P(E_j|\not{X})$$
or
$$P(E_j|X) \geq P(E_j) \geq P(E_j|\not{X})$$

for X = H, A, or $E_{i \neq j}$.

Similarly,

$$P(E_j|X,Y) \leq P(E_j|Y) \leq P(E_j|\not{X},Y)$$
or
$$P(E_j|XY) \geq P(E_j|Y) \geq P(E_j|\not{X},Y)$$

for X,Y = H, A, or $E_{i \neq j}$.

The derivation of these inequalities is straightforward and is provided in the appendix. These relationships can be thought of in terms of the matrix of conditional probabilities represented by Table 1, where the rows are either increasing or decreasing from left to right and the columns are either increasing or decreasing going from top to bottom.

Table 1. Matrix of related conditional probabilities.

	E_1	$\varnothing$[a]	$\not{E}_1$
H	$P(E_2\|E_1,H)$	$P(E_2\|H)$	$P(E_2\|\not{E}_1,H)$
$\varnothing$[b]	$P(E_2\|E_1)$	$P(E_2)$	$P(E_2\|\not{E}_1)$
A	$P(E_2\|E_1,A)$	$P(E_2\|A)$	$P(E_2\|\not{E}_1,A)$

[a]Independent of E_1 and $\not{E}_1$.
[b]Independent of H and A.

Given E_2 and H, the conditional dependence on a second item of evidence E_1 can be said to be "attractive" if $P(E_2|E_1,H) > P(E_2|H)$, "repulsive" if $P(E_2|E_1,H) < P(E_2|H)$, and "neutral" if $P(E_2|E_1,H) = P(E_2|H)$. Similarly, given E_1 and E_2, X "attracts" E_2 to the degree that $P(E_2|E_1,X) > P(E_2|E_1)$. Independence is equivalent to simultaneous neutrality with respect to X and its complement $\not{X}$ – i.e. the top and bottom rows of the matrix contain identical probability values.

These relations are symmetric with respect to exchange of E_1 and E_2.

As a consequence of these relationships, the various likelihood ratios obey the following relations:

If $P(E_j|X) \geq P(E_j)$ then $LR(E_j) = \frac{P(E_j|X)}{P(E_j|\not{X})} \geq 1$, and
If $P(E_j|X) < P(E_j)$ then $LR(E_j) = \frac{P(E_j|X)}{P(E_j|\not{X})} < 1$

If $P(E_i|E_j, X) \geq P(E_i|E_j)$ then $LR(E_i|E_j) = \frac{P(E_i|E_j,X)}{P(E_i|E_j,\not{X})} \geq 1$, and
If $P(E_i|E_j, X) < P(E_i|E_j)$ *then* $LR(E_i|E_j) = \frac{P(E_i|E_j,X)}{P(E_i|E_j,\not{X})} < 1$

Finally, note that for $LR(E_i|E_j) = LR(E_i)$, E_j must be "neutral" with respect to E_i for <u>both</u> X and $\not{X}$.

An alternative way to express the effect of conditional dependence is to write the Bayesian posterior probability as:

$$P(H|E_1E_2) = \frac{P(E_2|E_1H)}{P(E_2|E_1)}P(E_1|H) = \frac{P(E_1|E_2H)}{P(E_1|E_2)}P(E_2|H) \qquad (7)$$

Notice that when E_1 so strongly "attracts" E_2 that H has only a weak effect on the probability of E_2, then $P(E_2|E_1H) \approx P(E_2|E_1)$, and E_2 doesn't contribute to the posterior probability. Symmetry dictates that if E_2 strongly attracts E_1, then E_1 is similarly rendered irrelevant. When $P(E_2|E_1H) \approx P(E_2|E_1)$ it is also

Finally, we note that if $P(E_2|E_1X) \geq P(E_2|X)$ then $P(\not{E}_2|\not{E}_1X) \geq P(\not{E}_2|X)$. In fact, it is possible to derive a set of relations similar to those governing the Table 1 matrix, with $\not{E}_2$ replacing E_2.

Testing data for conditional dependence means establishing that $P(E_1|E_2X) = P(E_1|\not{E}_2X)$ for both X = H and A. Given a matrix of data E_1, $\not{E}_1$, E_2, and $\not{E}_2$ for exemplar samples drawn from the population generated under H or A, this can be done by showing that the null hypothesis $\frac{P(E_1E_2|X)}{P(E_1E_2|X)+P(\not{E}_1E_2|X)} = \frac{P(E_1\not{E}_2|X)}{P(E_1\not{E}_2|X)+P(\not{E}_1\not{E}_2|X)}$ cannot be rejected at some specified level of confidence.

3. An example of conditional dependence in microbial forensics

Consider a simplified scenario where the home of a suspected terrorist is investigated, seeking evidence that production of a quantity of anthrax agent was undertaken at that location. A small amount of liquid is found in a waste container, and a sample of it is tested for the presence of *B. anthracis* DNA using a PCR-based assay. In addition, one of the investigators perceives that the container has a faint odor of bleach, and a chemical test that can reveal the trace presence of sodium hypochlorite is performed.

We will denote the result of the hypochlorite assay as E_1 and the result of the PCR test as E_2. Each of these variables can have two states – positive or negative, depending on the result of the corresponding test. We want to consider how the test evidence bears on the hypothesis that the suspect did manufacture the agent and *may* have attempted to destroy residual contamination of the waste container with bleach. Let this hypothesis be H. The alternative hypothesis N is that no anthrax production took place (and bleach – if present – was simply used to de-odorize the waste bin.) It is, of course, necessary to differentiate between the hypotheses that bleach was or was not used (B versus $\cancel{B}$) and the hypochlorite test result being positive or negative (b or $\cancel{b}$). Table 1 provides a summary of the symbols used in the subsequent analysis.

Table 1. Summary of definitions for the PCR/bleach scenario

	Variable name/state	Meaning
Hypotheses	H	Anthrax agent was produced
	N	Anthrax agent was not produced
	B	Bleach was used
	$\cancel{B}$	Bleach was not used
Test results	$E_1 = b$	Bleach is detected
	$E_1 = \cancel{b}$	Bleach is not detected
	$E_2 = A$	PCR test for *B. anthracis* is positive
	$E_2 = \cancel{A}$	PCR test for *B. anthracis* is negative

Imagine that the performance of the PCR assay is tested on a large number of samples that have been obtained by running an equally large set of simulations in which surrogate terrorist teams operating under realistic conditions produce *B. anthracis* or some other biological agent. In some of those simulations bleach may be used to clean up waste containers, and in others not. Such tests could be used to estimate the values of the probabilities in Table 2.

In reality, of course, an extensive set of simulations as described is not practical. Instead, laboratory validation experiments at best produce estimates of $P(A|\cancel{B},H)$ and $P(A|\cancel{B},N)$ because they seldom, if ever, consider the effect of bleach. In principle, experiments where bleach is deliberately added to samples may be used

to estimate P(A|b,H), P(A|b,N), P(A|$\cancel{b}$,H), and P(A|$\cancel{b}$,N) by grouping samples according to whether they test positive or negative for hypochlorite.

Table 2. Conditional probabilities that could be estimated from an set of simulation exercises; note that in each case the probability of assay failure can automatically be derived from the equation P($\cancel{A}$|X,Y) = 1 - P(A|X,Y).

	B. anthracis was produced	*B. anthracis* was not produced
Bleach was used to clean up	P(A|B,H)	P(A|B,N)
Bleach was not used to clean up	P(A|$\cancel{B}$,H)	P(A|$\cancel{B}$,N)

As noted above, the hypochlorite test being positive or negative (b or $\cancel{b}$) is not strictly equivalent to the hypotheses that bleach was or was not used (B versus $\cancel{B}$). However, for simplicity in what follows we will assume that a positive test result implies that bleach was used with 100% probability and a negative test similarly implies that bleach was not used. Thus, we will freely substitute B for b and $\cancel{B}$ for $\cancel{b}$. Nonetheless, a more rigorous treatment would clearly entail relaxing this assumption. One could also imagine a different scenario where the finding of a number of empty bleach containers is the evidence for B, rather than a chemical test.

In order to illustrate how the conditional dependence of PCR testing on the presence of bleach, Table 3 provides some plausible values for various probabilities used in the calculations.

Table 3. Values of the conditional probabilities for the PCR/bleach scenario.

Quantity	Value	Comment
P(A|$\cancel{B}$,H)	0.99	
P(A|$\cancel{B}$,N)	1×10^{-6}	False positive rate
P(A|B,H)	0.05	Bleach reduces DNA target sequence copy number
P(A|B,N)	1×10^{-6}	Bleach doesn't affect false positive rate
P(B|H)	0.5	About 50% of terrorists try to eliminate incriminating evidence with bleach
P(B|N)	0.1	About 10% of ordinary households bleach their waste containers to eliminate odors
P(A|H)	0.52	Derived
P(A|N)	1×10^{-6}	Derived

The false positive rate for PCR reactions in modern laboratories is most typically the consequence of contamination. Clearly the rate will depend on the degree of quality control and other conditions of practice within the laboratory, but 10^{-6} appeared to be a plausible quantity for this illustration.

The quantities P(A|H) and P(A|N) represent the probabilities of getting a positive PCR result when it is not known whether bleach was used or not. Using the formal expansion of P(A|H) and the values in Table 3 we find that:

$P(A|H) = P(A|B,H)P(B|H) + P(A|\not{B},H)\,P(\not{B}|H) = 0.52$

This is interpreted to mean that if we tested samples generated by simulated terrorist red teams operating under realistic conditions, only slightly more than half of the PCR tests would be positive because half the red teams would choose to bleach their waste containers. Similarly we find:

$P(A|N) = P(A|B,N)P(B|N) + P(A|\not{B},N)\,P(\not{B}|N) = 1 \times 10^{-6}$

This is a direct consequence of our assumption that the probability of a false positive result is independent of whether bleach cleanup protocols are used or not.

Now let us consider the consequences of conditional dependence on estimating the strength of the findings E_1 and E_2. First, notice that our choices for P(B|H) and P(B|N) imply that a positive finding for bleach alone has significant diagnosticity with respect to the hypothesis H, since it is more likely to find bleach used to cover-up biological agent production than it is to find it in legitimate situations. The values in Table 3 imply that a positive bleach test increases the odds of H by a factor of 5.

Next consider the composite diagnosticity of finding positive results for both the PCR and hypochlorite tests. One plausible error an investigator could make is to use the laboratory based estimates $P(A|\not{B},H)$ and $P(A|\not{B},N)$ to estimate the diagnosticity (likelihood ratio) of the positive PCR finding, then simply multiplying this with the bleach diagnosticity noted above:

$$LR = \frac{P(A|\not{B},H)}{P(A|\not{B},N)} \times \frac{P(B|H)}{P(B|N)} = 9.9\times10^5 \times 5 = 4.95\times10^6. \quad \text{(Error)}$$

In contrast, taking into account the conditional dependence between B and A, the correct diagnosticity value of finding both tests positive is much smaller:

$$LR = \frac{P(A|B,H)}{P(A|B,N)} \times \frac{P(B|H)}{P(B|N)} = 5\times10^4 \times 5 = 2.5\times10^5. \quad \text{(Correct)}$$

This approximately 20-fold reduction of diagnosticity occurs because the presence of bleach severely reduces the probability that a positive PCR "hit" is actually due to the presence of target DNA, while the probability that an observed "hit" is a false positive remains the same. If the reduction in target DNA caused by bleach were assumed to be much larger, much larger reductions in diagnosticity will obtain. It can also be argued that this more than an order of magnitude reduction of

diagnosticity would have much higher consequence for a test whose false positive rate is much higher than we have quoted for PCR, and the likelihood ratio is much smaller to begin with.

Now suppose that the bleach test is positive, but the PCR test is negative. Using the fact that P(~~A~~|X,Y) = 1 - P(A|X,Y), the erroneous calculation based on P(A|~~B~~,H) and P(A|~~B~~,N) gives:

$$LR = \frac{P(\cancel{A}|\cancel{B},H)}{P(\cancel{A}|\cancel{B},N)} \times \frac{P(B|H)}{P(B|N)} = 0.01 \times 5 = 0.05. \quad \text{(Error)}$$

The correct calculation is:

$$LR = \frac{P(\cancel{A}|B,H)}{P(\cancel{A}|B,N)} \times \frac{P(B|H)}{P(B|N)} = 0.95 \times 5 = 4.75. \quad \text{(Correct)}$$

Thus, the likelihood ratio is only slightly smaller than that for the bleach finding alone. This agrees with the intuition that a negative PCR result has little diagnosticity when there is evidence that bleach has been used to clean up. In comparison, the erroneous calculation violates this intuition by converting a mild degree of support for H into a significant degree of support for N.

In the case that the PCR test is positive, but the hypochlorite test is negative, the use of P(A|~~B~~,H) and P(A|~~B~~,N) to calculate the diagnosticity of the PCR finding is correct, and:

$$LR = \frac{P(A|\cancel{B},H)}{P(A|\cancel{B},N)} \times \frac{P(\cancel{B}|H)}{P(\cancel{B}|N)} = 9.9\times10^5 \times 0.56 = 5.5\times10^5.$$

Finally, if both findings are negative,

$$LR = \frac{P(\cancel{A}|\cancel{B},H)}{P(\cancel{A}|\cancel{B},N)} \times \frac{P(\cancel{B}|H)}{P(\cancel{B}|N)} = 0.01 \times 0.56 = 5.6\times10^{-3}.$$

4. Other examples of conditional dependence

A. "Orthogonal assays"

Consider two nominally "orthogonal" assays for a particular pathogen with different performance characteristics. A concrete example might be a PCR based genetic signature assay and an immunoassay (IMA). Each assay has been characterized separately to establish its approximate limit of detection. Imagine that an extensive

series of reference samples has been created by spiking environmental samples at slightly more than the detection limit of the less sensitive assay. For example, these samples corresponding to H ("pathogen present") are those containing 100 or more CFU per gram of target pathogen. In addition, a second set of samples corresponding to the alternate hypothesis A ("pathogen not present") is constructed from a variety of unspiked environmental samples. (It is assumed that most background samples contain fewer than 100 CFU per gram of target pathogen.)

Each sample in the "H" and "A" sets are split into duplicate sub-samples, each of which is subjected to one of the two tests. These data provide estimates of the joint probability of occurrence of positive and negative tests for the two assays over the population of environmental samples, as shown in Table 4. The astute reader will notice that in order to produce accurate estimates of the order of magnitude quoted for the "A" group, at least 100,000 samples would be necessary. Outside of a program like BioWatch, it would be difficult to perform this large a validation exercise on environmental samples in practice.[13] Nonetheless, the values provided are arguably consistent with the findings of the BioWatch program, and could also be representative *mutatis mutandis* of multicenter clinical assay validation programs for common diseases like HIV infection or tuberculosis.

Table 4. Estimated joint probabilities for test results for two assays, PCR-based and immunoassay-based (IMA).

Test set	Estimated joint probabilities (X = H or A)			
	$P(E_1E_2\|X)$	$P(\cancel{E_1}E_2\|X)$	$P(E_1\cancel{E_2}\|X)$	$P(\cancel{E_1}\cancel{E_2}\|X)$
	(PCR + IMA +)	(PCR – IMA +)	(PCR + IMA –)	(PCR – IMA –)
X = H	0.67	0.001	0.3	0.029
X = A	1.0×10^{-5}	1.0×10^{-4}	1.0×10^{-5}	0.9999

The joint probabilities in Table 4 imply the values given in Table 5 for the conditional probabilities associated with this pair of assays. Presumably when applied to environmental samples the two assays have some causes for false positives and negatives in common, and other causes that are unique to each assay. In general, the values in Table 5 seem plausible for both PCR and IMA in environmental samples.

First of all, note that the values in Table 5 imply that both assays are highly diagnostic for the presence of the pathogen. The diagnosticity of a positive PCR assay finding is $\frac{P(E_1|H)}{P(E_1|A)} = 4.9 \times 10^4$, and the diagnosticity of a positive IMA finding is $\frac{P(E_2|H)}{P(E_2|A)} = 6.1 \times 10^3$. If one were to assume that the two assays are independent, the diagnosticity of a positive finding for both assays together would be the product of these two values, $\approx 3 \times 10^8$ – very strong support for the hypothesis that the pathogen was present.

However, the true diagnosticity of a "double positive" is only $\frac{P(E_1E_2|H)}{P(E_1E_2|A)} = 6.7 \times 10^4$,

not much higher than that of a positive PCR result alone! Nonetheless, in the real world it is unlikely that the experiment leading to the data in Table 4 would be performed, and more likely that each assay would be characterized individually on independent sample sets. An exaggerated sense of the diagnosticity of combining the two assays might easily arise from a naïve assumption of independence.

Table 5. Conditional probabilities derived from the joint probability estimates given in Table 4.

Quantity	Value	Interpretation
$P(E_2\|E_1,H)$	0.6907	In nearly 70% of the cases when the pathogen is present at greater than 100 CFU/gm and the PCR assay is positive , the IMA test is positive as well.
$P(E_2\|E_1,A)$	0.5000	When the PCR assay gives a false positive indication, IMA does too, about half the time
$P(E_2\|\cancel{E}_1,H)$	0.03333	In three percent of the cases when the pathogen is present at greater than 100 CFU/gm but the PCR test is negative, the IMA assay is nevertheless positive.
$P(E_2\|\cancel{E}_1,A)$	1.000×10^{-4}	In one case out of 10000, IMA gives a positive indication even though the pathogen is absent and the PCR test is negative.
$P(E_2\|H)$	0.6710	When the pathogen is present at greater than 100 CFU/gm the IMA assay is positive about 67% of the time.
$P(E_2\|A)$	1.100×10^{-4}	When the pathogen is absent the IMA assay is positive for about 1 in 10,000 samples.
$P(E_1\|H)$	0.9700	When the pathogen is present at greater than 100 CFU/gm the PCR assay is positive 97% of the time.
$P(E_1\|A)$	2.000×10^{-5}	When the pathogen is absent, the PCR assay is (false) positive for only 2 out of 100,000 samples.

It is tempting to argue that the causes of false positive and negative results for the two types of assay discussed in this example are, in the real world, primarily associated with differences in the way the two assays are performed either by two individuals or in two different laboratories – i.e. false negatives are primarily caused by chance mistakes in labeling or reagent addition and false positives are primarily caused by contamination with positive control standards. In this case the sources of error could be expected to be independent. Even if this were plausible it would remain to be demonstrated empirically by performing validation experiments like the imaginary one posited for this example.

In some operational settings one assay is used as a "presumptive" test and another as a "confirming" test, i.e. a second test that is done when the first test gives a positive value, in order to provide more confidence in the answer. This often makes sense when one assay is cheap and easy, but not very accurate and the other is more costly or complex but more accurate. In this context conditional probabilities are the natural way to express how diagnostic one assay is, given a positive (or negative) result for the other. It is clear from the following table (Table 6) that a

positive PCR test performed after the IMA test will increase the diagnosticity by an order of magnitude or more regardless of the results of IMA. However, if the first test is PCR, a positive result renders the IMA superfluous. This would imply that the IMA is not a good choice as a "confirming detection assay" for the PCR test, but the PCR test would be useful as a confirming assay for IMA. A more complete analysis of choosing presumptive and confirming assays is beyond the scope of this report. However, it should be clear that the potential conditional dependence of the tests is a key issue.

Table 6. Conditional diagnosticity expressions and values for the notional PCR and IMA data in Table 5.

Question	Expression	Value	Diagnosticity of first test	Total diagnosticity
How diagnostic is a positive PCR finding when IMA is negative?	$\frac{P(E_1\|\not{E}_2H)}{P(E_1\|\not{E}_2A)}$	9.12 x 10^4	0.329	3.00 x 10^4
How diagnostic is a positive PCR finding when IMA is positive?	$\frac{P(E_1\|E_2H)}{P(E_1\|E_2A)}$	11.0	6.10 x 10^3	6.70 x 10^4
How diagnostic is a positive finding for IMA when PCR is negative?	$\frac{P(E_2\|\not{E}_1H)}{P(E_2\|\not{E}_1A)}$	333	0.030	10.0
How diagnostic is a positive finding for IMA when PCR is positive?	$\frac{P(E_2\|E_1H)}{P(E_2\|E_1A)}$	1.38	4.85 x 10^4	6.70 x 10^4

B. Conditional dependence in multilocus PCR assays

Multilocus PCR assays are a mainstay of microbial forensic detection.[14] Are multiple PCR loci independent pieces of evidence for the presence of a pathogen of interest? Based on the discussion in section 2, it will depend on whether the false positive and negative rates of the two separate assays arise from factors in common. In many laboratories performing PCR assays contamination and the presence of PCR inhibitors is a common problem that can influence assays for different signatures equally. If the two assays are run simultaneously in the same sample as a duplex, competitive effects may cause correlation between the positive or negative results for the two signatures. Finally, depending on the genomic location of the two

signatures, the presence or absence of the two in background microbes could well be correlated. Hence there are many plausible reasons to suspect that the diagnosticity of multilocus assays are influenced by conditional dependence.

A simple example serves to demonstrate how conditional dependence might lurk in otherwise unremarkable data on the performance of a dual signature assay. In direct analogy to the discussion of "orthogonal" assays we will assume that the data set, displayed in Table 7, has been generated from assay results on a set of "H" samples containing 100 or more copies per gram of target pathogen genomic DNA and "A" samples containing less than 100 copies per gram, and that 100 copies per gram lies near, but above the detection limit of the assay. The two signatures are denoted S1 and S2. Notice that, unlike the "orthogonal" assays example, the performance of both assays is assumed to be equal. All probabilities are symmetric with respect to exchange of E_1 and E_2.

Table 7. Estimated joint probabilities for test results for two PCR signatures S1 and S2.

Test set	Estimated joint probabilities (X = H or A)			
	$P(E_1E_2\|X)$	$P(\cancel{E}_1E_2\|X)$	$P(E_1\cancel{E}_2\|X)$	$P(\cancel{E}_1\cancel{E}_2\|X)$
	(S1+ S2+)	(S1− S2+)	(S1+ S2−)	(S1− S2−)
X = H	0.975	0.01	0.01	0.005
X = A	1.0×10^{-5}	1.0×10^{-4}	1.0×10^{-4}	0.9998

Table 8. Conditional probabilities derived from the joint probability estimates given in Table 7.

Quantity	Value	Interpretation
$P(E_2\|E_1,H)$	0.9898	In nearly 99% of the cases when the pathogen genome is present at greater than 100 copies/gm both signatures test positive.
$P(E_2\|E_1,A)$	9.09×10^{-2}	When the S1 assay gives a false positive indication, S2 does too, about 10% of the time
$P(E_2\|\cancel{E}_1,H)$	0.6667	In about 67% percent of the cases when the pathogen genome is present at greater than 100 copies/gm but the assay for S1 is negative, the assay for S2 is nevertheless positive.
$P(E_2\|\cancel{E}_1,A)$	1.000×10^{-4}	In one case out of 10000, S2 gives a positive indication even though the pathogen is absent and the PCR test for S1 is negative.
$P(E_2\|H)$	0.985	When the pathogen is present at greater than 100 copies/gm the S2 assay is positive about 99% of the time.
$P(E_2\|A)$	1.100×10^{-4}	When the pathogen is absent the S2 assay is positive for about 1 in 10,000 samples.
$P(E_1\|H)$	0.985	When the pathogen is present at greater than 100 copies/gm the S1 assay is positive about 99% of the time.
$P(E_1\|A)$	1.100×10^{-4}	When the pathogen is absent, the PCR assay is (false) positive for only 2 out of 100,000 samples.

The conditional probabilities for this scenario are displayed in Table 8. Each signature by itself has a diagnosticity of 8.95×10^3. If one assumes independence between the results for the two signatures, the diagnosticity value is a hefty 8.02×10^7. However, the actual diagnosticity value is only 9.75×10^4, almost a factor of 1000 lower.

Unlike the case of the "orthogonal" assays, data sets generated from multilocus PCR validation exercises are much more likely to be available for quantifying dependences. Currently the BioWatch data set represents a large but inaccessible resource for study. However, smaller publicly available data sets may be available and we may see some assessments of this question in the future.

C. Two reference laboratories, one case laboratory

Suppose that a government agency decides that it will support an analytical capability for (say) ricin detection in two laboratories at separate institutions, but during an investigation will only provide a sample to one of them for analysis. (This is the actual policy of some US government agencies.) To establish the performance of the methodology used by the two labs they request performance data from each lab and pool the results in order to establish the overall diagnosticity of the method. The reasoning is that by pooling the results they are increasing the statistical precision as well as averaging over potential differences in the way the analyses are performed in different labs. The agency specifies that the samples in each lab are prepared in the same way, and the measurements are done using the same SOP. The number of positive results observed in spiked and unspiked samples is shown in Table 9.

In this case the probability of observing a positive detection is clearly conditional on which laboratory is doing the testing. In the case shown, the diagnosticity of the test implied by the pooled data is 127. However, if the agency quotes this value while sending a case sample to only one of the laboratories, it is misleading, because the diagnosticity of the test performed in each laboratory alone is actually much lower – in one case by a factor of two.

Table 9. True and false positive detections (positives/number tested) H = spiked samples, A = unspiked negative controls.

	Lab 1	Lab 2	Pooled
H	498/503	29/51	527/554
A	1/63	3/473	4/536
Diagnosticity expression	$\frac{P(E_1\|L_1H)}{P(E_1\|L_1A)}$	$\frac{P(E_1\|L_1H)}{P(E_1\|L_1A)}$	$\frac{P(E_1\|H)}{P(E_1\|A)}$
Diagnosticity value	62	89	127

This is an example of "Simpson's paradox", a well-known phenomenon that can occur when the diagnosticity of a test on a pooled population is assumed to apply to a subpopulation, and there is an unrecognized or ignored conditional dependence of the test results on the identity of the subpopulation.[15]

5. Conditional dependence relations complicate rigorous validation

Pearl has suggested that it is possible to organize knowledge in domains of expertise to promote the accuracy of independence assumptions.[10] However, it is not clear that the test developer is actually free to use this strategy in regard to the scientific tests devised for forensic applications. This means that when the evidence of more than one test is combined in pursuit of stronger diagnosticity, dependencies must be determined empirically. Unfortunately, this may not always be practical. Gustafson has discussed the practical difficulties involved with extracting estimates of conditional dependence from empirical data when even a modest number of variables are involved:

> "Assume you are attempting to determine whether three symptoms S1, S2, and S3 are independent given that the patient has a specific diagnosis D. A chi square test of conditional independence checks only for pairwise independence; more complex interactions that involve S1, S2, and S3 may exist even though S1 and S2 are pairwise independent given D. Even if higher-order interactions were ignored, the data collection and analysis necessary for testing two dimensional conditional independence of a large number of symptoms would be prohibitive." (Reference 1, page 63)

In the domain of microbial forensics, nowhere does this problem loom larger than in proposals to use Bayesian network-based inference systems to extract detailed growth process information from multiple types of chemical and physical analyses on biological agents. For example, consider a 2008 study by Jarman et. al., which suggested the construction of a Bayesian network for combining various kinds of mass spectrometric data in order to infer the culture medium of *B. anthracis* samples.[16] Two such data sources are isotope ratio mass spectrometry (IRMS) and Electrospray Ionization mass spectrometry (ESI-MS). IRMS determines the ratios of various isotopic forms of the elements carbon and nitrogen in a sample, while ESI-MS detects the presence of a mass fragment characteristic of agar. In constructing the inference network the authors assert:

> "First, we make the *standard assumptions* that any two child nodes [the probabilities of observing peak intensity values for ESI-MS and

> IRMS respectively] are conditionally independent of one another, given that we know the state of the parent node [representing the probability that agar has or has not been added]. This means, for example, that the ESI-MS peak intensities and the IRMS peak intensities are orthogonal (and statistically independent) of one another if we know whether agar was added to the culture medium." [Reference 16, page 3575; Emphasis added.]

To assess the possibility that this assumption may not hold, suppose that E_1 is the finding that the ESI peak intensity for agar is greater than or equal to a certain cutoff value, and $\not{E}_1$ is the finding that the intensity is less than the cutoff. Similarly let E_2 and $\not{E}_2$ be the findings that the IRMS peak intensities lie within and outside a stated domain, respectively. Let H be the state of the parent node in question when agar is added and A its state otherwise. Independence requires both that $P(E_1|E_2H) = P(E_1|\not{E}_2H)$ and $P(E_1|E_2A) = P(E_1|\not{E}_2A)$. Imagine we draw samples from the "H" population – all possible samples made using agar. This population is sure to contain samples that have been washed with organic solvents that both change E_2 to $\not{E}_2$ and in doing so also alter the probability of observing E_1. Among "A" samples, where agar has not been added, it is plausible that other organic additives that alter the C/N isotope ratios might also act as interferences in the ESI-MS spectrum, altering the probability of observing E_1. Clearly an inferential model that invokes independence in this way requires empirical validation.

A minimum experiment would require creation of reference samples that are *simultaneously* subjected to MS-ESI and IRMS measurements so that joint probabilities of observing the states of E_1 and E_2 can be estimated. When these are not defined to be dichotomous but have more possible states, the number of samples required to produce reasonable joint probability estimates increases rapidly. Validating conditional independence assumptions in more complex Bayesian networks begins to look like a daunting task.

6. Concluding remarks

It is important to appreciate that the examples given above are not "concocted" to exhibit strong conditional dependence phenomena. Instead, just a little playing around with numbers will convince the reader that one has to do something special to find *independence* in a data set. In fact, the only reliable way to produce a set of joint probabilities that exhibit conditional independence is to assume it outright – i.e. choose values for the $P(E_j|X)$ and then multiply them together. Thus it appears to be much more likely that real data sets from actual tests and real items of evidence will generally exhibit conditionally dependence to some degree. For those who develop or utilize multiple lines of evidence in forensic investigations this shifts the burden of proof considerably, since the standard approach has been to naively assume independence when dealing with different tests or different types of evidence.

The assumption of independence is a heuristic – and it is entirely possible that in "real life" such assumptions, while strictly incorrect, also do not generally matter because they seldom invalidate a basic conclusion. It has long been recognized that heuristics do surprisingly well in some cases simply because they are "ecologically rational", i.e reflect some actual properties of the real world.[17] One of those properties might be that in a complex world independence assumptions are usually "close enough" to the truth. On the other hand, there are well-documented legal cases where the naive assumption of statistical independence has lead to convictions later judged to be invalid. (Joseph B. Kadane, 2008)

The treatment in this report was deliberately kept simple in order to make it more suitable as a primer and thought-stimulator. Of course, much more complex problems arise when more than two hypotheses and variables with more than two states are involved. With regard for Pearl's observation about "knowledge organization", one of the modes of organization that the scientist can systematically employ is to develop assays and perform validation studies that are dichotomous by definition and design. Thus, "orthogonal" assays that ask "was this sample made using agar or not?" are much simpler to validate than those that ask "how was this sample made?"

One of the explicitly stated goals of the national research program in microbial forensics is to "[develop] orthogonal methods for conducting forensic comparisons between samples…."[19] In light of the discussion in this report one might conclude that rigorous investigation of conditional dependence through empirical studies ought to be an intrinsic feature of forensic assay development.

References

1. David H. Gustafson, Ward Edwards, Lawrence D. Phillips, and Warner V. Slack, "Subjective Probabilities in Medical Diagnosis", IEEE Transactions on Man-Machine Systems, Vol. MMS-10, NO. 3, September 1969.

2. Dennis G. Fryback, "Bayes' Theorem and Conditional Non-independence of Data in Medical Diagnosis", Computers and Biomedical Research 1978; **11**: 423-434.

3. Pamela M. Vacek, "The Effect of Conditional Dependence on the Evaluation of Diagnostic Tests", Biometrics 1985; **41**: 959-968.

4. Ian A. Gardner, Henrik Stryhn, Peter Lind, and Michael T. Collins, "Conditional dependence between tests affects the diagnosis and surveillance of animal diseases", Preventive Veterinary Medicine **45** (2000).

5. Richard O. Lempert, "Modeling Relevance", Michigan Law Review **75**(56), 1977, 1021-1057.

6. Franco Taroni, Colin Aitken, Paolo Garbolino, and Alex Biedermann, Bayesian Networks and Probabilistic Inference in Forensic Science", (John Wiley & Sons, West Sussex, England, 2006). Chapter 7: *Aspects of the Combination of Evidence*.

7. Steve Velsko and William Thompson, "Modeling Context Bias", Lawrence Livermore National Laboratory Technical Report, LLNL-TR-644404, September 30, 2013

8. A useful text that covers these topics is: Joseph B. Kadane, Principles of Uncertainty (Chapman & Hall/CRC Texts in Statistical Science, 2011). Available as a free pdf file at: http://lesswrong.com/lw/5v0/free_stats_textbook_principles_of_uncertainty/
9. Shawn P. Curley, J. Frank Yates, and Mark J. Young, "Seeking and Applying Diagnostic Information in a Health Care Setting", Acta Psychologia 1990; **73**: 211-223.

10. Judea Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, (Morgan Kaufmann Publishers, Inc., San Mateo, California, 1988).

11. Margaret S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction* (Oxford Statistical Science Series 28; Oxford University Press, New York, 2003)

12. Stanley Gudder, "Do Good Hands Attract?", Mathematics Magazine 1981; **54**(1): 13-16.

13. Sarah A. Lister and Dana A. Shea, "The BioWatch Program: Detection of Bioterrorism", Congressional Research Service Report RL 32152, November 19, 2003.

14. John J. Maurer, "Rapid Detection and Limitations of Molecular Techniques", Annual Review of Food Science and Technology 2011; **2**: 259-279.

15. Yao Hua Ooi, "Simpson's Paradox – A Survey of Past, Present and Future Research", Wharton Research Scholars Journal, Wharton School of Business, University of Pennsylvania, 2004; Posted at ScholarlyCommons. http://repository.upenn.edu/wharton_research_scholars/15

16. Kristin H. Jarman, Helen W. Kreuzer-Martin, David S. Wunschel, Nancy B. Valentine, John B. Cliff, Catherine E. Petersen, Heather A. Colburn, and Karen L. Wahl, "Bayesian-Integrated Microbial Forensics", Applied and Environmental Microbiology 2008; **74**(11): 3573–3582

17. Gerd Gigerenzer, Peter M. Todd, and the ABC research Group, Simple Heuristics that Make Us Smart, (Oxford University Press, 1999)

18. Joseph B. Kadane, Statistics in the Law, (Oxford University Press, 2008) Part 5 B: "Criminal Convictions and Statistical Independence".

19. Office of Science and Technology Policy: *National Microbial Forensics Research Strategy to Support Biological Attack Attribution Investigations and National Security*, Implementation Plan, 2010.

Appendix: Modeling conditional dependence.

In this section we consider a simple, few-parameter model for conditional dependence that can reproduce the two intuitive limiting cases for the likelihood ratio $L_{2|1} = \frac{P(E_2|E_1H)}{P(E_2|E_1A)}$. In the first limit, the items of evidence E_1 and E_2 are conditionally *independent*, and

$$L_{2|1} = L_2 = \frac{P(E_2|H)}{P(E_2|A)}. \tag{1a}$$

In the other limit, E_2 is completely and exclusively *dependent* on E_1 and

$$L_{2|1} = \frac{P(E_2|E_1)}{P(E_2|E_1)} = 1. \tag{1b}$$

Conditional dependence can be modeled using approaches developed to describe the interaction between non-independent medical tests. For two items of evidence E_1 and E_2, a hypothesis H, and its alternative A (A = Not-H), we write:

$$P(E_1E_2|H) = P(E_2|E_1H)P(E_1|H) = P(E_2|H)P(E_1|H) + \gamma_1 \tag{2a}$$

$$P(E_1\not{E}_2|H) = P(E_2|\not{E}_1H)P(\not{E}_1|H) = P(E_2|H)P(\not{E}_1|H) - \gamma_1 \tag{2b}$$

$$P(E_1E_2|A) = P(E_2|E_1A)P(E_1|A) = P(E_2|A)P(E_1|A) + \gamma_2 \tag{3a}$$

$$P(E_1\not{E}_2|A) = P(E_2|\not{E}_1A)P(\not{E}_1|A) = P(E_2|A)P(\not{E}_1|A) - \gamma_2 \tag{3b}$$

When E_1 and E_2 are the results of "binary" tests, i.e. tests that have only two possible results ("positive and "negative"), only the two parameters γ_1 and γ_2 are necessary to describe the degree of conditional dependence of the tests.

The values of the parameters γ_1 and γ_2 are constrained by the relations

$$0 \le P(E_1|E_2H) \le 1 \tag{4a}$$

and

$$0 \le P(E_1|E_2A) \le 1 \tag{4b}$$

and by the identity:

$$P(E_2|E_1) = P(E_2|E_1H)\, P(H|E_1) + P(E_2|E_1A)P(A|E_1). \tag{5}$$

Using equation (5) it is simple to show that $P(E_2|E_1H) = P(E_2|E_1A)$ if and only if $P(E_2|E_1H) = P(E_2|E_1)$. Moreover,

$$\text{if } P(E_2|E_1H) < P(E_2|E_1) \text{ then } P(E_2|E_1A) > P(E_2|E_1) \tag{6a}$$

and

$$\text{if } P(E_2|E_1H) > P(E_2|E_1) \text{ then } P(E_2|E_1A) < P(E_2|E_1). \tag{6b}$$

These relations are sometimes referred to as expressing "attraction" and "repulsion" between states of the two evidence items in the sense that if E_1 makes E_2 more probable under H (E_1 "attracts" E_2) then it makes it less probable under A (E_1 "repels" E_2), and vice versa.

Thus, either

$$P(E_2|E_1A) \leq P(E_2|E_1) \leq P(E_2|E_1H), \tag{7a}$$

or

$$P(E_2|E_1H) \leq P(E_2|E_1) \leq P(E_2|E_1A), \tag{7b}$$

and it follows that the likelihood ratio $L_{2|1} = \frac{P(E_2|E_1H)}{P(E_2|E_1A)}$ is greater than 1 when E_1 "attracts" E_2 under H and less than 1 when E_1 "attracts" E_2 under A. Note that these relations depend on the particular states of the evidence variables E_1 and E_2 and can change as these states change.

The likelihood ratio $L_{2/1}$ approaches 1 when

$$\gamma_1 = [P(E_2|E_1) - P(E_2|H)]P(E_1|H) \tag{8a}$$

and

$$\gamma_2 = [P(E_2|E_1) - P(E_2|A)]P(E_1|A). \tag{8b}$$

Equation (8b) can be derived from (8a) using the constraining relation (5).

$L_{2/1}$ is equal to $\frac{P(E_2|H)}{P(E_2|A)}$ when $\gamma_1 = \gamma_2 = 0$, however equation (5) implies that this cannot occur unless

$$P(E_2|E_1) = P(E_2|H)P(H|E_1) + P(E_2|A)P(A|E_1) \tag{9}$$

If we define a third parameter λ by

$$\lambda = P(E_2|E_1) - P(E_2|H)P(H|E_1) - P(E_2|A)P(A|E_1) \quad \text{where } -1 \leq \lambda \leq 1 \tag{10}$$

we can relate the three parameters that govern the degree of conditional dependence by:

$$\lambda P(E_1) = \gamma_1 P(H) + \gamma_2 P(A) \tag{11}$$

Thus, conditional *independence* is true only under much more stringent conditions than are required for diminished relevance caused by conditional dependence. In particular, while the fact that $P(E_1|E_2H) = P(E_1|E_2)$ is sufficient to ensure that $P(E_1|E_2A) = P(E_1|E_2)$, the fact that $P(E_2|E_1H) = P(E_2|H)$ does not imply $P(E_2|E_1A) = P(E_2|A)$.

The model can be used to explore the transition of the likelihood ratio to the limits expressed by equations (1a) and (1b) in the following way:

(a) Assume values for $P(E_1|H)$, $P(E_1|A)$, $P(E_2|H)$ and $P(E_2|A)$, and $P(H)$; Note that $P(A) = 1 - P(H)$ and $P(E_1) = P(E_1|H)P(H) + P(E_1|A)P(A)$.

(b) Make γ_1 a free parameter, choose values from the range:

$$\gamma_{1max} = \mathrm{Max}(q_1, s_1) \le \gamma_1 \le \gamma_{1min} = \mathrm{Min}(r_1, t_1), \text{ where:}$$

$q_1 = -\, P(E_2|H)\, P(E_1|H)$

$r_1 = P(E_2|H)\, [1 - P(E_1|H)]$

$s_1 = -\, [1 - P(E_2|H)]\, [1 - P(E_1|H)]$

$t_1 = [1 - P(E_2|H)]\, P(E_1|H)$

(c) Choose values of γ_2 from the range $\gamma_{2min} \le \gamma_2 \le \gamma_{2max}$, where

$\gamma_{2min} = \mathrm{Max}(q_2, s_2)$, $\gamma_{2max} = \mathrm{Min}(r_2, t_2)$, and

$q_2 = -\, P(E_2|A)\, P(E_1|A)$

$r_2 = P(E_2|A)\, [1 - P(E_1|A)]$

$s_2 = -\, [1 - P(E_2|A)]\, [1 - P(E_1|A)]$

$t_2 = [1 - P(E_2|A)]\, P(E_1|A)$

The choices of γ_1 and γ_2 determine λ through equation (11). (Note that choosing λ is equivalent to choosing a value for $P(E_1|E_2)$.)

(d) Use equations (2) and (3) to generate the likelihood ratio $L = \frac{P(E_1E_2|H)}{P(E_1E_2|A)} = \frac{P(E_1|H)P(E_2|H) + \gamma_1}{P(E_1|A)P(E_2|A) + \gamma_2}$.

(e) If the likelihood ratio $L_{2|1}$ (i.e. the likelihood ratio for E_2 conditioned on E_1) is desired, one can use the equation:

$$\frac{P(E_2|E_1H)}{P(E_2|E_1A)} = \frac{P(E_1E_2|H)P(E_2|A)}{P(E_1E_2|A)P(E_2|H)} \tag{12}$$

The parameters γ_1 and γ_2 can be chosen to change the "balance" in the top and bottom rows of the matrix of Table 1 (reproduced below for convenience.)

Table A1. Matrix of related conditional probabilities.

	E_1	∅[a]	$\not{E}_1$
H	$P(E_2/E_1,H)$	$P(E_2/H)$	$P(E_2/\not{E}_1,H)$
∅[b]	$P(E_2/E_1)$	$P(E_2)$	$P(E_2/\not{E}_1)$
A	$P(E_2/E_1,A)$	$P(E_2/A)$	$P(E_2/\not{E}_1,A)$

[a]Independent of E_1 and $\not{E}_1$.
[b]Independent of H and A.

For example, suppose that when $\gamma_1 = \gamma_{1min}$ $P(E_2/E_1,H) > P(E_2/\not{E}_1,H)$. As γ_1 increases, it eventually crosses zero whereupon, $P(E_2/E_1,H) = P(E_2/H) = P(E_2/\not{E}_1,H)$. As it continues to increase from zero towards γ_{1max}, $P(E_2/E_1,H) < P(E_2/\not{E}_1,H)$. The value of γ_2 has a similar effect on the bottom row.

Under certain conditions it is possible to find combinations of γ_1 and γ_2 that "balance" the first or third columns. Let γ_1^* be the value of γ_1 that makes $P(E_2/E_1,H) = P(E_2/E_1,A)$ and γ_2^* be the value of γ_2 that makes $P(E_2/\not{E}_1,H) = P(E_2/\not{E}_1,A)$. Then, using equations (2a,b) and (3a,b) we obtain:

$$\gamma_1^* = \frac{[P(E_2|H) - P(E_2|A)]P(E_1|H)P(\not{E}_1|H)}{[P(E_1|H)P(\not{E}_1|A) - P(E_1|A)P(\not{E}_1|H)}$$

and

$$\gamma_2^* = \frac{[P(E_2|H) - P(E_2|A)]P(E_1|A)P(\not{E}_1|A)}{[P(E_1|H)P(\not{E}_1|A) - P(E_1|A)P(\not{E}_1|H)}$$

Clearly, for these conditions to result in L =1 we must have

$$\gamma_{1min} \leq \gamma_1^* \leq \gamma_{1max}$$

and

$$\gamma_{2min} \leq \gamma_2^* \leq \gamma_{2max}$$

The parameters corresponding to the PCR-bleach example in Table 3, section 3 and the other examples in section 4 are provided below. They are given to 4 significant figures, but the reader is cautioned that some of the conditional probabilities are very sensitive to the precise value of these parameters. Some adjustment beyond the 4th decimal place may be necessary to reproduce the cited probabilities exactly.

Table A2. Model parameters corresponding to the example scenarios.

Scenario	Section	γ_1	γ_2
PCR - Bleach	3	-0.2350	-4.889×10^{-10}
Two orthogonal assays	4A	0.01914	9.998×10^{-6}
Two PCR signatures	4B	0.004775	9.992×10^{-6}
Two laboratories	4C	0.06939	0.01646