Advances in Pharmacoepidemiology & Drug Safety

Robson, Adv Pharmacoepidem Drug Safety 2012, 1:6
DOI: 10.4172/2167-1052.1000123

Review Article Open Access

Towards New Tools for Pharmacoepidemiology

Barry Robson*

Professor of Epidemiology, Biostatistics and Evidence Based Medicine, St. Matthew's University School of Medicine, Distinguished Scientist, Mathematics and Computer Science, University of Wisconsin-Stout, Wisconsin, Chief Scientific Officer, Quantal Semantics Inc., North Carolina, and Chair, The Dirac Foundation, Oxfordshire, UK

Abstract

Here the data analysis challenges for pharmacosurveillance and pharmacogenomics are reviewed. Four interrelated tools to tackle the ubiquitous problems of high dimensionality and sparsity are outlined: the contribution of negative or contrary evidence in the four pillars of evidence established in the early days of epidemiology, information and decision theory, the zeta function, and hyperbolic-complex algebra. These four tools are described in a fairly integrated way, and this order basically reflects the degree of novelty and degree of acceptance in biomedicine, the most recent and controversial coming last. The zeta function is essentially an estimate in information theory, an extension to express the expected information in a system, the amount of available to the observer via the data, and in one guise has been in use in bioinformatics since the early 1970s. Hyperbolic-complex algebra is concerned with encoding information in two directions of conditionality, of potential importance in inference about etiology, considerations derivable mathematically from the zeta function. It takes on importance when many zeta function terms as estimates of information terms are used in an inference network. Its usefulness remains to be seen, although it essentially represents the method of inference due to Dirac and already established in quantum field and particle theory. Including negative evidence in inference using estimates based on multiple factors, however, requires that we are careful in interpretation.

Keywords: Hyperbolic-complex algebra; Clinical data; Pharmacosurveillance; Pharmacogenomics

Introduction

Benefits and challenges of pharmacosurveillance

Monitoring the outcomes for patient health of marketed drugs is an important tool to raise alerts and analyze adverse drug reactions. It also has multiple quantitative purposes related to that. It can establish the pharmacogenomics and relevant environmental factors to stratify the population into those who will be benefited and those who are at risk, which is important in repurposing and pricing drugs. In many ways it is an ongoing clinical trial on a vast scale, facilitated by the rise of the digital patient record to capture the essential data. Those and the following considerations have been reviewed by the author elsewhere [1]. Without doubt, the primary challenge for the data analyst *per se* in pharmacogenomics and pharmacosurveillance is the multiplicity of factors involved in expressing disease and adverse reactions, and that even govern diagnosis and how therapies are prescribed and used, and in drawing actionable inference from it. The multifactorial aspect shows up well in the epidemiological-geographical aspects. Whilst sickle cell anemia is due to a single mutation and originally followed the geographical distribution of malaria very well (since it conferred protection against it), the distribution of mutations of the serotonin transporter gene show little relation to the distribution of anxiety and depression in which anomalies of serotonin are strongly implicated, because anxiety and depression are polygenic diseases and strongly influenced by environment. Even if it did, prescription of appropriate drugs and their outcomes would not be good markers since the distribution of diagnosis of anxiety and depression is confounded by psychological and social factors and shows rather poor correlation with prescription even in highly industrialized welfare states.

The scope of the multifactorial problem

Whilst admittedly behavioral disorders are extreme examples in that the brain is an exceptionally complex organ with correspondingly complex genetics, pathologies, and therapies, other complex diseases such as cardiovascular diseases and cancers have easily some n=100 relevant causative factors. These cannot completely be broken down into $n(n-1)/2 \approx 5000$ pair-wise independent and in some way additive contributions, but can theoretically represent up to some 2^n potential quantitative parameters or rules to be discovered and combined for inference, prediction, and decision making [2,3]. For n=100, 2^n is an astronomic 10^{30}. The involvement of multiple factors with combinatorial considerations on such scales means that the consideration of the data must be very high-dimensional, and this high dimensionality is the dragon, essentially entropy, that guards the gold that the data otherwise promises [2,3]. Whilst classical statistics massively reduces dimensionality by focus on specific data features and testing hypotheses that inevitably represent some prior knowledge of what the answer is likely to be [2,3], data mining deliberately does not allow itself such luxuries, and focuses on finding what is not expected. Consequently, the challenge rises explosively with the dimensionality that needs to be considered, and special techniques have to be developed to handle the higher dimensional cases [4-8].

The sparsity problem

In a curious irony, however, the very high dimensionality of the data removes much of the problem from our hands. We do not have enough computer power to handle this kind of big data and the combinatorial explosion of its factors. However, we do not have enough data to run

***Corresponding author:** Barry Robson, Professor of Epidemiology, Biostatistics and Evidence Based Medicine, St. Matthew's University School of Medicine, Distinguished Scientist, Mathematics and Computer Science, University of Wisconsin-Stout, Wisconsin, Chief Scientific Officer, Quantal Semantics Inc., North Carolina, and Chair, The Dirac Foundation, Oxfordshire, UK; E-mail: brobson@smu.ky

Received November 26, 2012; **Accepted** December 14, 2012; **Published** December 17, 2012

Citation: Robson B (2012) Towards New Tools for Pharmacoepidemiology. Adv Pharmacoepidem Drug Safety 1:123. doi:10.4172/2167-1052.1000123

Copyright: © 2012 Robson B. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

very far into that barrier. In data mining 667,000 medical records, there was simply no data for parameters or rules with more than some 5-10 factors [8], depending of course on the individual abundances of the factors. Even data mining some 6.7 million records (albeit in this case relating to chemical patents) does not greatly alleviate that problem [9]. There is no benefit in the data analyst anguishing over what he or she does not yet have. On the other hand, there is a problem at the ever-growing boundary between plentiful data and none. The combinatorial character of the problem means that the great majority of occurrences that are seen involve just some one or two observations. These can only be ignored at the diligent data analyst's peril. They contain information, and though it is individually weak information, many such contributions can in principle add up to outweigh a prediction or decision made on the basis of parameters or rules with few factors and many observed occurrences. This, along with the matter of rules with zero occurrences, may be described as the *sparsity problem*. It is addressed below, but specific number-theoretic tools, providing algorithmic tools for coping with it, are described elsewhere [6].

The four pillars of evidence

Strangely enough, inference from the quantitative rules obtained by data mining often neglects evidence for which data is typically abundant, and which is fundamental in a court of law: the evidence that something is not the case. Whilst it is true that probability P(A) itself implies 1−P(~A) where '~' can be read as 'not', and as indicating the commentary state to A, specifically neglecting probabilities involving ~A can dramatically change things, especially in necessarily making approximation and assumptions due to data sparsity and when ~X itself is supported by a chain of probabilistic reasoning involving many factors. Pharmacoepidemiology is the study of the use of and the effects of drugs in large number of people [10,11], rooted in the epidemiology that did recognize this principle from the outset. In the 1830s, John Snow considered himself as the leading expert on the biological effects of gasses, and used his pioneering experience in the quantification of the anesthetic effects of ether and chloroform, and careful use of maps and statistics, to disprove the miasma or "bad gas" theory as a basis for cholera. However, the importance of his contributions to principles of "negative evidence" runs deeper than that. If it be historically argued that there were degrees of awareness of the following considerations before now, it certainly appears to be the first time on record that such considerations (and statistics expressed on maps) swayed skeptical government authorities. Snow traced the origins of cholera to the water supply, in that case a specific pump in Broad Street, Soho, London [12] and because his partner in investigation, the Reverend Whitehead, objected that he had drunk from the Broad Street pump and did not get cholera, and evidently many who did not live close to the pump did, the issue soon became one of establishing four kinds of evidence. These were the number that actually drank from the pump and got cholera, and the number that actually drank from the pump and did not got cholera, and no less importantly the number that did not drink from the pump and got cholera, and the number that did not drink from the pump and did not get cholera. These numbers appeared to differ significantly from what would be expected on a chance basis, though quantifying that notion of significance would have had to await the development of the chi-squared test. This thinking underlies our perception of the four pillars of probabilistic evidence, P(A, B), P(A, ~B), P(~A, B), and P(~A, ~B) respectively. It is of course from these that we now obtain multiple measures: odds as likelihood ratios (such as relative risk), predictive odds, odds ratios, and absolute risk reduction, and Number Need to Treat and Number Needed to Harm. They also underlie the measures of predictive power of a theory, a diagnostic or statistical test, or a decision support system (accuracy, sensitivity, specificity, LR+, LR-). The notion of tradeoff between aspects of performance highlighted by these measures, such as sensitivity and specificity, lead in the Second World War to the idea of "tuning" to optimal performance, e.g. the use of the ROC curve [13] initially to improve the ability of radar to distinguish British and German aircraft.

Issues addressed here

Despite the widespread use of the above as measures, we do not typically see inference networks of any degree of elaborateness that specifically include the kind of information (relating to the four pillars of evidence) that these measures contain. Here some basic principles are outlined that address that neglect, including the impact of the sparsity problem. In addition, there is the problem that in combining all the four pillars of evidence in a classical way, information about direction of conditionality, important in inference about potentials causes or *etiology*, is lost. This is also addressed. The present report is concerned with highlighting some fairly new and relevant tools in introductory overview, tools that should help fix these issues. Detailed application will be reported elsewhere.

Theory

Basic principles: information theory

For completeness and to establish the current notation, we may start on familiar territory. Though the word "information" is often used rather qualitatively in the above context, it is well known that it is quantifiable from the probabilities by counting in the classical ("frequentist") way, providing the amount of data is large. For example, the probability of an adverse drug reaction A conditional upon being administered drug B is P(A|B)=n(A, B) / n(B) where n(A, B) is the number of joint occurrences A and B and n(B) is the number of occurrence of B with or without A. Then I(A|B) = −logeP(A|B). Typically, however, it is useful to start as much as possible from Fano mutual information I(A; B), which is the logarithm of an association constant K(A; B)=P(A, B)/P(A)P(B), and has the following useful definitions and equalities.

$$\begin{aligned} I(A;B) &= \log_e P(A,B) - \log_e P(A) - \log_e P(B) \\ &= \log_e P(A|B) - \log_e P(A) \\ &= \log_e P(B|A) - \log_e P(B) \\ &= I(A) + I(B) - I(A,B) \end{aligned}$$

An example measure derived from mutual information terms is as follows. It includes the evidence I(~A; B) against A as information which, like I(A; B), may be positive or negative, but which is subtracted in either case. It has the following useful definitions and equalities.

$$\begin{aligned} I(A:\sim A;B) &= I(A;B) - I(\sim A;B) \\ &= \log_e P(A,B) - \log_e P(\sim A,B) - \log_e P(A) + \log_e P(\sim A) \\ &= \log_e P(B|A) - \log_e P(B|\sim A) \\ &= I(B|\sim A) - I(B|A) \\ &= I(A:\sim A,B) - I(A:\sim A) \end{aligned} \tag{1}$$

Nonetheless, this is not, at first, an obvious choice of starting point. It is a log likelihood ratio, which is frequently applied as a relative risk. We might therefore perhaps have expected to see the log of P(A|B)/P(A|~B) with, say, A as a new condition of ill health and B and ~B as administration of the drug and not, respectively. In addition, the second last equality I(B |~A)−I(B | A) looks more like a prediction of the etiology or cause B of A, given A and ~A. The point is that this measures abstracts the information I(A: ~A)=logeP(A)+logeP(~A). Restoring the abstracted information, possibly as a new value, puts

I(A: ~A) in the role of prior information, in which case I(A: ~A, B)=I(A: B) – I(~A | B) plays the role of posterior information. Hence the exponential of the above reflects use of Bayes equation, posterior-probability=likelihood×prior-probability, extended to include the contrary evidence from ~A. In this kind of usage, Equation 1 demonstrably has predictive power in clear-cut test cases where we have plentiful well defined data for input and outcomes. It was essentially the basis of the GOR (Garnier-Osguthorpe-Robson) method [14] based on a theory of expected information first developed by the present author [15-16], widely used in protein bioinformatics and more recently appearing with applications in clinical informatics [8]. The method [15,16] is now seen as Bayesian [14], and represented an early use and possibly first use in molecular biology and biophysics when Bayesian methods were unpopular, then being considered subjective. The real strength as illustrated in the publications [6,7] is that it is easy to add in further evidence from different sources within the information-theoretic formalism. A very simple example is

$$\begin{aligned} I(A:\sim A;\{X\}) &= \sum_{i=1,2,3,\ldots}[I(A:\sim A;X_i)-I(A:\sim A)] \\ &= \sum_{i=1,2,3,\ldots}[\log_e P(A,X_i)-\log_e P(\sim A,X_i)-\log_e P(A)+\log_e P(\sim A) \end{aligned} \quad (2)$$

Strictly speaking, the summed terms require an added decision constant to be optimized for a complicated system [14-16]. It is a first order estimate of the accumulative evidence from various contributions {X} of which B was just one contribution. I(A: ~A) as prior information was replaced by an optimized decision constant in the bioinformatics application. The following is an example of an exact expansion.

$$I(A:\sim A;\{X\}) = I(A:\sim A;B)+I(A:\sim A;C|B)+I(A:\sim A,D|B,C)\ldots\ldots-I(A:\sim A) \quad (3)$$

It can be rewritten in many ways, e.g. starting with I(A: ~A; D), and with I(A: ~A; F|D) as the second term. We can actually include all possible terms for which we have data, which did not explicitly appear in some of these expansions. This is done by adding up all the possible expansions, and generating an average expansion in which each term I is weighted down by a simple combinatorial factor, or more generally how many times it repeats in all the terms available. It must suffice to say that the final effect is essentially the same idea as that used conceptually in Bayes Nets, prior to dismissing and simplifying terms to make approximations. Equation 2 is really an example of such an approximation in log form and extended to include contrary evidence.

Even just considering A and B, i.e. I(A: ~A; B) this does not include all the information that is for and against A. It uses only two of the four pillars of evidence. P(A, ~B) like P(~A, B) is also contrary evidence. The "double negative" P(~A, ~B) is evidence in favor of A. Possibly counterintuitive, the latter is hinted in the semantic equivalence of "All A are B" and "All non-B are non-A" by the logical law of the contrapositive, and relate to P(B | A) and P(~A | ~B) respectively. These two statements are guaranteed to be equivalent in probability, however, only when the probabilities are 1, i.e. under certainty, and a simple "proof" is that otherwise P(~A, ~B) would not be needed and would represent redundant information in an odds ratio. The log odds ratio is

$$\begin{aligned} I(A:\sim A;B:\sim B) &= I(A:\sim A,B)-I(A:\sim A,\sim B) \\ &= I(A,B)-I(\sim A,B)-I(A,\sim B)+I(\sim A,\sim B) \\ &= \log_e P(A,B)-\log_e P(\sim A,B)-\log_e P(A,\sim B)+\log_e P(\sim A,\sim B) \end{aligned} \quad (4)$$

Some inspection will show that the expansion and act of combining evidence is more difficult here, because the negation ~B then relates to the joint probability of all the states combined, ~(B, C, D,…). However, this is usually immaterial. That is because, if we have a lot of alternatives to B to introduce in combining evidence, noting that P(A, ~B)=P(A)–P(A, B) and P(~A, ~B)=P(~A) – P(~A, B), then P(A, ~B) ≈ P(A) and P(~A, ~B) ≈ P(~A). This therefore reduces to the log likelihood ratio of Equation 1.

$$\begin{aligned} I(A:\sim A;B:\sim B) &= I(A:\sim A,B)-I(A:\sim A,\sim B) \\ &= I(A,B)-I(\sim A,B)-I(A,\sim B)+I(\sim A,\sim B) \\ &\circledR\ \log_e P(A,B)-\log_e P(\sim A,B)-\log_e P(A)+\log_e P(\sim A) = I(A:\sim A;B) \end{aligned} \quad (5)$$

Although symbol® was originally a typesetting error, the author decided to retain it! It can be read as "reaches", and the problem otherwise is that symbols for "approximates" or "converge to" do not quite capture the full story! I(A : ~A; B : ~B) approximates and converges to the log probability expression calculated classically from numbers of observations when data becomes indefinitely large, and when B is really a complex event such as (B, C, D,…) with increasingly many factors B, C, D,…. Plus, (for full understanding of the present meaning) it more smoothly converges when information I is redefined in terms of zeta functions as described below, all the above being the case at the same time. There is symmetry, however, because a priori, in looking in both directions of conditionality, and making no mathematical distinction between A and B, a similar argument to the above could apply to ~A.

$$\begin{aligned} I(A:\sim A;B:\sim B) &= I(A:\sim A,B)-I(A:\sim A,\sim B) \\ &= I(A,B)-I(\sim A,B)-I(A,\sim B)+I(\sim A,\sim B) \\ &\circledR \log_e P(A,B)-\log_e P(A,\sim B)-\log_e P(B)+\log_e P(\sim B) = I(A;B:\sim B) \end{aligned} \quad (6)$$

Sparsity theory: finite data and expected information

Before proceeding, an important digression is that we never have complete information such as I(A; B) but only its estimate E(I(A; B)|D(A, B)) based on the data D(A, B) about A and B. It should be less available information if we have less data, converging to zero information when we have no data, and converging rapidly to I(A; B) as we would calculate it classically via P(A, B), P(A) and P(B) when data becomes indefinitely large. It should correspond to the expected value over all the values of the information measure according to the differing degree of belief that we hold, according to the data available [16]. This implies an integration [16] that leads to results later described in terms of the "incomplete" Riemann Zeta Function [6].

$$\zeta(s,n) = \sum_{i=1,2,3,\ldots n} i^{-s} \quad (7)$$

Here s=1 is of most interest, which relates to the natural logarithm, so that the sum is 1+ ½ + ⅓ +… 1/n, and n>0 on the understanding that with no summation and hence n=0, then ζ(s,n)=0. It emerges that simply adding or subtracting from n allows prior belief to be included. Putting that aside, it appears to reflect the amount of information in a system that is actually available to the observer via the data. It is thus expressed as the unqualified measure I(A; B) redefined, rather than by writing E(I(A; B) | D(A, B)) which suggests an arbitrary estimate. The redefinition is

$$I(A;B) = \zeta(s=1,o[A,B])-\zeta(s=1,e[A,B]),\quad \to \log_e(P(A,B)/P(A)P(B),\quad o[\,]\to\infty \quad (8)$$

Where o[ ] is the observed frequency (number of observations) and e[ ] is the expected frequency, as these senses are also used in the chi-squared test. For example, e[A, B]=o[A]o[B]/N=o[A]P(B)=P(A)o[B]. It should be declared that we ignore the Euler-Mascheroni constant γ defined as γ=ζ(s=1, n)–logen, n →∞, on two grounds. First it effectively cancels in Equation. 8, and second, and more to the point, it is here the definition of I(A; B) in Equation 8. I(A: ~A; B) thus becomes

$$\begin{aligned} &I(A:\sim A;B)=\zeta(s=1,o[A,B])-\zeta(s=1,e[A,B]) \\ &-\zeta(s=1,o[\sim A,B])+\zeta(s=1,e[\sim A,B]), \\ &\rightarrow \log_e\left(P(A,B)\,P(\sim A)/P(\sim A,B)\,P(A)\right), \quad o[\ ]\rightarrow\infty \end{aligned} \tag{9}$$

Theory of directional conditionality

Though physics is beyond present scope, causality theory, conditional directionality, and handedness are the core elements of quantum field theory. The current importance is that an algebraic entity, here called ι, relates to the Dirac spinor (essentially a linear operator with eigensolutions 0 and 1), with the following simple application.

For any real-valued scalar which has a symbolic adjoint such as for example conditional probabilities P(A|B) and P(B|A), it may be rendered as having an algebraic complex adjoint by writing say ι*P(A|B) +ιP(B|A) such that (ι*P(A|B) + ιP(B|A))* = ι*P(B|A) + ιP(A|B).

The asterisk indicates complex conjugation which changes the sign of the imaginary part. Here one sees empirical probabilities replacing the exponentials that represent statistical weights in quantum mechanics, which are in contrast calculated ab initio. We could have written ιP(B|A)+ι*P(A|B) as long as we internally consistent. As a tool for biomedical inference, the above thinking was introduced by the present author [17-19], who initially used the choice ι√P(B|A) +ι*√P(A|B) [17]. The form ι*P(B|A)+ιP(A|B) is preferred for consistency with physics. Briefly, the older choice would arguably lead to non-physicality in applying a recipe due to Dirac (ket normalization) when A and B are conjugate variables, and extension of that recipe to the current algebra also suggests that square roots are not required. To avoid such issues, we can start from several points, physical or better still, purely mathematical. As an excuse for addressing an area of research into counting, sampling, and data analytics that could emerge as of importance; we may start from our zeta function, which also has the arguable benefit of being purely a mathematical approach. In ζ( ), s=1 is not the only choice of s, though we would not in general speak of it as expected information, but rather as some measure of surprise, or as internal moments of the expected information [6] The most surprising choice of interest is however one that involves an imaginary number, and not just the more familiar i such that ii=−1, but rather the hyperbolic number h such that hh=+1. In the guise of Dirac's linear operatorι σ with eigen values −1, and +1, and γtime, and γ5, such hyperbolic imaginary numbers are fundamental to theoretical physics. Distinguish h above from Planck's constant; if that coexists in physical equations, our h is written ***h***. Note that:-

$$\begin{aligned} \zeta(s=\sigma+ht,n) &= \sum_{x=1\ldots n} X^{-(\sigma+ht)} \\ &= \sum_{x=1\ldots n} X^{-\sigma}e^{+ht\log e\,x} \\ &= \sum_{x=1\ldots n} x^{-s}\ \tfrac{1}{2}[\iota^* e^{-t\log e\,x}+\iota e^{+t\log e\,x}] \\ &= \iota^*\sum_{x=1\ldots n} X^{\acute{o}-t}+\iota\sum_{x=1\ldots n} X^{\acute{o}+t} \\ &= \iota^*\zeta(s=\sigma+1,n)=+\iota\zeta(s=\sigma-t,n) \\ &= \tfrac{1}{2}(\zeta(s=\sigma+t,n)+\zeta(s=\sigma-t,n)+\tfrac{1}{2}h(\zeta(s=\sigma-t,n)-\zeta(s=\sigma+t,n)) \end{aligned} \tag{10}$$

Here ι=½ (1+h), ι*=½ (1 + h), is seen as also a convenient notation. It is thus simple to find values of the incomplete (n<∞) and complete (n=∞) zeta function with h-complex values of s of interest, from the values for real values s=σ + t, s=σ−t. For example, ζ(s=1+h,n)=½(π2/6 −1/2) + ½h (ι1/2−π26)=ζ(s=1+ h, n)=(π2−3)/12−h (π2+3)/12, n →∞. From the above it can be seen that there are whole classes of functions that can be similar decomposed into components that are coefficients of ι* and ι, most notably for any real value of σ, t, w, x, y, z,

$$(\sigma+ht)^x = \iota^*(\sigma-t)^x+\iota(\sigma+t)^x$$

$$x^{\sigma+ht} = \iota^* x^{\sigma-t}+\iota x^{\sigma+t}$$

$$\iota^*(w+x)+\iota(y+z)=(\iota^* w+\iota y)+(\iota^* x+\iota z)$$

$$\iota^* wx+\iota yz=(\iota^* w+\iota y)\,(\iota^* x+\iota z)$$

Maclaurin series expansions of trigonometic function show that cos(hx)=cos(x), sin(hx)=hsin(x), tan(hx)=htan(x), and note that hi=−ih where ii=−1 Importantly, for any function f, note f(x + hy)*=*f*(x−hy) and *f*(i*x+iy)*=f(ix+i*y). Some above seem exotic cases, but throughout, whatever probabilistic system is being modeled and however modeled, ι and ι* are important as putting causes on an equal footing with consequences, and of monitoring their mutual consistency and coherence (see below) with respect to available data, all of great concern in making pharmacological sense out of pharmacosurvelliance,

Directionality in terms of zeta functions

The follow Sections give a brief flavor of the kind of research being carried out with the use of the above kind of thinking. Although an odds ratio converges to a likelihood ratio in the sense described earlier above, it does not preserve information regarding conditionality between A and B. To help establish etiology (cause) we need to handle two directions of conditionality and keep them distinct. For brevity it must be stated simply that attempts to build suitable expressions from the bottom up either end in containing a difference between I(A; B) and I(~A; ~B) where as discussed above the "double negative" should be supportive evidence and of same sign as I(A; B), whereas simply replacing the terms of the log odds ratio itself leads to cancellations of directional components so that the result is a purely real value and the original log odds. In the latter case we have

$$\begin{aligned} &\{A:\sim A\,|\,B:\sim B\}=\{B:\sim B\,|\,A:\sim A\} \\ &=\zeta(s=1,o[A,B])+\zeta(s=1,o[\sim A,\sim B])-\zeta(s=1,o[A,\sim B])-\zeta(s=1,o[\sim A,B]) \end{aligned} \tag{11}$$

Seeking the right form is thus not trivial. However, Equations 5 and 6 provide the clue. We can indeed capture these as two directions of conditionality. It is easy to show that we can redefine our real valued log odds ratio as

$$\begin{aligned} \{B:\sim B|A:\sim A\} &= \zeta(s=1,o[\sim A,B])] \\ &-\iota^*[\zeta(s=1,o[\sim A,B])+\zeta(s=1,o[A])-\zeta(s=1,o[\sim A])] \\ &+\iota\ [\zeta(s=1,o[A,\sim B])+\zeta(s=1,o[B])-\zeta(s=1,o[\sim B])] \end{aligned} \tag{12}$$

The term with [A, B] stays real because ι*x + ιx=x. Note that since we have both directions of conditionality encoded, we could define either direction as the key one of interest. In any event, it is subject to complex conjugation: {A: ~A | B: ~B}={B: ~B | A: ~A}*. Note also that the value of the above Equation 12 is purely real only if the contents of ι*[ ] and ι[ ] are equal in value.

The above is in some respects merely one example. For one thing, we could add in the real value ζ(s=1, o[~A, ~B]) to recapture that double negative contribution, but this would require renormalization, a need detectable by the fact that no zeta function is subtracted from it. The following appears to be a good approximation in many cases.

$$\begin{aligned} \{B:\sim B|A:\sim A\} &= \zeta(s=1,o[A,B])+\zeta(s=1,o[\sim A,\sim B])-\zeta(s=1,e[\sim A,\sim B]) \\ &-\iota^*[\zeta(s=1,o[\sim A,B])+\zeta(s=1,o[A])-\zeta(s=1,o[\sim A]) \\ &-\iota[\zeta(s=1,o[A,\sim B])+\zeta(s=1,o[B])-\zeta(s=1,o[\sim B]) \end{aligned} \tag{13}$$

Note that ζ(s=1, o[A, B]) + ζ(s=1, o[~A, ~B])ι− ζ(s=1, e[~A, ~B]) is

not itself directional, as it is purely real valued: it is the single negations that here provide directionality. By itself it is not normalized, though ζ(s=1, o[A, B])ι− ζ(s=1, e[~A, ~B])ι+ ζ(s=1, o[~A, ~B]) − ζ(s=1, e[~A, ~B]) would be.

Inference with negative evidence

The above directional form {B: ~B | A: ~A} is of little practical interest in isolation, but becomes very important in bidirectional inference networks, allowing inference about etiology. That form can be used in an inference network in the conceptual place of a conditional probability as it appears in a Bayes Net [20]. A very simple example is the chain rule of epidemiology, re-expressed with negative evidence. We see this in the mortality rate for an infectious disease.

P(death|complications)×P(complications|symptoms)×P(symptoms|infected)×P(infected|exposed)×P(exposed) which more generally, and in log form with negative evidence, may be represented

$$\{Z:\sim Z|\ Y:\sim Y\}+\{Y:\sim Y\,|\,X:\sim X\}\ldots+\{C:\sim C\,|\,B:\sim B\}+\{B:\sim B\,|\,A:\sim A\}$$

This uses Eqn. 13 for the terms { | }, but it applies to a number of similar treatments that will be described elsewhere. To include priors in both directions of conditionality and to give the counterpart of a joint probability estimated by a Bayes Net, the counterparts of prior probabilities need to be included, and the joint probability turns out to be as follows.

$$\iota I(Z;Z)+\{Z:\sim Z|\ Y:\sim Y\}+\{Y:\sim Y|\ X:\sim X\}\ \ldots+\{C:\sim C|\ B:\sim B\}+\{B:\sim B|\ A:\sim A\}+\iota^{*}I(A;\sim A)$$

An inference computation can be said to be coherent if it reflects Bayes theorem P(A|B)P(B)=P(B|A)P(A), which ironically is not represented in a traditional Bayes Net because it considers only forms in one direction of conditionality exemplified by P(B|A)P(A). Without negative evidence included this means that the resulting value of a network based on the above ideas and seeking to estimate a joint probability would be purely real valued. The imaginary part is the degree to which coherence is not satisfied, including by the way in which we may estimate negation. The simplest fix is therefore to take just the real part as the joint probability estimate, which corresponds to taking the simple arithmetic mean or average of the values of the network in each direction of conditionality. We can see this in ι *P(A|B) +ιP(B|A)=½[P(A|B)+P(B|A)] + ½ h[P(B|A) – P(A|B)], i.e. ½[P(A|B) +P(B|A)] is the real part Re(ι*P(A|B)+ιP(B|A)), whilst ½[P(B|A)–P(A|B)] is the imaginary part Im(ι*P(A|B)+ιP(B|A)). This "average direction" recipe also applies to Re(ι *P(B |~A)+ιP(A |~B)), which is a good approximation if P(B |~A) ≈ P(A | ~B). It reflects the logical law of the contrapositive, e.g. P("If x is not an odd number, then x is divisible by 2")=P("If x is not divisible by 2, then x is an odd number")=1, i.e. true in the case of certainty, and approximately true in near-certainty. But note "If there was no adverse reaction A, then drug B was used by the patient", and "If drug B was not used by the patient, then there was an adverse reaction A." Evidently, in the second statement the patient may have used drug C, which did not have adverse reactions, or no drug at all. Combining as inference from multiple contributing factors with inclusion of many independence assumptions, as a Bayes Net typically does, will run into this complication when ways to include negative evidence are found.

Non-Linear inference networks

Again, a more detailed analysis will be given elsewhere; for some degree of completeness the following taste of the considerations needed must suffice. Branches need more extensive explanation because Bayes Nets make certain independency assumption that shows have impact when we are considering information in both directions of conditionality. For example, P(A| B, C) P(B | D) P(C | E) considers B and C as interdependent in one direction of conditionality, to the right of the conditional bar '|' in P(A | B, C), but independent in the product P(B | D) P(C | E) where they appear to the left of the conditional bar. To give balance overall may either make a correction that makes the independence assumption interdependent, or vice versa. The latter is the better choice, since the data is evidently available to make it, and it does not lose information. Briefly, correction is implemented by ι*+ιeI(B; C) in the directional counterpart of a Bayes Net, as will be described elsewhere, and by due attention to the term I(B; C) in the present case. The procedure of taking the real part discussed above should be done after this "due attention to the term I(B; C)" and similarly throughout, else we discard information in one direction that is explicit or implicit in the other, and hence available.

The above, as cautioned at the outset, is only an outline, and barely a declaration of the importance and interrelationships of the tools with some detail added to convey the flavor. Details will be described elsewhere as they involve considerable discussion. Not least, cyclic paths in an inference network do not appear to constitute a problem, which is not the case in a traditional Bayes Net [20]. Like joint probabilities, truly cyclic paths comprised of h-complex probabilities or their logarithms are ideally purely real valued,

$$\begin{aligned}\iota^{*}\log_e P(A|B)P(B|C)P(C|D)P(D|A)+\iota\log_e P(A|D)P(D|C)P(C|B)P(B|A)\\ =\log_e P(A,B)P(B,C)P(C,D)P(D,A)\,/\,P(A)P(B)P(C)P(D)\end{aligned}\qquad(14)$$

although the inclusion of an *estimate* of the negative evidence may prevent that. Note that the real part −ιζ(s=1, o[B])+ζ(s=1, o[~B])] in Equation 13 reflected the need to remove the arguments of factors that occur more than once, and the above normalization by P(A)P(B)P(C) P(D) in Equation 14 plays a similar role.

Discussion and Conclusions

A broader unified approach may be possible involving the four pillars of evidence, information and decision theory, the Riemann Zeta function, and hyperbolic-complex algebra. Though much remains to be done, it at least seems clear that there is evidently still space for exploring some ideas founded in 19th century epidemiology. When discussing such theoretical matters in a specialized journal as this, the question arises as to whether the tools they imply, if accepted, are especially relevant and possibly even exclusive in importance to the field represented. As to some anecdotal evidence of direct and specific relevance, all references to the author's recent work given here have been specifically concerned with tackling genomic, proteomic and clinical data issues. Such areas are certainly major drivers of research and development of data mining and inference from it [1-3,6,21,22]. However, it may be big money and life-and-death issues that combined put them, or should put them, amongst the most critical considerations in pharmaceutical science and related biomedical research [6].

References

1. Robson B, Baek OK (2009) The Engines of Hippocrates: From the Dawn of Medicine to Medical and Pharmaceutical Informatics. John Wiley & Sons, USA.
2. Robson B (2004) The Dragon on the Gold: Myths and Realities for Data Mining in Biomedicine and Biotechnology using Digital and Molecular Libraries. J Proteome Res 3: 1113-1119.
3. Robson B, Vaithiligam A (2010) Drug Gold and Data Dragons: Myths and Realities of Data Mining in the Pharmaceutical Industry. In: Balakin KV (Ed). Pharmaceutical Data Mining. John Wiley & Sons 25-85.
4. Robson B (2003) Clinical and Pharmacogenomic Data Mining. 1. The

generalized theory of expected information and application to the development of tools. J Proteome Res 283-302.

5. Robson B, Mushlin R (2004) Clinical and Pharmacogenomic Data Mining. 2. A Simple Method for the Combination of Information from Associations and Multivariances to Facilitate Analysis, Decision and Design in Clinical Research and Practice. J Proteome Res 3: 697-711.
6. Robson B (2005) Clinical and Pharmacogenomic Data Mining: 3. Zeta Theory As a General Tactic for Clinical Bioinformatics. J Proteome Res 4: 445-455.
7. Robson B (2008) Clinical and Pharmacogenomic Data Mining: 4. The FANO Program and Command Set as an Example of Tools for Biomedical Discovery and Evidence Based Medicine. J Proteome Res 7: 3922-3947.
8. Mullins IM, Siadaty MS, Lyman J, Scully K, Garrett CT, et al. (2006) Data mining and clinical data repositories: Insights from a 667,000 patient data set. Comput Biol Med 36: 1351-1377.
9. Robson B, Li J, Dettinger R, Peters A, Boyer SK (2011) Drug discovery using very large numbers of patents. General strategy with extensive use of match and edit operations. J Comput Aided Mol Des 25: 427-441.
10. http://en.wikipedia.org/wiki/Pharmacoepidemiology
11. http://en.wikipedia.org/wiki/International_Society_for_Pharmacoepidemiology
12. Johnson S (2007) The Ghost Map: The Story of London's Most Terrifying Epidemic--and How It Changed Science, Cities, and the Modern World. Penguin group, USA.
13. http://en.wikipedia.org/wiki/Receiver_operating_characteristic
14. http://en.wikipedia.org/wiki/GOR_method
15. Robson B, Pain RH (1971) Analysis of the Code Relating Sequence to Conformation in Proteins: Possible Implications for the Mechanism of Formation of Helical Regions. J Mol Biol 58: 237-256.
16. Robson B (1974) Analysis of the Code Relating Sequence to Conformation in Globular Proteins: Theory and Application of Expected Information. Biochem J 141: 853-867.
17. Robson B (2007) The New Physician as Unwitting Quantum Mechanic: Is Adapting Dirac's Inference System Best Practice for Personalized Medicine, Genomics and Proteomics? J Proteome Res 6: 3114-3126.
18. Robson B (2009) Towards Intelligent Internet-Roaming Agents for Mining and Inference from Medical Data. Stud Health Technol Inform 149: 157-177.
19. Robson B (2009) Links between Quantum Physics and Thought. Stud Health Technol Inform 149: 236-238.
20. http://en.wikipedia.org/wiki/Bayesian_network
21. Svinte M, Robson B, Hehenberger H (2007) Biomarkers in Drug Dvelopment and Patient Care. Burrill 2007 Person. Med. Report 6: 3114-3126.
22. Robson B, McBurney R (2012) The Role of Information, Bioinformatics, and Genomics. 77-94 in Drug Discovery and Development. Technology in Transition, Churcill Livingstone, Elseveir.