ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Approximate Reasoning
39 (2005) 211–233
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

Benefits of embedding structural constraints in coherent diagnostic processes

Andrea Capotorti

Dipartimento di Matematica e Informatica, Università di Perugia, Italy

Received 1 February 2004; accepted 1 October 2004
Available online 8 December 2004

Abstract

This paper reviews recent results applicable to medical diagnosis, obtained by adding structural constraints to a coherent inference process. Such further considerations turn out to be useful whenever a basic lower–upper conditional probability assessment induces extension bounds too vague to motivate an informed decision. Three general types of qualitative judgements are proposed and fully described. They do not constitute a "panacea" to solve every problematic situation, but their application can considerably improve inferences results in specific cases, as two practical applications show.

Keywords: Coherent inference; Conditional exchangeability; Qualitative constraints; Diagnosis procedures

1. Introduction

In many practical applications, particularly in the medical field, there is a problem that the information at hand is not so fully detailed to allow us to adopt standard statistical tools. This happens especially whenever information is based on data collected from different sources or by heterogenous samples. In these cases, current

E-mail address: capot@dipmat.unipg.it
URL: http://www.dipmat.unipg.it/~capot

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2004.10.010

results in *genuinely* probabilistic reasoning can help reach useful inferences about relevant statements. In this approach, answers differ from usual *uniquely determined* statistical results, having, in general, interval-based conclusions. Common practice relies on *artificial* assumptions, such as the use of specific parametric distributions or unmerited assumptions of stochastic independence, to support computational shortcuts. However, such practices introduce the risk of misleading inferences. It is true that in situations of limited information, results can be so vague that it is impossible to make any reasonable decision. Hence, it is natural to search for further properties that can help to reach sharper conclusions. This can be obtained by a deeper analysis of the problem and also by further structural judgements. Of particular importance are conditional exchangeability assumptions, which are more general and can be more reasonable than those of conditional stochastic independence; comparisons between conditional probabilities, which are apt to capture expert convictions not fully numerically expressible; and restrictions on the admissible class of agreeing conditional measures, which are induced by indirect considerations on some statements not considered at the beginning.

In this paper we will explicitly show how such further considerations can be formalized and operationally adopted in general inference processes. Moreover we will have an idea of their relevance by applying them to two medical diagnostic procedures: a *median decision* process for the diagnosis of the asbestosis (fibrosis of the lung associated with friable asbestos exposure) based on X-ray film readings and a reliability judgement of a *gastrointestinal stromal tumor* (GIST) diagnosis based on histochemical results.

2. Coherent inferences with limited information

As already sketched out in the Introduction, when a problem does not allow a description by usual statistical models, a simple probabilistic approach can often be adopted to compute probabilistic bounds induced by the available information.

As prototypes of these situations we have chosen two medical applications with extremely weak information sources. In the first application to a *median decision* process (i.e., the diagnosis is made according to the majority of diagnoses by a certain number of experts) we have available only the single expert diagnostic sensitivities, the percentage of positive median diagnoses, and the proportion of those that are not obtained unanimously. With so few elements of information, particularly with the lack of median diagnostic specificities, it is almost impossible to adopt usual statistical decision models. In the second application to the diagnosis of GIST, we face the common problem of comparing the validity of a preliminary and relatively simple diagnostic technique with a new promising more sophisticated procedure. Here the scarcity of information is due to the extremely high costs of the new procedure, so that only few experiments are possible, and to its novelty, so that only preliminary and contradictory studies are available in the literature.

We shall see that even if the formalizations of these problems are extremely basic, it is possible to reach reasonable conclusions using the new methods reviewed here.

The sharper results are achieved by embedding the problem at hand into a coherent setting, i.e. representing the relevant entities through conditional events endowed with numerical values or bounds and looking for some class of conditional measures agreeing with them. Once a class has been detected, it can be used to make inference on relevant quantities (usually called "indexes").

For example, suppose we have to represent the situation that a disease D has a prevalence in the population between 5/100,000 and 10/100,000 and that there is a quite good clinical test T to detect it, with absence of false negatives and a specificity estimated to range between 90% and 95%. This can be formalized by the probabilistic constraints reported in Table 1 where the logical operator $\neg$ denotes the negation.

Such constraints implicitly restrict the set of probabilistic models that can be used to represent the problem. Using this set it is possible to compute the consistent bounds for any other relevant statement. For example, the constraints of Table 1 induce for the positive predictive value of the test $P(D|T)$ a lower bound of 5/10,000 and an upper bound of 2/1000. These bounds can be easily computed by Bayes' theorem applied to the extreme values listed in Table 1.

Of course the previous example is extremely simplified just to give an idea of the way to proceed. Things becomes interesting in more complex situations, like those reported in the real applications at the end of the paper.

In such an approach, we have both the peculiarity of a direct introduction of conditional probability assessments (i.e., they are not derived as sub-products of joints and marginal evaluations), and the direct awareness of working with imprecise tools (interval assessments, classes of distributions, bounds for conclusions, etc.). The wide range of subjects covered in the previous ISIPTA symposia ([17,18]), whose inspiration mainly refers to Walley [26], testifies to the meaningfulness and soundness of the latter aspect. The appropriateness and usefulness of the former aspect, both from a theoretical and a practical point of view, are developed in the work started in [11] and recently fully described in Coletti and Scozzafava's book [15].

2.1. Preliminaries

Let us now introduce a proper formalization to operate with the framework depicted before. For the sake of simplicity we will use conditional and unconditional events, but everything can be easily generalized to (finite) random variables, conditional or not. (See for example what has been done with conditional previsions in [8]). The initial information, usually a knowledge "and/or" rule base, is represented

Table 1
Probabilistic bounds for the prevalence of a disease D, the sensitivity of a test T and its specificity, respectively

Relevant entity	Description	Probability range
$P(D)$	Prevalence	[0.00005,0.0001]
$P(T|D)$	Sensitivity	1
$P(\neg T|\neg D)$	Specificity	[.9,.95]

through a conditional lower–upper probability assessment $(\mathscr{F},\mathscr{L}_C,\mathbf{p})$. The first component of an assessment is a generic list of n conditional events $\mathscr{F} = (S_1|C_1,\ldots,S_n|C_n)$, where each $S_i|C_i$ represents some macro-situation S_i (i.e. some single event or a combination of events) considered in some particular hypothetical circumstances C_i. (Usually the C_i's represent different scenarios and they can also be single or compound events.) Note that some $S_i|C_i$ could be actually unconditional (i.e. the situation S_i is considered without reference to any specific circumstance) and in such case C_i will coincide with the sure event Ω. In the following we will also refer to the set $\mathscr{U}_F = \{S_1,\ldots,S_n,C_1,\ldots,C_n\}$ of unconditional events appearing as components of the elements of $\mathscr{F}$.

For example, referring again to the simple assessment reported in Table 1, in that case we have the family $\mathscr{F} = (D,T|D,\neg T|\neg D)$ of cardinality three with components

$$\mathscr{U}_F = \{S_1 = D, S_2 = T, S_3 = \neg T, C_1 = \Omega, C_2 = D, C_3 = \neg D\}. \tag{1}$$

Incompleteness of the information can have two origins: firstly the S_i's might not describe all possible combinations of situations; secondly the different circumstances C_i's might *overlap* or might not cover all possibilities. For this, it is crucial to know which are the relationships of incompatibility, implication, equivalence or whatever, among the events in $\mathscr{U}_F$.

For example, among the events in (1) there are six logical relations:

$$\begin{aligned}
&S_1 \subset S_2 \ (D \text{ implies } T \text{ because of the absence of false negatives});\\
&C_2 \equiv S_1 \ (D \text{ is considered both as situation } S_1 \text{ and as circumstance } C_2);\\
&S_2 \wedge S_3 \equiv \phi \quad \text{and} \quad S_2 \vee S_3 \equiv \Omega \ (T \text{ and } \neg T \text{ form a partition});\\
&C_2 \wedge C_3 \equiv \phi \quad \text{and} \quad C_2 \vee C_3 \equiv \Omega \ (D \text{ and } \neg D \text{ form a partition}),
\end{aligned} \tag{2}$$

where ϕ, $\wedge$ and $\vee$ denote the impossible event, the logical conjunction operator, and logical disjunction, respectively.

In general, the list $\mathscr{L}_C$ of the logical constraints among $\mathscr{U}_F$, like (2), will appear as the second component of an assessment.

Such relationships $\mathscr{L}_C$ represent constraints that any model must fulfill and they limit which are the possible *atoms*. [1] The atoms A_r, with $r = 1,\ldots,a \leqslant 2^{2n}$, are elementary events (i.e. they form a partition) obtained by full combinations of affirmed or negated events in $\mathscr{U}_F$. Hence a general atom A_r is obtained by an expression like

$$A_r = \widetilde{S_1} \wedge \cdots \wedge \widetilde{S_n} \wedge \widetilde{C_1} \wedge \cdots \wedge \widetilde{C_n}, \tag{3}$$

where each component, say $\widetilde{C_i}$, can be either the affirmed event C_i or its negation $\neg\, C_i$. With the events in (1) we will have all possible instances of

$$A_r = \widetilde{D} \wedge \widetilde{T} \wedge \widetilde{\neg T} \wedge \widetilde{\Omega} \wedge \widetilde{D} \wedge \widetilde{\neg D}. \tag{4}$$

[1] In some discipline atoms are called *possible worlds*.

Table 2
Characteristic vectors of the events $\mathscr{U}_F$ listed in (1)

Label	A_1	A_2	A_3
$\mathbf{s}_1$	(1	0	0)
$\mathbf{s}_2$	(1	1	0)
$\mathbf{s}_3$	(0	0	1)
$\mathbf{c}_1$	(1	1	1)
$\mathbf{c}_2$	(1	0	0)
$\mathbf{c}_3$	(0	1	1)

It is immediate to see that not all the combinations are possible. For examples, (4) reduces to ϕ if $\widetilde{\Omega}$ is taken as $\neg\Omega$, and similarly if $\widetilde{T} \wedge \widetilde{\neg T}$ are taken as $T \wedge \neg T$. More precisely, by using all the logical relations (2), it turns out that of the 2^6 potential atoms that could be generated by (4), only the following three are possible:

$$\begin{aligned}
A_1 &= S_1 \wedge S_2 \wedge \neg S_3 \wedge C_1 \wedge C_2 \wedge \neg C_3 = D \wedge T \wedge T \wedge \Omega \wedge D \wedge D;\\
A_2 &= \neg S_1 \wedge S_2 \wedge \neg S_3 \wedge C_1 \wedge \neg C_2 \wedge C_3 = \neg D \wedge T \wedge T \wedge \Omega \wedge \neg D \wedge \neg D;\\
A_3 &= \neg S_1 \wedge \neg S_2 \wedge S_3 \wedge C_1 \wedge \neg C_2 \wedge C_3 = \neg D \wedge \neg T \wedge \neg T \wedge \Omega \wedge \neg D \wedge \neg D.
\end{aligned} \tag{5}$$

Moreover, as is usual in conditional contexts (see [13] and [15, Section 11.3]), we will refer only to atoms spanned by $\mathscr{U}_F$ and inside the disjunction $\bigvee_{i=1}^{n} C_i$, because only elementary situations contemplated in some of the considered scenarios must be involved to check the consistency of the assessment [2]. Hence the proper upper bound for the number of atoms a is 3^n.

In the sequel we will also need to use the characteristic vectors of the events. These are vectors whose components are 1 or 0 depending on whether the corresponding atom implies the event or not. We will denote such vectors with the same letter as the event, but in boldface lower-cases. Hence, $\mathbf{s}_i$ and $\mathbf{c}_i$ will denote the characteristic vectors of S_i and C_i, respectively, while their juxtaposition $\mathbf{s}_i\mathbf{c}_i$ will represent the characteristic vector of the conjunction S_iC_i. (For the sake of simplicity in the following we will omit the usual conjunction operator $\wedge$.)

Referring again to the events in (1) and to the atoms in (5), we obtain the characteristic vectors [3] listed in Table 2.

Introducing a vector of variables $\mathbf{x} = (x_1 \ldots x_a)$, where each component x_r is associated with possible values for the probability of the atom A_r, it is possible to *rebuild* the possible values of probability for any event in $\mathscr{U}_F$, say S_i, simply by

$$P(S_i) = \sum_{A_r \subseteq S_i} P(A_r) = \mathbf{s}_i \cdot \mathbf{x}, \tag{6}$$

where $\cdot$ represents the row–column matrix product.

[2] For those familiar with Walley's notation, a similar motivation is used to introduce the consistency property of *Avoiding Uniform Loss* rather than the *Avoiding Sure Loss* for conditional previsions. See [26, Section 7.1.3] and [27, Section 2.1].

[3] In Lad [19], the matrix whose rows are the characteristic vectors is called the *realm matrix*.

Note that the atoms, and consequently the characteristic vectors, are not a part of the conditional probability assessment, i.e. they are not given directly by the analyst, but they are implicitly defined by the first two components $\mathscr{F}$ and $\mathscr{L}_C$. Nonetheless, they are important because they are the main operational tool involved in the inferential process.

The last component of an assessment is represented by a vector of numerical bounds $\mathbf{p} = ([lb_1, ub_1], \ldots, [lb_n, ub_n])$. Each closed interval $[lb_i, ub_i]$ represents lower and upper bounds associated with probabilities for the corresponding conditional event $S_i|C_i$. These are usually estimated by expert beliefs, by literature reports or by collected data.

Note that some of the numerical bounds $[lb_i, ub_i]$ may degenerate to a single value p_i, representing a precise assessment (e.g. the second constraint in Table 1).

2.2. Coherence

If we cannot adopt a unique probabilistic model for the assessment $(\mathscr{F}, \mathscr{L}_C, \mathbf{p})$, it is still possible to search for the class $\mathbb{P}_{\mathscr{F}}$ of full conditional probability distributions that are compatible with the assessments we can make. It is possible to ask various properties of $\mathbb{P}_{\mathscr{F}}$: in the present paper we look for a class such that $\mathbf{p}$ coincides with the convex envelope of $\mathbb{P}_{\mathscr{F}}$ restricted to $\mathscr{F}$, i.e. such that

$$\forall P \in \mathbb{P}_{\mathscr{F}} \; lb_i \leqslant P(S_i|C_i) \leqslant ub_i \quad \text{for all } S_i|C_i \in \mathscr{F}; \tag{7}$$

$$\forall S_i|C_i \in \mathscr{F} \; \exists P', \quad P'' \in \mathbb{P}_{\mathscr{F}} \text{ s.t } P'(S_i|C_i) = lb_i \quad \text{and} \quad P''(S_i|C_i) = ub_i. \tag{8}$$

Practically speaking, the third component $\mathbf{p}$ of the assessment represents a set of numerical constraints that all the admissible models (the conditional probabilities $P \in \mathbb{P}_{\mathscr{F}}$) must satisfy (inequalities (7)). Such constraints must be tight enough so that their bounds can be actually reached by some of the admissible models (equalities (8)).

In the following, such a class $\mathbb{P}_{\mathscr{F}}$ of probability distributions will be said to *agree* with the assessment $(\mathscr{F}, \mathscr{L}_C, \mathbf{p})$ and its existence guarantees about the *coherence* of the assessment.

As has been already stated in [12,14], and in particular in [15, Section 15.2], the existence of $\mathbb{P}_{\mathscr{F}}$ can be checked *operationally* by the satisfiability of a *class of sequences* of linear systems $\{\mathscr{S}^j_\alpha\}$, with $j = 1, \ldots, 2n$ and $\alpha = 1, \ldots, \alpha_j$. Note that *sequences* of linear systems are necessary to allow conditioning events C_i's to have induced probabilities that are not bounded away from 0. This procedure partitions $\mathscr{F}$ in different *zero layers* indexed by α. (For a deeper exposition of this aspect refer again to [15], in particular to Sections 12 and 15).

Such linear systems reflect an attempt to determine unconditional probability distributions through which to construct the agreeing class $\mathbb{P}_{\mathscr{F}}$ (i.e. a set of conditional probabilities satisfying (7) and (8)). Hence, for each event $S_i|C_i \in \mathscr{F}$ there will be associated two sequences of linear systems $\mathscr{S}^{2i-1}_\alpha$ and $\mathscr{S}^{2i}_\alpha$. This to ensure that, according to (8), the bounds lb_i and ub_i can be actually attained. Of course whenever the bounds degenerate to a single value p_i, the two sequences coincide.

The set of all possible solutions to the class of linear systems sequence $\{\mathcal{S}^j_\alpha\}$ implicitly induces the searched class $\mathbb{P}_{\mathcal{F}}$. Hence, if such set of solutions is not empty, the assessment $(\mathcal{F},\mathcal{L}_C,\mathbf{p})$ is said to be coherent, otherwise not. Note that this coherence notion is almost the same as those usually adopted in imprecise probabilities frameworks. That is, $\mathbf{p}$ coincides with its *natural extension*. (See [26] and [27, Section 3.2] property (d).) The only difference occurs in the proper treatment of conditional events $S_i|C_i$ whose conditioning C_i can have probability not bounded away from zero.

For the sake of simplicity, in the following we shall neglect to specify in which zero layer α we are operating. Hence, each time a conditional probability will be expressed as ratio of unconditional probabilities, it must be intended in the proper zero layer where this ratio is possible.

2.3. Extension

In practical applications when information comes from different sources, it turns out that checking the coherence of the assessment $(\mathcal{F},\mathcal{L}_C,\mathbf{p})$ is a compulsory step. This will be seen in the second medical application we discuss here.

Once coherence has been assured, it is possible to perform inference on any conditional event $H|E$ judged important to reach conclusions in the problem. An example would be the positive predictive value of the test, $P(D|F)$, which is required for the simplified example of Table 1. Generally, H represents some hypothesis to judge whenever there should be some evidence E.

In this context, inference reduces to compute the coherent extension of $\mathbf{p}$ to $H|E$, obtainable as the closed interval $[lb_{H|E},ub_{H|E}]$ of the values $P(H|E)$ with $P \in \mathbb{P}_{\mathcal{F}}$.

Although this is theoretically simple, from the practical point of view it is more subtle. In fact, following a method similar of that depicted in the previous subsection, we are required to perform *sequences of optimizations* $\{\mathcal{O}^j_\alpha\}$. Thanks to the possibility of exploiting zero probabilities and thanks to proper normalization conditions, all the optimizations problems in $\{\mathcal{O}^j_\alpha\}$ are reduced to be linear programs. (Once more, for a full description of the technique refer to [15, Section 14.1].)

The number of sequences is the same as that of the linear systems $\{\mathcal{S}^j_\alpha\}$ for the check of coherence: there are two sequences with $j = 2i - 1$ and $j = 2i$ for each conditional event $S_i|C_i \in \mathcal{F}$. Hence they are actually "at worst" $2n$, on account of the already stated consideration about the coincidence of the pair of sequences whenever coherence requires a precise value for $p_i \in \mathbf{p}$.

What is important to note here is that each sequence of optimizations ends with a pair of optimal values $lb^j_{H|E}$ and $ub^j_{H|E}$. They represent the minimal and maximal value, respectively, for the searched probability $P(H|E)$. Those with an odd index $j = 2i - 1$ will be computed with the first equality in (8) as further specific constraint, while those with an even index $j = 2i$ with the second one.

The final coherent interval $[lb_{H|E},ub_{H|E}]$ will result from the convex combination of all the intervals $[lb^j_{H|E},ub^j_{H|E}]$, i.e. $lb_{H|E} = \min_{j\in\{1,\dots,2n\}} lb^j_{H|E}$ and $ub_{H|E} = \max_{j\in\{1,\dots,2n\}} ub^j_{H|E}$.

The main difficulty of such procedures is the usually huge number a of atoms. In fact, it has been show that already the problem of checking the coherence for unconditional precise assessments is NP-complete (see for example [1]). It is for this reason that heuristic procedures have been developed. Without entering into the details that would be outside the scope of this paper, we can mention that in the *unconditional* framework, a promising procedure [1,2] has been proposed based on variable elimination in the style of the Davis–Putnam procedure for satisfiability. It has been shown [3,4] that such a procedure solves in polynomial average time instances where the logical constraints $\mathscr{L}_C$ can be expressed by clauses with at most two literals.

For the *conditional* framework, thanks to a smart use of null probabilities and to the notion of *locally strong coherence*, in [6,9] the complexity problem has also been faced. Abstract problems have been solved with $O(n^3)$ logical satisfiability tests in place of solving the linear systems and the optimization problems that have $O(3^n)$ number of unknowns. Even with such promising results, a systematic complexity study of this last procedure is still missing.

3. Results improvement by structural constraints

Extension bounds $[lb_{H|E}, ub_{H|E}]$ are what our information implies for $H|E$ from a pure probabilistic point of view. Sometimes however, they can result in probability intervals too wide to support an informed decision. Even in such cases, it may be possible to shrink the agreeing conditional probability class $\mathbb{P}_{\mathscr{F}}$ while maintaining a *model free* approach. This will be reached by adding structural considerations (i.e. that require specific properties) to the numerical constraints $\mathbf{p}$. Of course there are several possible different kinds of constraints to introduce, but we will focus on few of them that are quite natural and have yielded satisfactory results in our examples.

3.1. Conditional exchangeability vs. independence

As already mentioned, a common method of restricting the variability of the conclusions is to adopt an assumption of stochastic independence. This is actually a powerful restriction, and is not always really appropriate. Specifically, when information is based on judgments made by several experts, the presumed independence of experts is often based on the fact that they each make their judgment without knowing the judgments of the others. But this does not really imply stochastic independence. Stochastic independence of their assessments would mean that we, the probability assessors, would not change our probability assessment for a positive judgment by one expert when we hear the judgment of another expert. This is not really the case, because we explicitly regard them all as experts.

For example, in the median decision problem of Section 4.1 there is the information that three X-ray readers assess *independently* one from the others their judgment about the presence or not of the fibrosis in the patients. Denoting with D_i, $i = 1, 2, 3$, the events of positive diagnosis by the readers and by F the event of actual presence of the fibrosis, the situation is usually modeled by the following conditional stochastic independence conditions

$$P(D_i|D_jF) = P(D_i|F) \quad \text{and} \quad P(D_i|D_j\neg F) = P(D_i|\neg F). \tag{9}$$

But such conditions reflect quite a different information. In fact they express that our probability assessment for each reader's judgement would remain the same, *even knowing* the judgment already given by one of the other experts. This contrasts with the fact that all the readers are experts with similar skills. Hence a positive answer given by one of them should reasonably influence our assessment of probability that another would make a positive diagnosis too.

What should be modeled is the fact that the judgments are thought to be given in *similar* circumstance and, mainly, by people with the same background. Hence, in the presence of such strong symmetries it is more suitable to introduce some kind of *exchangeability*. (For another similar situation, refer to Lad and Di Bacco [20].)

In fact exchangeability reflects information of perfect permutability among a set of events, that usually represent judgments or experiments, and it is appropriate whenever it is relevant to consider *how many* instead of *which particular* events hold.

More technically, exchangeability should be used whenever it is possible to identify a *sum* as a sufficient statistic (for a detailed explanation refer to [19, Section 3.9]). In particular, whenever the assessment is mainly conditional, *conditional exchangeability* could be the more suitable. For example, going back to the median decision problem, one reasonable assessment of conditional exchangeability could be (the full list will be given in Section 4.1)

$$P(D_1D_2\neg D_3|F) = P(D_1\neg D_2D_3|F) = P(\neg D_1D_2D_3|F) \tag{10}$$

that express equivalence, in the presence of fibrosis, of the chances to have joint judgments with two expert giving positive answers and one negative, irrespective of *who* is the expert in disagreement.

Formally, conditional exchangeability can be formulated as follows:

Definition 1. k events $E_1,\ldots,E_k$ are regarded as exchangeable *under a specific scenario* C_j if any conjunction of the E_i's with the same number of affirmed and negated events is evaluated identically when conditioned upon C_j. In other words, for any fixed number $s \in \{0,\ldots,k\}$ the probabilities

$$P(E_{i_1}\ldots E_{i_s}\neg\, E_{i_{s+1}}\ldots\ \neg E_{i_k}|C_j) \tag{11}$$

are assessed to be equal for any permutation of the indexes $i_1,\ldots,i_k$.

Conditions like (11) actually reduce the "degree of freedom" for the unknowns in the sequences of linear systems for the check of coherence and in the sequence of linear programs for the extension. This restricts "de facto" the admissible class of conditional measures $\mathbb{P}_{\mathscr{F}}$ and, possibly, it implies a shrinking for some extension bounds.

Since (11) refers to a fixed conditioning event C_j, restriction of this type are easily reported as linear constraints. In fact, let us denote with $\boldsymbol{\pi}_s$ and $\boldsymbol{\pi}'_s$ the characteristic vectors of two different permutations of the combination $E_{i_1}\dots E_{i_s}\neg E_{i_{s+1}}\dots\neg E_{i_k}$ and with $\mathbf{x}$ a generic vector of variables of the jth sequence of optimization problems. Hence extensions with the further conditional exchangeability requirement (11) follow by adding to the constraints of the linear programs in $\{\mathcal{O}_\alpha^j\}$ pairwise equalities of the form

$$(\boldsymbol{\pi}_s c_j - \boldsymbol{\pi}'_s c_j)\cdot\mathbf{x} = 0 \tag{12}$$

for each pair of permutations $\boldsymbol{\pi}_s$ and $\boldsymbol{\pi}'_s$ and each $s = 1,\dots,k-1$. (Note that the extreme cases $s = 0$ and $s = k$ do not actually constitute any constraint, because only one arrangement of "all 1's" or "all 0's" is possible.)

Note that equalities like (12) influence only the number of constraints in the linear programs, while they influence neither the dimension of the variables $\mathbf{x}$ nor the number of sequences of linear programs. The number of further constraints is $\sum_{s=1}^{k-1}\binom{k}{s}$.

Moreover, in [10] operational shortcuts to simplify the whole procedure in the presence of conditional exchangeability assessments have been introduced.

3.2. Conditional probabilities comparison

Sometimes there are conditional events which an expert believes more than some other, but he/she can express neither precise nor imprecise probability assessments, being only capable to compare them.

This is immediately interpretable as

$$P(S_i|C_i) \geqslant kP(S_l|C_l) \tag{13}$$

for some constant k. (Of course for a pure qualitative comparison it is enough to put $k = 1$.)

Continuing, if none of the conditional probabilities present in (13) is uniquely constrained, its direct representation in the optimization problems $\{\mathcal{O}_\alpha^j\}$ would be, if the two conditional events belong to the same zero layer

$$\frac{\mathbf{s}_i\mathbf{c}_i\cdot\mathbf{x}}{\mathbf{c}_i\cdot\mathbf{x}} \geqslant k\frac{\mathbf{s}_l\mathbf{c}_l\cdot\mathbf{x}}{\mathbf{c}_l\cdot\mathbf{x}} \tag{14}$$

or, more generally, in each zero layer

$$(\mathbf{x})^{\mathrm{T}}\cdot[(\mathbf{s}_i\mathbf{c}_i)^{\mathrm{T}}\cdot\mathbf{c}_\mathbf{l} - (k\mathbf{s}_l\mathbf{c}_l)^{\mathrm{T}}\cdot\mathbf{c}_i]\cdot\mathbf{x} \geqslant 0. \tag{15}$$

Computationally, such constraints have the drawback of being quadratic.

For example, going back to the extremely simplified example at the beginning of Section 2, we can introduce a vector of variables $\mathbf{x} = (x_1\ x_2\ x_3)$. If we would state the further constraint [4] $P(T) \geqslant P(D|T)$, by the characteristic vectors listed in Table 2 and the correspondence of events in (1), we would have

$$P(T) = \frac{\mathbf{s}_2\mathbf{c}_1\cdot\mathbf{x}}{\mathbf{c}_1\cdot\mathbf{x}} = \frac{x_1+x_2}{x_1+x_2+x_3} \geqslant P(D|T) = \frac{\mathbf{s}_1\mathbf{s}_2\cdot\mathbf{x}}{\mathbf{s}_2\cdot\mathbf{x}} = \frac{x_1}{x_1+x_2}, \tag{16}$$

or, more correctly,

[4] This specific constraint is actually redundant, but it could be useful for a better understanding of the proposed technique.

$$(x_1 + x_2)(x_1 + x_2) - (x_1)(x_1 + x_2 + x_3) = x_2^2 + x_1x_2 - x_1x_3 \geqslant 0. \tag{17}$$

The fact that (15) is in general quadratic increases the difficulties for the computation of the extension bounds. In fact, to deal with quadratically constrained linear programs there are specific Operational Research's techniques, like interior-point algorithms [24] or duality bound methods [25]. However, in our computational experience to date, they are not yet as reliable and stable [5] as those available for linear programming problems.

That is why we propose an approximation of (13) that, even being a weaker constraint, has the advantage of leaving the extension problem in a linear form. The idea is to express (13) in a parametric way and to introduce further unknowns which can capture the basic structure of the parametrization.

If we focus our attention on one of the two conditional probabilities in (13), let us say $P(S_l|C_l)$, we can take it as an *inference target* and compute its extension bounds $[lb_{S_l|C_l}, ub_{S_l|C_l}]$ as it has been illustrated in Section 2.3. We can now introduce new variables y_r, $r = 1,\ldots,a$, representing the quantities $P(S_l|C_l)x_r$, so that the inequality (13) can be represented by

$$\mathbf{s}_i\mathbf{c}_i \cdot \mathbf{x} - k\mathbf{c}_i \cdot \mathbf{y} \geqslant 0 \tag{18}$$

the link between new and old variables by

$$\mathbf{s}_l\mathbf{c}_l \cdot \mathbf{x} - \mathbf{c}_l \cdot \mathbf{y} = 0; \tag{19}$$

while the variability bounds for $P(S_l|C_l)$ imply the constraints

$$lb_{S_l|C_l}x_r \leqslant y_r \leqslant ub_{S_l|C_l}x_r \qquad \text{for } r = 1,\ldots,a. \tag{20}$$

If we apply this method to the comparison (16), we have to set $y_r = P(D|T)x_r$, $r = 1,2,3$, obtaining

$$\begin{aligned} &x_1 + x_2 - y_1 - y_2 - y_3 \geqslant 0;\\ &x_1 - y_1 - y_2 = 0;\\ &lb_{D|T}x_r = 5/10{,}000x_r \leqslant y_r \leqslant 2/1000x_r = ub_{D|T}x_r \quad r = 1,2,3, \end{aligned} \tag{21}$$

where the numerical values of the coherent bounds $lb_{D|T}$ and $ub_{D|T}$ were already explained in Section 2 after the assessment of Table 1.

Eqs. (18)–(20) are all implied by (13), while the reverse implication does not hold in general. Hence, if they are added as constraints in the linear programs sequences $\{\mathcal{O}_\alpha^j\}$ to obtain bounds for $P(H|E)$, we are not guaranteed to obtained an extension interval $[lb_{H|E}, ub_{H|E}]$ coherent with (13), but just an interval containing it.

However, once such bounds $lb_{H|E}$ and $ub_{H|E}$ are obtained, they can be substituted in (15) to check if that inequality holds. If not, the left-hand side of (15) will yield a negative value that can be adopted as a *measure of violation* of (13).

[5] We have tried to use some already implemented packages of non-linear programming obtaining dubious or inconsistent results.

Note that it is not needed to add sequences of optimizations to cyclically impose equalities in (18) and (20) because they must be fulfilled as they are by each $P \in \mathbb{P}_{\mathscr{F}}$.

At any rate, (18), (19) and (20) increase significantly the space complexity of the optimization procedure. In fact they double the number of variables and introduce among the constraints $2a$ new inequalities (those in (20)) plus one inequality and an equality (those in (18) and in (19)) for each conditional probability comparison. Hence, before adopting them it is better to check whether they are redundant, i.e., that they are not already implied within the agreeing class $\mathbb{P}_{\mathscr{F}}$.

3.3. Indirect restriction of the admissible class $\mathbb{P}_{\mathscr{F}}$

In this subsection it will be described a technique that works in reverse with respect to the previous two: the further constraints will be purely numerical and they will implicitly generate structural restrictions on the class $\mathbb{P}_{\mathscr{F}}$.

Since this technique will include an arbitrary choice, it should be used carefully. Moreover, it will require an *interpretation process* before being useful to represent knowledge of an expert in an applied field.

Analyzing the inference procedure for some conditional event $H|E$, it could happen that results are greatly influenced by variability of the probability for another *auxiliary* conditional event, $K|F \notin \mathscr{F}$.

Such auxiliary $K|F$ does not belong to the initial list of conditional events. Thus we would not have expressed prior bounds for $P(K|F)$, either because we do not have direct access to the data on which $\mathbf{p}$ can be based, or because there is not direct information on $K|F$. But, if we compute the coherent extension for $P(K|F)$ we may discover the range $[lb_{K|F}, ub_{K|F}]$ to be surprisingly wide. Hence we can think to restrict the admissible range for $P(K|F)$. This will *indirectly* shrink the agreeing class $\mathbb{P}_{\mathscr{F}}$ and consequently the range $[lb_{H|E}, ub_{H|E}]$ for the original inference target $H|E$.

The problem is, in absence of "a priori" information, how to restrict the range for $P(K|F)$. An accurate analysis of the way $[lb_{K|F}, ub_{K|F}]$ is obtained could be helpful.

In fact, recall that variability range $[lb_{K|F}, ub_{K|F}]$ results from the convex combination of all the single optimization bounds $\{lb^j_{K|F}, ub^j_{K|F}\}$, with $j \in \{1, \ldots, 2n\}$, obtained in the different sequences of linear programs in $\{\mathcal{O}^j_\alpha\}$. Moreover, note that differences among optimal values obtained in different sequences derive from the restrictions (8) applied each couple of sequences to a different conditional event $S_i|C_i \in \mathscr{F}$.

It could happen that, as we noticed in some practical applications, some of the intervals $[lb^j_{K|F}, ub^j_{K|F}]$ are much narrower than the others. This sheds some light on the different roles played by the agreeing distributions inside $\mathbb{P}_{\mathscr{F}}$.

These particular intervals $[lb^j_{K|F}, ub^j_{K|F}]$ can guide us, together with additional considerations (like literature hints, expert's behavior, etc.), to impose informative lower–upper bounds $\{lb^*_{K|F}, ub^*_{K|F}\}$ for $P(K|F)$. In this way the initial assessment $(\mathscr{F}, \mathscr{L}_C, \mathbf{p})$ can be updated in a new one $(\mathscr{F}^*, \mathscr{L}^*_C, \mathbf{p}^*)$ with $\mathscr{F}^* = \mathscr{F} \cup \{K|F\}$, $\mathscr{L}^*_C = \mathscr{L}_C \cup \{\text{log.rel. among } K,\, F \text{ and } \mathscr{U}_F\}$ and $\mathbf{p}^* = \mathbf{p} \cup \{[lb^*_{K|F},\ ub^*_{K|F}]\}$, so that a new inference on $H|E$ can be computed.

The following proposition gives a result that can help in choosing coherent restrictions for $P(K|F)$:

Proposition 1. *Let* $(\mathscr{F}, \mathscr{L}_C, \mathbf{p})$ *be a coherent conditional assessment,* $K|F \notin \mathscr{F}$ *a new conditional event and* $\{lb^j_{K|F}, ub^j_{K|F}\}_{j=1}^{2n}$ *the set of optimal values obtained by linear programs* $\{\mathcal{O}^j_\alpha\}$ *as described in Section* 2.3 *and performed on* $K|F$.

If the interval

$$[lb^*_{K|F}, ub^*_{K|F}] = \bigcap_{j=1}^{2n} [lb^j_{K|F}, ub^j_{K|F}]$$

is not empty then the conditional assessment $(\mathscr{F}^*, \mathscr{L}^*_C, \mathbf{p}^*)$*with*

$$\mathscr{F}^* = \mathscr{F} \cup \{K|F\};$$

$$\mathscr{L}^*_C = \mathscr{L}_C \cup \{\textit{logical constraints among } K,\ F \textit{ and } \mathscr{U}_F\};$$

$$\mathbf{p}^* = \mathbf{p} \cup \{[lb^*_{K|F}, ub^*_{K|F}]\},$$

is coherent.

The detailed proof of the proposition is given in [28], but is skipped here because of the limited scope of this paper. Anyhow we can briefly say that the existence of the agreeing class of conditional probabilities $\mathbb{P}_{\mathscr{F}^*}$ is obtained by proper linear combinations of the optimal solutions of the linear programs $\{\mathcal{O}^j_\alpha\}$ when $K|F$ is logically dependent on $\mathscr{U}_F$ (i.e. the atoms spanned by $\mathscr{F}^*$ are the same of those spanned by $\mathscr{F}$), otherwise by a peculiar redistribution of their masses.

Note that the range $[lb^*_{K|F}, ub^*_{K|F}]$ suggested in Proposition 1, even being very restrictive with respect to $\mathbb{P}_{\mathscr{F}}$, can be adopted without modifying the initial numerical evaluations $\mathbf{p}$.

In the second medical application about the GIST diagnosis in Section 4.2, as auxiliary conditional events we will choose the unconditional GIST's prevalence and we will make an even stronger choice for the rage of $P(\mathrm{GIST})$. In fact, with the help of further indications given by the pathologists, we will use only the lower bound of the intersections, i.e. the $\max_{j=1,\dots,2n} lb^j_{\mathrm{GIST}}$, as a "reasonable" value so to adopt a small coherent interval centered on it.

Of course, with such a technique it remains open the problem about the choice of the auxiliary conditional event $K|F$. There is not a general methodology. Its choice relies mainly on the context of the practical problem at hand and on the expertise of the analyst. However, $K|F$ is usually taken among the events, conditional or not, that are anyway judged relevant (e.g. disease's prevalence, positive predictive values, etc.) but that, for a lack of information, have been skipped in the initial domain $\mathscr{F}$.

4. Two medical applications

As already stated, we will show now how the procedures described before can be applied to practical problems. In particular we will illustrate the results we recently attained for two different medical diagnostic processes. The first problem will show feasibility and relevance of the conditional exchangeability assumptions and of the conditional probabilities comparisons as depicted in Section 3.1 and Section 3.2. In the second problem, we will show the importance of a preliminary check of coherence whenever information comes from different sources and the influence in the results of probability restrictions for an auxiliary event, in line with Sections 2.2 and 3.3, respectively.

4.1. Accuracy rates for an asbestosis median decision procedure

In [7] we re-examined the procedure of median decision making in the context of radiological determination of asbestosis (fibrosis of the lung associated with friable asbestos exposure). Median decision applies whenever there is a pool of experts, usually equivalent in skill, examining the same patients and each single case is finally diagnosed on the basis of the agreement of the majority of judgments.

In particular, in a recent paper [21], Tweedie and Mengersen analyzed a previous case-report about prevalence of asbestosis among a group of people with a similar history of asbestos exposure. Opinions of three radiologists were based on X-ray film readings, and the authors had rather limited information about the median decision procedure. Anyway, they were able to propose a tricky methodology to retrieve some conclusion about various probabilities associated with the correctness of the median diagnosis.

However, the authors' analysis deeply relies on the independence assumption for the experts' assessments. It was motivated because X-ray films were read *separately* by the three radiologists without intercommunication. As we have already motivated in Section 3.1, in such situations an assumption of conditional exchangeability would be more appropriate than independence would be. This choice will not result in a unique probability distribution as a solution such as proposed in [21], but merely a bounded interval for distributions.

To make a synthesis (a full description can be obviously found in the cited papers), we can formalize the problem as it follows.

First of all, in Table 3 we introduce events that refer to a generic patient with a X-ray film available:

On account of the similarity among radiologists' training, their *sensitivities* for the films' reading process $P(D_i|F)$, $i = 1,2,3$, are thought to be equal.

Table 3
Relevant events involved in the analysis of the asbestosis median decision procedure

Label	Description
F	Asbestosis (*fibrosis*) presence
D_i, $i = 1,2,3$	ith expert positive asbestosis judgment
D^*	Positive median decision diagnosis
S^*	Positive median decision with 2 positive and 1 negative judgement (in the following named *splitting vote*)

Table 4
Initial conditional probabilities values p given on $\mathscr{F}$

Statement	Prob. value
$D_i \mid F$	.82 $i = 1,2,3$
D^*	.12
$S^* \mid D^*$	.42

On the basis of recorded data on 642 patients and of specific literature references, the conditional probability values [6] $\mathbf{p}$ of Table 4 are given on $\mathscr{F} = (D_1|F, D_2|F, D_3|F, D^*, S^*|D^*)$. Although all the probability values shown in Table 4 are precise, this is not in contrast with an "imprecise probabilities" approach because the assessment is partial and hence the agreeing class $\mathbb{P}_{\mathscr{F}}$ does not reduce to a singleton. This will be evident when we will perform inferences and the results will be interval valued.

The first probability $P(D_i|F)$ comes from literature results on sensitivity analysis performed by comparing radiological and histopathological evaluations. The other two $P(D^*)$ and $P(S^*|D^*)$ derive from data reported in [21]. In particular, $P(D^*)$ is directly estimated by the ratio 77/642 of positive median diagnoses, while $P(S^*|D^*)$ is attained indirectly by the three individual proportions 82%, 86% and 90% of positive median diagnoses on which the single assessor gave a positive diagnosis, through the formula

$$\begin{aligned} P(S^*|D^*) &= P(\neg D_1|D^*) + P(\neg D_2|D^*) + P(\neg D_3|D^*) \\ &= (100 - 82)\% + (100 - 86)\% + (100 - 90)\% = 42\%. \end{aligned} \tag{22}$$

To complete the initial assessment we must explicitly state the possible logical relations $\mathscr{L}_C$ among the unconditional events $\mathscr{U}_F = \{F, D_1, D_2, D_3, D^*, S^*\}$. Apart from the obvious double role of D^* as unconditional situation (in $P(D^*)$) and as conditioning circumstance (in $P(S^*|D^*)$), according to the problem description we can pick out other logical dependencies as well. In particular, among the median decisions, with or without splitting vote, and individual experts' diagnosis there are two logical relations reported in Table 5. Such logical relations imply that the number of atoms is 16 instead of the potential 3^5 that could be spanned by an assessment with five conditional probabilities like Table 4. Moreover, since all the values of $\mathbf{p}$ are precise, there is actually only one sequence of linear systems $\mathscr{S}_\alpha$ and, since the probability of F is not bounded away from zero, there are two zero layers. Anyhow, thanks to the already mentioned procedure reported in [6], the check of coherence is reduced to the solution of only one linear system with eight unknowns and four equations. This system admits solutions, hence the assessment $(\mathscr{F}, \mathscr{L}_C, \mathbf{p})$ is coherent.

[6] In [7,21] several assessments with different sensitivity values are examined, here we report only the first one as a prototype.

Table 5
Logical relations among the events $\mathscr{U}_F$ of Table 3

Relation	Type
$S^* \equiv (D_1D_2\neg D_3) \vee (D_1\neg D_2D_3) \vee (\neg D_1D_2D_3)$	Equivalence
$D^* \equiv S^* \vee (D_1D_2D_3)$	Equivalence

Table 6
Conditional exchangeability assumptions

Scenario	s	Pairwise equalities
F	1	$P(D_1D_2\neg D_3\|F) = P(D_1\neg D_2D_3\|F) = P(\neg D_1D_2D_3\|F)$
F	2	$P(D_1\neg D_2\neg D_3\|F) = P(\neg D_1\neg D_2D_3\|F) = P(\neg D_1D_2\neg D_3\|F)$
$\neg F$	1	$P(D_1D_2\neg D_3\|\neg F) = P(D_1\neg D_2D_3\|\neg F) = P(\neg D_1D_2D_3\|\neg F)$
$\neg F$	2	$P(D_1\neg D_2\neg D_3\|\neg F) = P(\neg D_1\neg D_2D_3\|\neg F) = P(\neg D_1D_2\neg D_3\|\neg F)$

We can consider the assessment $(\mathscr{F},\mathscr{L}_C,\mathbf{p})$ as a partial knowledge base whose main omission is the absence of an estimate for the expert's *specificity* $P(\neg D_i|\neg F)$. Anyhow, using the conditional independence assumptions

$$P(D_i|D_jF) = P(D_i|F) \quad \text{and} \quad P(D_i|D_j\neg F) = P(D_i|\neg F) \tag{23}$$

and thanks to some algebraic manipulation involving Bayes' theorem, Tweedie and Mengersen uniquely determine probability values for the usual accuracy indexes *specificity, positive predictive value, negative predictive value* and estimate the *true positive proportion*. In Table 8 we can compare their results with what we obtained firstly without any assumption, secondly adopting the method of Section 3.1 to incorporate the conditions of conditional exchangeability [7] listed in Table 6 and finally using considerations of Section 3.2 by adding the conditional probability comparisons reported in Table 7. These comparisons arise from a preliminary analysis [8] of the influence of the knowledge of the answers of some expert on our assessment of behaviors of the others. They were introduced with the help of a physician extraneous to the rest of the work.

Note that the comparisons in Table 7, even being similar in structure (the first three actually reflect odds ratios comparisons), are distinguished by labels, between those of them that are actually linear constraints, since some quantity is uniquely determined, and those that are properly quadratic and need the proposed linear approximation.

Inferential computations based on various different assumptions are displayed in Table 8. Whenever conditional exchangeability cannot help in limiting vague inference bounds, the further probabilistic comparisons can be influential. In fact, apart

[7] With respect to the notation of Section 3.1 we have $k = 3$, $E_i = D_i$ and C_j equal at first to F and after to $\neg F$.

[8] In the final version of the cited paper the comparisons have been refined involving tens of them. They are not reported here because their description and motivations go beyond the present scope to illustrate the benefits of the introduction of simple structural constraints.

Table 7
Conditional probabilities comparisons

Comparison	Type
$\frac{P(D_3\|D_1D_2F)}{P(\neg D_3\|D_1D_2F)} \geqslant 3/2\frac{P(D_1\|F)}{P(\neg D_1\|F)}$	(linear)
$\frac{P(D_3\|\neg D_1\neg D_2F)}{P(\neg D_3\|\neg D_1\neg D_2F)} \leqslant 2/3\frac{P(D_3\|F)}{P(\neg D_3\|F)}$	(linear)
$\frac{P(D_3\|\neg D_1\neg D_2\neg F)}{P(\neg D_3\|\neg D_1\neg D_2\neg F)} \leqslant 2/3\frac{P(D_3\|\neg F)}{P(\neg D_3\|\neg F)}$	(quadratic)
$P(D_3\|D_1\neg D_2F) \in [.5,.5+(P(D_3\|F)-.5)]$	(linear)
$P(D_3\|D_1\neg D_2\neg F) \in [.5-(P(D_3\|F)-.5),.5]$	(linear)
$P(D_2\|D_1F) \geqslant P(D_2\|F)$	(linear)
$P(D_3\|D_2D_1F) \geqslant P(D_3\|D_1F)$	(quadratic)
$P(D_2\|\neg D_1\neg F) \leqslant P(D_2\|\neg F)$	(quadratic)
$P(D_3\|\neg D_2\neg D_1\neg F) \leqslant P(D_3\|\neg D_1\neg F)$	(quadratic)

Table 8
Different inferences performed on several accuracy indexes, specifying the particular assumptions adopted

Index	Description	Extension bounds under			
		cond. indep.	*no ass.*	*cond. exch.*	*qual. comp.*
$P(\neg D_i\|\neg F)$	experts' spec.	.957	$[0,1]$	$[.603,1]$	$[.820,.970]$
$P(F\|D^*)$	pos. predict. val.	.961	$[0,1]$	$[0,1]$	$[0,.779]$
$P(\neg F\|\neg D^*)$	neg. predict. val.	.988	$[.970,1]$	$[.971,1]$	$[.979,1]$
$P(F)$	asbest. prevalence	.126	$[0,.130]$	$[0,.130]$	$[0,.106]$
$P(D^*\|F)$	med. dec. sens.	.994	$[.730,1]$	$[.730,1]$	$[.820,.878]$
$P(\neg D^*\|\neg F)$	med. dec. spec.	.995	$[.880,1]$	$[.880,1]$	$[.954,.970]$

from the positive predictive value, all the intervals in the last column are tight enough to evaluate the reliability of the median decision procedure and to have an idea about the fibrosis prevalence. About the only "vague" interval $[0,.779]$ for $P(F|D^*)$, even though it does not bound the positive predictive value from below, it does give an interesting upper limit for this performance index.

Moreover, note that some intervals in the last column do not contain the corresponding values obtained by Tweedie and Mengersen. This holds because the introduced further constraints of Table 7 contradict implications of conditional independence, allowing some kind of correlation among individual diagnoses, but leaving "untouched" the conditional exchangeability framework.

What we obtained has been based on reasonable probabilistic statements, avoiding the introduction of arbitrary restrictions which are motivated mainly by the desire to derive single values instead of intervals.

4.2. *Reliability of GIST diagnosis based on partial information*

Other prototypes of applications of inference with a not fully detailed model are medical diagnostic procedures where there is not a *gold standard* protocol to follow.

Table 9
Relevant events for the GIST diagnosis.

Label	Description
SUSPECTED	Lesion is histologically suspected to be a GIST
GIST	Lesion is really a GIST
CD117	KIT protein expression
CD34	Hematopoietic progenitor cell antigen expression
SMA	Muscle actin expression
DESM	Desmin expression
S100	S-100 protein expression

This happens when new advances in the understanding of biology are made or new techniques are discovered. In such situations, different opinions appear in scientific literature and they are based on disparate case studies, each one with its peculiarity and heterogeneity of data.

In particular, in [5] we analyzed a diagnostic process for *gastrointestinal stromal tumors* (GISTs) where only recently a new and reliable phenotypic marker (the KIT protein CD117) for these neoplasm has been introduced. The KIT protein is not adopted systematically with all the gastrointestinal lesions, but only to those that, after a first analysis, are suspected to belong to the GIST family. This procedure is followed because the KIT protein is extremely expensive. Moreover, due to its novelty, it has not already been adopted as standard technique.

More specifically, the diagnosis path consists mainly of two stages: at first a histological analysis is done and later an immunohistochemical schema is adopted to confirm cases previously suspected to be GISTs.

What we have done was to numerically evaluate the quality of the first discrimination stage. This was possible by integrating the results observed in an empirical study [9] with the immunohistochemical behaviors reported in the relevant literature.

The problem can be synthesized as follows (refer to [5] for a detailed report): we have selected as relevant for a lesion the events listed in Table 9 where the first two distinguish the suspected tumors from those actually belonging to the GIST's family, while the others represent the positivity for specific immunohistochemical markers.

We had only the following logical restriction due to the extreme specificity of the KIT marker

$$\mathrm{CD117} \subseteq \mathrm{GIST}. \tag{24}$$

[9] The data set consists of 47 mesenchimal neoplasm analyzed at *Istituto di Anatomia e Istologia Patologica, Divisione di ricerca sul cancro, Università degli Studi di Perugia, Italy* during the period January 1998–September 2002.

Table 10
GIST knowledge base estimated by observed frequencies in the data set

Statement	Conditional probabilities
SUSPECTED	.510
CD117 CD34 ¬DESM ¬ S100|SUSPECTED	.308
¬SMA ¬ CD117 CD34 DESM ¬ S100|SUSPECTED	.077
¬SMA CD117 CD34 ¬ DESM S100|SUSPECTED	.077
SMA ¬ CD117 CD34 ¬ DESM ¬ S100|SUSPECTED	.077
SMA CD117 ¬ CD34 ¬ S100|SUSPECTED	.231
SMA CD117 ¬ CD34 ¬ DESM S100|SUSPECTED	.077
¬SMA CD117 ¬ CD34 ¬ DESM S100|SUSPECTED	.077

Table 11
GIST rule base derived by literature

Statement	Expected frequencies bounds
CD34|CD117	[.60, .70]
SMA|CD117	[.30, .40]
S100|CD117	[.096, .105]
DESM|CD117	[.01, .02]

Using our limited available data we estimated (by observed frequencies) the "knowledge base" reported in Table 10. However, this turned out to be incoherent with the "rule base" of Table 11, which we derived [10] by collecting different literature sources ([16,22,23]).

Incoherence here mainly results from the fact that if we take only the knowledge base of Table 10 as an initial assessment, it induces, by extension, that coherent values for the percentage of S100|CD117 should be between 13% and 70%—while it should be around 10% according to the rule base Table 11. A similar incoherence would incur if, for example, we had the simple data base of Table 1 along with an assumed positive predictive value of the test $P(D|T)$ expected to be around 1/100. (This latter expectation contrasts with the conclusion we had that it must lie between 5/10,000 and 2/1000.)

Incoherence was solved by a revision of the data base. Among the cases showing S-100 positivity, there were two very doubtful. In fact, after a deeper analysis their classifications were changed to S-100 negative cases. Such revision has modified the knowledge base to that shown in Table 12.

In this context, a patholigist's opinion has induced us to add to the whole assessment also the further information

$$P(\text{CD117}|\text{GIST}) \in [0.95,0.99] \tag{25}$$

for the sensitivity of the KIT marker.

[10] The intervals reported were obtained by using the minimum and maximum values when there were discrepant reports, while by rounding to the third decimal digit the vague statements "it should be *around* the 10%".

Table 12
A new GIST knowledge base obtained after revision of two doubt S-100 positive cases in the data base

Statement	Conditional Probabilities
SUSPECTED	.510
CD117 CD34 ¬ DESM ¬ S100|SUSPECTED	.380
¬SMA ¬ CD117 CD34 DESM ¬ S100|SUSPECTED	.077
SMA ¬ CD117 CD34 ¬ DESM ¬ S100|SUSPECTED	.077
SMA CD117 ¬ CD34 ¬ S100|SUSPECTED	.077
SMA CD117 ¬ CD34 ¬ DESM S100|SUSPECTED	.077
¬SMA CD117 ¬ CD34 ¬ DESM ¬ S100|SUSPECTED	.077

Table 13
Coherent ranges for accuracy indexes

Index	Description	Extension ranges
$P(\text{SUSPECTED}\|\text{GIST})$	Sensitivity	$[.47, .76]$
$P(\neg\text{SUSPECTED}\|\neg\text{GIST})$	Specificity	$[0, .88]$
$P(\text{GIST}\|\text{SUSPECTED})$	Positive predictive value	$[.85, .94]$
$P(\neg\text{GIST}\|\neg\text{SUSPECTED})$	Negative predictive value	$[0, 69]$

Taking as basic assessment all the information contained in Tables 11 and 12, along with (25) and (24), we obtain by extension the ranges for usual accuracy indexes listed in Table 13.

In this case we have 40 atoms and only one zero layer. Hence the optimization task $\{\mathcal{O}^j_\alpha\}$ consists of 10 single (i.e. sequences of length 1) linear programs (each one "associated" to the bounds of the four intervals in Table 12 and of the interval in (25)) with 18 constraints (the 17 induced by the assessment plus one of normalization).

Note that the rather broad interval results in Table 13, apart from the positive predictive value, reflect a weak influence of the constraints upon the assessment. Moreover, by adding the further probabilistic comparison $P(\text{SUSPECTED}|\ \text{GIST}) \geqslant P(\text{SUSPECTED}|\neg\text{GIST})$ we have not obtained appreciable improvements.

On the contrary, reasoning as described in Subsection 3.3, we have focused the attention on the "a priori" values of GIST's prevalence. In fact, its coherent extension requires only that $P(\text{GIST}) \in [.59, .97]$. Nonetheless, in one of the 10 optimization programs we obtained the more informative lower bound $lb^j_{GIST} = .81$. This has suggested us to ask the pathologist if it was reasonable to restrict the value of GIST's prevalence to be around 81%. Since the answer was positive, we added [11] to the whole assessment the restriction

$$P(\text{GIST}) \in [.806, .815], \tag{26}$$

which would cohere with the rest.

[11] This addition does not influence the number of atoms, which remains at 40. However, it requires the extensions to be performed by 12 linear programs with 20 constraints each.

Table 14
Improved ranges for accuracy indexes obtained by the addition of (26) to the initial assessment

Index	Description	Extension bounds
P(SUSPECTED|GIST)	Sensitivity	[.53, .59]
P(¬SUSPECTED | ¬GIST)	Specificity	[.58, .80]
P(GIST|SUSPECTED)	Positive predictive value	[.85, .93]
P(¬GIST | ¬SUSPECTED)	Negative predictive value	[.22, .32]

In such a way we obtained the sharper results reported in Table 14 that confirm a good positive predictive performance of the diagnostic procedure, although they express really poor reliability in the case of a negative diagnosis. This, in a way, reverses the role that the KIT marker should have. Instead of being used as a *confirmatory* tool in already suspected cases, it should have a crucial role for the right diagnosis of lesion at first not suspected to be GISTs. This very valuable insight was obtained from the methods we have highlighted in this paper.

5. Conclusions

In this paper we have shown that even in presence of very limited information, diagnoses based on model-free procedures are possible. This is due to the adoption of coherent conditional probability assessments. Reliability of the diagnoses can be improved adding to the initial assessments reasonable structural constraints. Through two medical applications it has been shown how conditional exchangeability judgements, conditional probability comparisons and indirect restrictions of the admissible classes of conditional distributions bring considerable benefits to our understanding of diagnostic processes, shrinking the coherent bounds for the probabilities of informative indexes.

Acknowledgment

I am very grateful to Frank Lad for clarifying discussions and precious suggestions. My appreciation extends to referees for their valuable comments.

References

[1] M. Baioletti, A. Capotorti, S. Tulipani, B. Vantaggi, Elimination of Boolean variables for probabilistic coherence, Soft Comput. 4 (2) (2000) 81–88.

[2] M. Baioletti, A. Capotorti, S. Tulipani, B. Vantaggi, Simplification rules for the coherent probability assessment problem, Ann. Math. Artif. Intell. 35 (2002) 11–28.

[3] M. Baioletti, A. Capotorti, S. Tulipani, Polynomial average time complexity of an algorithm for 2CPA. Inform Computat (submitted).

[4] M. Baioletti, A. Capotorti, P. Tiberi, S. Tulipani, An empirical complexity study for a 2CPA solver, in: Proceedings of the 10th International Conference IPMU 2004, Perugia, Italy, vol. III, 2004, pp. 1857–1864.
[5] A. Capotorti, S. Fagundes Leite, Reliability of GIST diagnosis based on partial information, in: M. Di Bacco, G. D'Amore, F. Scalfari (Eds.), Applied Bayesian Statistical Studies in Biology and Medicine, Kluwer Academic Publishers, Boston/Dordrecht/London, 2003.
[6] A. Capotorti, L. Galli, B. Vantaggi, How to use locally strong coherence in an inferential process based on upper–lower probabilities, Soft Comput. 7 (5) (2003) 280–287.
[7] A. Capotorti, F. Lad, Reassessing accuracy rates from median decision procedures, The American Statistician (submitted).
[8] A. Capotorti, T. Paneni, An operational view of coherent conditional previsions, in: S. Benferhat, P. Besnard, (Eds.) Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQUARU 2001, LNAI 2143, 2001, pp. 132–143.
[9] A. Capotorti, B. Vantaggi, Locally strong coherence in an inference process, Ann. Math. Artific. Intell. 35 (1-4) (2002) 125–149.
[10] A. Capotorti, S. Gardi, Conditional exchangeability and locally strong coherence, in: Proceedings of the of 6th Workshop on Uncertainty Processing in Expert Systems WUPES'03, Hejnice, Czech Republic, 2003, pp. 13–24.
[11] G. Coletti, R. Scozzafava, Characterization of coherent conditional probabilities as a tool for their assessment and extension, Int. J. Uncertain., Fuzz. Know.-based Syst. 4 (2) (1996) 103–127.
[12] G. Coletti, R. Scozzafava, Conditional measures: old and new, in: New Trends in Fuzzy Systems, World Scientific, Singapore, 1998, pp. 107–120.
[13] G. Coletti, R. Scozzafava, Conditioning and inference in itelligent systems, Soft Comput. 3 (1999) 118–130.
[14] G. Coletti, R. Scozzafava, The role of coherence in eliciting and handling imprecise probability and its application to medical diagnosis, Inform. Sci. 130 (2000) 41–65.
[15] G. Coletti, R. Scozzafava, Probabilistic logic in a coherent setting, Series "Trends in Logic", Kluwer, Dordrecht, 2002.
[16] C.D.M. Fletcher, J.J. Berman, C. Corless et al., Diagnosis of gastrointestinal stromal tumors: a consensus approach, Human Pathol. 33 (5) (2002) 459–464.
[17] ISIPTA'99, in: G. de Cooman, F.G. Cozman, S. Moral, P. Walley, (Eds.), Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications, Imprecise Probabilities Project, Universiteit Gent, Belgium, 1999.
[18] ISIPTA'01, in: G. de Cooman, T.L. Fine, T. Seinfeld, (Eds.), Proocedings of the Second International Symposium on Imprecise Probabilities and Their Applications, held at the Robert Purcell Community Center of Cornell University Ithaca, NY, USA, 26–29 June 2001, Shaker Publishing, The Netherlands, 2001.
[19] F. Lad, Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction, John Wiley, New York, 1996.
[20] F. Lad, M. Di Bacco, Assessing the value of a second opinion: the role and structure of exchangeability, Ann. Math. Artif. Intell. 35 (2002) 227–252.
[21] R. Tweedie, K. Mengersen, Calculating accuracy rates from multiple assessors with limited information, Am. Statist. 53 (1999) 233–238.
[22] M. Miettinen, M. Sarlomo-Rikala, J. Lasota, Gastrointestinal stromal tumors: recent advances in understanding of their biology, Human Pathol. 30 (10) (1999) 1213–1219.
[23] M. Miettinen, J. Lasota, Gastrointestinal stromal tumors: definition, clinical, histological, immunohistochemical, and molecular genetic features and differential diagnosis, Virch. Arch. 438 (2001) 1–12.
[24] P. Walley, R. Pelessoni, P. Vicig, Direct algorithms for checking coherence and making inferences from conditional probability assessments, Journal of Statistical Planning and Inference 126 (1) (2004) 119–151.
[25] N.V. Thoai, Duality bound method for the general quadratic programming problem with quadratic constraints, J. Optim. Theory Applic. 107 (2) (2000) 331–354.
[26] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.

[27] P. Walley, R. Pelessoni, P. Vicig, Direct algorithms for checking coherence and making inferences from conditional probability assessments, Quaderni del Dipartimento di Matematica Applicata alle Scienze Economiche, Statistiche e Attuariali ‘‘Bruno de Finetti’’, 6/99, 1999. Available from <http://www.univ.trieste.it/matappl/quaderni.htm>.

[28] M. Zagoraiou, Restrizioni coerenti di estensioni di probabilità coerenti. Tesi di Laurea del corso in Matematica, Università degli Studi di Perugia (supervisor A.Capotorti), 2004.