N-Tuplets in computer diagnosis


Birmingham

TRADITIONALLY all clinical assessment is based on the elucidation of sufficient signs, symptoms and tests to provide a cluster, pattern or N-Tuplet of descriptive criteria which can be matched with a complementary N-Tuplet or complex of signs, symptoms and tests appropriate to one specific diagnosis or diagnostic group with a high probability that it is correctly assigned.

An N-Tuplet consists therefore of a cluster of basic descriptive criteria. In this context, N-Tuplets are specific for some medical phenomenon whether an undefined clinical situation, syndrome or formal diagnosis.

However, appropriate clinical action can still be initiated without benefit of a specific diagnosis in much the same way that all animals, including man, make appropriate behavioural responses to environmental situations without the benefit of an intellectual assessment of that situation in terms of the specific factors which are interacting to produce it. The behavioural response is initiated directly by an appropriate perceptive pattern. This perceptive pattern is analogous to the sign-symptom N-Tuplet or complex in the diagnostic process and therapeutic action is analogous to appropriate behaviour. The indirect, intellectual establishment of the N-Tuplet of interacting factors in any problem situation is analogous to the specific diagnosis in terms of morphology, aetiology and pathology. The indirect establishment of the interacting factors in life situations not only enables a larger range of different environmental situations to be analysed, and different behavioural responses to be generated than could be done without this intermediate step, but it also results in an enhanced probability that the most appropriate behavioural response will be correctly assigned to each perceived environmental situation compared with the more simple mechanism of directly linked perceptive and behaviour patterns. Natural language is the basis for this indirect mechanism which distinguishes man from all other animals. In a similar way, a diagnostic system based on the intermediate establishment of a formal diagnosis or N-Tuplet, whose items are in terms of morphology, aetiology and pathology enhances the range of clinical situations which can be diagnosed as well as increasing the range of therapy and the specificity of the actual therapeutic action chosen in any given situation.

In this paper, some of the problems involved in the formal diagnostic process will be examined. A later paper will explore the simpler direct system where therapeutic action is initiated without the benefit of a prior formal diagnosis. The upper respiratory tract has been chosen for this first study, for reasons which will be elaborated later.

Traditionally, the complex of signs, symptoms and tests (N-Tuplets) appropriate to any specific diagnosis or at any rate the assessment end-point for any particular class of clinical problems, has been determined for good empirical and economic reasons, by the aberrations recognized by the patient himself, supplemented by additional signs and symptoms elicited by the practitioner. The process of eliciting these additional signs and symptoms is guided by the restricted range of all possible diagnoses with which the presenting symptoms may be associated. This in turn guides the subsequent search for elicited data. In this the step-like process of presenting data, all possible associated

*Material based in this study was submitted by Mr K. Dobell, as part requirement for the degree of M.Sc. (operational research), Department of Engineering Production.

diagnoses, elicited data, a second, more restricted set of possible diagnoses (though obviously sometimes an expanded set) and of further elicited data, proceeds until an N-Tuplet of sufficient appropriate data has been accumulated to indicate one specific diagnosis or diagnostic grouping with a high probability of correct assignment and to have eliminated all possible serious alternatives with degrees of probability related to their respective seriousness and treatability.

The segregation process relies on the absence of specific signs, symptoms and positive tests, as well as on their presence, possible alternative diagnoses being eliminated by failure to elicit some individual sign, symptom or test, or complex of such signs, symptoms or tests which have a high probability of being associated with that disease.

This step-like process is necessary for reasons of economy of skilled physicians' time. If the diagnosis of every common cold could only be established by working through a check list of all possible signs, symptoms and tests, few diagnoses of any sort would ever be made. For the same reason, when the aim of clinical assessment is mainly the elimination of more serious possibilities, as in the example of the common cold, the final diagnosis may indicate only a broad diagnostic grouping and the elimination of alternative possibilities be only implied.

Until recently, biochemical and other machine derived indices have been linked to and based on the primary clinical features of disease. Biochemical tests, whether as elicited or screening data, in diabetes, renal disease and cardiorespiratory failure are examples. However, in the recent exploitation of screening systems based on machine-derived data, the establishment of further relationships between abnormal ranges of these indices, singly or in complex patterns or N-Tuplets, must rationally be based on the frequencies with which these indices are associated or vary with clinical abnormality. These frequencies can only be established from a representative sample of all individuals, normal presymptomatic and symptomatic for any given clinical entity.

However, the signs and symptoms which are the units with which clinical practice traditionally operates, have never been evaluated numerically in this way and the specific patterns or N-Tuplets used for matching have evolved by the evolutionary process of trial and error using clinical tradition as the memory device for the system. Machine-derived indices have tended to be tacked on as an afterthought. The human brain, in problem solving, has a preference for data sequentially associated rather than for data associated statistically.

A rational assessment system must not only treat all relevant indices whether they be signs, symptoms or machine-derived, whether they have been presented by the patient, resulted from primary screening, or were elicited by a clinician, with equal weight, but clinicians must get used to thinking in an integrated way about all indices. This rational approach implies that the whole basis of clinical assessment must be reviewed and the total incidence of signs and symptoms numerically evaluated in the way suggested for machine-derived data. Machine-derived data can include far more than biochemical estimations. The answers to screening questionnaires must obviously be included.

The integration process has progressed to a stage where computers are used to scan the machine-derived primary screening data and these indices individually or in N-Tuplets or patterns, identified which are known from past experience to have a high probability of being associated with some clinical abnormality in its presymptomatic or symptomatic phase. The physician, with this information, then meets the patient and conducts a more or less traditional clinical interview.

In this brief analysis of clinical-problem solving we have identified various main components. The most important are probably the mechanisms of selection which underlie the process of maximizing the probability of reaching the correct solution to the clinical problem with the minimum of clinical effort. The solution may not even involve
the making of a formal diagnosis. However, underlying this and providing the basis for the whole of clinical problem solving are the relative frequencies or probabilities with which accessible data in the form of signs, symptoms or tests, singly or in interacting N-Tuplets are associated with appropriate clinical action in the form of advice or treatment. Clinical medicine, however, is based on an empirical tradition of establishing a diagnosis, or at least classifying the clinical problem, however broadly, according to its aetiology, pathology and morphology. This is traditionally a first link in a chain which can then move logically to appropriate treatment. This logical approach, certainly in general practice, where 90 per cent of all clinical decisions are taken, is more often than not honoured in the breach. This approach is only logical in that the preference for sequentially rather than statistically associated data dictates such intermediary steps if a diagnosis is to be made consistently. By using a completely systematic technique employed by an unthinking computer, inconsistencies in disease classifications and diagnoses will be identified which might otherwise be concealed by allowances the intelligent doctor would automatically make. This initial study, which concerns only the upper respiratory tract, will be extended to cover the whole respiratory tract; any overlap which exists in the symptomatology between the respiratory tract and all other illness, and the treatment given or clinical action taken.

Outline of problem and method

This study consists of two essential independent sections. A computer diagnosis of clinical problems involving the upper respiratory tract was attempted, as an interesting exercise in its own right, and as a method for examining possible underlying fundamental inconsistencies, in clinical diagnostic processes. If symptoms could be assumed to be independent, and all the knowledge about a case was included in our data, the statistical technique used for diagnosis is known to be self-consistent. In addition, an investigation of symptom interactions was carried out. It seemed likely to the authors that clinicians use N-Tuplets (sign-symptom patterns or complexes) implicitly in diagnosis, where the combination of signs, symptoms, and tests is more specific for a given disease than would be expected from examination of the single items. Determination of these interactions and the diseases in which they occur would confirm the utility of this approach, and is necessary for investigating the validity of the computer diagnostic procedure.

The study attempts to answer these questions:
1. Can computer diagnosis be applied successfully to upper respiratory tract infections?
2. Does a systematic diagnostic procedure suggest possible inconsistencies in clinical classification and diagnosis?
3. Does symptom interaction exist, and if so in what disease classes does the same interaction exist?

If these questions can be answered usefully in this rather ill-defined disease area, it should be possible to apply similar techniques in other more precisely-defined disease areas with equal or greater success.

Method of enquiry

A form listing 43 signs and symptoms was completed by 20 general practitioners, members of the Royal College of General Practitioners interested in research. They were asked to complete a form for every new patient who presented during the study period with an upper respiratory tract illness or any other illness in which any of the listed symptoms occurred. Apart from the symptoms, they also recorded a numerical indication of the certainty of their diagnosis:

Category 1 was used where there was no diagnosis, where the diagnosis took the form of a symptom complex, was tentative only, or where there were multiple possibilities.

Category 2 was used where the diagnostic process was eliminative, that is as in 1 above, but where all serious possibilities have been eliminated.

Category 3 was used only where a firm diagnosis was made. Multiple forms were requested for patients suffering from multiple illnesses.
They did not, in this study, check the entire list of symptoms, nor were all diseases, other than those in the upper respiratory tract, where any of the listed signs or symptoms occurred, systematically recorded. No check was possible on the diagnosis or recorded symptoms; the study is based on acceptance of the data. This is an inherent difficulty in any study dealing with minor illness.

The college classification of morbidity was used. In this study, no attempt was made to establish and use standardized definitions of the diseases in terms of specific signs and symptoms.

**Statistical methods**

*Computer diagnosis*

Bayes formula is used for computing the probability that a set of data or N-Tuple will be associated with a particular hypothesis (disease group or diagnosis) when the probability of occurrence of that N-Tuple or set of data given the hypothesis is known, and the unconditional probability associated with each hypothesis is known. In this case, independence of symptoms is assumed, and the applicable formula is:

$$p(D_j/s_1,s_2,...s_n) = \frac{p(D_j) \cdot p(s_1,s_2,...s_n/D_j)}{\sum_j p(D_j) \cdot p(s_1,s_2,...s_n/D_j)}$$

where

- $p(D_j/s_1,s_2,...s_n)$ is the probability of disease $j$ given the symptoms $s_1,s_2,...s_n$
- $p(s_1,s_2,...s_n/D_j)$ is the probability of the symptoms $s_1,s_2,...s_n$ occurring given disease $j$
- $p(D_j)$ is the unconditional probability of occurrence of disease $j$

By recording case histories, these probabilities can be built up from symptom frequencies in disease groups and disease group frequencies in the total sample. If we know the frequencies, $f$, of symptoms 1, 2, 3 and 4 in disease group 1, and we have "n" cases of disease 1 making up 20 per cent of our sample, then

$$p(D_j) = .2$$

$$p(s_1,s_2,s_3,s_4/D_j) = f_1f_2f_3f_4/n^4$$

The assumption of independence is known to be invalid. In clinical medicine, the items of data which constitute each N-Tuple are signs, symptoms and tests. Many such items, for example 'earache' are symbols representing a true N-Tuple consisting of more basic data, where these have been noted intuitively to occur together so frequently that a single symbol can be used to identify them. They are already complex N-Tuples, and some of them, for example earache, toothache, dyspepsia (or more specifically, ulcer dyspepsia), angina of effort, intermittent claudication, are so complex that they almost defy description in terms of individual signs and symptoms as units and require a conversational type of description for economic classification.

Notwithstanding these reservations, it was still considered worth while to investigate the relative specificity of diagnostic systems using the Bayes formula, as outlined above, since investigations suggested that the effect of symptom interaction might not be too severe. Warner using congenital heart disease as the subject, was the first to apply Bayes theorem to medical diagnosis.

*Tests for symptoms interaction*

All symptom doublets were tested for interaction by the use of a $2 \times 2$ contingency table and a $\chi^2$ test. Each doublet (an N-Tuple of two sign-symptom items) was considered over disease groups 1–7 (table I). The frequencies of all triplets was examined, and all triplets with a frequency in any disease group 1–7 higher than expected on
independence was recorded, though no statistical tests were carried out.

Computer diagnosis

Two approaches to the problem were attempted: the first considered only the symptom present when applying the Bayes formula. The second, assumed that absence of the symptom on the survey form could, for specified reasons, be taken to mean absence of that symptom in the patient. The absence of one of these symptoms—pyrexia, running nose, sore throat, enlarged tonsils, pus on tonsils, red pharynx, enlarged adenoids, and cold—was then regarded as a symptom in its own right. The results from both these approaches are shown in table I; bracketed figures refer to the second approach. Five cases were undiagnosed, since the exclusion of that particular case from the store resulted in a zero probability in all disease groups for that combination of symptoms.

Before inclusion of symptom absence, the medical diagnosis was reproduced in 60.5 per cent of cases. About 18 per cent of category 1 (febrile common cold) cases were placed in category 2 (pharyngitis), about 18 per cent of category 2 in category 1 and about 46 per cent of category 5 cases in either 1 or 2. With symptom absence, 67.5 per cent of diagnoses were reproduced and, as shown in the table I, the number of cases incorrectly

<table>
<thead>
<tr>
<th>Medical diagnosis</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>U</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Febrile common cold</td>
<td>64</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>79</td>
</tr>
<tr>
<td>2. Pharyngitis</td>
<td>49</td>
<td>195</td>
<td>(58)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>261</td>
</tr>
<tr>
<td>3. Sinusitis</td>
<td>12</td>
<td>11</td>
<td>(1)</td>
<td>(17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>4. Tracheitis</td>
<td>16</td>
<td>7</td>
<td>(2)</td>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43</td>
</tr>
<tr>
<td>5. Non-febrile common cold</td>
<td>48</td>
<td>20</td>
<td>(0)</td>
<td>(22)</td>
<td>(10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>145</td>
</tr>
<tr>
<td>6. Otitis media</td>
<td>1</td>
<td>9</td>
<td>(9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>7. Bronchitis</td>
<td>7</td>
<td>1</td>
<td>(7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>8. Otitis externa</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>9. Adenitis</td>
<td>1</td>
<td>9</td>
<td>(1)</td>
<td>(7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>10. Other disease of Eustachian tube, middle and outer ear</td>
<td>1</td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>11. Hay fever</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>199</td>
<td>254</td>
<td>13</td>
<td>16</td>
<td>521</td>
<td>169</td>
<td>39</td>
<td>14</td>
<td>42</td>
<td>7</td>
<td>10</td>
<td></td>
<td>662</td>
</tr>
</tbody>
</table>

placed in category 1 fell considerably. The improvement in diagnosis occurred in categories 2, 3, 4 and 5 (pharyngitis, sinusitis, tracheitis and non-febrile cold); this suggests that these groupings are defined by the absence of contradictory symptoms as well as the presence of corroborative symptoms. The absence of pyrexia will obviously discriminate between categories 5 and 1; the absences of other symptoms appear to play a similar rôle between categories 2 and 1, 3 and 1 and 4 and 1.
Category 4 (tracheitis) is the main example of disagreement between the subjective methods of the clinician, and statistical diagnosis; even with the incorporation of symptom absence, only in nine cases out of 43 was the medical diagnosis reproduced. This disagreement probably arises from the ill-defined nature of tracheitis; because the trachea lies anatomically between the upper and lower respiratory tract, inflammatory processes involving the trachea are almost always secondary to primary disease elsewhere. Similar factors play a rôle in the poor statistical diagnosis of cervical adenitis (category 9), though the limited number of cases of cervical adenitis in the store and its relatively low frequency overall, are also partly responsible. An analysis of inter-doctor inconsistencies is shedding further light on these anomalies, and will be discussed in a later paper.

**N-Tuplets and negative doublets**

Statistically significant positive association between symptoms was discovered in 68 doublets; this association occurred in more than one disease in only eight of these doublets. In the triplets search, 356 triplets were found to occur more frequently than would be expected on a hypothesis of complete independence. Negative correlation between symptoms was found, but the significance level of these interactions was in general much lower than that found for positive associations.

Obvious doublets such as red pharynx, sore throat (8–23), pain on swallowing, sore throat (4–8), and enlarged tonsils, pus on tonsils (21–22) were found to be significant in certain disease categories. In these examples, p (doublet/appropriate disease) would be underestimated by approximately a factor of two if independence was assumed.

**Discussion**

In a discussion of the results obtained in this study, the problem of data reliability is of considerable importance. In any field of medicine, diagnosis is more or less uncertain; in this particular area, virtually no confirmatory evidence is available. Diseases considered here are generally self-limiting; in many illnesses treatment is similar, so that an imprecise diagnosis provides a guideline for treatment, and is clinically acceptable. This inherent uncertainty must be taken into account in the evaluation of these results. It seems likely that doctors continue examination only until enough information to indicate appropriate therapeutic action is obtained.

The occurrence of statistically-significant doublets, negatively and positively associated, has been discussed. However, many possible doublets were tested in each of the first seven disease categories; some spurious associations are bound to arise. The perennial question of the significance of 'statistical significance' appears in this context.

Most of the positively associated doublets are extremely significant and many can be justified on physiological grounds; it seems quite reasonable to ascribe genuine significance to these.

**Future work**

This study has shown the existence of significant doublets; it would be desirable to extend this to higher order interactions. A multi-dimension contingency table should provide the information required.

The problem of narrowing the total area to be considered was discussed in the introduction; this problem is relevant for any mechanical diagnostic system, since consideration of N-Tuplets and singlets over even a limited number of symptoms and disease multiplies storage requirements and computation considerably. Since the Bayes approach especially as modified in the manner outlined above, seems likely to provide good results, a hierarchical series of Bayes calculations is a reasonable approach to this problem. Different symptoms would be required in the initial stages of the search; they would of course be more general, and would be chosen for their specificity to a given
N-TUPLETS IN COMPUTER DIAGNOSIS

area of illness. If data could be obtained for lower respiratory tract illnesses, and a sample of all illnesses in which the symptoms occurred and a Bayes model set up, a two layer model could be attempted. Designing an over-lapping symptom set, and incorporating a decision process to branch into either upper or lower respiratory tract would be an interesting and useful exercise.

This project has concerned itself primarily with reproducing medical diagnoses. However, when a clinician makes a diagnosis as a basis for prescribing treatment, he considers one more factor—the value of a diagnosis. If a disease is untreatable, diagnosis is of limited value to the patient; if a disease is easily cured, diagnosis is essential. Similarly, the importance of diagnosis varies with the seriousness of the disease. It should be possible to attach numerical values to—say—the importance of a missed diagnosis in each disease category. By obtaining the distributions of the estimates in the Bayes equation, it then becomes possible to determine a diagnosis which minimizes the total missed diagnosis value. This again would be an interesting and useful project.

Conclusions and summary

In the diagnosis of diseases affecting the upper respiratory tract it is concluded that:

1. Computer diagnosis can be applied in this disease area with moderate success; the accuracy of diagnosis (60.5 per cent before inclusion of symptom absence and 67.5 per cent after inclusion) compares with that obtained by computer in other disease areas.
2. Doctors use absence as well as presence of symptoms.
3. Significant symptom doublets do exist, and significant higher order interactions seem likely; a given doublet is usually significant in only one disease category.
4. The greatest inconsistency between clinical and statistical diagnosis was found in tracheitis and cervical adenitis.

Probabilities on independence, observed probabilities and significance levels for doublets as tables, are available on request to Dr D. L. Crumbie, general practice research unit, The Royal College of General Practitioners, c/o Birmingham Regional Hospital Board, 146 Hagley Road, Birmingham, 16.

REFERENCES


After several years of experimental schemes involving individual practices, a scheme was introduced in Southampton in 1968, whereby all district nurses, midwives and health visitors were totally attached to general practices in the area. In the case of the health visitors considerable administrative difficulties were encountered, but these have been overcome and the scheme is now running smoothly. An initial finding is a sharp increase in the amount of work undertaken by the district nurses. Routine visits by nurses to individual doctors' surgeries are not necessary. "The essence of establishing real and effective doctor–nurse teams lies in each doctor knowing the names of the community nurses who will work with him in caring for the patients on his list, and not in argument about where records are to be kept or communications passed."