Communication of Uncertainty in
Clinical Genetics Patient Health Communication Systems

Nancy Green, Tami Britt, and **Karen Jirak**
Department of Mathematical Sciences, Computer Science Division
University of North Carolina at Greensboro
Greensboro, NC 27402 USA
nlgreen@uncg.edu

Abstract

This paper presents an overview of communication of uncertainty in genetic counseling patient letters, and its implications for the design of patient-tailored communication systems. A system designer should take into account the prevalence of uncertainty expressions. Many can be analyzed as giving conditional probabilities. Often, the probability values are qualitative or imprecise. Thus, we discuss use of a hybrid Bayesian/Qualitative probabilistic network domain model for this type of system.

Introduction

This paper presents an overview of communication of uncertainty in a corpus of patient letters written by genetic counselors to their clients. The typical patient letter summarizes (possibly quite complex) information about a case discussed during the counselor's meeting with a client (Baker et al. 2002). The goal of our study is to inform development of intelligent multimedia systems for patient-tailored health communication in clinical genetics. As a first step, we are designing GenIE, a prototype system that will generate the first draft of a patient letter as a tool for genetic counselors. In addition, GenIE could be used to brief physicians on communicating with a patient about a genetics-related diagnosis or risk assessment. (Note that the role of GenIE would not be to make diagnoses or assess risk, but to help communicate that kind of information to the lay person.) In future work, we would like GenIE to create patient-tailored supplementary materials in other media. For example, spoken dialogue with animation of genetic mechanisms might be more effective than a letter alone for low-literacy and/or low-numeracy clients.

Our work takes the writings of authors considered experts at patient communication in this domain (i.e. the letters in our corpus) as a starting point. For analytical purposes, we have encoded letters from the corpus using a Bayesian network (BN) based coding scheme for

representing their biomedical content (Green 2003). One goal was to help design GenIE's knowledge base, which we planned to implement as a BN. Given information about a client's case entered into GenIE by the counselor (e.g., a counselor could input symptoms, test results, and diagnosis through user interface menu selections), the system would use general domain knowledge stored in the BN when drafting a letter documenting and explaining the clinic's diagnosis.

The other goal in encoding the letters was to provide an abstraction of the letters' contents to enable linguistic regularities to be analyzed without reference to particular genes, proteins, health problems, etc. Knowledge gained from that linguistic analysis is needed to build the natural language generation component (Reiter and Dale 2000) of GenIE. Linguistic analysis of the corpus has shown the prevalence of uncertainty expressions in communication between genetic counselors and their clients. Many uses of uncertainty arise from the probabilistic nature of this domain, as in many other areas of medicine. Also, the analysis shows a more frequent use of qualitative and imprecise expressions of uncertainty than use of exact numeric probabilities.

In the first part of the paper, we summarize sources of uncertainty in clinical genetics and survey general problems in the communication of uncertainty. Then we categorize uses of uncertainty in the corpus. The second half of the paper investigates the applicability of qualitative probabilistic networks (QPNs) to our goals.

Scope of Problem

Sources of Uncertainty in Clinical Genetics

When genetic counselors meet with their clients, a common topic of discussion is the *diagnosis* of a genetic basis for a patient's health problems. The diagnosis is the result of evidential reasoning, where the patient's pedigree, symptoms, other medical history, and test results may provide probabilistic evidence for or against various diagnoses. For example, the patient's symptoms could be indicative of a problem with a certain gene even

though the patient's test result for some specific mutation of that gene was negative; i.e., the test result has not ruled out other mutations of that gene. Or, even if a patient's test result is positive, the test may give a certain proportion of false positive results. Or, evidence from a patient's medical history may include his mother's ethnic background, and it may be known from a previous study that 11% of persons of the same ethnicity were carriers of a particular genotype (thus, increasing the chance that the patient has the genotype, inherited from his mother).

Another common topic is *prediction of the effects* of a particular genotype on patient health. Such predictions are probabilistic due to biological factors such as

- penetrance, the probability that a particular gene will result in any observable traits (at any time in an individual's life or at a particular age of onset),
- expressivity, i.e., the severity of a genetic disorder may differ between individuals,
- co-dominant inheritance, where the abnormal allele has an additive effect, e.g. the effect on an individual who has two abnormal copies of an allele is greater than the effect on someone who has only one copy,
- multifactorial disorders, where effects of one gene depend on other genes and/or environmental factors.

The third major source of uncertainty in this domain is due to the *uncertainty in predicting inheritance* itself. In Mendelian inheritance patterns, the probability that an offspring will have a certain genotype is a function of the parents' genotypes. Other sources of uncertainty include:

- non-paternity, i.e., where the biological father is not the alleged father,
- new mutations, e.g., as parental age increases there is increased risk for a new mutation resulting in a genetic disorder in the offspring; mutations can occur in the parent's germline as well as in the developing child after conception.

Lastly, some types of testing may have risks to the patient or fetus. Thus, a patient's decision to undergo testing may require knowledge of the likelihood of possible *complications* and their seriousness.

Communication Problems Involving Uncertainty

In describing the following problems, we shall distinguish two types of communication problems, linguistic and conceptual, and problems faced by two audiences, expert and lay. (In the context of genetic counseling, we assume that in addition to the difference in training, another significant difference between the expert and lay audience is that the lay audience may have a high degree of emotional involvement in the information communicated.)

As for linguistic problems in conveying uncertainty to experts, research on the generation of qualitative explanations from BNs has addressed the problem of converting numeric probability values to non-numeric expressions such as adjectives and adverbs; e.g., see

previous work surveyed in (Druzdzel 1996). According to Druzdzel (1996) empirical studies, where the subjects were experts, found some within-subject consistency in interpretation of expressions of qualitative probability, but lower between-subject consistency; and that interpretation may be influenced by context.

Grimes and Snively, citing (Bloch et al 1979), claim that "Recipients of genetic counseling [on risk of birth defects] prefer numeric to nonnumeric information" (1999, p. 913). This is problematic for a lay audience, since Grimes and Snively report that previous studies have shown limitations in the general public's understanding of "fractions, large numbers, or percentages (Dewdney 1993)". For example, they cite research showing that proportions with larger denominators may be assessed mistakenly as greater than those with smaller denominators (Walker 1997). Furthermore, Grimes and Snively performed an experiment showing that patients' ability to compare magnitude of risk was more successful when risks were expressed as rates than when expressed as proportions. Although rates (e.g. three per 1000) are the usual form given in epidemiologic studies, Grimes and Snively state that genetic counselors typically communicate risk information as a proportion (a fraction with a numerator of one, e.g., one in 333) in the belief that proportions are easier for clients to understand.

Psychologists have shown conceptual problems in assessing probability, both for experts and non-experts (e.g. see papers in Kahneman, Slovic, and Tversky 1982). Assessments may be inaccurate or biased due to use of cognitive heuristics. For example, subjects overestimate the likelihood of events that are more salient, or more typical, or which can be made to fit a causal scenario. In addition, comprehension by non-experts in a clinical setting may be impeded by the client's negative emotional response to the information. Sarangi et al. (2003) identify six discourse strategies in a corpus of taped genetic counseling sessions; the strategies are used to present risk information in different ways depending on whether the speaker's goal is informational or psychosocial.

Corpus Study

We have collected a corpus of patient letters, currently composed of letters written by three genetic counselors from different clinics and one published letter designed to illustrate writing guidelines for genetic counselors (Baker et al. 2002). In this section, we report our preliminary qualitative analysis of uses of expressions of uncertainty in the corpus. We shall address the following questions:

- Form: How is uncertainty expressed?
- Function: When (why) is uncertainty used?
- How does form relate to function?

Before turning to these questions, we briefly describe a coding scheme we developed that enables the biomedical

content of a letter to be modeled as a BN (Green 2003). For general information on Bayesian networks see, e.g., (Korb and Nicholson 2004).

Coding Scheme

Phrases of a letter are annotated with tags referring to nodes of a BN representing the relations among concepts discussed in the letter. A tag is composed of the node type, a numeric subscript, and a suffix indicating the individual described by the tagged phrase. The following types of nodes are defined: History, Genotype (a pair of alleles of a gene), Mutation-Event, Biochemistry, Physiology, Symptom, and test Result. In addition, a Probability tag is used to annotate quantitative or qualitative phrases that express or imply belief in the past, current, or future possibility, frequency or probability of an event. (The current coding scheme does not distinguish different forms of numeric values.) Qualitative expressions may belong to a number of lexical categories including determiners, adjectives, modals, verbs, and adverbs.

From the tagged phrases, a coder can construct a BN graph representing the causal structure of the information conveyed in the letter (or just probabilistic dependencies between variables). For example, the following phrase in letter NF of the corpus, *Approximately 80% of individuals affected with NF have mild to moderate symptoms*, is shown in Figure 1 in annotated form with tags in bold. The tags, e.g., **genotype-7** and **symptom-7.1**, refer to nodes in the BN that represent the concepts denoted by the tagged phrases. (In this example, the *population* suffix indicates that the *genotype* and *symptom* nodes describe beliefs about a generic individual. Also, probability tags do not bear suffixes since they do not refer to nodes of the BN.)

---

(7) <**probability-7** Approximately 80%> of individuals affected with <**genotype-7/population** NF> have <**symptom-7.1/population** mild to moderate symptoms>; …

Figure 1. Annotated sentence (letter NF)

In addition, the tags are used to encode conditional probability statements in the letter. For example, the phrase in Figure 1 conveys that the conditional probability is approximately 80% that an individual in the population studied has mild to moderate symptoms of NF, given that the individual has the genotype for NF; the conditional probability statement would be encoded as shown in Figure 2.

P(symptom-7.1 | genotype-7) = probability-7
Features: progressive order, retrospective

Figure 2. Encoded probability statement

Two other features of this probability statement are encoded. First, we classify statements such as the one in Figure 1 as *progressive* when the order of presentation of variables in the text matches the causal direction; a statement is classified *regressive* when the variables are given in the reverse of the causal direction. Distinguishing regressive and progressive statements is a non-subjective judgment given the causal relationships of this domain. We wish to track this aspect of usage since presentation order may influence comprehension; this is plausible since psychology experiments show that reasoning from cause to effect is easier than in the opposite direction (Tversky and Kahneman 1980).

Second, following Eddy (1982), we distinguish retrospective and predictive statements. As in the above example, a *retrospective* statement provides the reported frequency of an observable variable's value in a population whose "true diagnosis" is known; this is the form one would expect to find in the research literature. On the other hand, *predictive* statements provide the probability of a hypothesis, e.g. the genotype, given knowledge of observables such as symptoms; this is the form of most relevance to the patient. Although Bayes Theorem can be used to convert a predictive to a retrospective statement and vice versa, the probability values may be different, i.e., in general $P(A \mid B)$ does not have the same probability value as $P(B \mid A)$. However, even medical experts confuse the two types of statement (Eddy 1982).

Table 1 shows the results of an earlier study of probability statements in the corpus (Green 2003). In the three letters (identified as VCF, HL, and NF), the ratio of tagged probability statements (PS) to the number of sentences (SC) was high. Also, the table shows the count of retrospective (Retro.) versus Predictive (Pred.) statements, and Regressive (Reg.) versus Progressive (Prog.) in those letters.

Letter	PS/SC	Retro. PS	Pred. PS	Order	
				Reg.	Prog.
VCF	19/24 (79%)	10	9	8	9
HL	34/40 (85%)	10	23	9	19
NF	20/38 (53%)	10	10	7	10
Total	73/102 (72%)	30	42	24	28

Table 1. Results from preliminary study

Form

Table 2 (at the end of the paper) presents the distribution of different types of expressions used to explicitly convey uncertainty in a representative letter from the corpus. It illustrates a variety of ways of expressing uncertainty. Also qualitative uses (29), qualified numeric uses (3), and numeric range expressions (2), in all (34) greatly outnumber uses of precise quantitative expressions (5), which is typical of the other letters as well.

Function as Probability Value

In the corpus, quantitative and qualitative uncertainty expressions are frequently used to express the probability value in a conditional probability statement. The types of conditional probability statements we have found in the corpus are listed below. (To make the examples more readable, we have given variables self-descriptive names instead of the using the scheme shown in Figure 2.)

Evidence for hypothesis from population data. The writer makes a retrospective probability statement based on data about a population, $P(X | Y) = p$, where p may be expressed quantitatively or qualitatively, and X is evidence for a hypothesis Y. Example: *Individuals with VCF often have heart defects and learning problems*. Encoded form:
P(symptom-defects-learning-problems | genotype-VCF) = p, where p is "often".

Evidence for hypothesis from patient data. The writer makes a predictive probability statement about the patient, $P(Y | X) = p$, where p is expressed qualitatively, and X is evidence for a hypothesis Y.
Example: … *the results of the collagen studies is consistent with your having an altered gene for type I collagen..* Encoded form: *P(genotype-OI-I | result) = p*, where p is "is consistent with".

Prediction of effect based on model. The writer makes a retrospective probability statement based on a Mendelian inheritance model, $P(X | Y) = p$, where p is realized quantitatively or qualitatively, and Y directly or indirectly causes X. Example: *A couple who are both carriers have a one in four (25%) chance that a child they conceive will inherit both altered alleles.* Encoded form:
P(genotype-child | genotype-mother, genotype-father) = 25%.

Prediction of effect based on population data. The writer makes a retrospective probability statement based on data about a population, $P(X | Y) = p$, where p may be expressed quantitatively or qualitatively, and Y directly or indirectly causes X. Example: *Individuals ... can have a genetic condition called ... OI which can range from a relatively mild form to ...* Encoded form:
P(symptoms-OI-mild | genotype-OI) = p, where p is "can".

Comparison of likelihood of hypotheses from population data. The writer compares the probability of one hypothesis to the probability of another hypothesis (in the following, a less specific hypothesis) based on data about a population. Example: *The most common change in this gene [GJB2] causing deafness is called 35delG...* Encoded form:

P(genotype-GJB2-mutation-35delG | genotype-GJB2-mutation) = p, where p is "most common".

Base rate from population data. The writer gives the prior probability of a variable X based on data about a population, $P(X) = p$ or $P(X | Y) = p$. Example: *... an individual in the general Caucasian population has about a 2 to 3% (2 to 3 out of 100) chance of being a carrier of ...* Encoded form: *P(genotype | history-Causasian) = p*, where p is "2 to 3%".

Rhetorical Functions

In addition to the above informational goals, some uses may be motivated by rhetorical goals. Except for *restatement*, these uses require qualitative expressions of uncertainty.

Restatement. Having previously given the risk, p, of an event X, the writer states that the probability of $\sim X$ is *1-p*, e.g. a 25% risk of inheriting a mutation and a 75% risk of not inheriting it. This writing strategy is in accordance with genetic counseling guidelines (Baker et al. 2002) for helping clients understand risk information, since a lay person may not be acquainted with the probability calculus.

Avoiding commitment. Although the writer may be able to provide a more precise estimate of likelihood, in some cases the writer may deliberately hedge (Hyland 1998), e.g., to avoid false hopes or misinterpretation.

Avoiding numeric overload. It is plausible that the writers follow a strategy of using numeric probabilities sparingly in order to make a letter less intimidating to the lay reader. Also, by following this strategy, the few numeric probabilities that *are* provided would become more salient to the reader (Hyland 1998).

Mitigating negative impact. The writer can use a weaker expression of uncertainty as a way of mitigating the effect of information that could evoke negative feelings in the client (Baker et al. 2002).

Qualitative Probabilistic Networks

To build a BN, the conditional probability of each variable, given its direct predecessors, must be specified. Acquiring all of the numerical probabilities for a BN can be an engineering problem, but the corpus shows that patient letters do not always provide precise numeric probability values. Qualitative probabilistic networks are abstractions of BNs in which the numeric probabilities are replaced by qualitative constraints (Wellman 1990, Druzdzel and Henrion 1993). Thus we are investigating implementing the domain model for our system using a

hybrid approach in which both qualitative constraints and some numeric probability values are represented. (Note that the role of the BN/QPN in our system is *not* to provide decision support functions such as to diagnose genetic disorders. However, in order for the system to generate information for the patient, it must be able to model to some extent the reasoning that the human experts have employed, e.g., the influence of the available evidence on the expert's diagnosis.)

Qualitative Explanation for Experts

Druzdzel (1996) describes an explanation generator that provides qualitative verbal explanations as part of the user interface to decision support systems based on BNs. The goal was to enable an expert to evaluate the system's recommendations. An explanation of the system's reasoning, i.e., how it derived the posterior probability of a variable, is composed of three types of elementary explanations: causal inference (*cause to effect*), diagnostic inference (*effect to cause*), and intercausal inference (such as *explaining away*). To generate the explanation, first a BN is converted automatically to a QPN. Then, a qualitative belief propagation algorithm generates all qualitative inferences (i.e. the above three types of inference) from the QPN. The explanation presented to the user consists a list of all of the elementary inferences and a conclusion; for example

> *Qualitative influence of greasy engine block on worn piston rings: Greasy engine block is evidence for oil leak. Oil leak and excessive oil consumption can each cause low oil level. Oil leak explains low oil level and so is evidence against excessive oil consumption. Decreased likelihood of excessive oil consumption is evidence against worn piston rings. Therefore greasy engine block is evidence against worn piston rings* (Druzdzel 1996).

Comparison to Sample Letter from Corpus

Recognizing that the purpose and intended audience of Druzdzel's system's explanations differs from that of a genetics counseling patient letter, nevertheless we informally analyzed a letter from the corpus in terms of the components of an explanation listed above for his system. To aid the discussion, a simplified BN/QPN diagram representing the structural relationships between concepts covered in the letter is shown in Figure 3. (For readability, Boolean-valued variables are used, and variables have been given self-descriptive names instead of names consistent with our coding scheme.) Numbers written inside the nodes give the relative order in which nodes were discussed in the letter.

This letter first describes an initial (but no longer believed) diagnosis that was based on symptoms alone (see nodes numbered 1-3), then the general relationship between collagen abnormalities and possible symptoms (4-6), then the currently believed most likely diagnosis (7-8), then the currently believed second most likely diagnosis (9-11), then a possible consequence of either of the two currently believed most likely diagnoses (12), then a general causal explanation for the second most likely diagnosis (13-14) and its possible consequences to future children (15), then (repeated) the argument for the most likely diagnosis (16-18), and lastly a discussion of possible consequences of the most likely diagnosis (19-25).

Not surprisingly, the organization of information in the human-authored letter is much more complex than a list of inferences. Natural language generation techniques will be used in GenIE to decide what information from the knowledge base to include, how to organize it, and how to express it (Reiter and Dale 2000).

QPN Relationships for Clinical Genetics

Previous work on QPNs has defined domain-independent qualitative relationships: qualitative influence S, additive synergy Y, and product synergy X. Formal definitions can be found in (Druzdzel and Henrion 1993; Wellman and Henrion 1993). Informally speaking, a node A has a *positive qualitative influence* on a node B in a QPN, written as $S^+(A,B)$, if observing a high value for A makes a high value for B more likely; similarly, negative (S^-) and zero (S^0) qualitative influence can be defined. If nodes A and B are parents of node C, A and B have *negative product synergy* with respect to a value c of C, written as $X(\{A,B\},c)$, if the proportional increase in the probability of c on raising the value of B is smaller for higher values of A; positive (X^+) and zero (X^0) product synergy can be defined similarly.

To give an example based on (Wellman and Henrion 1993), the relationship between A, the event that a patient has a cold, and C, the event that he sneezes is $S^+(A,C)$; similarly for an event B that he has an allergic reaction, $S^+(B,C)$. Also, $X(\{A,B\},c)$ holds, where c denotes $C=true$. The event D, that he has a cat allergy and E, that a cat is present are related to B by $S^+(D,B)$ and $S^+(E,B)$, but $X^+(\{D,E\},b)$ where b denotes $B=true$. The rest of this section illustrates some qualitative relationships relevant to clinical genetics.

Qualitative influence. For any nodes A, B, and C representing a genotype of the two parents and their child, respectively, A and B each have a positive influence (S^+) on C. For example, in Figure 3, the nodes labeled *OI-genotype-Autosomal-Dominant-form-in-mother* and *OI-genotype-Autosomal-Dominant-form-in-father* each have a positive influence on *OI-genotype-Autosomal-Dominant-form* (of their child).

Noisy-Or. A NOISY-OR connection can be used to model the causal relationship of nodes A and B, representing competing independent possible diagnoses, to their child node C, representing the effect. A and B each have a positive influence (S^+) on C. Given $C=true$, increasing belief in one of the hypotheses reduces belief

in (*explains away)* the other. This is a kind of inter-causal reasoning as described earlier, which has been shown to be an instance of negative product synergy, i.e. $X^-(\{A,B\},C=true)$ (Wellman and Henrion 1993). There are three examples of this type of convergent connection in Figure 3 labeled NOISY-OR.

Additive Synergy. An example (not illustrated in Figure 3) of positive additive synergy (Y^+) is the relationship between the genotype of each parent, A and B, and the genotype of their child C for the autosomal recessive inheritance pattern. (One copy of a gene is inherited from each parent. In an autosomal recessive disorder, someone having two copies with the mutation will be affected.) Informally speaking, this is positive additive synergy since the joint influence of each parent's genotype on their child's genotype is more than the sum of their individual effects. For discussion purposes, let the domain of each of the genotype variables be represented as {0, 1, 2}, where the value represents the number of mutated copies of a gene. Applying the formal definition of product synergy to the probability table for this inheritance pattern, it can be shown that $X^0(\{A,B\},C=2)$. Therefore, explaining away (an X relation) is *not* licensed for this connection. For example, given that the child is affected (C=2) and then increasing belief that one of the parents is a carrier (A=1) should not decrease belief that the other parent is a carrier (B=1). Also this is not a case of positive product synergy (X^+). That is, given that the child is affected (C=2) and then increasing belief that one of the parents is a carrier (A=1) should not *increase* belief that the other parent is a carrier (B=1).

Enables. An ENABLES connection can be used to model the relationship of nodes A and B, and their child node C, where A represents a variable that may enable B to cause C. This is like the example of $X^+(\{D,E\},b)$ presented earlier (where *D* and *E* represent cat-allergy and cat-present variables, respectively, and *b* means that the patient is sneezing). An example (not shown in Figure 3) is where a mutation-event enables someone who was born with one abnormal copy of a BRCA gene to develop a mutation in the other copy of the gene in the same cell, resulting in breast cancer. Thus, explaining away is not licensed. Instead, given belief that the patient has two mutated copies of BRCA, increasing belief that the patient was born with one copy of the mutation should increase, not decrease, belief that a mutation causing event was experienced by the patient.

Inhibits. An INHIBITS connection (not illustrated in Figure 3) can be used to model the relationship of nodes A and B, and their child node C, where A represents a variable that may diminish the effect of B on C. For example, taking an antihistamine (A) can diminish the effect of an allergic reaction (B) on sneezing (C). According to (Druzdzel and van der Gaag 1995), this kind of relationship can be described as negative additive synergy: $Y^-(\{B,A\},C)$. (In their example, condom use inhibits the contraction of HIV from intercourse.) An example of this relationship from clinical genetics is where *A* is a history node representing whether the patient had a prophylactic masectomy (i.e. removal of the breasts), *B* is a BRCA genotype node, and *C* is the symptom of having breast cancer. Again, explaining away is not licensed in this kind of connection. Instead, given that the patient has breast cancer, finding out that the patient had the masectomy should not decrease belief that the patient has two mutated copies of BRCA.

Related Work

Because of the complex causal domain model, design of genetics counseling patient communication systems faces issues not addressed in most previous patient-tailored natural language generation systems for healthcare (Cawsey et al. 1997). One exception is RAGS, a decision support program to assess genetic risk of breast cancer and explain its risk assessment to the physician in qualitative terms (Coulson et al. 2000). However, the focus of their research was on the decision support methods rather than natural language generation methods. Also, RAGS' decision support component did not use a BN-based domain model. Although not designed for a healthcare application, NAG (Korb et al. 1997) generates natural language arguments from a BN-based domain model. NAG's arguments are designed to address several typical human biases and fallacies in reasoning about probability, but have not been tested in a biomedical domain.

Conclusions

In the first half of the paper, we presented an overview of communication of uncertainty in a corpus of genetic counseling patient letters. The goal of the study is to inform development of intelligent multimedia systems for patient-tailored health communication in clinical genetics. As a first step, we are designing GenIE, a prototype natural language generation system to generate the first draft of a patient letter. GenIE could be used to assist genetic counselors or to brief physicians on how to communicate with patients on genetics-related issues.

From analysis of the corpus, we identified different functions of uncertainty expressions. Informational uses of uncertainty were categorized in terms of various types of conditional probability statements. Also, we noted the prevalence of qualitative and imprecise probability expressions in the corpus. Thus, we have been investigating the applicability of qualitative probabilistic networks to our goals. In the second half of the paper, we defined several relationships in clinical genetics that can be defined in terms of QPN constraints. Currently, we are

implementing a hybrid BN/QPN domain model for GenIE.

Acknowledgments

References

Baker, D.L., Eash, T., Schuette, J.L., and Uhlmann, W.R. 2002. Guidelines for Writing Letters to Patients. *Journal of Genetic Counseling*, 11:5, Oct 2002, 399-418.

Bloch, E.V., DeSalvo, M., Hall, B.D., Epstein, C.J. 1979. Alternative Ways of Presenting Empiric Risks in Birth Defects. *Birth defects: Original article series XV(5c)*. New York: Alan R. Liss, Inc., pp. 233-44.

Cawsey, A.J., Webber, B.L., and Jones, R.B. 1997. Natural Language Generation in Healthcare. *JAMIA*.

Coulson, A.S., Glasspool, D.W., Fox, J., and Emery, J. 2000. Computerized Genetic Risk Assessment and Decision Support in Primary Care. *Informatics: the Journal of Informatics in Primary Care*, May 2000.

Dewdney, A.K. 1993. *200% of Nothing.* New York: John Wiley and Sons, Inc.

Druzdzel, M. J. 1996. Qualitative Verbal Explanations in Bayesian Belief Networks. *Artificial Intelligence and Simulation of Behavior Quarterly*, 94:43-54.

Druzdzel, M.J. and van der Gaag, L.C. 1995. Elicitation of Probabilities for Belief Networks: Combining Qualitative and Quantitative Information. *Proc. 11th Annual Conf. on Uncertainty in AI (UAI-95)*, 141-148.

Druzdzel, M.J. and Henrion, M. 1993. Efficient Reasoning in Qualitative Probabilistic Networks. *Proc. 11th Nat. Conf. on AI (AAAI-93)*, 548-553.

Eddy, D. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In Kahneman, D., Slovic, P., and Tversky, A. (eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press. Chapter 18 (pp. 249-267).

Green, N. 2003. Towards an Empirical Model of Argumentation in Medical Genetics. *Workshop Proceedings of CMNA 2003: IJCAI 2003 Workshop on Computational Models of Natural Argument*, pp. 39-44.

Grimes, D.A., and Snively, G. R. 1999. Patients' Understanding of Medical Risks: Implications for Genetic Counseling. *Obstetrics and Gynecology*, pp. 910-914.

Hyland, K. 1998. *Hedging in Scientific Research Articles*. John Benjamins, Amsterdam.

Korb, K.B., McConachy, R., and Zukerman, I. 1997. A Cognitive Model of Argumentation. *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, pp. 400-405.

Korb, K.B., and Nicholson, A.E. 2004. *Bayesian Artificial Intelligence*. Boca Raton, Florida: Chapman & Hall/CRC.

Reiter, E. and Dale, R. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.

Sarangi, S., Bennert, K., Howell, L., and Clarke, A. 2003. 'Relatively speaking': relativisation of genetic risk in counselling for predictive testing. *Health, Risk and Society*, 5:2, pp. 155-170.

Walker, A.P. 1997. Genetic Counseling. In Rimoin, D.L, Connor, J.M., Pyeritz, R.E., (eds.). *Emery and Rimoin's Principles and Practices of Medical Genetics*. 3rd ed. New York: Churchill Livingstone, pp. 595-618.

Wellman, M.P. 1990. Fundamental Concepts of Qualitative Probabilistic Networks. *Artificial Intelligence* 44(3): 257-303.

Wellman, M.P. and Henrion, M. 1993. Explaining "explaining away". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):287-292.

Type	percen-tages	propor-tion	rate	qualified numeric	range	modal	adverb	det.	adjective, verb
Number	2	2	1	3	2	15	3	1	10
Examples	50%,	1 in 2	2 in 200	about 0.5%, as high as 50%	0.8-1%, 1.5 to 2 in 200	might, could, can, may	usually, typically, possibly	many	possible, expected, consistent with, at risk, do not know, predict, attempt to determine, depends upon, difficult to assess, most common

Table 2. Types of uncertainty expressions used in letter OI. (Total number of sentences = 46)
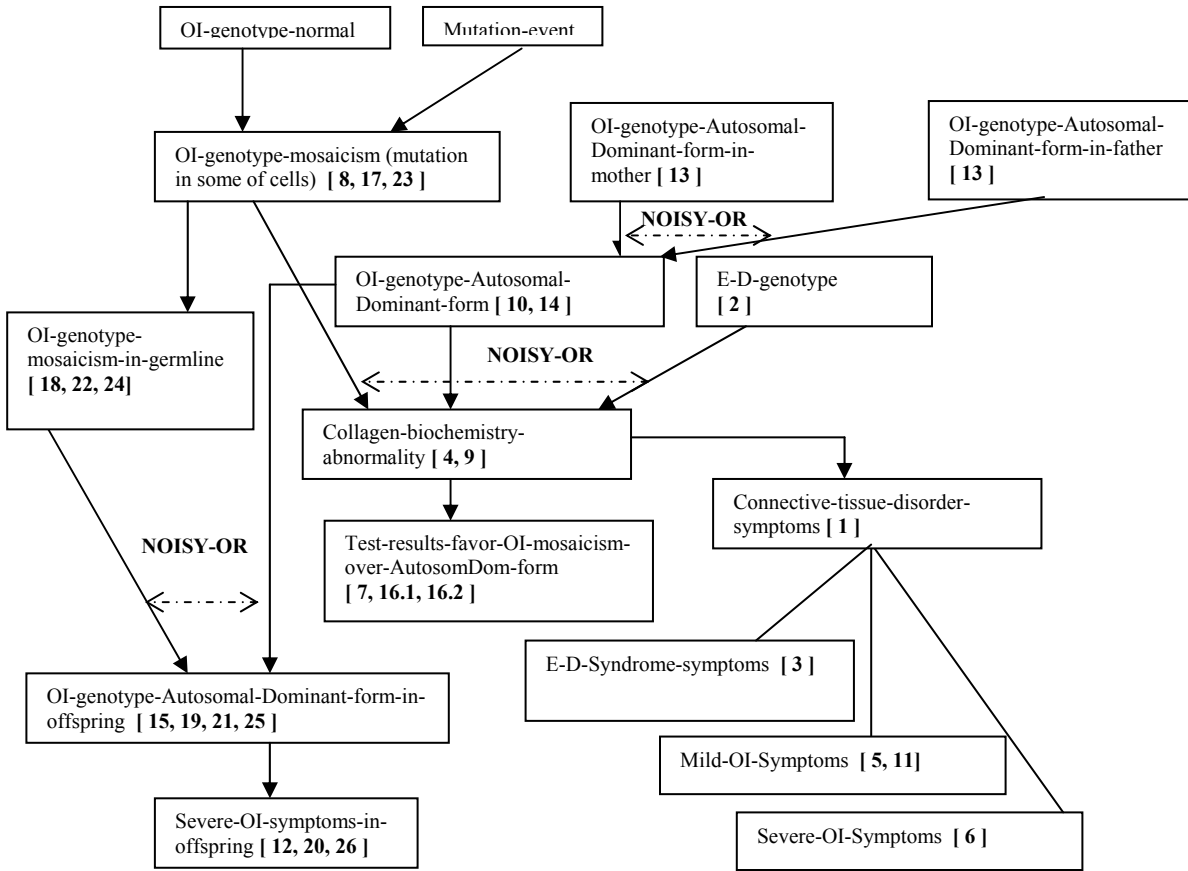
Figure 3. Simplified QPN for letter in corpus on OI. Numbering shows order in which nodes were discussed in letter.