Towards an Empirical Model of Argumentation in Medical Genetics

Nancy Green
Department of Mathematical Sciences
University of North Carolina at Greensboro
Greensboro, NC 27402 USA
nlgreen@uncg.edu

Abstract

We present a coding scheme, based on a Bayesian Network (BN) formalism, for describing probabilistic and causal information in arguments in medical genetics. The scheme was applied to a corpus of genetic counseling letters and evaluated for intercoder reliability. Results show that the model is highly relevant to the corpus while intercoder reliability of the coding scheme is good. We plan to use the coding scheme in an empirical study of argument strategies. Since the coding scheme refers only to BN concepts and general concepts in medical diagnosis, it may be useful to other researchers for empirical studies of natural language corpora in medicine.

1 Introduction

Genetic counselors in the U.S.A. meet with their clients, typically, to discuss testing for, diagnosis of, management of, and/or risks of genetic disorders (Baker et al., 1998). A genetic counselor may explain general concepts in human genetics, define medical and scientific terminology, provide reasons for testing, explain the basis for diagnoses, and provide estimates and explanations of genetic risk. After the meeting, it is customary practice for the counselor to write a one- to two-page letter summarizing what was discussed for the benefit of the client, as well as other family members and health providers. Summary letters are used in other areas of medicine as well (Waterson and Lazero, 1994). One of our goals is to analyze this genre in order to develop natural language generation tools to assist genetic counselors.

Medical genetics professionals have provided informal, prescriptive writing guidelines describing topics to cover (Baker et al., 1998, p. 238) and preferred wording (Baker et al., 2002). However, these guidelines are not sufficient for developing natural language generation technology. Our preliminary, informal analysis of the genre showed that it makes considerable use of argumentation, i.e., discourse that weighs evidence and presents multiple points of view. Thus, as a step towards developing computational tools for genetics counselors, we plan to develop computational models of argumentation in a corpus of genetic counseling letters.

In this paper we present a coding scheme, based on a model of medical genetics encoded in a Bayesian Network (BN) formalism, that we developed to describe the use of probabilistic and causal information in the corpus. Our plan is to so encode the corpus as a step towards identifying argument strategies that make use of this type of information. Our quantitative analysis shows that the model is highly relevant to the corpus. Also, a preliminary evaluation shows that the coding scheme has good intercoder reliability. Since the coding scheme refers only to BN concepts and general concepts in medical diagnosis, it may be useful to other researchers for empirical studies of natural language corpora in medicine. In the next section we present our coding scheme. In section 3 we describe the corpus analysis and intercoder reliability evaluation and our plans for future work. In section 4, we survey related work in empirical and computational approaches to natural language argumentation.

2 Bayesian Network Model for Medical Genetics

A Bayesian Network (BN) is a graphical model whose nodes represent discrete random variables and whose arcs represent dependencies of conditional probability between variables, e.g. see (Jensen, 2001). In this section, we focus on our use of a BN formalism for modeling technical information presented in genetics counseling letters. Figure 1 shows a BN based partly on a letter in our corpus; however, in order to illustrate more of our scheme, some nodes were removed, others were added, and some variables were given fictitious values. BN nodes are depicted with rectangles.

BNs often are used to represent causal relations between variables diagrammatically. For example, in Figure 1 arcs show the causal relation between the genotypes of the parents (labeled as variable types *Genotype/mother*, *Genotype/father*) and the genotype of their child (labeled *Genotype/proband*). (The term *genotype* is defined as a pair of *alleles* of a gene and *proband* as the subject of genetic analysis.)

According to classical Mendelian genetics, the genotype can consist of two normal alleles, one normal and one abnormal allele, or two abnormal alleles (which we denote as *AA*, *Aa*, and *aa*, respectively). A conditional probability table for the node labeled *Genotype/proband* in Figure 1 would specify the likelihood of each of these three

genotypes given the genotype of the parents. According to classical genetics, for example, the probability that the proband inherits two abnormal alleles (*aa*) when one parent is *homozygous* (*aa*) and the other is *heterozygous* (*Aa)* is 50%. Also Figure 1 models a causal chain relating genotype to biochemical effects (represented with the *Biochemistry* variable type), which in turn may have physiological effects (the *Physiology* variable type), which in turn may account for the proband's symptoms (the *Symptom* variable type). (In some letters in the corpus, the writer discusses the link from genotype to symptoms without discussing intermediate states of biochemistry or physiology.) Another causal link is shown in Figure 1 relating the proband's genotype to physical traits that may be but are not necessarily related to the proband's symptoms (the *Finding* variable type).

In addition to causal relations, BN variables may be related by association. For example, in Figure 1 the node labeled by the variable type *History/proband* is needed to model an association between onset of hearing loss in childhood and a particular genotype. (The causal and associational interpretations of arcs are depicted by solid and dashed arcs in the figure, respectively.) More generally, we use *History* type variables for demographic and predispositional factors that are associated with, although not directly the cause of, other variables such as genotyope. Finally, we include *Test* symbols, depicted with diamonds, to show the occurrence of a test, e.g. a test for a genetic condition; these have a constant value identifying a test that can be performed and have no associated probability table. We use these symbols to depict the causal relation between two variables such as *Genotype* and *Result* (a variable type representing the outcome of a test) given the occurrence of the *Test* event.

When used in a probabilistic reasoning application, a BN is created with initial conditional probability tables for each node. Then as new information is acquired, e.g., by observing the result of a test, posterior probabilities are recomputed. Changes in probability may propagate through the network based upon its topology. For example in Figure 1, a positive test result may increase the likelihood that the proband is homozygous (has genotype *aa*), which in turn may increase the likelihood that the mother is heterozygous (*Aa*), which in turn may increase the likelihood that future offspring of the two parents (see the node of type *Genotype/sibling*) will be heterozygous (*Aa*) or homozygous (*aa*). For computational tractability, constraints on topology and the number of nodes in a BN are recommended (Jensen, 2001). However, since we use BNs purely as descriptive devices, we are not subject to these constraints.

In summary, use of a BN formalism enables us to represent the complex informational content of the arguments in our corpus in a unified framework. The arguments use causal explanation, assumptions, and probability statements in justifying diagnoses and in supporting predictions about possible future outcomes. Moreover, the diagnoses and predictions themselves may be stated in probabilistic terms. In the next phase of our corpus study, we plan to use the elements identified through this BN-based analysis as components in the description of argument strategies. This higher level description will address selection and presentation order of information and argumentative function.

For the current corpus analysis, we defined the set of eight BN variable types given in Table 1. This set was designed originally to cover classical Mendelian inheritance, representing over 4500 disorders (Wilson, 2000). The type *Mosaicism* was added to handle a class of non-Mendelian disorders in the corpus. Types have been classified according to whether they represent observable or non-observable properties. The former has been subclassified as evidential or predispositional according to whether the relation to non-observable variables (i.e. the two types of arcs shown in Figure 1) is causal or associational, respectively.

3 Corpus Analysis

3.1 Corpus

The portion of the corpus used in the study reported here consists of two letters provided by a genetic counselor and a third published in (Baker et al., 2002), a total of 102 sentences (2016 words). Each letter discusses one of the following possible diagnoses: Velocardiofacial syndrome (VCF), sensorineural hearing loss, and Neurofibromatosis (NF). The first two letters discuss autosomal recessively inherited mutations and the third discusses both autosomal dominant and non-Mendelian inheritance patterns.

3.2 Procedure

The coders for the study reported here consisted of the author and a sociolinguist. Neither had previous training in medical genetics. The sociolinguist had no previous familiarity with BNs. Before the corpus analysis, the coders read and briefly discussed background readings on medical genetics and BNs. The sociolinguist was given the author's trial encoding of the VCF, hearing loss, and NF letters. Then the coders met and discussed whether they agreed upon the trial encoding. In the few cases where they differed, they discussed their reasons and modified the analysis to reach consensus. The results reported in the next section are based on the consensus analysis.

The following procedure is used to code a letter. While reading the letter, the coder draws one or more BNs representing his or her interpretation of the causal and associational relations among concepts. The coder is allowed to assist his or her interpretation of the text using background knowledge about medical genetics and reference books. The variables in a coder's reconstructed BN are restricted to the types in Table 1 and their possible causal or associational relations described in Section 2. Figure 2 shows one of the two BNs reconstructed from the

text of the NF letter. This BN only includes nodes reconstructed on the basis of items in the letter explicitly discussing the proband's family. A second BN was reconstructed (not shown here) from items in the letter expressed as beliefs about the general population.

While reconstructing a BN from a letter, the coder must tag any phrases in the letter that provide evidence for the coder's reconstruction of the BN. For example, the first column of Table 2 shows the coding of sentence (3) in the NF letter, corresponding to the nodes labeled *Symptom-3.1, Symptom-3.2, and Genotype-3* in Figure 2. To distinguish variables of the same type, each tag is numbered with the sentence number and, if more than one tag for the same type of variable occurs in the sentence, then decimal numbering (e.g. 3.1) is assigned sequentially. Also, each variable tag is assigned an *owner modifier* describing the individual that the variable was intended to describe (e.g. *proband, parent, sibling*, etc.). As shown in the coding for sentences (4) and (7) in Table 2, phrases not describing the proband's case explicitly are tagged with the owner modifier *population*. These tag naming conventions are designed to provide unique variable names that also convey information useful to the analyst. In addition to assigning BN variable tags, coders tag phrases conveying probability assessments, e.g. *probability-4* in Table 2. (Those tags will be described below.)

It is not deemed important for coders to agree on the precise phrase boundaries of tagged text; what is considered important is agreement on the reconstructed underlying BN. After tagging the text, the coder provides an analysis of the following relationships between tags:

- *coreference*: two tags refer to the same node in the same BN.
- *subtype*: two tags refer to nodes *A* and *B*, where *A* represents a concept subsumed by *B*; e.g., in Table 2 *genotype-4* represents a more specific diagnosis (NF type 1) than *genotype-7* (NF).
- *analog*: two tags are analogs if one refers to a node in a BN describing the proband (and possibly his or her family) while the other refers to a node of the same type in a BN describing the general population; e.g., in Table 2 *genotype-3* is an analog of *genotype-7*.

The second column of Table 2 shows the coders' analyses of the above relations as well as analyses of the probability statements. Each probability statement is represented using BN variable names and an indicator of the writer's assessment of probability or frequency given in the text (e.g. "about 1 in every 3000"). In addition to quantitative values, qualitative indicators of probability, possibility, or frequency are used to encode probability statements, e.g., modal auxiliaries ("may") and adverbs ("often"). In cases where the coder believes that the writer's probabilistic judgement is implicit, e.g., in sentence (3) of Table 2, the coder tags the probability value as "[IMPLIED]".

After coding was completed, we also compiled the following statistics on the types of probability statements that had been tagged in the text. First, we counted whether the information in the sentence was presented in the same order as the node ordering of the BN (*progressive*) or in the reverse ordering (*regressive*). This distinction will be used to describe presentation order in argument strategies. Second, we counted statements giving the probability of a non-observable or observable variable as *predictive* or *retrospective,* respectively (Eddy, 1982). *Predictive* statements provide the probability of a particular diagnosis (e.g. having the BRCA1 genotype for breast cancer) given knowledge of observables such as the patient's test results (e.g. a normal mamogram), i.e. P(Genotype=BRCA1 | Result=Normal); *retrospective* statements provide the reported frequency of an observable variable's values in a population whose "true diagnosis" is known, e.g., P(Result=Normal | Genotype= BRCA1). For example in Table 2, the probability statement expressed in (3) is classified as *predictive* since it asserts the probability of a non-observable variable (the genotype) given that the symptoms are known, and as *regressive* since the variables are presented in the sentence in the order of symptoms before genotype, which is the reverse of their causal ordering. This distinction will be used in classifying argumentative function.

3.3 Results and Future Work

Table 3 shows the results of analyzing three letters in the corpus by the procedure described in Section 3.2. The table shows that the BN model presented in Section 2 is of practical significance in terms of two relevance metrics. The basic measurements shown in Table 3 for each letter include the word count (WC), sentence count (SC), the number of unique BN nodes reconstructed from the letter (BN), and the number of probability statements (PS) reconstructed from the letter. The first relevance metric is the ratio of unique BN nodes to sentences (BN/SC), and the second is the ratio of probability statements to sentences (PS/SC). For example, in the VCF letter every two sentences contribute a unique node to the BN and out of a total of 24 sentences, 19 probability statements are given. In short, the results show that our model is highly relevant to the corpus. The results are especially striking considering that SC and WC include text not relevant to the technical argument content such as opening and closing paragraphs. (We expect such text to play a role in the emotional impact of the letters, which we will address in future work.) Table 3 also shows the counts for Progressive vs. Regressive and Retrospective vs. Predictive probability statements, showing that on each dimension there is significant use of both options.

Having created a vocabulary for describing the basic informational elements of technical argumentation in the corpus, our next goal is to identify argument strategies occurring in the corpus. For example, Zukerman et al. (2000) propose a number of argument strategies (e.g. Premise to Goal, Reasoning by Cases, Reductio ad absurdum) which they used in a BN-based natural language argument generation system. However, they did

not perform a corpus analysis to identify these strategies. For an example of part of an argument strategy occurring in our corpus, consider the following annotated excerpt:

(5) <*test-5.2/proband* a special analysis of the long arm of chromosome 22> was done to test for <*genotype-5/ proband* Velocardiofacial syndrome (VCF)>.

(6) Individuals with <*genotype-6/population* VCF> <*probability-6* often> have <*symptom-6/population* [TYPE OF BIRTH DEFECT] and learning problems>.

While sentence (5) implicitly conveys the claim that the health care provider believes that there is some chance that the proband has VCF, (6) provides justification for that claim. The linkage between the two is characterized in our coding scheme by annotation of analog relations between *genotype-5/proband* and *genotype-6/population* and between *symptom-3/proband*, and *symptom-6/population*. A preliminary, qualitative survey of the strategies that we have identified so far is presented in (Green, 2003).

3.4 Preliminary Intercoder Reliability Study

We performed a preliminary evaluation of the intercoder reliability of the coding scheme using the following procedure. A third coder, who has a Masters degree in Computer Science, background in probability and statistics, but no familiarity with Bayesian Networks, and no previous training in linguistics or medical genetics, was asked to code the VCF letter. The coder was given a copy of the information presented in sections 2 and 3.2 of this paper (including tables and figures) and a copy of the consensus coding of the hearing loss letter to read. After being given a chance to ask questions, the coder was instructed to encode the VCF letter in the format shown in Table 2. Then this coder's work was compared to the consensus coding of the VCF letter.

The coder's tagging reconstructed essentially the same BNs as those of the consensus coding, with the following differences: the coder added a history node not identified by the other coders, encoded the tests discussed in the letter as result nodes rather than as test nodes, and coded each conjunct of conjoined noun phrases separately (the consensus version had coded some of these with a single node). In addition, there were only the following differences in the encoding of probability statements. First, the coder omitted two statements encoded in the consensus version. However, the corresponding sentences in the text did not contain explicit indicators of probability. Second, a probability statement encoded in the consensus version in the form of P(B|A) was encoded in the form P(A|B). These minor differences suggest that, with better instructions, reconstruction of the BN and probability statements can be performed with very good intercoder reliability. Therefore, since the intercoder reliability study reported here, we have written a manual for coders and plan to conduct a more formal evaluation of intercoder reliability soon.

4 Related Work

GENINFER uses BNs to calculate risk of inheriting a genotype (Szolovits and Pauker, 1992). However, only the pedigree (i.e. the genotypes of the "family tree") is represented in the BN; and no natural language generation is performed. The RAGs project developed a decision support tool for doctors that assesses the patient's risk and explains its reasoning by listing reasons for and against the risk assessment (Emery et al. 1999). Banter (McRoy et al., 1998) provides a natural language dialogue interface to a Bayesian reasoner used for training doctors.

NAG (Zukerman et al., 2000) is a natural language argument generation system. Given a goal proposition NAG uses argument strategies to select propositions from a BN. However, the strategies are not based upon a corpus analysis. Carofiglio and de Rosis (2003) propose using BN to represent both logical and emotional reasoning in a unified framework for argument generation. Non-Bayesian argumentation systems based on informal argumentation have been developed for changing health-related attitudes (Grasso et al., 2000; Reiter et al., 1999).

The only argumentation coding scheme we know of that has undergone intercoder reliability evaluation is based on a model of the argument structure of scientific research articles (Teufel et al., 1999). However, a preliminary attempt to apply their scheme to our corpus, which represents a different genre, showed that it was not useful for our goals.

5 Conclusion

We have presented a coding scheme for describing use of probabilistic and causal information in arguments in medical genetics. The scheme was applied to a corpus of genetic counseling letters. Results show that the model is highly relevant to the corpus and a preliminary evaluation shows that the coding scheme has good intercoder reliability. In future work, we plan to perform a more formal evaluation of intercoder reliability, and then use the coding scheme as part of an empirical study of argument strategies occurring in the corpus. Since the coding scheme refers only to BN concepts and general concepts in medical diagnosis, it may be useful to other researchers interested in performing empirical studies of natural language corpora in medicine.

Acknowledgments

We thank the coders (Dr. Jeutonne Brewer and Mr. A. Huber) and two genetic counselors, Ms. Myra Roche (UNC-CH) and Ms. Nancy Callanan (UNCG), who provided us with helpful information. This material is based upon work supported by the National Science Foundation under Grant No. 0132821.

References

Baker, D.L., Schuette, J.L., and Uhlmann, W.R. (1998). *A Guide to Genetic Counseling*. New York: Wiley-Liss.

Baker, D. L., Eash, T., Schuette, J.L., and Uhlmann, W.R. (2002). Guidelines for writing letters to patients. *Journal of Genetic Counseling*, Vol. 11, No. 5, p. 399-418.

Carofiglio, V. and de Rosis, F. (2003). Combining logical with emotional reasoning in natural language argumentation. To be presented at *Workshop on Affective User Modeling at User Modeling 2003*.

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).

Emery, J., Walton, R., Coulson, A., Glasspool, D., Ziebland, S., and Fox. J. (1999). Computer support for recording and interpreting family histories of breast and ovarian cancer in primary care (RAGS): qualitative evaluation with simulated patients. *British Medical Journa*l, 319, 32-36.

Grasso, F., Cawsey, A., and Jones, R. (2000). Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition. *Int. J. Human-Computer Studies* 53, 1077-1115.

Green, N. (2003). Argumentation in genetic counseling summary letters. To be presented at IPrA 2003.

Jensen, F. V. (2001). *Bayesian networks and decision graphs*. New York: Springer-Verlag.

McRoy, S. Liu-Perez, A., and Ali, S. (1998). Interactive computerized health care education. *Journal of the American Medical Informatics Association*, 5(4), 76-104.

Reiter, E., Robertson, R., Osman, L. (1999). Types of knowledge required to personalise smoking cessation letters. In Werner Horn et al. (eds.). *Proc. of the Joint European Conf. on AI in Medicine and Medical Decision Making*. Berlin: Springer-Verlag, 389-399.

Szolovits, P. and Pauker, S.P. (1992). Pedigree analysis for genetic counseling. In Lun, K.C. et al. (eds.), *MEDINFO 92: Proceedings of the Seventh Conference on Medical Informatics*, pp. 679-683. Elsevier (North Holland).

Teufel, S., Carletta, J., and Moens, Marc. (1999). An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of EACL 1999*.

Waterson, T. and Lazaro, C. (1994). Sending parents outpatient letters about their children. *Quality in Health Care*, 3: 142-146.

Wilson, Golder N. (2000). *Clinical Genetics: A Short Course*. NY: Wiley-Liss, Inc.

Zukerman, I., McConachy, R., and Korb, K. (2000). Using argumentation strategies in automated argument generation. In *Proceedings of INLG 2000*, 55-62.

Category	Subcategory	Variable type	Examples of variable with sample values
observable	predispositional	History	age: {Child, Adult}
	evidential	Symptom	hearing loss: {Yes, No}
	evidential	Finding	freckling: {Yes, No}
	evidential	Result	FISH test result: {Positive, Negative}
non-observable		Genotype	GJB2: {AA, Aa, aa}
		Mosaicism	germline mutation: {Yes, No}
		Biochemistry	Connexin26: {Normal, Altered}
		Physiology	chemical equilibrium: {Normal, Abnormal}

Table 1. BN variable types used to model medical genetics corpus

(3) … Philip was diagnosed as having a <symptom-3.1/proband pseudoarthrosis> of the left tibia and also was noted to have <symptom-3.2/proband several café-au-lait spots> both of which are features of <genotype-3/proband neurofibromatosis>.	regressive order, predictive: P(genotype-3/proband | symptom-3.1/proband, symptom-3.2/proband) = [IMPLIED]
(4) <genotype-4/population Neurofibromatosis 1> (NF) is a progressive disorder of the skin and nervous system that occurs in <probability-4 about 1 in every 3000> individuals.	P(genotype-4/population) = "about 1 in every 3000" genotype-4 is a subtype of genotype-7
(7) <probability-7 Approximately 80%> of individuals affected with <genotype-7/population NF> have <symptom-7.1/population mild to moderate symptoms>;…	progressive order, retrospective: P(symptom-7.1/population | genotype-7/population) ="approximately 80%" genotype-3 is an analog of genotype-7

Table 2. Examples of coding for sentences 3, 4, and 7 in NF letter

Letter	WC	SC	BN	BN/ SC	Probability statements (PS)				
					PS/SC	Retrospec-tive	Predic-tive	Order	
								Regressive	Progressive
VCF	446	24	12	12/24 (50%)	19/24 (79%)	10	9	8	9
hearing loss	756	40	29	29/40 (73%)	34/40 (85%)	10	23	9	19
NF	814	38	10	10/38 (26%)	20/38 (53%)	10	10	7	10
Total	2016	102	51	51/102 (50%)	73/102 (72%)	30	42	24	28

Table 3. Results of Corpus Analyis

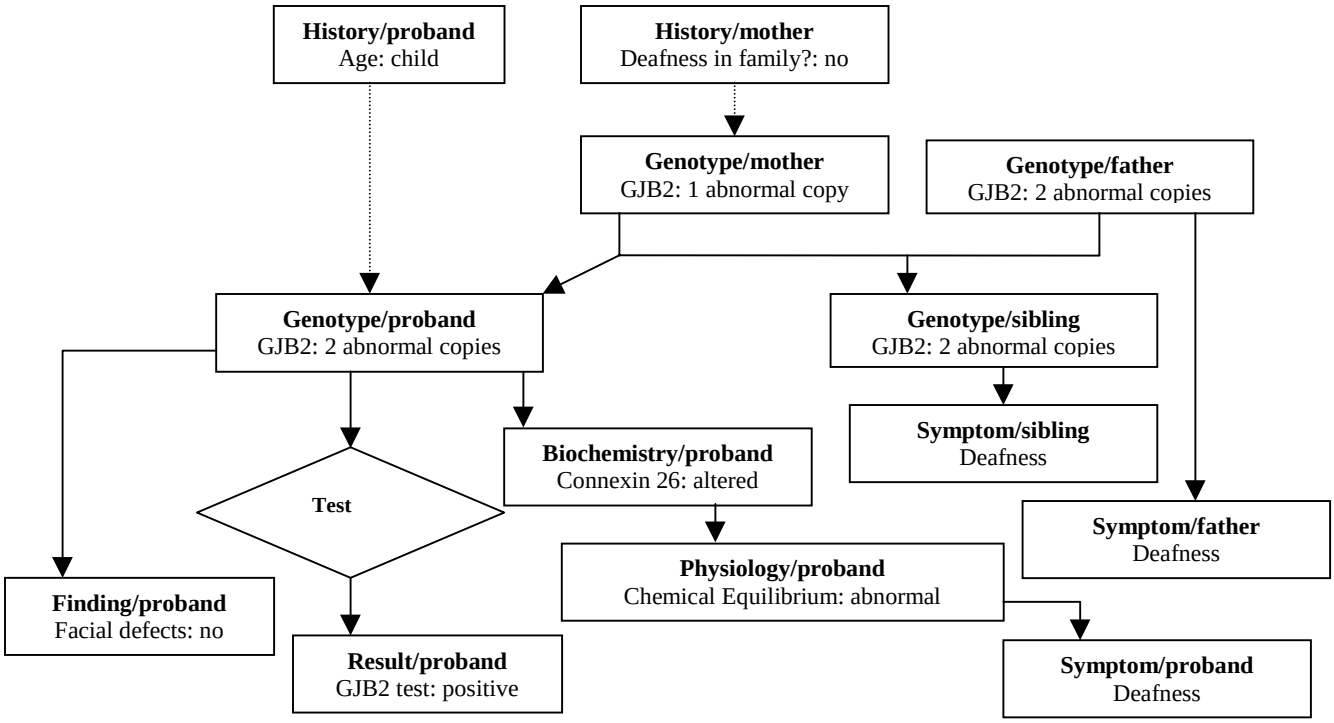

Figure 1. Sample BN for Medical Genetics

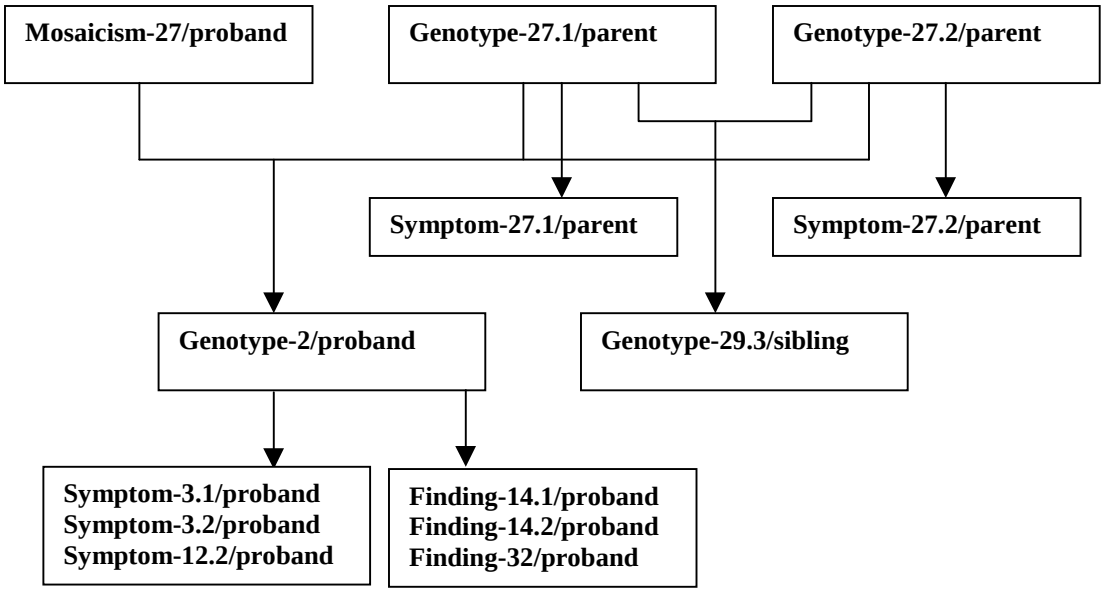

Figure 2. One of two BNs for NF letter. This BN shows discussion about the proband's family and includes tag numbers.