Generation of Biomedical Arguments for Lay Readers

Nancy Green
Department of Computer Science
University of North Carolina Greensboro
Greensboro, North Carolina 27402-6170 USA
`nlgreen at uncg.edu`

Abstract

This paper presents the design of a discourse generator that plans the content and organization of lay-oriented genetic counseling documents containing arguments, and an experiment to evaluate the arguments. Due to the separation of domain, argument, and genre-specific concerns and the methodology used for acquiring a domain model, this approach should be applicable to argument generation in other domains.

1 Introduction

The goal of our research is to develop methods by which intelligent systems can help lay audiences to understand biomedical and other kinds of scientific arguments. We have been studying how one type of lay-communication and biomedical-domain expert, the genetic counselor, presents written arguments in *patient letters*, standard documents summarizing information and services provided to the client (Baker et al., 2002). Clinical genetics involves causal probabilistic reasoning, e.g., diagnosis of a genetic basis for a health problem or prediction of inheritance risks. The patient letter is designed to document the experts' reasoning for medical and legal purposes, as well as to provide an explanation that a lay client can understand.

This paper presents, for the first time, the design of a discourse generator that plans the content and organization of genetic counseling patient letters containing arguments; and an experiment that we performed to evaluate the arguments. The discourse generation process involves three modules: a qualitative causal probabilistic domain model, a normative argument generator, and a genre-specific discourse grammar. In (Green, 2005), we reported a corpus study that produced a reliable biomedical coding scheme. In subsequent workshop papers (Green et al., 2004; 2005), we introduced our use of qualitative probabilistic constraints and provided a brief description of the biomedical domain model. We have also provided informal descriptions of argument patterns in the corpus (Green, to appear; 2006). However, we have not previously published the design of the discourse generator, including the discourse grammar and argument generator, and their relationship to the domain model.

The theoretical significance of this work is three-fold. First, it is empirically based, i.e., based on analysis of arguments in a corpus of genetic counseling patient letters, since the goal is to produce the same kinds of normative arguments as are used in expert-lay communication. Second, the normative argument generator creates an intentional-level representation of the arguments in the text, which provides a foundation for an intelligent system's ability to engage in follow-up discussion about the arguments that have been presented. Finally, due to the separation of domain, argument, and genre-specific concerns in the design, and due to the methodology used to acquire a domain model, it should be possible to apply this approach to lay-oriented argument generation in other domains. The practical significance of this work is that it is major step in the design of a deployable system to generate the first draft of genetic counseling patient letters. As genetics plays an increasingly important role in medicine, there is a need for tools to aid in dissemination of patient-tailored information.

In the next section, we give an overview of a prototype generation system, whose main components are described in more detail in sections 3-5; an example of the generation process is given in section 6; an experiment to evaluate the generated arguments is presented in section 7; and related work is summarized in section 8.

2 System Overview

We are developing a prototype system for genetic counselors that will synthesize the first draft of a patient letter. The deployed system will consist of a graphical user interface for the genetic counselor, a domain model/reasoner, an argument generator, a discourse grammar, and a linguistic realizer. Prototypes of all components except the linguistic realizer have been implemented. Although this paper focuses on discourse generation and its relationship to the domain model, as background we now describe the flow of information through the system.

The domain model (section 3) is initialized with generic information on clinical genetics. Through a user interface providing menus and other non-free-text input devices, the counselor will provide standard clinical information such as a patient's symptoms and information about his family tree; test results; preliminary diagnosis (before testing); and final diagnosis (after test results are known). The system uses this information to transform its generic domain model into a specialized domain model of the patient and his family.

In this genre, a patient letter must provide not only the above information, but arguments for the diagnosis and other inferences made by the medical experts. The discourse generation process works as follows. A discourse grammar (section 4) encodes the high-level topic structure of letters in this genre. The discourse grammar rules generate a derivation instantiated from the domain model with information specific to a patient's case. For each of the writer's claims about the case for which a normative argument must be provided according to standard practice, the discourse grammar invokes the argument generator.

The argument generator (section 5) uses non-domain-specific argument strategies that are instantiated with information from the domain model. The argument generator returns a structured representation of an argument in which the communicative function of information, e.g., as data or

warrant, is identified. As illustrated in section 6, in future interactive systems knowledge of communicative function could be used to support follow-up discussion. In the current prototype, this knowledge is used to determine presentation order, e.g., that data supporting a claim is to be presented before that claim. One of the goals of the experiment described in section 7 was to evaluate this ordering. In the final system, the output of discourse generation will be transformed by a linguistic realizer into the first draft of a letter.

3 Domain Model

In a previous study of the corpus (Green, 2005), we identified a small set of categories (e.g. *genotype, test result, symptom*) with good inter-rater reliability that can be used to describe the biomedical content of a genetic counseling letter as a causal probabilistic network (Korb and Nicholson, 2004). A prototype domain model has been manually constructed covering representative genetic disorders using only these categories of variables. By restricting a domain model to these categories, the result should reflect the simplified conceptual model of genetics used by genetic counselors in communication with their lay clients; this facilitates generation since the generator will not have to distinguish what information in the domain model is appropriate to communicate to a lay audience. Another benefit of restricting a domain model in this way is that it reduces the knowledge acquisition effort of choosing variables and determining network topology; any genetic disorder in the scope of the coding scheme (over 4500 single-gene autosomal disorders) would be modeled in terms of a small number of variable types and a standard topology. Thus, it should be straightforward to semi-automatically construct a domain model covering many different genetic disorders.

Figure 1 shows part of a domain model after it has been updated with information about a particular patient's case. The nodes labeled *GJB2 (mother)*, *GJB2 (father), GJB2 (child)* are genotype variables, representing the mother's, father's, and child's GBJ2 genotype, respectively. (A genotype is a pair of alleles of a gene; one allele is inherited from each parent. An individual who has two mutated alleles of the GJB2 gene usually experiences hearing loss.) The nodes labeled *hearing loss (child)* and *non-syndromic (child)* are vari-

ables representing the child's symptoms. The node labeled *test result (child)* is a variable representing the results of testing the child's GJB2 genotype.

The most likely states of the variables are shown beside the nodes in Figure 1; T_1 and T_2 represent the time at which the (experts') belief is held, before or after the child's genetic test results are known, respectively. The information recorded in the network about this particular case is that the child was observed to have hearing loss and no features of a genetic syndrome; the preliminary diagnosis, i.e. before testing, was that the cause of hearing loss is having two mutated alleles of GJB2; the test results were negative, however; thus, the current diagnosis is some other (unspecified) autosomal recessively inherited genetic condition, represented by the genotype variable labeled *other genotype (child)*. In addition, the parents are hypothesized to be carriers (i.e. to each have one mutated allele) of that genotype, represented by the variables labeled *other genotype (mother)*, *other genotype (father)*.

Although a causal probabilistic network used to perform diagnosis or risk calculation would require specification of numeric probabilities, the role of the network in our system is to qualitatively model the reasoning that the medical experts have performed outside of the system. Also, we found that in the corpus numeric probabilities were provided only when citing epidemiological statistics or risks calculated according to Mendelian inheritance theory (which does not require Bayesian probability computation). Thus, instead of using numeric probabilities for domain reasoning, the domain model uses qualitative constraints based upon formal relations of qualitative influence, product synergy, and additive synergy (Druzdzel and Henrion, 1993).

In addition to being adequate for natural language generation, this approach greatly reduces knowledge acquisition effort; it should be straightforward to semi-automatically acquire the qualitative constraints of a full-scale domain model due to regularities in this domain and the use of a restricted set of variable types as described above. For example, qualitative constraints between genotypes of parents and child would be determined by whether a genotype follows an autosomal dominant or recessive inheritance pattern.

We now describe some of the qualitative domain constraints. An influence relation holds between a node in a causal graph and its direct descendant. $\mathbf{A}$ has a *positive qualitative influence* on $\mathbf{B}$, written $\mathbf{S^+(state(A,V_A), state(B,V_B))}$, if the state of $\mathbf{A}$ reaching a threshold value $\mathbf{V_A}$ makes it more likely that the state of $\mathbf{B}$ reaches value $\mathbf{V_B}$. For example, if having two mutated alleles of a genotype A normally results in the appearance of a symptom B, this could be described as $\mathbf{S^+(state(A,2), state(B,yes))}$. Each arc in Figure 1 implicitly represents an $\mathbf{S^+}$ relation.

Product and additive synergy describe converging connections, i.e., the relation between a set of variables $\{\mathbf{A, B}\}$ and their direct descendant $\mathbf{C}$ in a graph. $\mathbf{A}$ and $\mathbf{B}$ have *negative product synergy* with respect to state $\mathbf{V_C}$ of $\mathbf{C}$, written $\mathbf{X^-(\{state(A,V_A), state(B,V_B)\}, state(C,V_C))}$, if either the state of $\mathbf{A}$ reaching a threshold $\mathbf{V_A}$ or the state of $\mathbf{B}$ reaching a threshold $\mathbf{V_B}$ makes it more likely that the state of $\mathbf{C}$ reaches $\mathbf{V_C}$. This type of relationship characterizes mutually exclusive alternative diagnoses that could account for the same symptom; it also characterizes autosomal dominant inheritance, an inheritance pattern where inheriting one mutated allele of a genotype (from either parent) is usually sufficient to cause health problems. In Figure 1, the possible alternative causes of the symptoms are indicated by the $\mathbf{X^-}$ annotations.

On the other hand, autosomal recessive inheritance, an inheritance pattern where inheriting two mutated alleles (one from each parent) is usually necessary to cause health problems, is characterized by zero product synergy ($\mathbf{X^0}$); $\mathbf{A}$ and $\mathbf{B}$ have *zero product synergy* with respect to state $\mathbf{V_C}$ of $\mathbf{C}$, $\mathbf{X^0(\{state(A,V_A), state(B,V_B)\}, state(C,V_C))}$, if the state of $\mathbf{A}$ reaching a threshold $\mathbf{V_A}$ *and* the state of $\mathbf{B}$ reaching a threshold $\mathbf{V_B}$ makes it more likely that the state of $\mathbf{C}$ reaches $\mathbf{V_C}$. For example, if the mother's, father's, and child's genotype are represented by variables $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, respectively, then $\mathbf{X^0(\{state(A,1), state(B,1)\}, state(C,2))}$ can represent the constraint that if the child's genotype $\mathbf{C}$ has two mutated alleles, then one mutated allele must have come from each parent. In Figure 1, the autosomal recessive inheritance pattern of GJB2 and the other hypothesized genetic disorder are indicated by the $\mathbf{X^0}$ annotations.

Other qualitative constraints used in the domain model are based on negative qualitative influence ($\mathbf{S^-}$), positive product synergy ($\mathbf{X^+}$), and negative additive synergy ($\mathbf{Y^-}$). In addition, the domain model stores epidemiological statistics as probabil-

ity statements composed of variables used in the network, e.g., the frequency of hearing loss due to GJB2. This type of information can be used as backing in an argument (see section 5) but does not play a role in domain reasoning.

4 Discourse Grammar

A discourse grammar was written based upon our analysis of the corpus and a description of standard practice in genetic counseling (Baker et al., 2002). The current grammar is intended to cover letters on single-factor autosomal genetic disorders. Thanks to the regularities in this domain and in this genre, the grammar consists of a small number of rules. The starting rule of the grammar represents the main sections of a letter in their standard order: opening, referral, preliminary diagnosis, testing, final diagnosis, origin of genetic condition, inheritance implications, prognosis/treatment, and closing. One or more grammar rules describe each of these sections.

Grammar rules may request the domain reasoner for case-specific information to be included in the letter. In addition, when the grammar provides a choice of rules, rule selection is based upon case-specific information provided by the domain reasoner. For example, one rule for reporting the final diagnosis handles cases in which the patient's test results confirm the preliminary diagnosis, and another rule those cases where the preliminary diagnosis has been disconfirmed by test results; the domain reasoner returns the information needed to choose between those two rules.

The process described so far creates an initial outline of the information to be presented (in non-linguistic form), including various claims requiring an argument. Each of those claims is passed to the argument generator described in the next section. For example, the letter shown in Figure 2 contains seven claims labeled C_1 to C_7; argument generation adds information labeled D_1 to D_7, W_1 to W_7, and B_1 to B_4. The information returned by the argument generator is added to the outline, completing the structure that will be transformed by the linguistic realizer into text.

5 Argument Generation

Given a claim, the argument generator uses argument strategies to construct a normative argument for the claim from information provided by the domain reasoner. The strategies are non-domain-specific in the sense that they refer to formal properties of the qualitative causal probabilistic domain model rather than to genetics.

According to Toulmin's model of normative argument structure (1998), an argument for a claim can be analyzed in terms of various functional components: the data, warrant, and backing. The *data* are the facts used to defend a claim. The *warrant* is a principle that licenses the claim given the data. An optional *backing* may be used to justify the warrant, e.g., by giving the facts upon which the warrant is based. To derive the argument strategies used in the system, we analyzed the arguments in the corpus in terms of Toulmin's model; the resulting strategies describe mappings from formal properties of the domain model to the data and warrant supporting a claim and to the backing of a warrant. Several strategies are paraphrased below for illustration.

Strategy 1. *Argument for belief in causal claim, based on effects*: An argument for the *claim* that it is believed to some extent at time T_i that state(A,V_A) holds and that state(A,V_A) is responsible for the states of variables $B_1..B_i$, i.e., state(B_1,V_{B1}) .. state(B_i,V_{Bi}), consists of the (presupposed) *data* that state(B_1,V_{B1}) .. state(B_i,V_{Bi}) hold, and optionally other *data* that state(B_j,V_{Bj}) .. state(B_k,V_{Bk}) hold, where the *warrant* is a positive influence relation S^+(state(A,V_A), state(B_p,V_{Bp})) for each B_p in $\{ B_1 .. B_i , B_j .. B_k\}$.

Strategy 2. *Argument for decrease in belief to unlikely that state of causal variable is at or over threshold value, based on absence of predicted effect*: An argument for the *claim* that there has been a decrease in belief, from time T_1 to T_2, to the belief at T_2 that it is unlikely that state(A,V_A) holds, consists of the (newly acquired) *data* that it is unlikely that state(C,V_{Ci}) holds for all $V_{Ci} \geq V_C$, where the *warrant* is a positive influence relation S^+(state(A,V_A), state(C,V_C)).

Strategy 3. *Argument for increase in belief in causal claim, based on decrease in belief in alternative cause:* An argument for the *claim* that there has been a increase in belief, from time T_1 to T_2, to the belief at T_2 that it is believed to some extent that state(A,V_A) holds and that state(A,V_A) is responsible for the states of variables state(B_1,V_{B1}) .. state(B_i,V_{Bi}), consists of the (presupposed) *data* that state(B_1,V_{B1}) .. state(B_i,V_{Bi}) hold, and the

(newly acquired) *data* that it is unlikely that state(Alt,V_{Alt}) holds for all $V_{Alt} \geq V_{threshold}$, where the *warrant* is a negative product synergy relation $X^-(\{state(A,V_A),state(Alt,V_{threshold})\},state(B,V_B))$ for each B in $\{B_1 .. B_i\}$.

Strategy 4. *Argument for belief in joint responsibility, based on effect.* An argument for the *claim* that it is believed to some extent at time T_i that state(A,V_A) and state(B,V_B) hold and that state(A,V_A) and state(B,V_B) are jointly responsible for state(C,V_C), consists of the (presupposed) *data* that state(C,V_C) holds, where the *warrant* is a zero product synergy relation $X^0(\{state(A,V_A), state(B,V_B)\}, state(C,V_C))$.

6 Example

This section gives an example of discourse generation for the case in section 3. An outline created by application of the discourse grammar to the domain model in Figure 1 would contain, in addition to basic information about the case not requiring an argument, several claims requiring further support to be provided by the argument generator.

First, the claim that it was believed, before testing, that the child's hearing loss could be due to having two mutated alleles of GJB2 would be supported by an argument constructed using Strategy 1. The data of the argument is the presupposition that the child has hearing loss and the additional finding that she has no syndromic features. The warrant is the positive influence relations (S^+) linking the variable representing the child's GJB2 genotype to each of the two variables representing the child's symptoms. Note that if a reader questioned this argument, an interactive system could provide information on the source of the data or epidemiological statistics backing the warrant.

Second, the claim that it is currently believed, after testing, that it is unlikely that the child's GJB2 genotype has two mutated alleles would be supported by an argument constructed using Strategy 2. The data of the argument is that the child's GJB2 test results were negative. The warrant is the positive influence relation (S^+) from the child's GJB2 genotype to the child's GJB2 test results, which predicts that if the child had this mutation, then the test results would have been positive. If a reader questioned this argument, an interactive system could provide information on the source of the

data or back the warrant by providing information about the rate of false negatives.

Third, the claim that it is currently believed, after testing, that it is possible that the child has some other genetic condition that is responsible for her hearing loss would be supported by an argument constructed using Strategy 3. The data of the argument is that she has hearing loss and the current belief that GJB2 is not likely responsible. The warrant is the negative product synergy relation (X^-) between the child's GJB2 genotype and another genotype to hearing loss. If a reader questioned this argument, an interactive system could provide information on the proportion of cases of hearing loss that are due to other genetic conditions as backing for the warrant.

Fourth, the claim that it is currently believed, after testing, that it is possible that the parents are carriers (i.e., each has one mutated allele) of the unspecified genotype claimed to be responsible for the child's hearing loss would be supported by an argument constructed using Strategy 4. The data of the argument is the presupposition that the child has two mutated alleles of the other genotype. The warrant is the zero product synergy relation (X^0) between the two parents' genotype for this alternative to GJB2 and the child's genotype for this same alternative. If a reader questioned this argument, an interactive system could provide an explanation of the warrant, which is based on the theory of Mendelian inheritance; or it could provide the argument for the data, i.e., the belief that the child has two mutated alleles of the other genotype.

Finally, the claim that it is currently believed, after testing, that assuming they are both carriers there is a 25% probability that each future child that the two parents have together will inherit two mutated alleles of the other genotype would be supported by an argument constructed by a strategy not shown in section 5. The data is the assumption that the parents are both carriers, and the warrant is the same zero product synergy relation (X^0) used in the argument for the fourth claim. If a reader questioned this argument, an interactive system could provide an explanation of how the probabilities are determined by zero product synergy.

7 Experiment

Argument generation was evaluated in the following experiment. Five biology graduate students,

screened beforehand for writing ability in biology, were shown two patient letters. The letters were created by the experimenter by paraphrasing the output of discourse generation that would be input to the realizer. The paraphrases are similar in syntax and lexical style to letters in the corpus, but the genetic disorders covered in the experiment's letters differ from those covered in the corpus. One letter concerns a child confirmed to have cystic fibrosis (CF); the other a child whose test results for Waardenburg syndrome (WS) were negative. The text of letter CF is given in Figure 2. The first column contains annotations describing the communicative function of the information: C for claim, D for data, W for warrant, and B for backing. Each label is subscripted with an integer referring to the argument. (The row labeled C_2/D_3 functions as both the claim of argument 2 and the data of argument 3.) Annotations were not shown to the experiment's participants. Communicative function was used to determine presentation order within each argument. Letters CF and WS had 23 and 25 segments, respectively, where a *segment* is defined as a unit fulfilling one of the above functions, or a non-argument-related function.

The goal of the experiment was to conduct a preliminary evaluation of the acceptability of the arguments in terms of content, explicitness, and presentation order within arguments. The participants were asked to revise each letter as needed to make it more appropriate for its intended recipients, the biological parents of a patient. Participants were told they could reword, reorder, and make deletions and additions to a letter. The results are summarized in Table 1, which includes the average number of segments to/from which information was added (New) or deleted (Delete), and reordered (Reorder). (Rewordings are not tabulated since it was not our goal to evaluate wording.) New and Delete are measures of acceptability of argument content and explicitness. Reorder is a measure of acceptability of ordering. On average, the number of New, Delete, and Reorder revisions were low: less than two per letter, with most revisions in the category of Reorder. This is encouraging since the system to be built for genetic counselors should provide acceptable arguments requiring a minimum of revision.

To provide more details about the results, first, the only segments to which participants added information are warrants. The deletions of data consist of information presumably already known to the recipients, e.g. D_6 in letter CF; other deletions are of part or all of a warrant or all of a backing. The only deletions of claims consist of information duplicated in another part of the letter; there were no cases where a claim was deleted even though it could be inferred from data and warrant. The reorderings were across-argument, which violates conventional topic structure in the genre, or within-argument. In the latter, half repositioned a claim from final position in an argument to a position before the warrant or backing; the other half repositioned the warrant or backing before the data.

8   Related Work

Due to space limitations, this section focuses on research on generation of *normative* arguments (as opposed to behavior-change and evaluative arguments), and arguments designed for text rather than dialogue. Zukerman et al. have presented several papers on argument generation from Bayesian network domain models (e.g., 2000). The type of domain model used in our work differs in two respects. First, it is based on empirical research since it is intended to represent the simplified conceptual model presented to a lay audience in this genre. Second, it uses qualitative probabilistic constraints. One difference in argument generation is that our system's argument strategies are based on analysis of the corpus. Also, our system creates an intentional-level representation of an argument.

Teufel and Moens (2002) present a coding scheme for scientific argumentation in research articles that is designed for automatic summarization of human-authored text. Thus, it would not be sufficient for generation from a non-linguistic knowledge base. Also, it does not make the finer-grained distinctions of the Toulmin model.

Branting et al. (1999) present the architecture of a legal document drafting system. In it, a discourse grammar applies genre-specific knowledge, while a legal reasoning module creates the illocutionary structure of legal arguments. Branting et al. argue for maintaining a distinct intentional-level representation of arguments to support interactive follow-up discussion. We agree, but our design further distinguishes domain reasoning from argument generation.

As for work on ordering and explicitness, Reed and Long (1997) propose ordering heuristics for

arguments of classical deductive logic. Fiedler and Horacek (2001) present a model for deciding what can be omitted from explanations of mathematical proofs. Carenini and Moore (2000) present an experiment to determine how much evidence is optimal in an evaluative argument.

9   Conclusions

This paper presents the design of a discourse generator that plans the content and organization of genetic counseling letters containing arguments. A preliminary evaluation of the arguments was promising. The most important contribution of this work is the design of a non-domain-specific normative argument generator that creates an intentional-level representation of an argument. From the corpus, we formulated argument strategies that map formal properties of qualitative causal probabilistic models to components of Toulmin's model. Due to the separation of domain, argument and genre-specific concerns and the methodology used for acquiring the domain model, this approach should be applicable to lay-oriented normative argument generation in other domains.

Acknowledgments

References

Baker DL, Eash T, Schuette JL, Uhlmann WR. 2002. Guidelines for writing letters to patients. *J Genetic Counseling,* 11(5):399-418.

Branting LK, Callaway CB, Mott BW, Lester JC. 1999. Integrating Discourse and Domain Knowledge for Document Drafting. *Proc ICAIL-99*, 214-220.

Carenini G, Moore J. 2000. An empirical study of the influence of argument conciseness on argument effectiveness. *Proc Ann Meeting of ACL*, 150-7.

Druzdzel MJ, Henrion M. 1993. Efficient reasoning in qualitative probabilistic networks. *Proc 11th Nat Conf on AI,* 548-553.

Fiedler A, Horacek H. 2001. Argumentation in Explanations to Logical Problems. *Computational Models of Natural Language Arguments. Proc ICCS 2001*. Springer LNCS 2073, 969-978.

Green N. 2005. A Bayesian Network Coding Scheme for Annotating Biomedical Information Presented to Genetic Counseling Clients. *J Biomed Inf,* 38: 130-144.

Green N. 2006. Representing Normative Arguments in Genetic Counseling. *AAAI SSS: Argumentation for Consumers of Healthcare.*

Green N. To appear. Argumentation in a Causal Probabilistic Humanistic Domain. *Int J Intell Sys*.

Green N, Britt T, Jirak K. 2004. Communication of Uncertainty in Genetic Counseling Patient Education Systems. *AAAI FSS: Dialog Sys for Health Commun*.

Green N, Britt T, Jirak K, Waizenegger D, Xin X. 2005. User Modeling for Tailored Genomic E-health Information. *User Modeling 2005 Workshop: Personalisation for eHealth*,

Korb K, Nicholson AE. 2004. *Bayesian artificial intelligence.* Chapman Hall/CRC, Boca Raton, Florida.

Reed C, Long D. 1997. Content ordering in the generation of persuasive discourse. *IJCAI-97*, 1022-27.

Teufel S, Moens M. 2002. Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status. *CL*, 28(4):409-445.

Toulmin SE. 1998. *The uses of argument*. 9th ed. Cambridge Univ. Press, Cambridge, England.

Zukerman I, McConacy R, Korb K. 2000. Using argumentation strategies in automated argument generation. *Proc INLG-2000*.

letter CF	New	Delete	Reorder	letter WS	New	Delete	Reorder
	0	2	1		0	2	4
	0	2	0		0	0	1
	0	0	1		0	1	1
	0	0	6		2	1	1
	0	2	1		1	0	0
AVG	0	1.2	1.8	AVG	0.6	0.8	1.4
STDEV	0	1.1	2.4	STDEV	0.9	0.8	1.5

Table 1. Number of revisions in letters CF and WS. (N=5)

T1: 1 mutated alleles
T2: 0 mutated alleles

T1: 1 mutated alleles
T2: 0 mutated alleles

T1: 0 mutated alleles
T2: 1 mutated alleles

T1: 0 mutated alleles
T2: 1 mutated alleles

GJB2 (mother)   GJB2 (father)

Other genotype (mother)   Other genotype (father)

X^0

X^0

T1: 2 mutated alleles
T2: 0 mutated alleles

GJB2 (child)

Other genotype (child)

$X^.$ $X^.$

T1: 0 mutated alleles
T2: 2 mutated alleles

Test result (child)   Hearing loss (child)   Non-syndromic (child)

T1: unknown
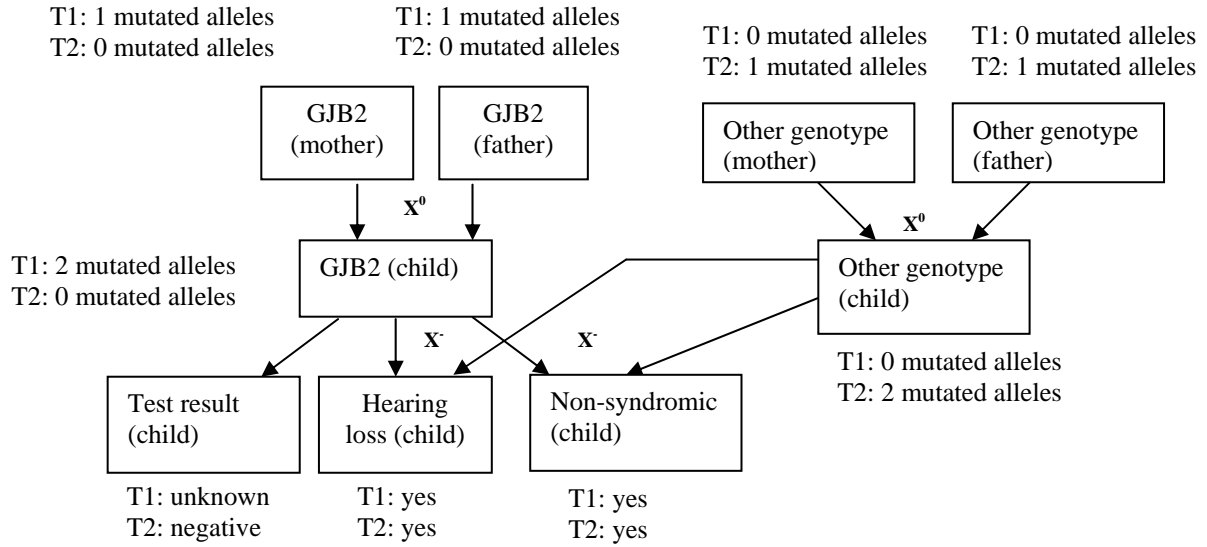T2: negative

T1: yes
T2: yes

T1: yes
T2: yes

Figure 1. Qualitative causal probabilistic network for hearing loss case.

	[Patient] was referred by [doctor] to [clinic] on [date] for evaluation.
D_1	She has had frequent respiratory infections.
W_1	A genetic condition known as cystic fibrosis (CF) can cause respiratory problems.
B_1	Eighty percent of CF patients have chronic respiratory complaints.
C_1	[Doctor] suspected that CF could be the cause of her respiratory problems.
	Patient was given a sweat chloride test for CF.
D_2	The test showed an abnormal sweat chloride level (75 mmol/L).
W_2	A result over 60 mmol/L is considered positive for CF.
C_2 / D_3	It is very likely that [patient] has CF. This means that cells in [patient's] body contain two altered copies of a gene called CFTR.
W_3	This alteration affects organs that secrete mucous, such as the lungs. The alteration causes excessive secretions, resulting in frequent lung infections.
C_3	This alteration of the CFTR gene is most likely the cause of [patient's] respiratory problems.
D_4	Both of you, [patient's] parents, are of Northern European ancestry.
B_4	One in twenty-five people of N. European ancestry carry one altered copy of the CFTR gene.
C_4	Each of you could carry this alteration.
W_5	Our cells contain two copies of each gene. One copy is inherited from each parent. A child who inherited two altered copies of a gene must have gotten one ... from the mother and one … from the father.
D_5	Since [patient's] cells contain two altered copies of CFTR,
C_5	it is likely that she got one altered copy of CFTR from each of you.
C_6	This is likely
D_6	even though neither of you have cystic fibrosis.
W_6	When a parent has one altered copy and one normal copy of a gene such as CFTR, he or she is not usually affected. Someone who has only one altered copy is called a "carrier". A child who inherits two altered copies will be affected since she has no normal copy.
W_7	A couple in which both are carriers will have a one in four (25%) chance that each child that they conceive will inherit two altered copies and be affected. This also means that they have a three in four (75%) chance that the child will inherit at least one unaltered copy from one parent and not be affected.
D_7	Assuming that you are both carriers,
C_7	the chances for each child that you conceive together is 25% that the child will have CF, and 75% that the child will not have CF.

Figure 2. Letter used in experiment (column 2) with argument annotations (column 1).