Chapter 3

Research Strategy and Methodology

This part covers the research strategy and methodology; including (1) population and sampling design, (2) data collection, and (3) data analysis. This study has utilized the Bayesian Belief Network (BBN) as quantitative and qualitative instruments for the data analysis.

3.1 Population and Sampling Design

3.1.1 Population

The target population of this study includes ethnic minority groups residing scattered throughout the mountainous region along Thailand-Myanmar borders in Chiang Mai, Chiang Rai, and Mae Hong Son, including Akha, Hmong, Karen, Lahu, Lisu, and Yao ethnicities with a total population ofapproximately 378,000 persons or 1,200 villages (for more information, see appendix A).

3.1.2 Sampling Design

To determine the appropriate villages; it requires background information of the outbreak of Trichinosis provided by the Office of Disease Prevention and Control 10, the Bureau of Epidemiology, and the interview with knowledgeable individuals including the officers from the Department of Livestock

Development in Chiang Mai, Chiang Rai, and Mae Hong Son. The reported cases of Trichinellosis by sub district during 2003-2012 are mapped using Google Earth. Two highlanders' villages in Mae Ai district, Chiang Mai Province were selected to conduct the in-depth study, including one that experienced an outbreak namely Huai Chan Si village and one that has never experienced an outbreak namely Huai Ma Fueang. There are a total of 84 households in Huai Chan Si village and 118 households in Huai Ma Fueang village. Twenty-six households from Huai Chan Si village and 28 households from Huai Ma Fueang village were selected using simple random selection (See Figure A-3 and Figure A-4).

3.2 Data Collection

The survey instruments including questionnaire, environmental survey, in-depth interview, evaluation form and focus group are used in this study.

3.2.1 Institution Survey

To understand the roles of institutions, we conducted in-depth interviews with staff working at the Department of Livestock Development, the Tambon Health Promoting Hospital, the Bureau of Epidemiology, and the Office of Disease Prevention and Control 10 in reducing parasitic zoonoses transmission.

3.2.2 Household Survey

A questionnaire was developed for the household survey based on the Trichinellosis risk factors deriving from experts' opinion. Twelve enumerators including 8 students from the faculty of Veterinary Medicine and 4 students from the faculty of Economics, Chiang Mai University were trained on how to conduct the questionnaire in the selected villages and at the same time the questionnaire is tested.

3.2.3 Environmental Survey

An environmental survey form was developed by an expert to investigate environmental factors related with Trichinellosis risk. To help the enumerators to understand the transmission of the disease, they were trained by the experts to understand One Health Approach. This form considers the interaction of highlanders with the pigs they grow and their environment as a single system.

3.2.4 Focus Groups

After conducting the household survey, we developed a set of data preparing for the experts to evaluate the Trichinellosis risk circumstance in the selected villages using experts' meeting. Seven experts are invited to join the focus groups, including;

a) Animal Health Experts

Assist.Prof.Panuwat Yamsakul	Faculty of Veterinary Medicine, Chiang Mai University
Dr.Veerasak Punyapornwithaya	Faculty of Veterinary Medicine, Chiang Mai University
Ms. Pornpen Tablerk	Department of Livestock Development, Nan Province

b) Disease Ecologist

Prof. Bruce A. Wilcox	Integrative Research & Education Program, Faculty of Public Health, Mahidol University and Tropical Disease Research Laboratory, KhonKaen University

c) Human Health Experts

Assoc.Prof.Dr.Pichart Uparanukraw	Faculty of Medicine, Chiang Mai University
Assoc.Prof.Dr.Nimit Morakote	Faculty of Medicine, Chiang Mai University
Mr. Adulsak Wijit	The Office of Diseases Prevention and Control 10

3.3 Data Analysis

3.3.1 Descriptive Statistics

Descriptive statistics were used to quantitatively describe the collected data. They are divided into 4 sub-systems, including: animal husbandry, food chain, environment, and economic condition.

3.3.2 Modeling

To conduct an in-depth household study within the limited time and financial resources, only 54 households were randomly selected. The complexity of these circumstances has led to model-based approaches for investigating the interconnections and for predicting management outcomes (Jakeman et al., 2006). A probabilistic graphical model for qualitative instrument called BBN is applied for this analysis since it does not need large amounts of empirical data. The conceptual transdisciplinary framework of Trichinellosis risk is developed by experts based on the existing knowledge and the experience from the field study to explain interconnection of the risk factors. It is also applied to solve decision problems related with management of the relevant institutions attempting to reduce the risk. When the net is solved, a decision rule which indicates choices for making a certain decision for each possible condition will be found for the node that optimizes the expected utility.

a) Purposes of the Modeling

The purposes of this modeling are to, first, gain a better understanding of the transmission of Trichinellosis, second, solve decision problems in management systems related to pig production and the public health situation to reduce Trichinellosis, and, finally, develop a universal Trichinellosis risk model explaining the circumstance in other areas.

b) Developing the Models

The Trichinellosis risk framework was developed based on the opinions of veterinarians, disease ecologists, medical doctors and public health officers (See Figure B-1). There are a total of 77 variables to be studied categorized into four subsystems to investigate Trichinellosis risk, including: animal husbandry, food chain, environment and economic condition.There are two kinds of variables in this study, including discrete data and continuous data. These variables are associated with probabilistic functions (the states of each variable are explained in the Appendix E). There are two sources of information to feed in the model, including the data from the field study and the data from experts' opinions. Netica, a powerful and easy-to-use program for working with BBN and influence diagrams are applied to analyze this set of data.

c) Specifying Modeling Context

We broke the Trichinellosis risk framework into two parts based on the decision problems that we attempted to investigate. There are two main decision problems in management systems related with pig production and public health situation to reduce Trichinellosis, including the decision to switch from the original pig production mode of keeping pigs in pens and the decision to stop consuming raw or undercooked meat.

a. Institution's Decision to Encourage Villagers to Switch to Keep Pig Pen

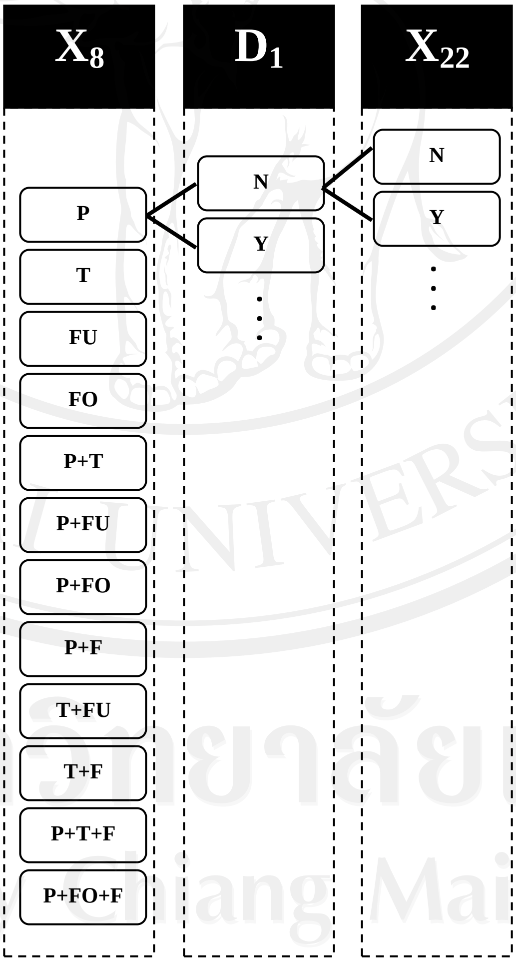

Figure 3-1: Decision tree representing the institution's decision to encourage people to switch to keep pig in pen

$$U_1 = f(X_8, D_1, X_{22})$$

Given, U_1 = Benefits from switching to keep pigs in pen

X_8 = Pig production modes

D_1 = Institution's decision to construct pig pen

X_{22} = Household's decision to keep pigs in pens

In order to reduce the possibility of getting infected by Trichinella and other parasites in animals, a complex set of issues must be considered, mainly social and economic trade-offs. In considering a campaign to change the original pig production mode to keeping pigs in pens, the benefit of doing this is the reduction in the possibility of getting infected by Trichinella and other parasites in animals which in turn yields a higher productivity and reduces the risk of getting Trichinellosis and other parasitic zoonoses in humans. However, keeping pigs in pens bears a huge cost to the farmer. The cost of construction is seen as a small portion if we take the opportunity costs into consideration (more details of cost structures and revenue streams of different pig production modes can be seen in Appendix G). Those who do crop farming as a primary career need to devote their time in a field whichis located far away from the village. Many of them decide to let their pigs roam freely because they do not have to prepare feed for them which takes hours to prepare.

Assuming that an institution, for example, a governmental institution, has unlimited money to construct pig pens for villagers and that will not affect to its utility. The institution's decision whether to provide money to construct pig pens (3,981.43 Baht each, the average cost derived from the field study) for the pig growers is based on the satisfaction of a household from switching the practices and that we considered only the average gains (or losses) a household will face if it changes the practice. If the institution does not want to support the money to construct pig pens for the pig growers, they have to bear this cost by themselves. We calculated the average gains of each pig production mode and compared those with the average gain from raising pigs in pens (see Table C-1). We ignore the possible benefits from the reduction in the risk that pigs will be infected by Trichinella from keeping pigs in pens.

b. Institution's Decision to Encourage People to Stop Consuming Raw or Undercooked Meat

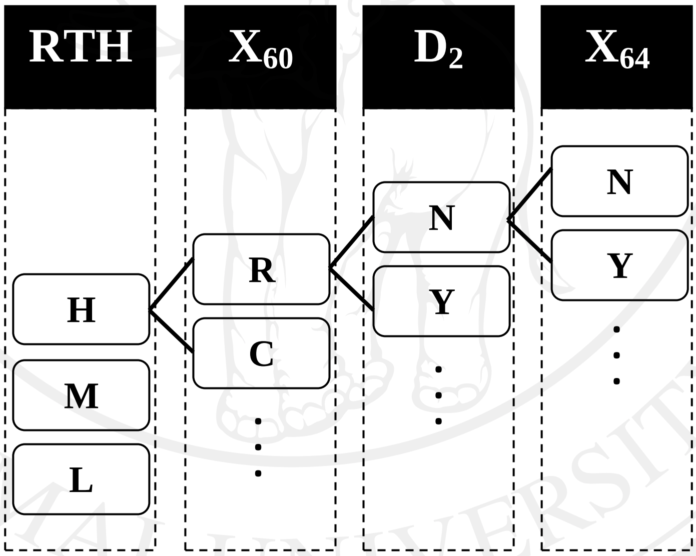

Figure 3-2: Decision tree representing the institution's decision to encourage people to stop consuming raw or undercooked meat

$$U_2 = g(RTH, X_{60}, D_2, X_{64})$$

Given, U_2 = Benefits from decision to stop eating raw/undercooked meat

RTH = Risk of getting Trichinellosis in human

X_{60} = Meat preparation

D_2 = Institution's decision to encourage people to stop consuming raw/undercooked meat

X_{64} = Individual's decision to stop eating raw/undercooked meat

The second decision is to decide whether the government should go to the field to encourage people to stop consuming raw or undercooked meat. In each year, the public health officers in both local offices and provincial offices have put in effort trying to encourage people, especially those who live in the country side, to stop consuming raw or undercooked meat by providing them knowledge about the danger of consuming raw or undercooked meat. Even though these people are educated about the harm of consuming raw or undercooked meat, they still insist on consuming it. This means that no matter how much public health officers put in effort to encourage people to stop consuming raw or undercooked meat, if they are not aware of the danger, they still will not change their behavior. Therefore, if the public health officers understand the behavior and attitude of the villagers very well, they can decide whether they should keep educating them of the danger of consuming raw or undercooked meat or should stop and rather put the effort on other issues instead. On the other hand, if the decision makers know that some villagers are undereducated about the danger of consuming raw or undercooked meat and if they are educated they tend to change behavior, therefore, the effort that public health put will be quite effective and worth the money and time.

Contrary to the previous decision, with this decision we assume that an institution, for example, a public health organization, has limited money to encourage people to stop eating raw or undercooked meat and that is its decision to allocate resources wisely. A local institution spends approximately 6,000 Baht each time it visits village providing knowledge about hygiene.

The decision of an institution whether to launch a campaign to encourage people to stop eating raw/undercooked meat depends on benefits a household will receive from stopping eating raw/undercooked meat. These benefits can be calculated from the reduction in the burden of illness or the loss from death from Trichinellosis. In so doing, we consider the risk that an individual can get Trichinellosis that is evaluated by human health experts. We also take the severity of the illness in to consideration. However, since we do not have enough information to calculate the severity of getting Trichinellosis, we assume that those who have higher risk may face higher severity. The higher severe case bears higher economic losses. There are three levels of severity including high, medium and low levels. We derived the data on the economic losses from the illness from the case outbreak in Nan Province. Assoc.Prof.Dr.Pichart Uparanukraw, a human health expert determined the levels of severity of the illness from the case outbreak. In addition, we also take the individual's decision whether to stop eating raw/undercooked meat or not and the costs that the institution bears in visiting a village in order to provide the knowledge to villagers into account. The economic losses of illness and death from Trichinellosis in human can be seen in Table D-1.

c) Model Structure and Parameters

a. Decision to Switch to Keep Pigs in Pens

Posterior Probability Equation of TIP

Based on Bayesian statistics, the posterior probability for this model is defined as;

$$p(TIP_i | X^A) = \frac{p(X^A | TIP_i)}{p(X^A)} \cdot p(TIP_i)$$

Where, i = Levels of the risk that pigs will be infected by Trichinella

= {H,M,L}

X^A = All risk factors associated Trichinella infection in pigs

= $\{X_1, X_2, \ldots, X_{57}\}$

$p(TIP_i|X^A)$ = Posterior probabilities (or probabilities of the parametersTIP$_i$) given evidence X^A

$p(X^A|TIP_i)$ = Likelihood functions (or the probabilities of evidence X^A) given the parametersTIP$_i$

$p(TIP_i)$ = Prior probability probabilities (or the probabilities of risk that pigs will be infected by Trichinella based on the subjective assessment of experienced experts)

$p(X^A)$ = Probability of all evidences in set X^A, regardless of any other information

From the law of total probability,

$p(X^A)$ = $p(X|X_1{=}H)p(X_1{=}H)+p(X|X_1{=}M)p(X_1{=}M)+p(X|X_1{=}L)p(X_1{=}L)+$
$p(X|X_2{=}H)p(X_2{=}H)+p(X|X_2{=}M)p(X_2{=}M)+p(X|X_2{=}L)p(X_2{=}L)+\ldots+$
$p(X|X_{57}{=}H)p(X_{57}{=}H)+p(X|X_{57}{=}M)p(X_{57}{=}M)+p(X|X_{57}{=}L)p(X_{57}{=}L)$

Expected Utility Function of the Decision to Switch to Keep Pig in Pen

$$E\left[U_1(D_1)\middle|p(X^A)\right] = \sum_{j,k,l} U_1\left(X_8^j, X_{22}^k, D_1^l\right) p\left(X_8^j\right) p\left(X_{22}^k\right) p(D_1^l)$$

Where, $E\left[U_1(D_1)\middle|p(X^A)\right]$= Expected utility or expected consequences from a decision making of supporting pen construction for pig growers under uncertainty about Trichinellosis risk.

j = {P,T,FU,FO,P+T,P+FU,P+FO,P+F,T+FU,T+F,P+T+F,P+FU+F,P+FO+F}

k = {N,Y}

l = {N,Y}

However, X_{22} = f(TIP), and we are interested to see the effect of TIP on the expected outcome (U_1). The posterior probability of X_{22} is defined as,

$$p\left(X_{22}^k \middle| TIP_i\right) = \frac{p(TIP_i|X_{22}^k)}{p(TIP_i)} \cdot p(X_{22}^k)$$

$$\therefore\ p\left(X_{22}^k\right) = \frac{p\left(X_{22}^k \middle| TIP_i\right) p(TIP_i)}{p(TIP_i|X_{22}^k)}$$

From the posterior probability of TIP,

$$p(TIP_i \middle| X^A) = \frac{p(X^A|TIP_i)}{p(X^A)} \cdot p(TIP_i)$$

$$p(TIP_i) = \frac{p(TIP_i \middle| X^A)\ p(X^A)}{p(X^A|TIP_i)}$$

$$p\left(X_{22}^k\right) = \frac{p\left(X_{22}^k \middle| TIP_i\right) p(TIP_i \middle| X^A)\ p(X^A)}{p(X^A|TIP_i) p(TIP_i|X_{22}^k)}$$

$\therefore\ E[U_1(D_1) \middle| p(X^A)]$

$$= \sum_{i,j,k,l} U_1\left(X_8^j, X_{22}^k(TIP_i), D_1^l\right) p\left(X_8^j\right) p(D_1^l) \frac{p\left(X_{22}^k \middle| TIP_i\right) p(TIP_i \middle| X^A)\ p(X^A)}{p(X^A|TIP_i) p(TIP_i|X_{22}^k)}$$

An institution will make a decision whether to support pen construction for pig growers or not based on expected utility maximization.

However, in statistics, we normally use loss function instead of utility function;

$$L_1\left(X_8^j, X_{22}^k(\ TIP_i), D_1^l\right) = -U_1(X_8^j, X_{22}^k(TIP_i), D_1^l)$$

Expected loss:

$\rho_1(p(TIP_i|X^A), D_1)$

$$= \sum_{i,j,k,l} L_1\left(X_8^j, X_{22}^k(TIP_i), D_{1_l}\right) p\left(X_8^j\right) p(D_{1_l}) \frac{p\left(X_{22}^k \middle| TIP_i\right) p\left(TIP_i \middle| X^A\right) p(X^A)}{p(X^A|TIP_i) p(TIP_i|X_{22}^k)}$$

The objective is to make a decision (choose whether to support pen construction for pig growers or not) that minimizes the expected loss ($\rho_1{}^*$) based on the posterior probability of TIP.

$$\rho_1(p(TIP_i|X^A), D_1{}^*) = \rho^*(p(TIP_i|X^A))$$

b. Institution's Decision to Encourage People to Stop Consuming Raw or Undercooked Meat

Posterior Probability Equation of RTH

Based on Bayesian statistics, the posterior probability for this model is defined as;

$$p(RTH_m|X^B) = \frac{p(X^B|RTH_m)}{p(X^B)} \cdot p(RTH_m)$$

Where, m = Levels of the risk that humans will be infected by Trichinellosis

= {H,M,L}

X^B = All risk factors associated Trichinella infection in pigs

= $\{ X_1, X_2, \ldots, X_{69}\}$

$p(RTH_m|X^B)$ = Posterior probabilities (or probabilities of the parametersRTH$_m$) given evidence X^B

$p(X^B|RTH_m)$ = Likelihood functions (or the probabilities of evidence X^B) given the parameters RTH_m

$p(RTH_m)$ = Prior probability probabilities (or the probabilities of risk that human will be infected by Trichinellosis based on the subjective assessment of experienced experts)

$p(X^B)$ = Probability of all evidences in set Y, regardless of any other information

From the law of total probability,

$p(X^B)$ = $p(X|X_1{=}H)p(X_1{=}H)+p(X|X_1{=}M)p(X_1{=}M)+p(X|X_1{=}L)p(X_1{=}L)+$ $p(X|X_2{=}H)p(X_2{=}H)+p(X|X_2{=}M)p(X_2{=}M)+p(X|X_2{=}L)p(X_2{=}L)+\ldots+$ $p(X|X_{69}{=}H)p(X_{69}{=}H)+p(X|X_{69}{=}M)p(X_{69}{=}M)+p(X|X_{69}{=}L)p(X_{69}{=}L)$

Expected Utility Function of the Decision to Stop Consuming Raw or Undercooked Meat

$$E[U_2(D_2)|p(X^B)] = \sum_{m,n,p,q} U_2(RTH_m, X_{60}^n, X_{64}^n, D_2^q)p(RTH_m)p(X_{60}^n)p(X_{64}^n)p(D_2^q)$$

Where, $E[U_2(D_2)|p(X^B)]$= Expected utility or expected consequences from decision making to encourage people to stop consuming raw/undercooked meat under uncertainty about Trichinellosis risk.

m = {H,M,L}

n = {R,C}

p = {N,Y}

q = {N,Y}

However, we are interested to see the effect of RTH on the expected outcome (U_2).

From the posterior probability of RTH,

$$p(RTH_m|X^B) = \frac{p(X^B|RTH_m)}{p(X^B)} \cdot p(RTH_m)$$

$$p(RTH_m) = \frac{p(RTH_m|X^B)\, p(X^B)}{p(X^B|RTH_m)}$$

$\therefore E\left[U_2(D_2)\middle|p(X^B)\right]$

$$= \sum_{m,n,p,q} U_2\left(RTH_m, X_{60}^n, X_{64}^n, D_2^q\right) p(X_{60}^n) p(X_{64}^n) p\left(D_2^q\right) \frac{p\left(RTH_m \middle| X^B\right) p\left(X^B\right)}{p\left(X^B \middle| RTH_m\right)}$$

An institution will make a decision whether to encourage people to stop consuming raw/undercooked meat or not based on expected utility maximization.

However, in statistics, we normally use loss function instead of utility function;

$$L_2\left(RTH_m, X_{60}^n, X_{64}^n, D_2^q\right) = -U_2\left(RTH_m, X_{60}^n, X_{64}^n, D_2^q\right)$$

Expected loss:

$\rho_2(p(RTH_m|Y), D_2)$

$$= \sum_{m,n,p,q} L_2\left(RTH_m, X_{60}^n, X_{64}^n, D_2^q\right) p(X_{60}^n) p(X_{64}^n) p\left(D_2^q\right) \frac{p(RTH_m|X^B) p(X^B)}{p(X^B|RTH_m)}$$

The objective is to make a decision (choose whether to encourage people to stop consuming raw/undercooked meat or not) that minimizes the expected loss (ρ_2^*).

$$\rho_2(p(RTH_m|X^B), D_2^*) = \rho_2^*\left(p(RTH_m|X^B)\right)$$

d) Testing the Modeling

The objective of this test is to evaluate the quality of the Bayesian Networks (Appendix B) using a set of real cases using Netica. This test will illustrate how well the models match the actual cases by considering the actual belief levels of the states in determining how well they agree with the value of the case file. We first incorporate 60% of the cases into the model. Then, the nodes in which we wish to find their inferences, including, TIP and RTH nodes were selected. We used 40% of the samples to verify the validity of the model. When the Netica was done, it printed a report called scoring rule results of each of the selected nodes (see Table C-3 and Table C-4). The reports included error rate, logarithmic loss score, quadratic (Brier score), and spherical payoff score.

Error rate determines how many times the classifier misclassifies a case divided by the number of classifications. It is only with respect to the probability distribution of the test cases.

Logarithmic loss values are calculated using the natural log. The values are between zero and infinity. Zero indicates the best performance.

$$\text{Logarithmic loss} = \text{MOAC}\ [-\log(P_c)]$$

Quadratic loss values or the Brier score are between zero and two. Zero indicates the best performance.

$$\text{Quadratic loss} = \text{MOAC}\ [1 - 2(P_c) + \textstyle\sum_i^n p_i^2]$$

Spherical payoff values are between zero and one. One represents the best performance.

$$\text{Spherical payoff} = \text{MOAC}\ \left[\frac{p_c}{\sqrt{\sum_i^n p_i^2}}\right]$$

Where, P_c = Probability predicted for the correct state

P_i = Probability predicted for state i, n is the number of states

MOAC = Mean (average) over all cases (Norsys Software Corp., 2013)

Another way to verify the validity of the models is to use Netica to pass through the case file by processing cases one-by-one. For each case, the software reads the case except the nodes that we wish to find their inferences. After that, the software will revise the actual value for those nodes and compare them with the beliefs the model generated. Netica accumulates all the comparisons as illustrated in Table C-4 and Table D-4. The models were selected based on the values of sum square error (SSE). The less SSE, the best the model is.

For animal health perspective, the values of logarithmic loss and quadratic loss of the model 2 were slightly smaller than the model 1, while the value of spherical payoff value of the model 2 was slightly larger. Though, the model 2 yielded slightly larger of sum square error (SSE) than the model 1, the model 2 was selected since it yielded a lot less error rate of only 20%.

For the human health perspective, the scoring rule results and the error rate yielded no difference values between the model 1 and the model 2. Though, model 2 yielded a slightly larger value of SSE, it was selected since it was much less complicated than model 1.

After we derived the models, we incorporated all the data into the selected models. As a result, we would see the learned probability distributions appeared in each node (see Figure C-1, C-2, D-1 and D-2).

In order to solve the decision problems, we augmented the decision node and the utility node into the models. For the human health perspective, we augmented the institution's decision whether to encourage people to stop consuming raw or undercooked meat or not. For animal health perspective, we augmented the institution's decision whether to support the constructing cost of pens to pig growers or not. Netica would attach a deterministic function which provided a value for the decision node for each possible configuration of parent values. The links into a decision node indicate what the decision maker will know when he is about to make the decision. For the human health perspective, we assume that the institution may know the knowledge of food-preparing persons and their attitudes to change the eating habits. On the other hand, for the animal health perspective, we assume that the institution may know the pig production mode that pig growers apply and their attitudes towards changing the practices. The decision function from the decision node will maximize the expected value of the sum of the utility node (see Figure C-3 and Figure D-3).