Food Research International 89 (2016) 463–470

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

A holistic approach to food safety risks: Food fraud as an example

Hans J.P. Marvin *, Yamine Bouzembrak, Esmée M. Janssen, H.J. van der Fels- Klerx, Esther D. van Asselt, Gijs A. Kleter

RIKILT - Wageningen UR, Akkermaalsbos 2, 6708 PD Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 25 May 2016
Received in revised form 27 July 2016
Accepted 22 August 2016
Available online 25 August 2016

Keywords:
Food fraud
Holistic approach
Bayesian network
Food safety risks
Prediction

ABSTRACT

Production of sufficient, safe and nutritious food is a global challenge faced by the actors operating in the food production chain. The performance of food-producing systems from farm to fork is directly and indirectly influenced by major changes in, for example, climate, demographics, and the economy. Many of these major trends will also drive the development of food safety risks and thus will have an effect on human health, local societies and economies. It is advocated that a holistic or system approach taking into account the influence of multiple "drivers" on food safety is followed to predict the increased likelihood of occurrence of safety incidents so as to be better prepared to prevent, mitigate and manage associated risks. The value of using a Bayesian Network (BN) modelling approach for this purpose is demonstrated in this paper using food fraud as an example. Possible links between food fraud cases retrieved from the RASFF (EU) and EMA (USA) databases and features of these cases provided by both the records themselves and additional data obtained from other sources are demonstrated. The BN model was developed from 1393 food fraud cases and 15 different data sources. With this model applied to these collected data on food fraud cases, the product categories that thus showed the highest probabilities of being fraudulent were "fish and seafood" (20.6%), "meat" (13.4%) and "fruits and vegetables" (10.4%). Features of the country of origin appeared to be important factors in identifying the possible hazards associated with a product.

The model had a predictive accuracy of 91.5% for the fraud type and demonstrates how expert knowledge and data can be combined within a model to assist risk managers to better understand the factors and their interrelationships.

1. Introduction

Against a background of previous food safety incidents, such as BSE and dioxins, the last two decades have witnessed the establishment of increasingly sophisticated and elaborated food safety control systems, dedicated institutions, and public awareness of food safety. Despite these preventive measures, incidents do still occur. While some of the incidents were due to unintended or unforeseen consequences of practices or processes, others were linked to fraud and criminal activities, such as the adulteration of foods with non-food-grade materials such as Sudan dyes and dioxin-containing oils. Several studies have indeed mentioned fraud and criminal attacks as new threats to food safety (Spink & Moyer, 2011).

In order to prevent incidents from happening in the future, various researchers have explored the possibilities to forecast, including the timely identification of trends and events that might eventually give rise to, such food safety incidents. Generally, various international and local developments may directly or indirectly influence the performance of food-producing systems, among them climate change, economy and trade, human behaviour, and new technologies (Boland et al., 2013; GO-Science, 2011; Godfray et al., 2010; Miraglia, De Santis, Minardi, Debegnach, & Brera, 2005). The European Food Safety Authority (EFSA) defined a driver as "a driver may act as modifiers of effect on the onset of emerging risks, namely they can either amplify or attenuate the magnitude or frequency of risks arising from various sources" (EFSA, 2010). The key drivers to food safety and nutrition risks were identified recently in a scoping study on food safety and nutrition (FCEC, 2013): i) global economy and trade, ii) global cooperation and standard setting, iii) governance, iv) demography and social cohesion, v) consumer attitudes and behaviour, vi) new food chain technologies, vii) competition for key resources, viii) climate change, ix) emerging food chain risks and disasters, and x) new agri-food chain structures.

A holistic or system approach taking stock of the forces that act upon the food chain (from farm to fork) and their effect on food safety has been adopted by FAO (2003). Such an approach has also been proposed to address climate induced food safety risks (GO-Science, 2011; Marvin et al., 2009; Miraglia et al., 2005). A holistic approach includes a full host

* Corresponding author at: RIKILT Wageningen UR, Akkermaalsbos 2, PO BOX 230, 6700 AE Wageningen, The Netherlands.
E-mail address: Hans.marvin@wur.nl (H.J.P. Marvin).

http://dx.doi.org/10.1016/j.foodres.2016.08.028
0963-9969/

environment analysis of the whole food chain in which the driving forces of food safety risks are determined, including associated data sources.

Application of the holistic approach in a working system to identify known and emerging food safety risks needs a model that links all drivers and their dependencies, as well as underlying databases from various natures and origins and preferably allows scenario studies. The model should access, preferably real-time, data on the drivers, process these data, and perform calculations, such to provide predictions on (emerging) food safety risks.

To date, no modelling approach or system has been developed for the food production chain that is able to take into account underlying databases, interactions and feed-back loops of the drivers as encountered in a holistic approach. Here we advocate that a Bayesian Network (BN) approach is suitable for this purpose. BNs are a class of probabilistic models originating from the Bayesian statistics and decision theory combined with graph theory (Bonafede & Giudici, 2007; Nielsen, 2007), which are able to model dependencies between variables, manage non-linear interactions and integrate different kinds of information about the system such as expert knowledge, measurement data, feedback experience and information regarding the system behaviour (Buriticá & Tesfamariam, 2015). BNs have been applied in a number of diverse problem domains such as medical diagnosis (Wiegerinck et al., 1999), image classification (Malka & Lerner, 2004), financial fraud detection (Kirkos, Spathis, & Manolopoulos, 2007; Ngai, Hu, Wong, Chen, & Sun, 2011), nuclear waste disposal (Lee & Lee, 2006) and electrical power systems (Huo, Zhu, Zhang, & Chen, 2004).

For this study, fraud was chosen as a case particularly given its role in a number of landmark food safety incidents, such as the illegal use of Sudan dyes, admixture of PCB/dioxin containing industrial oils with edible oils, and addition of melamine to milk used for infant formula (Unnevehr et al., 2010), (Guan et al., 2009; Jia & Jukes, 2013). Food fraud is a collective term that is driven by economic gain and encompasses the deliberate substitution, addition, tampering, or misrepresentation of food, food ingredients or food packaging, or false or misleading statements made about a product (Spink & Moyer, 2011). Food fraud may thus cause food safety risks and is driven by different factors from within and outside the food supply chain (NSF International, 2014). Examples of factors contributing to food fraud opportunities were presented in the NSF international report (NSF International, 2014): (i) increase of the complexity of supply chain networks, (ii) the rapid development of technology (internet, printing, mobile phone etc.) provides powerful tools to fraudsters, (iii) the rapid growth of warehouse systems and refrigerated transport enabling the long term storage and transfer of large quantities of perishable food. The aim of this research was to demonstrate the usefulness of BNs in connecting different drivers, data sources and their interactions in a holistic approach in order to determine the main factors influencing food fraud.

2. Methods

The approach applied consisted of four steps: (i) collection of reported food fraud cases, (ii) identification of the main drivers that can affect food fraud and collect data from different data sources such as: literature (Everstine, Spink, & Kennedy, 2013; NSF International, 2014), food fraud databases, (EMA, 2014; RASFF, 2015) and food fraud expert knowledge; (iii) building the BN including nodes, arrows, states, and the parameters for each node in the form of Conditional Probability Tables (CPTs); (iv) and validating the model. Fig. 1 shows the steps followed.

2.1. Collecting reported food fraud cases

Two publicly available databases that publish (detected) food fraud cases in the EU (Rapid Alert for Food and Feed (RASFF)) (RASFF, 2015) and in the US (Economically Motivated Adulteration incidents database (EMA)) (EMA, 2014) were used as references to real detected cases. All notifications reported in the RASFF database under the hazard category "adulteration/fraud" were extracted from the period 01/01/2000 to 31/12/2015 (in total 1035 records). Each RASFF record contains the following information: month, year, country notifying, notification type, product, product category, hazard and origin country. All food fraud incidents reported in the EMA database were extracted from the period 01/01/2000 to 31/12/2015. The database contains 651 distinct food fraud incidents grouped into 20 food product categories. The food fraud incidents reported in RASFF and EMA (i.e. in total 1686 records) were divided into seven categories based on the description provided in these databases. The categories are defined in Table 1. For each food fraud report in RASFF and EMA, data on the identified drivers (see next section) from the time (i.e. month and year) of the particular food fraud incident were collected and stored in an underlying database. In this way, information recorded in RASFF and EMA (e.g. fraud type, product, country of origin, detecting country and date of detection) was expanded with available information retrieved from other data sources, as related to drivers (i.e. in total 15 different data sources) that are directly or indirectly linked to the occurrence of the food fraud incident.

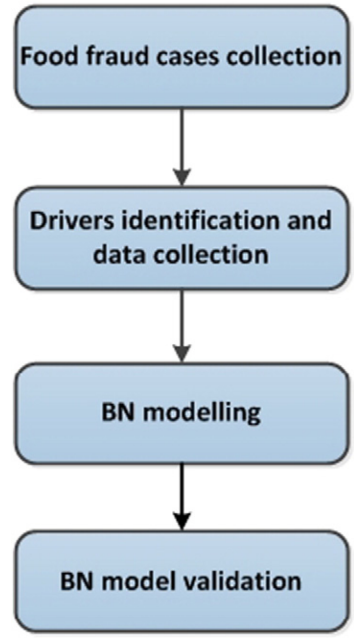

Fig. 1. Steps in the development of the BN model for food fraud.

Table 1
Food fraud types in RASFF and EMA.

Fraud type	Description
HC	Improper, fraudulent, missing or absent Health Certificate (HC)
Illegal importation	Illegal or unauthorized import, trade or transit
Tampering	Adulteration, fraud or tampering, substitution, counterfeit, artificial enhancement, transhipment, intentional distribution of contaminated product, dilution.
CED	Improper, expired, fraudulent or missing common entry document (CED), import declaration, or analytical report
Expiration date	Expiration date
Origin labelling	Mislabelling, origin labelling
Theft and Resale	Theft and resale

Table 2
Bayesian Network variables, definition and data sources.

Variables (nodes)	Definition	Data source
Year	Food Fraud incident year	https://webgate.ec.europa.eu/rasff-window/portal/ http://www.foodfraudresources.com/ema-incidents/
Month	Food Fraud incident month	https://webgate.ec.europa.eu/rasff-window/portal/ http://www.foodfraudresources.com/ema-incidents/
Data source	Food Fraud incident data source	https://webgate.ec.europa.eu/rasff-window/portal/ http://www.foodfraudresources.com/ema-incidents/
Product	Product name	https://webgate.ec.europa.eu/rasff-window/portal/ http://www.foodfraudresources.com/ema-incidents/
Product category	Product category	https://webgate.ec.europa.eu/rasff-window/portal/ http://www.foodfraudresources.com/ema-incidents/
Price (Y)	Product price (year)	http://ec.europa.eu/eurostat/en
Price (M)	Product price (Month)	http://ec.europa.eu/eurostat/en
Demand	Product demand increase	http://ec.europa.eu/eurostat/en
Fraud profitability	Profitability level of the fraud	(NSF International, 2014), Expert judgment
CPI(O)	The Corruption Perception Index (CPI) of the origin country	http://www.transparency.org/cpi2014
Governance (O)	The governance index of the origin country	http://info.worldbank.org/governance/wgi/index.aspx#home
Legal system (O)	Whether there is a legal system of the food in the origin country	http://data.worldbank.org/indicator/IC.LGL.CRED.XQ
GDP (O)	The GDP of the origin country	http://data.worldbank.org/indicator/NY.GDP.MKTP.CD
Economic growth (O)	The economic growth of the origin country	http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG
Country (O)	Product country of origin	https://webgate.ec.europa.eu/rasff-window/portal/ http://www.foodfraudresources.com/ema-incidents/
Supply chain (O)	The supply chain index of the origin country	http://www.fmglobal.com/page.aspx?id=04060000#!year =2015&idx=Index&handler=map
Trade volume	The trade volume per year	http://faostat3.fao.org/download/T/TM/E
Risk (O)	The political risk index (PRI) of the origin country	http://www.prsgroup.com/category/risk-index
Fraud detection	Whether the food fraud can be detected	(NSF International, 2014), Expert judgment
Country (N)	Control country name	https://webgate.ec.europa.eu/rasff-window/portal/ http://www.foodfraudresources.com/ema-incidents/
CPI (N)	The corruption level of the control country	http://www.transparency.org/cpi2014
Human development (O)	The human development index of the origin country	http://hdr.undp.org/en/data
Technology (O)	Global Innovation Index of the country of origin	https://www.globalinnovationindex.org/content.aspx?page=GII-Home
RASFF ratio (N)	RASFF ration of the control country	http://www.sciencedirect.com/science/article/pii/S0278691513001749
Press (O)	The press index of the origin country	https://index.rsf.org/#!/
Food fraud type	The type of food fraud	https://webgate.ec.europa.eu/rasff-window/portal/ http://www.foodfraudresources.com/ema-incidents/
Food safety (O)	The food safety level of the origin country	http://foodsecurityindex.eiu.com/
Food safety (N)	The food safety level of the control country	http://foodsecurityindex.eiu.com/
Fraud complexity	The complexity to fraud the food	(NSF International, 2014), Expert judgment
Price spikes	Whether the product price had spike in the market	http://ec.europa.eu/eurostat/en

N = notifying country; O = country of origin; M = month; Y = year.

2.2. Drivers identification and data collection

The first step for characterizing the food fraud risk of a product is to review the drivers known to be helpful in detection food fraud. These food fraud drivers are outlined in Table 2 and were determined by expert judgment using brainstorming methods (Wilson, 2010) with 4 food fraud experts, literature (Everstine et al., 2013; NSF International, 2014) and food fraud databases (EMA, 2014; RASFF, 2015). The included drivers are, among others, prices of the fraudulent product at the time of detection, trade volumes of the product between the country of detection and country of origin and the supply chain index of the product detected (i.e. it measures performance along the logistics supply chain within a country). Furthermore, it was determined whether or not there was a price spike of the fraudulent product around the period of fraud detection using a product price database (i.e. EUROSTAT). Characteristics of the country of origin and of the country that detected the incident were also included, such as indices for perceived corruption, food safety, governance, legal system, press, human development and technology. The owners of the databases accessed in this study varied greatly, from official governmental organisations (e.g. EUROSTAT, EFSA, FDA, World Bank) to independent private organisations such as Transparency International publishing the “corruption perception index” of a country. The purpose of this model was to demonstrate its applicability as a “holistic” approach. However, the flexibility of the BN approach allows an easy inclusion of additional data and/or expert knowledge when these become available. Next, publicly available databases were identified for each driver (Table 2). The data from these databases were stored in an Excel file and used to construct the BN model. The Excel file consists of one column for each node in the model. Each row in the file represents one case which consists of all available data for the various drivers.

2.3. BN model building

A BN is a directed graphical model that represents conditional probabilities among variables of interest. A BN contains: (i) a set of discrete or continuous variables $U = \{A_i, \ldots, A_n\}$ and a set of directed edges between variables; (ii) each discrete variable has a finite set of mutually exclusive states which simply explain the condition of a variable; (iii) the variables together with the directed edges form an acyclic directed graph (DAG). If there is an edge from A_i, to A_j, then we say that node A_i is the parent of A_j and A_j is the child of A_i. A variable A_i with its parents, $pa(A_i)$, specifies a conditional probability distribution, $P(A_i|pa(A_i))$. This is a CPT for a set of discrete variables (Nielsen, 2007). BN specifies a unique joint probability distribution of all variables, $P(U) = P(A_1, \ldots, A_n)$, given by the product of all conditional probability tables specified in BN:

$$P(U) = \prod_{i=1}^{n} P(A_i|pa(A_i)) \quad (1)$$

However, calculating P(U) for a large network is complex and intractable since P(U) grows exponentially with the number of variables. A BN approach provides a compact representation for calculating P(U) more efficiently by finding a marginal distribution of a variable, $P(A_i)$, or finding the conditional distribution of A_i given the evidence, e, $P(A_i|e)$. The notion of evidence means that some of the variables are observed and take values from their respective domains. Let $e_1, \ldots, e_m$ be findings. Then

$$P(U,e) = \prod_{i=1}^{n} P(A_i|pa(A_i)) \prod_{j=1}^{m} ej \tag{2}$$

and for A ∈ U we have

$$P(A|e) = \frac{\sum_{U\{A\}} P(U,e)}{P(e)} \tag{3}$$

For more details, an example of BN calculations is shown in the annex (see Appendix A).

To construct the BN and to calculate the CPTs of the model, we used approximately 83% of the collected data (i.e. 1393 cases), the software Hugin 8.3 (http://www.hugin.com/) and the expectation-maximization (EM)-algorithm (Denœux, 2010, 2011). In this setting, the relationship between all parameters was constructed. It does not provide cause-effect relationships but instead shows how the various parameters are influenced (statistically) by each other.

2.4. BN model validation

The BN model provides the probabilities for the states of all drivers of the model and these may be used to obtain an idea about the validity of the model by comparison to earlier findings by other researchers.

The validity of the BN model should also be judged by its ability to predict the right type of food fraud. Therefore, approximately 17% of the collected data (i.e. 293 cases) were used to validate the BN model. All variables (except the fraud type) were used as input parameters in the BN model to predict the "fraud type" as mentioned in the databases (RASFF and EMA). In case the country of origin was not known or when a combination of parameters was not seen before, these cases (i.e. 19 in total) were excluded in this validation (see Supplement 1). A score of 1 was given when the fraud type predicted by the BN model was similar (e.g. highest probability) to the food fraud type in the databases, and 0 was given when a wrong fraud type was predicted with the highest probability.

3. Results and discussion

3.1. Reported food fraud cases

All notifications reported in the RASFF database under the hazard category "adulteration/fraud" were extracted from the period 01/01/2000 to 31/12/2015 (in total 1035 records). Products of RASFF notifications linked to fraud in the period analysed were predominantly fish and seafood products (17.4%), nuts (16.3%) and meat products (15%), mixed food (9%), and fruits and vegetables (7%), which is similar to earlier reported findings, albeit for shorter time periods (Bouzembrak & Marvin, 2015; Tähkäpää, Maijala, Korkeala, & Nevas, 2015). It should be noted that one specific food fraud incident reported in RASFF may not all be unique cases, especially in the cases when a food item has crossed country borders in the EU, which may be the case when an alert is reported. In addition, origin country of the product not necessarily means its country of production, it could have been re-labelled in a free trade zone.

In the period analysed (01/01/2000 to 31/12/2015), the EMA database contained 651 distinct food fraud incidents grouped into 20 food product categories. The most cited food fraud incidents were fish and seafood (25.6%), fruits and vegetables (15.5%), meat (10.8%), dairy products (9.7%), oils and fats (7.2%), and alcoholic products (5.6%). The purpose and focus of EMA are somewhat different from RASFF. The EMA database focuses on the intentional adulteration of food for economic gain, or food fraud. Information for this database is compiled through literature and media searches of EMA incidents in food products since 1980. Sources include: LexisNexis, PubMed, Google, FDA Consumer and FDA recall records, state reports, and reports from RASFF. Therefore, some overlap may exist between the records collected in this study from EMA and RASFF. The presence of potential duplicates will influence the accuracy of the BN model derived from this data. However, as the purpose of this study was to demonstrate its potential, this was deemed irrelevant.

3.2. The BN model

A BN model was constructed using the food fraud records from RASFF and EMA supplemented with data from other data sources from 23 selected drivers (Table 2).

Fig. 2 shows the constructed BN model, which consists of nodes (i.e. drivers), and their interactions, and states (i.e. data) of the nodes. The model represents the interactions (direct and indirect) and their probabilities between all identified factors/drivers related to the detected food fraud in RASFF and EMA. The BN model shows that food fraud type "Tampering" occurs with the highest probability (35.9%), followed by "Health Certificate (HC)" (27.6%), "Illegal importation" (18.3%), "Origin labelling" (11.6%), "Missing common entry documents (CED)" (4%), "Expiration date" (1.7%) and "Theft and resale" (0.8%). Product category "Fish and seafood" (20.6%) has the highest probability to be fraudulent followed by "Meat" (13.4%) and "Fruits and vegetables" (10.4%) (Fig. 2).

3.3. BN model validation

Two ways were followed to verify the validity of the constructed BN model. Firstly, we compared the output of the BN model to results previously published by other researchers. Tähkäpää and co-workers (Tähkäpää et al., 2015) reported similar observations as found with the BN model, albeit that their study pertained to a shorter period (2008–2012). For example, these authors reported that fish and fish products, meat and meat products and nuts, nut products and seeds are among the most frequent fraudulent food categories, and this is also found with the BN model. Further, Tähkäpää and co-workers reported that the EU countries most frequently reporting food fraud in RASFF are in decreasing order: United Kingdom, Italy, France and Germany. The BN model confirms this finding with the following figures related to food fraud reporting countries in RASFF: United Kingdom (13.9%), Italy (13.2%), Germany (8.9%), Spain (5.1%) and Poland (5.1%).

Secondly, we used the BN model to predict the type of food fraud. The BN model, which is based on approximately 83% of the collected data from RASFF and EMA until 2015, was used to predict the food fraud type in the approximately 17% subset (i.e. 293 cases) that had been extracted randomly from the full data set. All variables (except the fraud type) were used as input parameters in the BN model to predict the "fraud type" as mentioned in the databases (RASFF and EMA). The BN model predicted the fraud type correctly in 91.5% of the cases (see Supplement 1). In total 19 cases were invalid in the validation because the combination of parameters of these cases were not in the data set and hence have not occurred before. These cases were excluded in the validation.

3.4. Sensitivity analyses

The contribution of each variable in the model to the probabilities of the type of food fraud can be determined by "sensitivity analysis". Sensitivity analysis aims to measure the effect of a specific variable into the output variable, which in this case is the food fraud type node (Cover & Thomas, 2006).

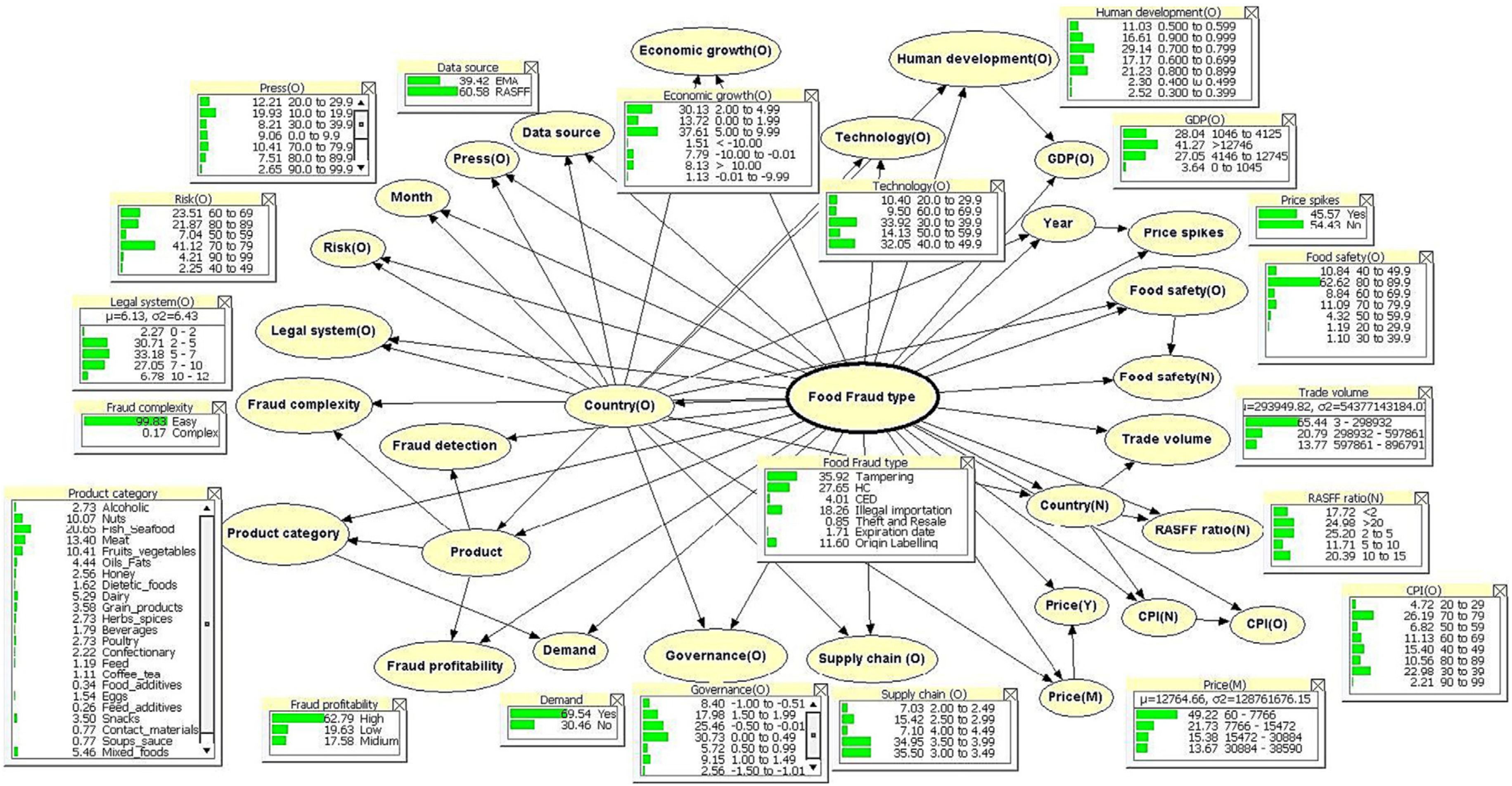

Fig. 2. Bayesian network model of food fraud detection showing the interdependencies and probabilities between the various drivers/factors.

In the present context, it is useful to determine which variables have the largest impact on the uncertainty in the type of food fraud. In the literature, the entropy has been highlighted for sensitivity analysis for BN (Cover & Thomas, 2006; Kjærulff & Madsen, 2013). Entropy is a measure of how much the probability mass is scattered over the states of a variable. It consists of calculating the entropy function $H(X) \in [0, \log(n)]$ of any variable X with n states characterized by a probability distribution P(X), as follows:

$$H(X) = -\sum_{X} P(X) \log P(X) \tag{4}$$

Based on the Entropy calculations (Table 3), the variables origin country (0.61) and product (0.52) are identified as having the greatest influence on the type of food fraud node in the model. Next most influential variables are notifying country (0.33), year (0.24), press index of the origin country (0.18) and CPI of notifying country (0.15). This result shows that we should prioritize data collection efforts on these variables rather than on other variables listed in the BN.

3.5. Application of the BN model

The results summarized above show that the BN model can be used to assess the food fraud dependency on all factors/drivers included in the model as well as the interrelationships between these drivers. For illustration of the potential use of the BN model, we explored some examples for meat.

A clear distinction could be made between the types of fraud reported for meat coming from different countries (Fig. 3), with the latter also showing a variegated pattern of occurrences (intense periods with intermittent calmer periods) over time (Fig. 4). These observations on food fraud can be helpful for governments and the private sector when designing monitoring and control measures to selectively target those products that may be at increased risk of fraud based on the product type, origin, pricing, and recent development in, for example, product demand. Alternatively, economic and political events within a country may affect its efficiency regarding food safety control, as reflected in a change of the food safety index.

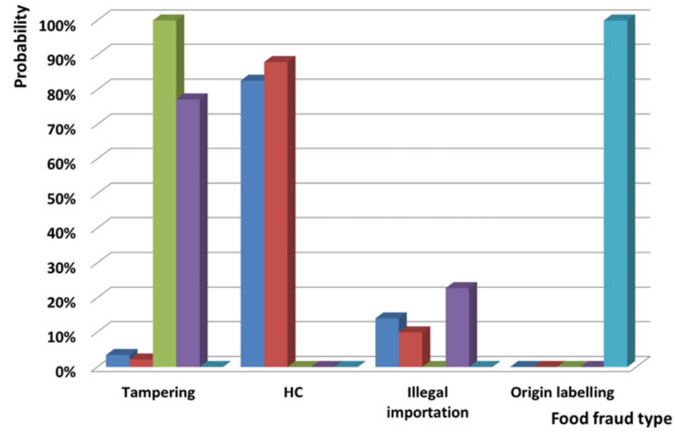

Fig. 3. The effect of country of origin on the type of food fraud for meat in the low price category originating from Brazil (dark blue), Argentina (red), Poland (green), Canada (purple) and Belgium (light blue).

3.6. Limitation of the model

The BN model is based on food fraud records reported in RASFF and EMA and is therefore depending on the information provided in these reports. Obviously, only those cases where violation of the law is found are reported which limits the use of the current BN model. However, if all information would be available than it is expected that a BN model based on these data will be able to predict food fraud (hence not only the type) of any product imported from any country. The RASFF database includes both intentional food fraud notifications (adulteration cases) and unintentional food fraud cases such as improper or missing documents. These unintentional fraud notifications are categorised as adulteration cases in RASFF, which is not the case in other food fraud databases (EMA, USP).

The current BN model predicted 91.5% of the validation cases. It is expected that an even better predictive performance may be obtained when more food fraud cases are available to build the model (Banko & Brill, 2001; Friedman & Yakhini, 1996). BNs are easily adaptable to

Table 3
Sensitivity analysis results.

Variables	Entropy
Country (O)	0.61
Product	0.52
Data source	0.45
Product category	0.41
Country (N)	0.33
Year	0.24
Press (O)	0.19
CPI (N)	0.15
Human development (O)	0.13
CPI (O)	0.12
Governance (O)	0.12
Economic growth (O)	0.11
Technology (O)	0.10
Risk (O)	0.09
Legal system (O)	0.08
Food safety (O)	0.08
Food safety (N)	0.07
Fraud detection	0.07
GDP (O)	0.07
Price (M)	0.06
Supply chain (O)	0.06
Price spikes	0.05
RASFF ratio (N)	0.05
Trade volume	0.05
Month	0.05
Fraud profitability	0.04
Price (Y)	0.01
Demand	0
Fraud complexity	0

N = notifying country; O = country of origin; M = month; Y = year.

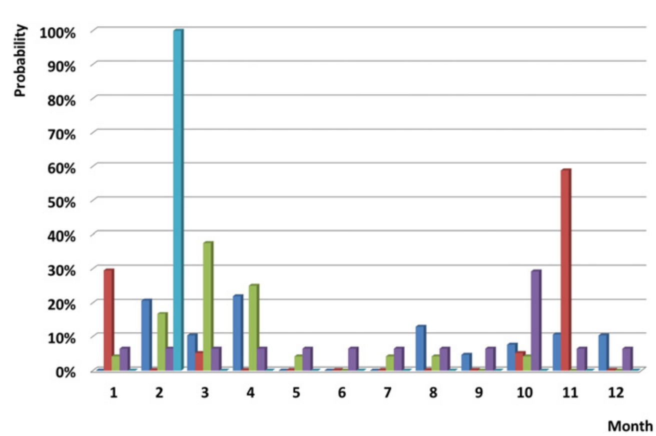

Fig. 4. Probability of food fraud for meat in the low price category depending on the month of the year for various countries of origin: Brazil (dark blue), Argentina (red), Poland (green), Canada (purple) and Belgium (light blue).

new data, and the model can continuously be updated to reflect any new information and further refine the model and enhance its predictive capacity. To anticipate on new developments, we therefore recommend adding new cases as they are reported in RASFF and EMA.

4. Conclusions

In this paper, we presented an application of the holistic approach in food safety to identify known and emerging food safety risks using Bayesian Network modelling and several data sources. This approach relies on the identification of influencing drivers, availability of data and expert judgements.

The BN model that was constructed for food fraud was able to confirm earlier findings published in the literature and could predict the type of food fraud with an accuracy of 91.5%. In more general terms, given the versatility of the BN approach to other fields of food safety and the availability of relevant data linked to the drivers identified, the approach developed within our research holds great potential for application to other kinds of safety issues. For example, the model can be helpful for authorities and the industrials when designing monitoring and control measures to select targeted products that may be at increased risk of fraud based on the origin, pricing, and recent development in product demand.

The BN model can still be developed further into a dynamic model that takes into account temporal variations in all food fraud drivers which could be useful in dynamic food fraud prediction. The inclusion of other data sources such as monitoring data and customs data could also improve accuracy of the future models. Improvement can also be made by incorporating more expert knowledge in the model.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.foodres.2016.08.028.

Acknowledgement

The research leading to this result has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 613688 and from the Dutch Ministry of Economic Affairs (KB-15). The authors would like to thank prof dr. M. Nielen (RIKILT Wageningen UR) for his valuable suggestions.

Appendix A

A.1. Bayesian network calculations

In this example (Fig. A.1), there are three nodes, A_1, A_2 and A_3. All of these nodes have two states (L: low or H: high). From the edges we can see that A_1 and A_2 may influence node A_3. We can also see the unconditional probability tables (UPTs) and the conditional probability table (CPT).

A.2. Unconditional probability tables

The unconditional probabilities and the conditional probabilities shown Fig. A.1 can be obtained through training from data (Li, Yin, Bang, Yang, & Wang, 2014); (Akhtar & Utne, 2014) or expert knowledge elicitation (Martin et al., 2012).

Parents variables (A_1 and A_2), are assigned marginal probabilities, which means the probability of being in a particular state is independent of the other variables in the model. For example, in case of equal weights, the UPs of node A_1 will be $P(A_1 = L) = \frac{1}{2}, P(A_1 = H) = \frac{1}{2}$. However, if the state of A_1 is determined to be H, the UPs will be $P(A_1 = L) = 0, P(A_1 = H) = 1$. In case of the unconditional probabilities are not known, $\frac{1}{n}$ probability assignment can be useful, where n is number of states of the parameter.

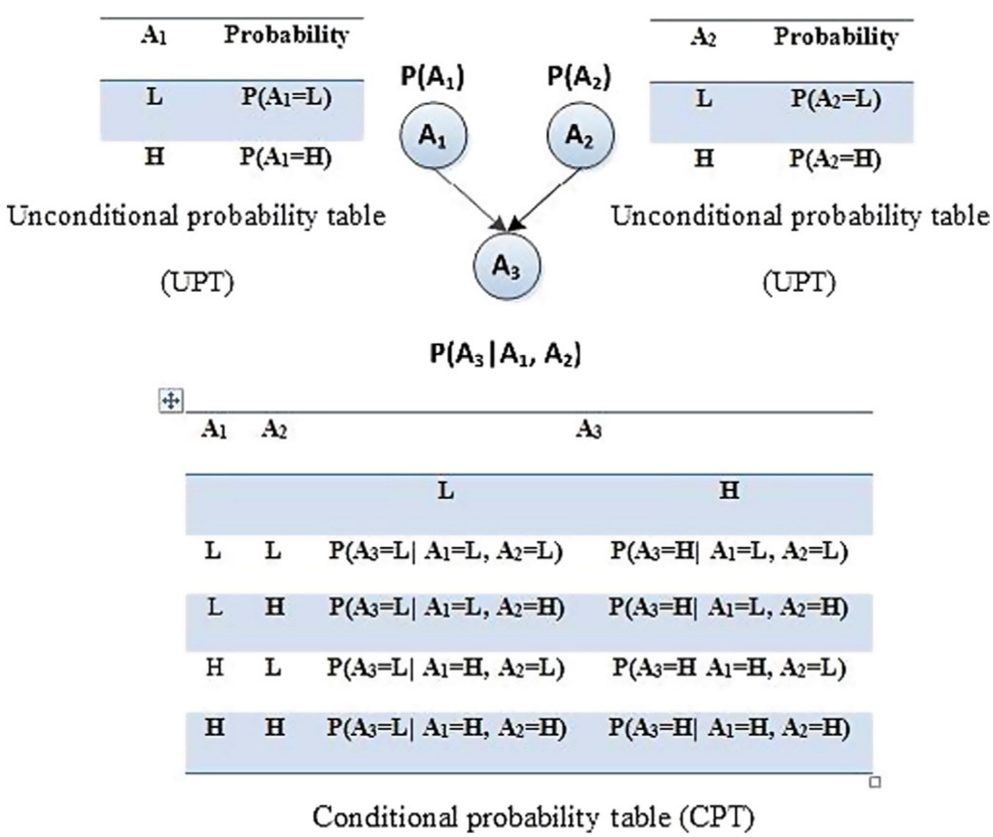

Fig. A.1. An example of Bayesian network structure with three variables. A_1 is a variable that has two states (L, H); A_2 is a variable with two states (L, H); A_3 is a variable with two states (L, H); the arrows indicate the conditional relationships between the variables, where A_1 and A_2 are considered *parent* nodes, and A_3 is the *child* node for this Bayesian network.

A.3. Conditional probability tables

Child variable (A_3), is conditionally dependent on the states of the parent variables (A_1 and A_2). In Fig. A.1, the CPT for the child variable (A_3), is a tables of all possible state combinations of the parent variables. The first probability in the CPT would be described as follows: given that A_1 is "L" and A_2 is "L", the probability that A_3 will be equal to "L" is the conditional probability $P(A_3 = L | A_1 = L, A_2 = L)$. In fact, the number of CPT entries increase exponentially with the number of parent nodes, and the number of states of the parent nodes. For a node with i states and j parent nodes and if each parent node has k states, $i \times j^k$ conditional probability values are required (Achumba, Azzi, Ezebili, & Bersch, 2013; Knochenhauer et al., 2013).

References

Achumba, I. E., Azzi, D., Ezebili, I., & Bersch, S. (2013). Approaches to Bayesian network model construction. In G. -C. Yang, S. -l. Ao, & L. Gelman (Eds.), *IAENG transactions on engineering technologies: Special volume of the world congress on engineering 2012* (pp. 461–474). Dordrecht: Springer Netherlands.

Akhtar, M. J., & Utne, I. B. (2014). Human fatigue's effect on the risk of maritime groundings – A Bayesian network modeling approach. *Safety Science*, *62*, 427–440.

Banko, M., & Brill, E. (2001). Scaling to very very large corpora for natural language disambiguation. *Proceedings of the 39th annual meeting on association for computational linguistics*. Toulouse, France: Association for Computational Linguistics.

Boland, M. J., Rae, A. N., Vereijken, J. M., Meuwissen, M. P. M., Fischer, A. R. H., van Boekel, M. A. J. S., ... Hendriks, W. H. (2013). The future supply of animal-derived protein for human consumption. *Trends in Food Science & Technology*, *29*(1), 62–73.

Bouzembrak, Y., & Marvin, H. J. P. (2015). Prediction of food fraud type using data from Rapid Alert System for Food and Feed (RASFF) and Bayesian network modelling. *Food Control*, *61*, 180–187.

Buriticá, J. A., & Tesfamariam, S. (2015). Consequence-based framework for electric power providers using Bayesian belief network. *International Journal of Electrical Power & Energy Systems*, *64*, 233–241.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory (Wiley series in telecommunications and signal processing)*. Wiley-Interscience.

Denœux, T. (2010). Maximum likelihood from evidential data: An extension of the EM algorithm. *Advances in intelligent and soft computing*, *vol. 77*. (pp. 181–188).

Denœux, T. (2011). Maximum likelihood estimation from fuzzy data using the EM algorithm. *Fuzzy Sets and Systems*, *183*(1), 72–91.

EFSA (2010). Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal*, *8*(6).

EMA (2014). The economically motivated adulteration (EMA) databases. http://ema.foodshield.org/

Everstine, K., Spink, J., & Kennedy, S. (2013). Economically motivated adulteration (EMA) of food: Common characteristics of EMA incidents. *Journal of Food Protection*, *76*(4), 723–735.

FAO (2003). FAO's strategy for a food chain approach to food safety and quality: A framework document for the development of future strategic direction. In M. A. T. Eession (Ed.), *Item 5 of the provisional agenda, committee on agriculture (COAG)*. Rome: Food and Agriculture Organisation (FAO).

FCEC (2013). *Scoping study delivering on EU food safety and nutrition in 2050 - Scenarios of future change and policy responses.* (European Commission).

Friedman, N., & Yakhini, Z. (1996). On the sample complexity of learning Bayesian networks. *Proceedings of the twelfth international conference on uncertainty in artificial intelligence*. Portland, OR: Morgan Kaufmann Publishers Inc.

Bonafede, C. E., & Giudici, P. (2007). Bayesian networks for enterprise risk assessment. *Physica A: Statistical Mechanics and its Applications*, *382*(1), 22–28.

Godfray, H. C., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., ... Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. *Science*, *327*(5967), 812–818.

GO-Science (2011). *The future of food and farming. Final project report.* London, UK: The Government Office for Science.

Guan, N., Fan, Q., Ding, J., Zhao, Y., Lu, J., Ai, Y., ... Yao, Y. (2009). Melamine-contaminated powdered formula and urolithiasis in young children. *New England Journal of Medicine*, *360*(11), 1067–1074.

Huo, L., Zhu, Y., Zhang, Z., & Chen, L. (2004). Bayesian networks application to reliability evaluation of electric distribution systems. *Diangong Jishu Xuebao/Transactions of China Electrotechnical Society*, *19*(8), 113–118.

Jia, C., & Jukes, D. (2013). The national food safety control system of China – A systematic review. *Food Control*, *32*(1), 236–245.

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, *32*(4), 995–1003.

Kjærulff, U. B., & Madsen, A. L. (2013). Sensitivity analysis. *Bayesian networks and influence diagrams: A guide to construction and analysis* (pp. 303–325). New York, NY: Springer New York.

Knochenhauer, M., Swaling, V. H., Dedda, F. D., Hansson, F., Sjoekvist, S., & Sunnegaerd, K. (2013). *Using Bayesian Belief Network (BBN) modelling for rapid source term prediction final report.* Denmark: Nordic Nuclear Safety Research (NKS), 81.

Lee, C. -J., & Lee, K. J. (2006). Application of Bayesian network to the probabilistic risk assessment of nuclear waste disposal. *Reliability Engineering & System Safety*, *91*(5), 515–532.

Li, K. X., Yin, J., Bang, H. S., Yang, Z., & Wang, J. (2014). Bayesian network with quantitative input for maritime risk analysis. *Transportmetrica A: Transport Science*, *10*(2), 89–118.

Malka, R., & Lerner, B. (2004). Classification of fluorescence in situ hybridization images using belief networks. *Pattern Recognition Letters*, *25*(16), 1777–1785.

Martin, T. G., Burgman, M. A., Fidler, F., Kuhnert, P. M., Low-Choy, S., McBride, M., & Mengersen, K. (2012). Eliciting Expert Knowledge in Conservation Science Obtención de Conocimiento de Expertos en Ciencia de la Conservación. *Conservation Biology*, *26*(1), 29–38.

Marvin, H. J. P., Kleter, G. A., Frewer, L. J., Cope, S., Wentholt, M. T. A., & Rowe, G. (2009). A working procedure for identifying emerging food safety issues at an early stage: Implications for European and international risk management practices. *Food Control*, *20*(4), 345–356.

Miraglia, M., De Santis, B., Minardi, V., Debegnach, F., & Brera, C. (2005). The role of sampling in mycotoxin contamination: An holistic view. *Food Additives & Contaminants*, *22*(Suppl. 1), 31–36.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, *50*(3), 559–569.

Nielsen, T. D., & Verner Jensen, F. (2007). *Bayesian networks and decision graphs.* Springer New York.

NSF International (2014). The new phenomenon of criminal fraud in the food supply chain. *NSF international report* (pp. 1–17).

RASFF (2015). Rapid alert system for food and feed (RASFF). *Directorate general for health and consumer protection.* Brussels: European Commission.

Spink, J., & Moyer, D. C. (2011). Defining the public health threat of food fraud. *Journal of Food Science*, *76*(9), R157–R163.

Tähkäpää, S., Maijala, R., Korkeala, H., & Nevas, M. (2015). Patterns of food frauds and adulterations reported in the EU rapid alert system for food and feed and in Finland. *Food Control*, *47*, 175–184.

Unnevehr, L., Eales, J., Jensen, H., Lusk, J., McCluskey, J., & Kinsey, J. (2010). Food and consumer economics. *American Journal of Agricultural Economics*, *92*(2), 506–521.

Wiegerinck, W. A. J. J., Kappen, H. J., ter Braak, E. W. M. T., ter Burg, W. J. P. P., Nijman, M. J. O. Y. L., & Neijt, J. P. (1999). Approximate inference for medical diagnosis. *Pattern Recognition Letters*, *20*(11−13), 1231–1239.

Wilson, C. (2010). CHAPTER 4 - Brainstorming. *User experience re-mastered* (pp. 107–134). Boston: Morgan Kaufmann.