5-13-2016

Bayesian Networks with Expert Elicitation as Applicable to Student Retention in Institutional Research

Jessamine Corey Dunn
Georgia State University

ACCEPTANCE

This dissertation, BAYESIAN NETWORKS WITH EXPERT ELICITATION AS APPLICABLE TO STUDENT RETENTION IN INSTITUTIONAL RESEARCH, by JESSAMINE COREY DUNN, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree, Doctor of Philosophy, in the College of Education and Human Development, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chairperson, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty.

____
William Curlette, Ph.D.
Committee Chair

____
Janice Fournillier, Ph.D.
Committee Member

____
Chris Oshima, Ph.D.
Committee Member

____
Kerry Pannell, Ph.D.
Committee Member

____
Date

____
William Curlette, Ph.D.
Chairperson, Department of Educational Policy Studies

____
Paul A. Alberto, Ph.D.
Dean
College of Education and Human Development

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education and Human Development's Director of Graduate Studies, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

____

Jessamine Corey Dunn

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Jessamine Corey Dunn
Educational Policy Studies
College of Education and Human Development
Georgia State University

The director of this dissertation is:

William Curlette, Ph.D.
Department of Educational Policy Studies
College of Education and Human Development
Georgia State University
Atlanta, GA 30303

CURRICULUM VITAE

Jessamine Corey Dunn

ADDRESS: 423 Clairemont Ave NE #12
Decatur, GA 30030

EDUCATION:

Ph.D.	2016	Georgia State University Department of Educational Policy Studies
Masters Degree	2007	Georgia State University Department of Educational Policy Studies – Educational Research
Bachelors Degree	1999	Mary Baldwin College - Economics

PROFESSIONAL EXPERIENCE:

2014-Present	Director of Institutional Research Agnes Scott College
2011-2014	Senior Decision Support Analyst Georgia Institute of Technology
2007-2011	Institutional Research Analyst III Georgia Institute of Technology
2006-2007	Analyst, Strategy Management Arthritis Foundation, Inc.
2000-2006	Finance & Operations Coordinator Arthritis Foundation, Inc.
1999-2000	Financial Services Accounting Analyst SNL Financial, Inc.

PRESENTATIONS AND PUBLICATIONS:

Dunn, J.C. (2014, October). *Using a Bayes Net to Model Accepted Student Enrollment Yield and Target Recruitment.* 2014 Annual Forum of the Southern Association for Institutional Research, Destin, FL.

Dunn, J.C. (2014, February). *Traditional Co-Op: A Sustainable Future. Using Outcomes to Make the Case.* 2014 Conference for Industry Education and Collaboration, Savannah, GA.

Dunn, J.C. (2008, September). *Mid-Term Progress Reports – Do They Really Work?* 2008 Annual Forum of the Southern Association for Institutional Research, Nashville, TN.

PROFESSIONAL SOCIETIES AND ORGANIZATIONS

2007-2016	Association for Institutional Research
2007-2016	Southern Association for Institutional Research

, BAYESIAN NETWORKS WITH EXPERT ELICITATION AS APPLICABLE TO STUDENT RETENTION IN INSTITUTIONAL RESEARCH

by

JESSAMINE COREY DUNN

Under the Direction of William Curlette, Ph.D.

ABSTRACT

The application of Bayesian networks within the field of institutional research is explored through the development of a Bayesian network used to predict first- to second-year retention of undergraduates. A hybrid approach to model development is employed, in which formal elicitation of subject-matter expertise is combined with machine learning in designing model structure and specification of model parameters. Subject-matter experts include two academic advisors at a small, private liberal arts college in the southeast, and the data used in machine learning include six years of historical student-related information (i.e., demographic, admissions, academic, and financial) on 1,438 first-year students. Netica 5.12, a software package designed for constructing Bayesian networks, is used for building and validating the model. Evaluation of the resulting model’s predictive capabilities is examined, as well as analyses of sensitivity, internal validity, and model complexity. Additionally, the utility of using Bayesian networks within institutional research and higher education is discussed.

The importance of comprehensive evaluation is highlighted, due to the study’s inclusion of an unbalanced data set. Best practices and experiences with expert elicitation are also noted, including recommendations for use of formal elicitation frameworks and careful consideration of operating definitions. Academic preparation and financial need risk profile are identified as key variables related to retention, and the need for enhanced data collection surrounding such variables is also revealed. For example, the experts emphasize study skills as an important predictor of retention while noting the absence of collection of quantitative data related to measuring students’ study skills. Finally, the importance and value of the model development process is stressed, as stakeholders are required to articulate, define, discuss, and evaluate model components, assumptions, and results.

INDEX WORDS: Bayes Theorem, Bayesian Networks, Expert Elicitation, Institutional Research, Retention

BAYESIAN NETWORKS WITH EXPERT ELICITATION AS APPLICABLE TO STUDENT RETENTION IN INSTITUTIONAL RESEARCH

by

JESSAMINE COREY DUNN

A Dissertation

Presented in Partial Fulfillment of Requirements for the

Degree of

Doctor of Philosophy

in

Educational Policy Studies – Research, Measurement, and Statistics

in

Department of Educational Policy Studies

in

the College of Education and Human Development
Georgia State University

Atlanta, GA
2016

DEDICATION

This dissertation is dedicated to my beautiful and smart daughter, Rowan Hughes Dunn.

.

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Dr. William Curlette, for his guidance, patience, and encouragement during this process. I would also like to thank Drs. Oshima and Fournillier for their willingness to serve on my committee and for providing excellent instruction during my time at Georgia State. I would also like to thank Dr. Pannell for serving on my committee and representing Agnes Scott College. I would also express special thanks to my professional colleagues at Georgia Tech and Agnes Scott for their encouragement, flexibility, and advice during this process. Finally, I am forever grateful to my family and friends for their constant encouragement and love throughout this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 A BAYESIAN APPROACH, EXPERT ELICITATION, AND BAYESIAN NETWORKS AS APPLICABLE TO INSTITUTIONAL RESEARCH: A REVIEW OF THE LITERATURE

While applications of a Bayesian approach to statistics are commonly practiced in a number of fields, examples of studies addressing and incorporating Bayesian statistics in educational research are less common. Narrowing the field of interest to institutional research, defined by Saupe (1990) as "research conducted within an institution of higher education to provide information that supports institutional planning, policy formation and decision making" (p.1), a Bayesian approach to research offers a tool box rich in resources for handling and modeling the uncertainty, complexity, and uniqueness of institutional data while also providing a formal mechanism for incorporating institutional memory, expertise, and prior data into analysis. With an eye towards the completion of a Bayesian research study within the field of institutional research, this manuscript provides a comprehensive review of the literature surrounding a Bayesian approach to institutional research. Beginning with a background of general Bayesian statistics, the review also discusses the elicitation of subjective probabilities and development and use of Bayesian networks.

Guiding Questions

The guiding questions shaping this review are as follows: How is a Bayesian approach relevant to institutional research? How can an institutional researcher leverage and incorporate expert information and experience into data analysis and modeling? How can Bayesian networks be used in institutional research, particularly those that predict an outcome of interest?

Introduction to Bayesian Statistics

A Bayesian approach to statistics is one in which statisticians attempt to describe a true state or event in probabilistic terms. Contrary to the classical or frequentist approach in which probability is defined as the proportion of successful outcomes to number of attempts, Bayesian statistics views probability as degree of belief. In other words, Bayesian probability is a measure of the degree of belief in the probability of specific outcome. This degree of belief represents prior knowledge pertaining to the likelihood of an event, which is then updated with data relevant to this event in order to form a new, or posterior, belief in the probability of the same event occurring. As Gill (2009) wrote, “Bayesians generally interpret probability as ‘degree of belief,’ meaning that prior distributions are descriptions of relative likelihoods of events based on the researcher’s past experience, personal intuition, or expert opinion, and posterior distributions are those prior distributions updated by conditioning on new observed data” (p.135).

The following sections address the core tenets of Bayesian inferential methods, including an explanation of Bayes’ theorem and its role in the function of combining observed data with prior knowledge, a discussion of the prior distribution and how it is formed, and consideration of model fit. Discussion of arguments surrounding the subjectivity of the Bayesian approach is also included, as well as a summary of the advantages and limitations of Bayesian methods, particularly within the context of social sciences, and educational and institutional research.

Bayes’ Theorem

Fundamentally, Bayesian methods provide a way to revise probabilities by incorporating new data. Equation 1.1 demonstrates Bayes’ theorem, in which the probability of event B given event A (the new data) is modeled as a function of the probability of event A given event B multiplied by the probability of event B alone and divided by the probability of event A.

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \tag{1.1}$$

Medical testing is commonly used as an illustrative tool when describing Bayes' theorem. Consider an example provided by Gigerenzer (2002) concerning the efficacy of mammogram testing for breast cancer. Within Gigerenzer's example, it is presented that .8% of women in the general population have breast cancer. The probability that a woman with breast cancer receives a positive mammogram is 90%, while the probability that a woman *without* breast cancer receives a positive mammogram is 7%. An accurate calculation of the probability that a woman with a positive mammogram actually has cancer necessitates the inclusion of all of the information presented: The incidence of breast cancer in the study's population (.8%), the probability of a correct mammogram test (90%), and the probability of an incorrect test (7%). For simplicity, first consider the given information in terms of frequencies (rather than probabilities and percentages) as presented in Table 1.1:

Table 1.1
Mammogram/Breast Cancer Frequency Table

		Cancer?		Total
		Yes	No	
Mammogram Result	Positive	7	70	77
	Negative	1	922	923
Total		8	992	1000

As Table 1.1 shows, .8% (8/1000) of women in the population actually have cancer. Of those eight women, 88% (7/8) will receive a positive mammogram. Seven percent of the remaining women in the population (70/992) will also receive a positive mammogram even though they don't have cancer. Therefore, a total of 77 women (seven of the women who actually do have cancer and 70 of the women who do not have cancer) will receive a positive mammogram. Importantly, only seven of these 77 women actually have cancer, meaning that the probability of actually having breast cancer after receiving a positive mammogram is only 9% (7/77).

This information is easily transferred into the variables presented in Bayes' theorem. Consider $p(B \mid A)$ to represent the probability of breast cancer given a positive mammogram. Bayes' theorem requires that in order to determine the probability of breast cancer given a positive mammogram, the overall probability of breast cancer within the population, $p(B)$ or .8%, be multiplied against the probability of a positive mammogram given the presence of breast cancer, $p(A \mid B)$ or 90%, and divided by the overall probability of a positive mammogram, $p(A)$ or 8%. Substitution of these values into Bayes' theorem results in a 9% chance that a woman receiving a positive mammogram actually has breast cancer. Essentially, Bayes' incorporation of the known prevalence of cancer within the given population along with test accuracy and sensitivity act to produce a probability of cancer given a positive mammogram. In order words, a Bayesian approach investigates how the probability of one event is affected by the probability of another – conditional probability.

This incorporation of conditional probability is a critical factor in Bayesian methods. Using the same mammogram example for illustration, there are two conditional probabilities – the probability of breast cancer given a positive mammogram, $p(B \mid A)$, and the probability of a positive mammogram given the presence of breast cancer, $p(A \mid B)$. The known prevalence of breast cancer in the general population, p (B), is termed a "prior" probability. The conditional probability of a positive mammogram given the presence of cancer, $p(A \mid B)$, is termed "likelihood" and introduces the incorporation of new information (a positive mammogram) into consideration. In other words, the likelihood portion of Bayes' theorem estimates the effect of a positive mammogram on a prior belief that a person has breast cancer. However, the mammogram's test sensitivity and accuracy are also considered in the equation, as reflected in the divisor. The overall probability of receiving a positive mammogram, *p* (A), is a function of the test's accuracy (in terms of the probability of false positive) and the test's sensitivity (in terms of probability of cancer detection given the presence of cancer), both of which are tempered by the known prevalence (prior) of breast cancer. The product of the prior and likelihood is considered the "posterior probability." Using Gigerenzer's (2002) mammogram example, Table 1.2 illustrates the components used in Bayes' theorem to estimate the posterior probability of breast cancer given a positive mammogram.

Table 1.2
Components of Bayes Theorem, Using Gigerenzer's (2002) Mammogram Example

Posterior Probability	Prior Probability		Likelihood		
$p(B \mid A)$ Probability of breast cancer given positive mammogram = $\frac{p(A\mid B)p(B)}{p(A)}$ = $\frac{90\% * .8\%}{8\%}$	p (B) Breast cancer prevalence in population	p ($B_{\text{no cancer}}$) Probability of not having breast cancer in general population	$p(A \mid B)$ Probability of positive mammogram given breast cancer	$p(A \mid B_{\text{no cancer}})$ Probability of positive mammogram given no breast cancer (false positive)	p(A) Probability of receiving positive mammogram: = (90% * .8%) + (99.2% * 7%)
9%	.8%	99.2%	90%	7%	8%

Without the information provided by the mammogram, and holding all other risk factors constant, the only way to estimate the probability of breast cancer is to simply quote the prevalence within the general population. Bayes' theorem allows for the introduction of the new mammogram information (including allowances for the mammogram's sensitivity and accuracy, in terms of false negatives and false positives) to adjust the prior belief and form a more accurate understanding of the probability of breast cancer given a positive mammogram. Understanding of the theorem and its components also allows for a more accurate assessment of what a positive mammogram really implies – without considering the actual prevalence of breast cancer in the general population, as well as the test's sensitivity and accuracy, one would mistakenly interpret a positive mammogram as a 90% chance that the patient has cancer. Incorporating all the information available reveals only a 9% chance that a patient with a positive mammogram has breast cancer. The incorporation of new information is a major factor of what sets the Bayesian approach apart from frequentist techniques.

Bayesian Priors

Another major difference between Bayesian and frequentist approaches involves the use of Bayesian priors. Put most simply, a Bayesian prior is a quantification of the researcher's *a priori* beliefs. In the previous example, the incidence of breast cancer in the population served as the prior probability. Although there is much variation in the literature regarding nomenclature, Bayesian priors can be broadly categorized as uninformative or informative. Within these categories are a number of subcategories, often depending on the weight assigned and source of the prior knowledge. The following sections provide a general discussion of these two broad categories of priors. More in depth discussion of the philosophical interpretations of probabilities that form the foundation of these priors follows.

Uninformative priors. Uninformative priors (also termed "objective," "noninformative," "flat," "vague," "diffuse," and "reference," among others, in the literature) provide little additional information or explanatory power, and are often employed to reflect objectivity (Gill, 2009). Gelman, Carlin, Stern, and Rubin (2004) suggested that the use of uninformative prior distributions is a way "to let the data speak for themselves" (p. 62), thereby limiting, or even eliminating any influence of prior information on current data and posterior distributions. Uninformative priors are also used in the case when more subjective prior distributions are unavailable or when resources involved with gathering prior information is deemed prohibitive (Ghosh, 2011).

The uninformative prior employed by the earliest Bayesians is the uniform prior, in which all possible outcomes are equally likely (Bayes, 1763; Laplace, 1825/1902; Syversveen, 1998). In response to the uniform distribution's problems with lack of invariance (variation in posterior distributions resulting from non-linear transformations of the same uniform distribution), Jeffreys (1961) proposed a prior that is invariant under reparameterization and incorporates Fisher's information (Data & Ghosh, 1996). Box and Tiao (1973), Lindley (1965), Press (1972), and Zellner (1971) expanded on Jeffrey's work, demonstrated Bayesian methods' ability to more efficiently address statistical problems, and set the stage for an extensive amount of literature exploring uninformative or objective priors (Bernardo, 2005).

The concept of a "reference prior," or a prior whose influence is subjugated to information provided by the data, is an important form of uninformative prior that emerged from these discussions (Berger & Bernardo, 1992). Importantly, reference priors are understood to represent formal, consensus-driven functions developed among a scientific community, ensuring "that the information provided by the data will not be overshadowed by the scientist's prior beliefs" (Bernardo, 2005, p. 3). This understanding of uninformative priors as providing minimal impact is also an important distinction from earlier ideas that uninformative priors are attempts to represent or express ignorance (Kass & Wasserman, 1996). One special case of uninformative priors, and one that results in posterior probabilities requiring careful interpretation, is an “improper” prior, or one which the sum of all possible values specified by the prior distribution does not result in a finite value (Gelman at al., 2004).

In the past fifty years, discussion of uninformative priors has dramatically expanded beyond uniform and Jeffrey’s priors. Yang and Berger (1997) provided a catalog of uninformative priors, while Kass and Wasserman (1996) offered a comprehensive guide to the selection of the many types of uninformative priors. The large amount of literature focused on the methods behind and selection of uninformative priors points toward a search for a default or generally agreed-upon uninformative prior that will address concerns of objectivity.

Informative priors. Informative priors intentionally include knowledge designed to influence posterior probabilities and, ultimately, statistical inference. Press (2003) outlined the advantages of informative priors as follows: Such priors are proper, act as supplementary data, capitalize on expert knowledge, and present an avenue for analysis when other information (“objective” Bayesian priors, or even a frequentist approach) is unavailable.

Informative priors can be derived from a number of sources, often including previous studies and results, researcher expertise, subject-matter expertise, and mathematical convenience (Gill, 2009; Gelman et al., 2004). For example, Ibrahim and Chen’s (2000) “Power Prior” is a form of informative prior built from historical data, in which the influence of the historical data is weighted based on the researcher’s belief in how closely the historical data can be tied to current data and inferences. This type of informative prior is most popular in clinical settings, as there are often large amounts of historical data available. A conjugate prior is an informative prior chosen due to its conjugacy (same distribution families) with the likelihood function, the use of which simplifies the calculations necessary to compute a posterior (Raiffa & Schlaifer, 1961). Note that the development of Markov chain Monte Carlo (MCMC) techniques has eased the computational burden on statisticians when dealing with non-conjugate priors (Hahn, 2006). In following sections, this manuscript will address the formal elicitation of subject-matter expertise in order to develop informative priors.

Other priors. The literature include a number of other prior forms that do not fall neatly into the "informative" or "uninformative" categories, most of which were developed in the interest of increasing flexibility and applicability. For example, hybrid priors combine informative and uninformative priors for use in hierarchical Bayesian models (Gill, 2009). Jaynes (1980) developed maximum entropy priors in an effort to increase flexibility in describing comparative levels of uncertainty. Nonparametric-priors were developed to respond to problems of determining appropriate model complexity (Gershman & Blei, 2012).

Prior Evaluation

Although the sources of priors often depend on a researcher's judgment, they should not be chosen cavalierly. Gill (2009) encouraged explicit explanations of prior choice and specification, as well as analyses of a resulting model's sensitivity to changes in the chosen prior. Reimherr, Meng, and Nicolae (2014) further emphasized the importance of evaluating the impact of an informative prior on the posterior distribution. In other words, it is important to measure how much prior knowledge influences inferences and conclusions. Additionally, the literature has suggested model checking (comparing observed data with model-generated data) regardless of informative or uninformative prior (Evans & Moshonov, 2006; Gelman, Meng, & Stern, 1996; Kelly & Smith, 2011; Rubin, 1984). For example, as Rubin pointed out, the sensitivity of conclusions to how a Bayesian model is set up exposes scientific uncertainty – if inferences change based on model specifications, researchers can conclude that more information and study is necessary to address the uncertainty revealed. In another form of sensitivity analysis, Berger

(2006) recommended a comparison of conclusions drawn from a subjective Bayesian analysis against those of an objective prior analysis, noting that large differences due to choice of priors should be discussed and further investigated and justified. Speigelhalter and Rice (2009), succinctly summarized the evaluative process: "In particular, audiences should ideally fully understand the contribution of the prior distribution to the conclusions, the reasonableness of the prior assumptions, the robustness to alternative models and priors, and the adequacy of the computational methods" (p. 5230).

The incorporation of prior knowledge is a critical advantage of Bayesian methodology. As Gill (2009) wrote, "priors are a means of systematically incorporating existing human knowledge, quantitative or qualitative, into the statistical specification" (p. 138). Ultimately, the selection of prior is based on a number of factors, including research question, availability of data, and the researcher's experience. More broadly, the decision of a researcher to employ Bayesian methodology has much to do with her interpretation of probability. A frequentist interpretation of probability concludes that the results of long-run, controlled, and repeated experiments can eventually be interpreted as representative of the short term as well. However, these types of experiments can be cost prohibitive and time consuming, and are typically unrealistic within behavioral and social sciences. The Bayesian approach to probability offers an alternative in which probability represents a degree of belief, and prior probabilities reflect this degree of belief *a priori* to any new evidence. As discussed in depth below, this degree of belief is also termed "subjective probability."

Subjective Probability

At the root of any discussion regarding the use of Bayesian inferential techniques lies the idea of subjective probability. In their discussions of the philosophical foundations of probability, both Weatherford's (1982) and Gillies' (2000) definitions of subjective probability are remarkably similar: Weatherford defined subjective probability as "the degree of belief of a given person at a given time" (p. 220), and Gillies stated that subjective probability theory "identifies probability with the degree of belief of a particular individual" (p. 1). These definitions are mostly based on the work of subjective theorist Bruno de Finetti, who provided the philosophical and mathematical groundwork for subjective probability with his definition of probability as "a measure of a degree of belief attributed to the various possible alternatives" (de Finetti, 1972, p. 147-148). Further, de Finetti proposed the modification of degree of belief should be the result of observation of prior events – in short, learning from experience (Cifarelli & Regazzini, 1996). Subjective probability was further explored through the work of Kyburg and Smokler (1964), Luce and Suppes (1965), Ramsey (1931), Savage (1954), and Savage, Hacking, and Shimony (1967).

It is important to note that de Finetti's and others' concept of subjective probability was not without limits. de Finetti likened the limitation of degree of belief to a gambling situation – "…the degree of probability attributed by an individual to a given event is revealed by the conditions under which he would be disposed to bet on that event" (de Finetti, 1937/1964, p. 101). Additionally, de Finetti included discussion of the necessary conditions under which degrees of belief could serve as probabilities: Degrees of belief are measurable and coherent, or rational, ensuring a grounding in reality (Weatherford, 1982). Other philosophers have expounded on the circumstances necessary for, and influencing, degrees of belief. Like de Finetti, Ramsey (1931) employed gambling allusions (avoidance of falling victim to a Dutch Book, in which a better irrationally agrees to a bet in which he is guaranteed to lose) to illustrate how adherence to the axioms of probability is critical in determining and acting on degree of belief. Bonjour (1985), Dawid (1982), and Lewis (1946) put forth that, in the instance of two or more pieces of information being used to form a belief, confidence in that belief increases according the congruence, or coherence, of the pieces of information. In their discussion of the epistemology underlying the use of probabilities in Bayesian methods, Bovens and Hartmann (2003) proposed three conditions thought to influence degree of belief: The degree to which the information forming the belief was expected, the reliability of the information, and the coherence of the information. In summary, while emphasizing the personalistic and subjective properties of probability, subjective theorists recognize the necessity of coherency and consistency in the formation of belief.

Within the realm of Bayesian statistical methods of inference, subjective probability is important when considering the initial probability assigned to a hypothesis (the Bayesian prior). As discussed in earlier sections of this manuscript, a researcher using Bayesian inference first establishes a belief surrounding the probability of an event occurring, and then uses available data to update that prior probability and form a posterior probability. Recall that there are generally two types of priors – uninformative and informative – and that uninformative priors are typically considered "objective." It can be argued that an informative prior can be considered fundamentally subjective, as it represents a degree of belief given current and situational knowledge, experience, reasoning, and logic. The subjective properties of the prior belief sound very similar to the ideas behind constructivism and phenomenology, such as the understanding that reality is constructed and meaning is made through individuals' first-person experiences and interactions with others (Bogdan & Biklen, 2007). Curlette (2006) suggested that this places Bayesian methods incorporating a subjective prior belief within a phenomenological framework.

This use of subjective priors introduces a controversial aspect of Bayesian methodology – the conflict over subjectivity.

Objectivity vs. Subjectivity

Arguments concerning the roles of objectivity and subjectivity within scientific research are not uncommon – for example, consider the qualitative versus quantitative "Paradigm Wars" in educational research (Eisner & Peshkin, 1990; Gage, 1989; Guba, 1990). A criticism of Bayesians' acceptance of subjective probability is that it defies the fundamental objectivity embraced by the mainstream understanding of the scientific method. In his article, *The Case for Objective Bayesian Analysis* (2006), Berger acknowledged the fact that statistical methods are understood to be a tool for producing unbiased, objective validation of scientific results, and proposes that wider acceptance and valuation of Bayesian methods is predicated on the appearance of objectivity, particularly within a regulatory climate. The vast amount of literature and studies addressing the choices of uninformative priors (see "Background – Uninformative Priors") speaks to this drive to legitimate Bayesian methodology through a focus on "objective" priors. There are even attempts to propose more objective informative priors. For example, Berger and Sun (2008) recommended a set of informative priors that, based on parameters of interest, can be used as default priors – prescriptive/standardized priors given the research question. Some would suggest that these prescriptive priors remove any researcher bias or subjectivity in the actual choice of prior (Lenk & Orme, 2009).

However, it can also be argued that, regardless of approach, there is no such thing as pure objectivity – researchers' choices regarding research questions, methodologies, and analysis techniques could all be considered subjective choices influenced by experience, habit, etc. (Berger, 2006; Gill, 2009; Hennig, 2009; Press & Tanur, 2001; Stevens & O'Hagan, 2002;

Weatherford, 1982). D'Agostini (2001) elaborated on this idea by referencing Bayesians' requirement for coherence in subjective probabilities:

> Once coherence is included in the subjective Bayesian theory, it becomes evident that 'subjective' cannot be confused with 'arbitrary', since all ingredients for assessing probability must be taken into account, including the knowledge that somebody else might assess different odds for the same events. Indeed, the coherent subjectivist is far more responsible (and more 'objective', in the sense that ordinary parlance gives to this word) than those who blindly use standard 'objective' methods. (p. 25)

In other words, the process of appropriately incorporating subjective Bayesian priors – ensuring coherence and adherence to the laws of probability, as well as accounting for the conditions affecting degree of belief – introduces even greater levels transparency and thoughtfulness than typical frequentist methods.

In summary, an important aspect of Bayesian inference is the idea that the prior belief in the probability of an event occurring is often subjective, as it reflects degree of belief based on specific, often individual, circumstances. This subjectivity is not only accepted, but valued in eliciting expert opinions to form prior probabilities. A Bayesian approach such as this serves two purposes: It accepts that the ideal of purely objective scientific research is unrealistic, and it does not waste any available, and potentially enlightening, information. Following a summarizing discussion contrasting and comparing Bayesian versus frequentist approaches, the process of formally eliciting informative, subjective prior probabilities for use in Bayesian inference will be described and reviewed.

Comparing Bayesian and Frequentist Approaches

In his address to a group of statisticians, particle physicists, astrophysicists, and cosmologists, Bradley Efron (2003), former President of the American Statistical Association, addressed the conflict between Bayesian and frequentist factions as follows:

> The Bayesian-frequentist argument is certainly a long-lived one, even by the standards of philosophy. It reflects, I believe, two quite different attitudes toward the scientific process: the cautious frequentist desire for objectivity and consensus, versus the individual scientist trying aggressively to make the best sense of past data and the best choice for future direction. (p. 1)

While Efron's summation was profound and cogent, there are a number of differences between the two methodologies influencing their different approaches to the scientific process. Spiegelhalter and Rice (2009) wrote that the main difference between frequentist and Bayesian approaches to inference is that "Bayesians make statements about the relative evidence for parameter values given a dataset, while frequentists compare the relative chance of datasets given a parameter value" (p. 5230). More concisely, Bayesians consider the probability of a hypothesis given data, while frequentists consider the data, given a hypothesis. In other words, Bayesians consider all information or evidence available to draw conclusions about a certain parameter, while frequentists evaluate how well certain data sets conform to a hypothesized parameter. Encompassed within this overall distinction are a number of other differences described below.

Interpretations of Probability

As discussed earlier, Bayesian methods employ an understanding of probability as a degree of belief. In the frequentist approach, probability represents the likelihood that an event will occur in a large number of repeated trials – as Gill (2009) wrote, "…frequentists see probability measure as a property of the outside world and Bayesians view probability as a personal internalization of observed uncertainty" (p. 27). This differing interpretation is particularly notable due to the use of prior probabilities in Bayesian approaches – the inclusion of prior probability distributions to represent a state of knowledge prior to the introduction of new data is in direct contrast to the frequentist idea that there is some fixed, unchanging probability of events that can be calculated through frequency counts of long-run experiments.

Hypothesis Testing/Inference

Frequentist hypothesis testing centers around the work of Neyman and Pearson (1933), in which researchers choose between a null and alternative hypothesis based on the calculation and acceptance or rejection of false-positive or false-error rates. Neyman and Pearson posited that a greater amount of objectivity is achieved by limiting error through replication and deductive reasoning, and alluded to a trade-off between objectivity and drawing conclusions from a single experiment when they wrote:

> …no test based upon a theory of probability can by itself provide any valuable evidence of the truth or falsehood of a hypothesis. But we may look at the purpose of tests from another viewpoint. Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behavior with regard to them, in following which we insure that, in the long run of experience, we shall not often be wrong. (p. 291)

Offered as a measure of evidence against the null, and not originally intended to be used in inference, Fisher's *p* value is often incorporated into classical hypothesis testing (Fisher 1925; Fisher 1935; Fisher 1956). There is a large body of literature addressing the problems with the widespread misinterpretation of *p* values in significance testing, in which the authors point out that incorrect use of *p* values often involves the conflation of *p* values with Type 1 error rates, or conceptual errors in which *p* values serve as probability statements describing the likelihood of a hypothesis (Carver, 1978; Cohen, 1994; Dixon, 2003; Gelman & Loken, 2014; Gigerenzer, 1993; Hubbard & Lindsey, 2008; Johansson, 2011; Royall, 2000; Wagenmakers, 2007, etc.). More relevant to this discussion, however, is why *p* values and hypothesis testing have been combined into this widespread hybrid method. Both hypothesis testing and *p* values were proposed in response to a culture valuing increased objectivity and rigorous quantitative methods (Marks, 1997; Matthews, 1995; Porter, 1995). However, with Neyman-Pearson hypothesis testing's focus on controlling error rates over the long-run and the limitation of Fisher's *p* value to only indicating evidence *against* a null hypotheses, Goodman (1999a) suggested that the coupling of the two approaches is the result of researchers' understandable desire to be able to draw conclusions from a single experiment using "objective" methodology. In other words, the combination of hypothesis testing and *p* values presents researchers with a seemingly viable, although often conceptually incorrect, platform for "evidenced-based" research.

Contrastingly, Bayesian methods *do* offer a formal avenue towards quantifying statistical evidence for *or* against a hypothesis. Unlike null hypothesis significance testing, a Bayesian approach can be used to calculate the probability that, given data or evidence, a hypothesis is true or untrue. This is done using the "Bayes factor," the likelihood ratio included in Bayes' theorem (Jeffreys, 1961; Robert, 2007). The Bayes factor is a ratio comparing the probability of data given one hypothesis ($D|H_1$) with the probability of data given an alternative hypothesis ($D|H_2$), As shown in Equation 1.2, the Bayes Factor essentially indicates the weight of the data in altering prior odds of a hypothesis into posterior odds.

$$Bayes\ Factor = \frac{p(D|H_1)}{p(D|H_2)} \tag{1.2}$$

How closely the prior resembles observed data will determine support or rejection of hypothesis being tested.

Using the familiar parlance of null hypothesis significance testing, the Bayes factor can be used to compute how much evidence (in the form of data) revises the probability that a null hypothesis is true – in essence, the Bayes factor evaluates the predictive accuracy of the null and alternative hypotheses. This is an important distinction between frequentist and Bayesian hypothesis testing: While the objective of frequentist hypothesis testing is to consider the probability of data given a null hypothesis (and accept or reject that null hypothesis based on a pre-determined threshold of acceptable risk that the observed data are due to chance alone), Bayesian hypothesis testing aims to evaluate a hypothesis given data (Kass & Raftery, 1995).

Kass and Raftery (1995) offered a comprehensive summation of the uses for, interpretations of, and advantages and disadvantages of Bayes factors in hypothesis testing. For examples of studies incorporating Bayesian hypothesis testing for comparing models see Goodman (199b), Li, Zeng, & Yu (2014), Morey & Rouder (2011), and Ranganathan, Spaiser, Mann, & Sumpter, (2014). Note that there is also substantial literature surrounding inconsistencies between conclusions drawn from Bayesian and frequentist hypothesis testing on similar data (Berger & Berry, 1988; Berger & Sellke, 1987; Casella & Berger, 1987; Moreno & Girón, 2006, Rocha, Loschi, & Franco, 2011; Samaniego & Reneau, 1994).

A final note on the differences in hypothesis testing using frequentist and Bayesian approaches: The oft mentioned argument against Bayesian's incorporation of subjective probabilities is highlighted when considering hypothesis testing. As Berger and Berry (1988) emphasized, there are a number of subjective choices made in hypothesis testing – on the frequentist side, acceptability of error rates, *p* values and statistical power are all subjective choices. Further, both Bayesians and frequentists make subjective choices regarding the alternative hypotheses with which to compare the null. However, as both Wagenmakers, Lee, Lodewyckx, and Iverson (2008) and Rouder, Speckman, Sun, Morey, and Iverson (2009) argued, at least these subjective qualities of hypothesis testing are openly acknowledged, and thus discussed and critiqued, within a Bayesian approach.

Treatment of Prior Information

As discussed earlier, a fundamental property of Bayesian methodology is the formal incorporation of prior information. Within the frequentist context, the influence of prior information is avoided in the interest of ensuring objectivity. Reviews of literature often cite the methodologies and conclusions of similar studies, and perhaps inform *a priori* hypotheses, methodology choice, and model design. Meta-analysis synthesizes information from multiple sources (Borenstein, Hedges, Higgins, & Rothstein, 2009), and there are information theory techniques (Akaike, 1992; Burnham and Anderson, 2001) that allow for the comparison and combination of a number of different models. However, these frequentist approaches do not allow for the explicit introduction of any prior information into analysis of any new data.

A large component of frequentist angst over Bayesian methodology is centered on the idea that there is no guarantee that the prior information used by separate researchers examining the same question is going to be identical or even similar. This is especially true within more "subjective" Bayesian analysis, as it allows for the incorporation of prior data that is not easily or universally quantifiable from actual prior experience. As Efron (2013) wrote, "the Bayesian/frequentist controversy centers on the use of Bayes' rule in the absence of genuine prior experience" (p. 133). The resulting threat to replicability and generalizability is contrary to frequentist approaches focusing on repeated tests over time in controlled environments. However, it could be argued that such controlled environments are often unrealistic and that there is as much subjectivity involved in the design of frequentist studies as there is in the use of Bayesian priors. As Poirer (1988) pointed out, it is the formal quantification and incorporation of prior beliefs that brings about increased levels of transparency:

> …I believe subjective prior beliefs should play a *formal* role so that it is easier to investigate their impact on the results of the analysis. Bayesians must live with such honesty whereas those who introduce such beliefs informally need not. (p. 130)

Additionally, recall that a number of Bayesian scholars recommend evaluation of the influence of prior information as a critical step in Bayesian methodology (Berger, 1994; Berger, 2006; Gelman, Meng, & Stern, 1996; Gill, 2009; Reimeherr et al., 2014; Rubin, 1984; Spiegelhalter & Rice, 2009). Through the analysis of the sensitivity to and robustness of Bayesian inference to the prior, valuable conversation is added to the literature surrounding a problem that is otherwise ignored in frequentist methods.

Even given the increased attention to and discussion of the Bayesian approach to statistics, frequentist approaches remain the dominant techniques first taught to students of statistics. In addition to the controversies over the interpretation of probability, prioritization and understanding of objectivity, and treatment of prior information, Oakes (1986), Schmidt (1996), and Tversky and Kahneman (1971) suggested additional reasons for the tenacity of researcher attachment to frequentist techniques like null hypothesis significance testing even in light of well-documented criticisms. The reliance on significance testing is particularly true within the social sciences (Gigerenzer, 2004; Harlow, Mulaik, & Steiger, 1997; Hoekstra, Finch, Kiers, & Johnson, 2012; Kline, 2004). While there does appear to be a general movement away from significance testing and *p* values in the more recent literature, the alternatives often offered are still based within the frequentist framework. For example, the latest version of the authority in social science research publication, the American Psychological Association's *Publication Manual* (2013) , encouraged authors to view null hypothesis statistical testing as "but a starting point" (p. 33) and to seek out and report other frequentist results such as effect sizes and confidence intervals. A number of authors have extolled the use of alternatives such as confidence intervals to null hypothesis significance testing within the social sciences (Cumming, 2014; Fidler & Loftus, 2009; Hoekstra, Morey, Rouder & Wagenmakers, 2014), and Gigerenzer (2004) and Finch et al. (2004) suggested that editorial support of alternatives to null hypothesis statistical testing is a necessary but insufficient environment to foster alternatives. The editors of the *Journal of Advanced Academics* recently revealed new editorial policies expressing preference for effect sizes and confidence intervals, as well as encouragement of replication of other studies, over significance testing (McBee & Matthews, 2014). Once again, these are all frequentist alternatives.

However, Bayes' popularity is growing in increasing numbers of disciplines – as Andrews and Baguley (2013) pointed out, Bayesian methods were present in 20% of articles in the most highly respected statistics journals. As its popularity grows, Bayesian methodology will face increased attention, scrutiny, and inevitable disagreements among its own practitioners. For example, within the current overall Bayesian camp, there are ongoing dialogs and philosophical conflicts surrounding "practical Bayesianism," the practice of using Bayesian techniques without a commitment to or adoption of the Bayesian philosophy of science (Boorsboom & Haig, 2013; Gelman & Shalizi, 2013; Kruschke, 2013; Morey, Romeijn & Rouder, 2013). Authors like Dennis (1996), while acknowledging advantages in Bayesian techniques, expressed reluctance in adopting an approach until more discipline-specific researchers analyze, discuss and test Bayesian methodologies. Further, the widespread consideration and discussions of articles like Gorard's (2014) and Ioannidis' (2005) indictments of research methodologies and publication bias hints that the scientific community is open to a diversity ideas and approaches. This openness, along with increased external and internal examinations of Bayesian methodology, can only help to improve and move forward scientific inquiry.

Bayesian Methods and Social Science/Educational Research

Bolstad (2007) succinctly summarized the advantages of a Bayesian approach, noting the following benefits: The formal consideration of prior information, easily interpretable results in the form of probability statements, and one universal tool (Bayes' theorem) that is applicable to every question or situation. As noted throughout earlier discussion, hypothesis testing and the use of *p* values is commonly misunderstood and misused. Additionally, hypothesis testing requires of researchers a number of judgments and decisions surrounding rejection/acceptance thresholds, model design, statistic used, etc. In contrast, Bayesian methods require only one decision in the choice of a prior. Outside of these advantages of simplicity and universality, however, Bayesian methods are particularly amenable to social science research. Gill (2009) extolled the suitability of Bayesian methods to social and behavioral research by noting that many of the overarching questions and topics within the field of human behavior simply don't fit within the frequentist (long run probability and replicability) paradigm: "Ideas like 'personal utility,' 'legislative ideals points,' 'cultural influence,' 'mental states,' 'personality types,' and 'principal-agent goal discrepancy' do not exist as parametrically uniform phenomena in some physically tangible manner" (p. 26). These questions of human behavior are difficult to generalize and measure at the individual level, thus there is a unique amount of uncertainty within social science research. As Montgomery and Nyhan (2010), Raftery (1996), Rubin (1984), and Western (1999) explained, Bayesian methods are well-adapted for handling and expressing uncertainty. Gill further pointed out that the influence of social norms is particularly important within social science research, especially in terms of biases, judgments and assumptions brought to the research by researchers, and that Bayesian's use of subjective probability and a formal prior is particularly suited to transparently addressing this influence. Among his arguments in favor of using Bayesian methods for social science research, Raftery (1995) noted that social science often uses large data sets especially sensitive to *p* values and subsequent rejection of null hypotheses, and Western (1999) highlighted Bayesian's handling of accounting for uncertainty as a fundamental reason of its compatibility to social science research. Berk, Western, and Weiss (1995) and Gorard (2014) pointed out that the underlying assumptions required for significance testing, particularly that of selecting and comparing truly random samples against a known population, are typically unmet by social science data. Further, due to ethical and logistical complications, generating random samples (and ultimately using analytical techniques assuming random sampling) within social science research is often impossible. Ranganathan, Spaiser, Mann, and Sumpter (2014) highlighted the efficiency in comparing Bayes factors in model selection in the social sciences, and numerous other authors (Bolstad, 2007; Gelman, 2008a; Gelman et al., 2004; Gill, 2009) discussed how Bayesian methods are suited for the types of hierarchical modeling often encountered in the social sciences.

Bayesian Methods and Institutional Research

Chapter Two of this manuscript involves the development of a predictive retention model using a Bayesian approach. In addition to being considered social science and educational research, this type of activity is more specifically categorized as “Institutional Research,” a specialized type of educational research performed within educational institutions used to inform decision- and policy-making within that specific institution or system of institutions. Institutional Research often deals with large sets of population data, as analysts have access to current and historical information databases, and while experimental designs are not unheard of within institutional research, the aforementioned difficulties of random selection and controlled trials within educational research make them difficult and rare. Additionally, the data available to institutional researchers often simply do not meet the assumptions required in the frequentist paradigm – as Luan and Zhao (2006) wrote, “Institutional researchers often feel frustrated as assumptions for valid statistical inferences are often violated with dealing with real institutional research problems and when messy, ambiguous, and incomplete data are present” (p. 117). Bayes’ comfort with uncertainty and the explicit use of formal priors within Bayesian methods provide an alternative avenue for institutional researchers – calling on the expertise and experience of teachers, administrators and others, institutional researchers can incorporate this knowledge into formal priors that can then be combined with institutional data to generate *a posteriori* conclusions. Additionally, the use of formal priors requires institutional researchers to unambiguously address and justify potential bias and assumptions.

A major function of institutional research is providing decision-support to administrators, and as Bernardo and Smith (2000) and Robert (2007) pointed out, Bayesian methods are particularly helpful in decision-making. Bayesian approaches towards decision-making and decision support, mostly through the use of Bayesian networks, have been widely employed in medical and engineering fields – see Berner (2006), Greenes (2014), Lucas, van der Gaag, Abu-Hanna (2004), and for examples of medical uses and Li, Han and Kang (2013), Rezaee, Raie, Nadi, and Ghidary (2011), Swiler, (2006), and Zhu and Deshmukh (2003) for examples within engineering. However, such approaches to institutional, and even more broadly, educational, research are less common in the literature. Bekele and Menzel (2005) used Bayesian networks to predict student performance, noting that they chose a Bayesian approach due to its specialized skill in handling and expressing uncertainty. Loeb (2003) incorporated Bayesian estimation in hierarchical linear modeling to explore gender equity in faculty salaries, and Laru, Naykki & Jarvela (2012) used Bayesian methods to identify predictors of learning outcomes. In the larger field of educational research, Vomlel (2004), Wainer, Wang, Skorupski and Bradlow (2005), and Ricketts and Moyeed (2011) incorporated Bayesian methods in the evaluation and improvement of educational testing, and a number of authors have explored the role of Bayesian methods in Item Response Theory (Gao and Chen, 2005; Johnson, 2013; May, 2006; Almond, Mislevy Steinberg, Yan, & Williamson, 2015). A more extensive review of Bayesian networks in institutional research is included in subsequent sections.

In summary, a Bayesian approach is particularly suited to institutional research for a number of reasons. The use of formal priors allows for the incorporation of additional sources of knowledge or experience, as well as the transparent acknowledgement and discussion of biases, subjectivity, and assumptions involved in the research process. Additionally, a Bayesian approach embraces the high levels of uncertainty and unmet frequentist assumptions frequently encountered in institutional research topics. Finally, the qualitative and social science research skills of institutional researchers are featured during the process of formal elicitation of Bayesian priors. Chapter Two will incorporate Bayesian methods through the creation and use of a Bayesian network using formal elicitation, and the following sections will provide a background and literature review of these two subjects, with a focus on applicability to institutional research.

Elicitation of Subjective Probabilities from Experts in Order to Form Bayesian Priors

> Often, we want to make use of the opinion of a person whom we regard as an expert. Does the weatherman think it will rain, the doctor that we shall soon get well, the lawyer that it would be better to settle out of court, or the geologist that there might be lots of oil at the bottom of a deep hole? (Savage, 1971, p. 795)

The main idea behind the Bayesian approach to statistics is that researchers revise their understands or beliefs of certain outcomes in light of new evidence – Bayesian statisticians combine both prior information and new data through the use of Bayes' theorem in order to estimate the probability of an outcome (Bolstad, 2007). One of the most exciting aspects of employing Bayesian methodology in institutional research is the use of experts' beliefs to quantify and use prior information. The elicitation of experts' beliefs introduces a qualitative aspect to research design, often in the form of interviewing, and the information gathered is then quantified and used in computing the probability of studied events occurring. Not only does this methodology enhance potential for increased insight into current educational and institutional research questions, but it also introduces an opportunity to further explore and realize the benefits of mixed-methods research. This section focuses on the gathering of experts' opinions using various elicitation techniques, including a discussion of the types of probabilities being elicited, elicitation best practices and challenges, and the role of subjective probability elicitation within institutional research.

Examples of expert elicitations used to form Bayesian priors can be found in a number of fields, including business (Bajari & Ye, 2003; Gustafson et al., 2003; Lagerström, Johnson, Höök, & König, 2009), clinical settings (Johnson, Tomlinson, Hawker, Granton, & Feldman, 2010; Prajna et al., 2013; Spiegelhalter, Abrams, & Myles, 2004; White, Carpenter, Evans, &

Schroter, 2007), engineering (Dogan, 2012; Jorgensen, Teigen, & Molokken, 2004; Kaplan, 1992), communications (Vogelgesang & Scharkow, 2009), information security (Ryan, Mazzuchi, Ryan, de la Cruz, & Cooke, 2012), politics (Buckley, 2004), public policy (Morgan, 2014), and, most commonly, ecology (Choy, O'Leary, & Mengersen, 2009; Kuhnert, Martin, Mengersen, & Possingham, 2005; Murray et al., 2009; O'Neil, Osborn, Hulme, Lorenzoni, & Watkinson, 2008). While less common, the literature does include expert elicitations within the educational field. For example, Bosworth, Gingiss, Porthoff and Roberts-Gray (1999), asked health education experts to estimate the likelihood of a program's successful implementation. In the narrower field of institutional research, Subbiah, Srinivasan, and Shanthi (2011) discussed the potential advantages of using expert-elicited Bayesian priors in enrollment management. However, the majority of educational research studies incorporating Bayesian methodologies do not include elicitation.

Elicitation and Uncertainty

Expert elicitation involves the quantification and transformation of experts' opinions into subjective probabilities used to inform a prior probability distribution which is then updated with new data using Bayesian techniques. Just as discussion of definitions of probability is fundamental to the differences between frequentist and Bayesian approaches, distinguishing between a frequentist and Bayesian definition of probability is important in elicitation. As a main goal of expert elicitation is to quantify uncertainty regarding some particular event or variable over which an expert can provide better information than any other source, it is understood that the expert is providing personal, or subjective, probabilities. Expert elicitation involves the gathering of experts' *degrees of belief* in some uncertain event or value. O'Hagan et al. (2006) explained this further by placing expert elicitation within the context of uncertainty: In contrast to aleatory, or random, uncertainty, expert elicitation deals with epistemic (imperfect knowledge) uncertainty, in that experts are asked to weigh in on the uncertainty related to a unique or specific event on which there is imperfect knowledge. In other words, experts are asked to make judgments or decisions under conditions of uncertainty. Ideally, these judgments and decisions under uncertainty adhere to normative theories of decision making as described by de Finetti (1973), DeGroot (1970), and Savage (1954), which basically state that an expert's degree of belief is a function of rationality, maximizing rewards, and adherence to the axioms of probability. Normative decision theory also states that probabilities are an adequate, or even ideal, avenue towards expressing uncertainty (Cooke, 1991). In short, the probabilities elicited in an expert elicitation are assumed to be a rational, coherent representation of an expert's uncertainty. However, as discussed in the upcoming discussion of best practices in elicitation, the assumption that elicited probabilities are actually those described by the normative theories of decision making is often challenged.

Elicitation Best Practices

Kadane and Wolfson's "Experiences in Elicitation" (1998) is used by many researchers incorporating expert elicited Bayesian priors as a guide to "best practices" (Garthwaite, Kadane, & O'Hagan, 2005; Gill, 2009; O'Hagan et al., 2006). The authors defined a successful elicitation as one in which the researcher assures the process is "as easy as possible for subject-matter experts to tell us what they believe, in probabilistic terms, while reducing how much they need to know about probability theory to do so" (p. 4). Other authors who have provided step-by-step guidance in the elicitation process in numerous settings include Clemen and Reilly (2001), Cooke and Goossens (2004), Garthwaite et al. (2005), Meyer and Booker (2001), Phillips (1999), Shephard and Kirkwood (1994), and Walls and Quigley (2001). A review of these works highlights the following major stages of the elicitation process: Preparation of the researcher/elicitor, selection of expert(s), training of expert(s), confirmed understanding or acceptance of the model for which judgments are being elicited, and the actual elicitation, including assessment and feedback. These practices serve to provide the expert ample opportunity to adequately express her beliefs while also allowing the researcher to gather as much helpful information as possible and verify her own understanding of what the expert is trying to communicate. Chapter Two of this manuscript will incorporate the guidance of Kadane and Wolfson and the later work of O'Hagan et al. (2006) in a formal elicitation of undergraduate retention experts.

In terms of the actual information being elicited, best practices center around the research question, the type of prior desired, and the expert(s). There exists a large amount of literature discussing individuals' ability (and, more often, lack thereof) to estimate or judge statistical quantities – see Beach and Swenson (1966), Erlick (1964), and Shuford (1961) as examples – that speaks to the necessity of careful consideration paid to the types of summary statistics being elicited. Findings reported by Wallsten, Budescu, Rapoport, Zwick, & Forsyth (1986) and Wardekker, van der Sluijs, Janssen, Kloprogge, & Petersen (2008) suggested that elicitations cannot simply rely on experts using words like "likely" or "unlikely" to qualify uncertainty – elicitations typically involve quantification. Elicitations involving the gathering of a single probability (see Spetzler & Staël von Holstein, 1975) can be enhanced by asking the experts to consider probability judgments within the context of gambling (Clemen & Reilly, 2001; Renooij, 2001), relative frequency (Price, 1998), and probability scales (Wang, Dash & Druzdel, 2002). Winkler (1967) provided an introduction to eliciting probability distributions, although elicitations within the social sciences seeking probability distributions more often use indirect methods (Gill, 2009). For example, Bedrick, Christensen, and Johnson (1997), Higgins, Huxley, Wapenaar, and Green (2014), and O'Hagan et al. (2006) asked experts to express their beliefs through quantiles or intervals which were retroactively fitted to probability distributions. Chaloner and Duncan (1987) discussed the elicitation of multinomial distributions, and Garthwaite and Dickey (1992) and Garthwaite, Al-Awadhi, Elfadaly, and Jenkinson (2013) addressed the elicitation of summary statistics related to regression. Garthwaite et al. (2005) provided recommendations for multivariate elicitations, noting that the joint probability distribution required in such situations is particularly challenging, while Goldstein (2004) and Oakley and O'Hagan (2007) explored nonparametric approaches to incorporating expert elicitations as Bayesian priors.

In many cases, information from multiple experts will be used to form a Bayesian prior, and a number of authors have addressed and evaluated methods of combining expert opinions in prior elicitation – see Clemen and Winkler (1999) and Hammitt and Zhang (2013) as examples. The aggregation of expert opinions is completed mathematically or behaviorally. For example, the mathematical combination of expert judgments can be accomplished through the use of averaging (Burgman et al., 2011; Cooke, 1991), pooling (French, 1985; Genest & Zidek, 1986) or even using Bayesian approaches (Albert et al., 2012; Roback & Givens, 2001). Behavioral approaches to combining expert judgments involve interaction and consensus building among the experts. Practices like the Delphi method (Dalkey, 1969) and Nominal Group Technique (Delbecq, Van De Ven, & Gustafson, 1975) are often used in an effort to counteract challenges of bullying and "group think" often encountered in group interactions (Clemen & Winkler, 1999). Each approach has drawbacks and challenges, which, as with many of the best practices related to expert elicitation, have much to do with the research question or topic and choice of expert.

A final best practice related to expert elicitation involves testing the elicited information's accuracy in terms of the extent to which it matches experts' true knowledge or beliefs as well as reality. Garthwaite et al. (2005) recommended examination of the coherence of elicited probabilities, as well as offering feedback to the experts in an effort to clarify or correct for inconsistencies. The authors also suggested analyses of the sensitivity of elicited probabilities or distributions to changes in assumptions or other model parameters. In cases where data reflecting actual events are available, scoring rules, in which experts are awarded a score based on the quality of calibration of their judgments with reality, can be used to judge and even improve the accuracy of elicited probabilities (Gneiting & Raftery, 2007; Matheson & Winkler, 1976; Savage, 1971). Building on the use of scoring rules, Cooke (1991) recommended gathering multiple expert opinions, weighting each expert's assessment based on the performance scores, and then producing a weighted synthesized score.

Challenges of Expert Elicitation

There are a number of challenges related to expert elicitation, the majority of which stem from the design of the elicitation protocol. As a goal of elicitation is to allow the expert to communicate her most accurate degree of belief, it is important to design an elicitation protocol that strives to minimize the introduction of bias or other confounding influences on an expert's true degree of belief. Articles designed to serve as guides to designing and carrying out elicitations (Choy et al., 2009; Jenkinson, 2005) emphasized the importance of carefully considering and articulating a research question and elicitation protocol as critical parts of the elicitation process. This is similar to the advice found in texts directed towards standard qualitative research (see Denzin & Lincoln, 2000, and Kvale & Brinkmann, 2009), and it is recommended that thoughtful and measured attention to the various considerations (costs, timing, expert statistical background, etc.) is given in the design of the elicitation protocol.

Authors studying calibration report that the quality of correspondence between elicited subjective probabilities and actual occurrences has much to do with the way the elicitation is conducted (Beach & Braun, 1994; Lichtenstein, Fischhoff, & Phillips, 1982; Ronis & Yates, 1987). Therefore, consideration of the findings of the larger field of calibration research into the design of an elicitation protocol is one helpful way to increase the likelihood of superior elicitation. For example, Gigerenzer, Hoffrage, and Kleinbölting (1991), Gigerenzer (1996) and Thomson, Önkal-Atay, Pollock, and Macaulay (2003) found that task characteristics, or how experts are asked to respond, can influence the extremity and over/under-confidence of elicited probabilities. Carlson (1993) and Wright and Ayton (1992) suggested that the timing of the event for which probabilities are being sought affects calibration, while Bornstein and Zickafoose (1999), and Jonsson and Allwood (2003), and West and Stanovich (1997) found that the knowledge domain of an elicitation influences judgment. Additionally, the definition of "expertise" is not always universal (Caley et al., 2013; Martin et al., 2012), and researchers should recognize that subject-matter expertise does not guarantee skill in expressing probabilistic beliefs. In their guide to expert elicitation, Kuhnert, Martin, and Griffiths (2010) differentiated between two styles of elicitation activities – direct, in which experts are asked to provide opinions in probabilistic terms, and indirect, in which experts are asked to provide opinions in less technical terms that may be more amenable to their field of expertise – and recommended that the choice of styles be dependent upon the experts' background. Fischhoff (1989), Murphy and Winkler (1984), and O'Hagan et al. (2006) offered strategies for avoiding or mitigating potential biases, including expert training, careful facilitation and feedback provided by the researcher, and paying attention to experts' coherence of beliefs. Additionally, the introduction of computer software used to elicit expert belief that can also quickly recognize incoherent, irrational, or contradictory information, mitigating some of the challenges associated with expert elicitation (Garthwaite & Dickey, 1992; James, Choy & Mengersen, 2010; Lau & Leong, 1999; Morris, Oakley, & Crowe, 2014).

Many of the aforementioned challenges to calibration stem from the use of heuristics in the formation of probability judgements. Hogarth (1987) and Kadane and Wolfson (1998), expanding on the work of Tversky and Kahneman (1974), provided a list of the common heuristics that can introduce unhelpful bias into experts' opinions: Availability bias (an easier-to- recall occurrence may incorrectly be deemed more important or likely), anchoring (experts may calculate probability based on an initial value), overconfidence, and hindsight (experts who have seen sample data may update their opinion). Heuristics are a particular threat to the assumption that elicited probabilities are those that can be described under the normative theory of decision making – the use of heuristics can interfere with the rationality and logic required under the normative theories. Kynn (2008) warned that psychological research concerning the bias introduced when experts use heuristics to make probability judgments has not kept up with the heightened attention paid to Bayesian methods. As Kynn wrote, "…we should be equally concerned with not only *what* we ask experts to assess, but *how* we ask it" (p. 240).

An equally important challenge to expert elicitation deals with bias on the part of the researcher. Expert elicitations (including the elicitation carried out in Chapter Two) commonly take the form of an interview, and references to and discussions of researchers' subjectivity and bias are common in the interviewing literature. For example, Scheurich (1995) termed the potential influences on a qualitative researcher (socioeconomic background, experience, expertise, funding sources, power) as "baggage," and recommended that interviewers thoughtfully consider and disclose these types of baggage. Kvale & Brinkmann (2009) pointed out that these subjective influences can even be found in the actual transcription of interviews, thus affecting the subsequent analysis.

Qualitative Research Lessons for Expert Elicitations

These issues of bias and calibration underscore the importance of a well-designed elicitation in which the expert(s) are carefully chosen and the questions are designed in a way to minimize the overuse of heuristics and other potential sources of bias affecting experts' confidence in their judgments. However, formal and informal review of the studies incorporating expert elicitation of subjective probabilities, including those cited in this manuscript, reveals little discussion of the psychological theories surrounding probability judgments or the potential for bias on the part of the experts or the researchers. Additional research regarding the use of heuristics and mitigation of resulting bias will be necessary as elicitations to form Bayesian priors become more common. As Hogarth (1975) wrote, elicitation of subjective probabilities "should be designed both to be compatible with man's abilities and to counteract his deficiencies" (p. 284). Additionally, knowledge about the basics of interviewing techniques is also recommended, as the majority of researchers employing Bayesian methodology appear to have little to no experience with qualitative techniques such as the responsive interviewing model described by Rubin & Rubin (2012). In the interest of exposing and exploring one's own baggage, as well as that of the experts, researchers using expert elicitation and Bayesian inference might consider adopting the type of discussion and disclosure often practiced by more qualitative-leaning peers.

Expert Elicitations and Institutional Research

As first pointed out by Gill and Walker (2005), a review of the literature reveals that expert elicitation is much more common within life and engineering sciences than in the social sciences. However, at least in the studies included as examples in this manuscript, the full advantages of expert elicitations are not as readily extoled in the sciences. For example, a few of the clinical and engineering studies reviewed reveal minimal information about the elicitation process, and even dismiss its value. Dogan (2012) simply mentioned that elicitation was used without providing any other detail, and White et al. (2007) used language painting the expert elicitation as a sort of consolation prize to use when other data are unavailable.

Two of the social sciences studies included here as examples addressed the idea that, due to its relative familiarity and experience with qualitative methods, social science research should be particularly responsive to the expert elicitation/Bayesian approach (Buckley, 2004; Vogelgesang & Scharkow, 2009). Buckley further pointed out that the adoption of Bayesian methodology within the social sciences will only catch on after clear and relevant guidance regarding the transformation of expert opinion into quantitative data is provided. Buckley's sentiment is echoed in Moyé's (2008) critique of Bayesian clinical research, where he called for researchers to "take a strong stand for disciplined research methodology" (p. 477) that rivals the well-articulated and accepted frequentist version.

The social science complexities and nuances of institutional research render the discipline an excellent candidate for Bayesian methods and expert elicitation. While large amounts of quantitative data are collected and available, there is also a legacy of qualitative approaches that can be leveraged and applied towards expert elicitation. In addition to an institutional researcher's familiarity with qualitative methods, consider the fact that a large number of experts stand at the ready. Based on their experiences, educators and administrators are experts within their individual classrooms and institutions. Expert elicitation provides an avenue from which the expertise of these individuals can be mined and combined with other types of data to produce informed and comprehensive conclusions. Furthermore, expert elicitation is often used to aid decision making or prediction under conditions of uncertainty, such as instances in which there are not empirical data available, or the introduction of unfamiliar or new situations or problems. Additionally, information provided by experts is a viable alternative source of information when resources are limited (Kuhnert et al., 2010). The ideas of uncertainty and limited resources are particularly relevant to institutional research, as the field typically operates within tightly managed budget and calendar situations with varying amounts of access to data and among unique and changing student populations.

However, before incorporating expert elicitations into institutional research, researchers should carefully select elicitation techniques that are appropriate to the specific research questions, experts' abilities, and the researchers' own capacity. The researcher should be knowledgeable in the benefits and challenges of qualitative methods and be prepared to evaluate and discuss heuristics and bias. Additionally, researchers should have a confident understanding of and justification for quantification of the elicited prior. As Buckley (2004) pointed out, the likelihood that the employment of expert elicitation and Bayesian inference in qualitative or mixed-method social science research will become commonplace is directly related to whether or not Bayesian scholars can effectively equip researchers to correctly gather, analyze and apply expert elicitations. In summary, through careful selection of prior type and application of formal elicitation best practices, institutional researchers can leverage and incorporate expert information into data analysis and modeling.

Elicitation Conclusions

Eliciting Bayesian priors serves as an excellent way to explore and better answer questions for which empirical data are either unavailable or insufficient. As Gill (2009) wrote:

> The core of this argument is the idea that if the prior contains more information that pertains to the estimation problem, then we are foolish to ignore it simply because it does not neatly fit into some familiar statistical process. (p. 28)

The anxiety introduced by the subjective nature of elicited priors can have benefits outside increased information and explanatory power. For example, research can only benefit from increased discussion and disclosure of underlying theories, assumptions, and subjectivities. Additionally, the elicitation of priors allows for greater communication and collaboration between researchers and experts, ultimately producing a feedback loop of knowledge (Garthwaite et al., 2005). The complexity and nuance inherent in institutional research, abundant experts, and a legacy of effective qualitative research methods highlight the discipline's suitability for expert elicitation and Bayesian methodology.

Bayesian Networks

Chapter Two of this manuscript describes a study using a Bayesian approach, in which expert judgment is elicited and used to design a model that predicts retention. The practice of modeling processes or systems in order to better understand them is not uncommon. However, when reasoning or trying to make decisions under conditions of uncertainty, such models need to account for this uncertainty. Probabilistic networks offer an approach to producing models that incorporate uncertainty through the use of probabilistic inference. These networks typically represent causality, illustrate the strength of relationships between variables using conditional probabilities, incorporate the numerical quantification of choices or preferences, and solve for maximum expected utility (Kjærulff & Madsen, 2008). Introductions to probabilistic networks typically involve medical diagnosis examples where a physician is tasked with making a medical diagnosis given symptoms and other information. In such cases, a probabilistic network is a formal way of representing the diagnosis process and conclusions, using probabilistic inference and graphical representations of relationships between and among symptoms, other information, and the presence of a medical condition. The following sections discuss Bayesian networks, a specific form of probabilistic networks, including such networks' development, applications, and role in educational and institutional research.

Very simply, a Bayesian network is a graph that models the probabilistic relationships between and among variables. Kjærulff and Madsen (2008, p. 3-9) presented a technical definition of Bayesian networks, describing the two main elements: A directed acyclic graph (DAG) forming the structure of the model where variables included in the model are represented by nodes (often squares, ovals, etc.), and the relationships of independence and dependence between and among the variables, which are represented by directed edges (arrow-ended lines) and are quantified by conditional probability distributions. Together, these two elements form a Bayesian network that models a joint probability distribution that is equal to the product of the conditional distributions of each node. Equation 1.3 describes this joint distribution:

$$P_a(x_1, \dots x_n) = \prod_{i=1}^{n} P_a(x_i | \prod x_i) \tag{1.3}$$

Figure 1.1 is an example of a very simple, hypothetical Bayesian network modeling the retention of an undergraduate student from one year to the next. The simplified Bayesian network implies that retention is dependent on whether or not the student is engaged in student life and whether or not the student's financial need is met. In this example, based on the relationships indicated by the directed links, "Student Engagement" and "Financial Aid Need

Met" are "parent" nodes of the "child node," "Student Retained." The lack of a link between "Student Engagement" and "Financial Aid Need Met" implies independence between these two variables. The Conditional Probability Table (CPT) (Table 1.3) summarizes the probability of retention given the states, or conditions, of the parent nodes. For example, the probability that a student will be retained if she is engaged and has her financial need met is 90%, compared to a 35% probability of retention given a lack of engagement and unmet financial need. Note that the parent nodes (Student Engagement and Financial Need Met) are not conditioned on other nodes, and thus do not have conditional probabilities – the probabilities associated with these nodes are considered the prior probabilities.

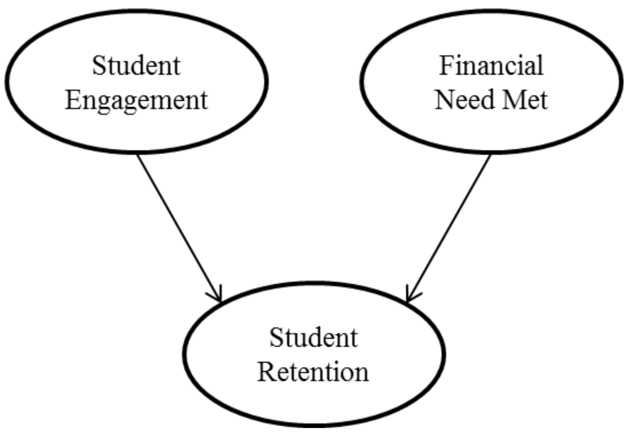

Figure 1.1 Simplified BN modeling student retention as dependent on student engagement and financial need met. "Student Engagement" and "Financial Need Met" represent parent nodes. The directed links represent the parent nodes' influence on the child node, "Student Retention."

Table 1.3
Conditional Probability Table (CPT) – Simple Retention BN

Parent nodes		Child node (Student Retained)	
Student Engagement	Financial Need Met	Retained (%)	Not Retained (%)
Engaged	Met	90	10
Engaged	Not Met	70	30
Not Engaged	Met	60	40
Not Engaged	Not Met	35	65

First defined by Kim and Pearl (1983), Pearl (2000) noted that the descriptor “Bayesian networks” was adopted in order to highlight the following defining characteristics: “(1) the subjective nature of the input information; (2) the reliance on Bayes’ conditioning as the basis for updating information; and (3), the distinction between causal and evidential modes of reasoning…” (p. 12). Note that multiple terms describing probabilistic networks with Bayesian applications exist in the literature, including “influence diagrams,” “belief networks,” “relevance diagrams,” and “knowledge maps,” but the term “Bayesian network” or “BN” will be used throughout this manuscript.

BNs were developed in response to other rule-based systems that failed to consistently represent and perform under conditions of uncertainty (Pearl, 1988). In short, Bayesian networks use probability to describe and incorporate uncertainty in a causal situation. Causality is a key property of BNs, and is addressed and explored in Druzdzel and Simon (1993), Heckerman and Shachter (1995), and Pearl and Verma (1991). Heckerman (1997) noted that as conditional dependence can be viewed in terms of causality, the directed links between variables in a BN typically imply cause and effect relationship between variables. In other words, direct influences on variables within a BN are represented by a directed edge or arrow between two of the variables. Through the representation of independence of variables in the DAG (unconnected nodes), a BN models conditional independence and allows for the “explaining away” (inter-causal inference) of less likely events using conditional probability. It is this incorporation of conditional probabilities that set BNs apart from other graphical models – using conditional probabilities and Bayes’ theorem (Equation 1.1), a BN models the change in probability of one event occurring given, or conditioned by, another event occurring (Pearl, 2000). The graphical representation of these conditional independencies in probabilistic terms provides users with a clear representation of the relationships between and among variables within a system or network.

BNs perform probabilistic inference to estimate an outcome on one or more variables given the state of other variables. Such inference is completed using applications of Bayes' theorem (Equation 1.1), where a prior distribution is updated via conditional probabilities represented in the model to form a posterior distribution. Although Cooper (1990) and Dagum and Luby (1993) concluded that probabilistic inference within Bayesian networks is an NP-hard undertaking, a number of researchers have developed inference algorithms to ease computation. Heckerman (1997) provided an overview of techniques for probabilistic inference in BNs, highlighting an algorithm endorsed by Howard and Matheson (1984), Olmstead (1983), and Shachter (1988), which applies Bayes' theorem to reversals in the directed links between variables, as well as the algorithm developed by Dawid (1992), Jensen, Lauritzen and Olesen (1990), and Lauritzen and Spiegelhalter (1988), which employs message passing in a tree structure. With the development of computing power, the ability to complete inference in increasingly complex BNs using these and other algorithms has grown (Jensen & Nielsen, 2007).

Kjærulff and Madsen (2008) highlighted the advantages of Bayesian networks, noting the efficiency in which these networks conduct inference and convey causal relationships, the ease in which the graphical representations can be understood by numerous audiences, and a firm foundation in decision theory. Speigelhalter, Dawid, Lauritzen and Cowell (1993) cited the ability of BNs to simultaneously "…be forgiving of limitations in the data but also exploit the accumulated data" (p. 221), and Heckerman (1997) acknowledges the ability of BNs to operate with incomplete data. Additionally, due to the use of Bayesian statistics, a BN can flexibly and efficiently incorporate additional information as it is gathered. Chapter Two of this manuscript will include a Bayesian network used for describing, and ultimately predicting, the likelihood of retention given certain demographic, academic and affinity variables. BNs are not without their limitations, however. Niedermeyer (2008) pointed out that novel events may threaten the predictive validity of BNs, and cautioned that, even with computing advances, a network with a large number of variables may require unreasonable computing and computational power. Pourret, Naim, and Marcot (2008) considered the requirement that BNs be acyclic to be a limitation as feedback loops are often found in reality. Additionally, the reliability and quality of prior information included in BNs affects a model's usefulness, although this can be explored through adequate model evaluation (Cowell, Dawid, & Spiegelhalter, 1993; Pitchforth & Mengerson, 2013).

Applications - General

Charniak (1991) and Henrion, Breese and Horvitz (1991) offered introductory overviews of BNs, while Darwiche (2009), Jensen and Nielsen (2007), and Koller and Friedman (2009) provided detail and instruction on the fundamental theories and applications of BNs. Kjærulff and Madsen (2008), Korb and Nicholson (2010), and Pourrett, Naim and Marcot (2008) included applied instruction and real-world examples. A number of applications of BNs in specific fields can be found throughout the literature. Spiegelhalter et al. (1993) illustrated the use of BNs in medical and other types of diagnosis, and Donald et al. (2012), Gao, Madden, Chambers, and Lyons (2005), and Neapolitan (2009) provided illustrations of the use of BNs within bioinformatics. Other fields employing BNs include marketing (Baesens, Viaene, van den Peol, Vanthienen, & Dedene, 2002; Cui, Wong, & Lui, 2006; Neapolitan, 2007), space flight (Horvitz & Barry, 1995), ecology (Landuyt et al., 2013; Marcot, 2008; Shenton, Hart, & Chan, 2014), and risk assessment in various disciplines such as business (Phillipson, Matthijssen, & Attema, 2014), engineering and construction (Leu & Chang, 2013; Zhang, Wu, Skibniewski, Zhong, & Lu, 2014), and health (Aussem, de Morais, Rodrigues, & Corbex, 2012).

Applications - Prediction

As Chapter Two of this manuscript will focus on the development of a BN used for prediction, particular attention is paid to the application of BNs for that purpose. In their discussion of the role of causality in BNs, Heckerman, Geiger, and Chickering (1995) wrote the following of authors supporting the development of causal formalisms within Bayesian networks:

> They argue that the representation of causal knowledge is important not only for assessment, but for prediction as well. In particular, they argue that causal knowledge – unlike knowledge of correlation – allows one to derive beliefs about a domain after intervention. (p. 213)

The key to using a BN for predictive purposes lies in the interpretation of the links between variables – within a causal Bayesian network, the nodes from which arrows descend are considered parent nodes and direct, immediate causes of the nodes at which they point. Friedman, Linial, Nachman, and Pe'er (2000) explained this idea further by pointing out that the directionality and causal interpretation of the links between nodes allows for observation of intervention effects – if a parent node is a direct cause of a child node, then a change (intervention) in the value of the parent node will effect change in the value of its child node. This causal interpretation of a BN also requires acceptance of the Causal Markov Assumption, which basically states that a variable is independent of all variables outside of its direct causes and effects (Hausman & Woodward, 1999; Spirtes, Glymour and Scheines, 2000). Pearl (2000) provided a comprehensive explanation of causal Bayesian networks.

Instances of BNs used for prediction are found throughout the literature. Axelrad, Sticha, Brdiczka, and Shen (2013) and Venkatesh, Cherurveettil, and Sivashanmugam (2014) employed BNs to predict risks to cybersecurity. Fenton, Neil, and Marquez (2008) built a BN to predict software defects, Stahlschmidt, Tausendteufel and Härdle (2013) used a BN to predict offender profiles, and Sun and Shenoy (2007) attempted to predict bankruptcy using BNs. Predictive BNs are also commonly used in the study of biological networks (Friedman, et al., 2000; Jansen et al., 2003), medicine (Cho, Park, Kim, Lee & Bates, 2013; Sandri, Berchialla, Baldi, Gregori and De Blasi, 2014; Jiang, Xue, Brufsky, Khan, & Neapolitan, 2014) and weather forecasting (Cano, Sordo & Gutiérrez, 2004).

Model Development

A review of the literature concerning best practices in network modeling reveals consensus on a number of steps that should take place, including comprehensive description of the model's principal function and assumptions, careful construction of the network's structure and underlying probabilities, assessment of the model's functionality, and discussion of the entire model development process (Chen & Pollino, 2012; Crout et al., 2008; Marcot, Steventon, Sutherland, & McCann, 2006).

Structure and relationships. Spiegelhalter et al. (1993) listed three stages of constructing a Bayesian network: A qualitative stage in which the author defines the relationships among and between variables in terms of conditional independence and develops a graphical model that reflects these relationships, a probabilistic stage in which the author considers the model's joint distribution, and a quantitative stage in which the author assigns values to the underlying conditional probability tables (CPTs). Approaches to each stage can be manual (theory- and expert-driven) or automatic (data-driven), or even a combination of both.

The decision regarding approach to model construction often depends on the field on which the model is based (Chen & Pollino, 2012; Uusitalo, 2007) or the availability of data (Pitchforth & Mengersen, 2013).

Manual approaches to the construction of a BN require expertise and advanced familiarity with the system being modeled – manual construction of Bayesian networks will almost certainly require domain knowledge input from experts or previous research. Neil, Fenton, and Nielsen (2000) described the manual process of designing the structure of a BN as "knowledge engineering," and offered a step-by-step approach to the manual construction of a BN structure that encourages developers to first focus on the relationships between smaller groups of variables before considering the relationships among variables in the entire network. The authors proposed five "natural and reusable patterns in reasoning" (p. 13), termed "idioms," that can be used as guidance in modeling directionality, causality, measurement accuracy, induction, and reconciliation of two competing factors or explanations. See Fenton, Neil, and Lagnado (2013) as an example of the use of the idioms in the design of network structure. In the same vein of beginning with the building blocks of networks, Helsper and van der Gaag (2002) and Fenz (2012) proposed an approach for manual BN development based on ontology, in which the anticipated model's operational definitions and assumptions are clearly defined, and recommended collaboration between the knowledge engineering and domain expert(s) in the creation of the ontology. More basic approaches to determining the BN's structure involve identification of the types of variables within the network (background, problem, mediating, or symptom) and recognition of the each variable type's role in a causal network (Kjærulff & Madsen, 2008). Edwards (1999), Blodgett and Anderson (2000), Fenton et al., (2013), Laskey and Mahoney (2000), and Xu, Liao and Li (2008) provided examples of the diversity of approaches to manual construction of a BN.

In cases where data are available, it is possible to learn a BN structure and parameters from such data. A number of algorithms have been developed to accomplish structure learning, most of which are either score-based (BN iterations are scored based on data fit) or constraint-based (incorporates *a priori* understandings of independence among variables) (Margaritis, 2003). Examples of score-based techniques include Naïve Bayes' (Duda & Hart, 1973), evolutionary programming (Larrañaga, Karshenas, Bielza, & Santana, 2013), and Tree Augmented Naïve Bayes' (TAN) learning (Friedman, Geiger, & Goldszmidt, 1997), while constraint-based approaches include the PC and SGS algorithms (Spirtes et al., 2000), the inductive causation (IC) algorithm (Pearl, 2000), and Necessary Path Condition (NPC) (Steck & Tresp, 1999). Algorithms such as Maximum likelihood (ML), the Expectation-Maximization (EM) algorithm (Lauritzen, 1995), and Active Learning (Tong & Koller, 2002) are used for parameter/CPT estimation in BNs. Kjærulff and Madsen (2008) and Neapolitan (2004) presented an in-depth discussion of the steps involved in structure and parameter learning from data, and Aitkenhead and Aalders (2009) and Cui et al. (2006) provided real-world examples of purely data-driven learning of BN structure and parameters.

Note that structure learning from data has been criticized for resulting in over-fitting of data (Clark, 2003), and difficulties in finding and training domain experts for manual construction are not uncommon. A hybrid approach to constructing BNs in which both expert guidance and data are incorporated in the determination of structure and parameterization is a potential solution. For example, Heckerman et al. (1995) proposed that an expert-generated BN can be subsequently updated and improved upon by observed data. Other hybrid approaches involve employing expert knowledge in identifying the presence and type of relationships between variables and subsequently introducing related constraints in the structure and parameter learning process – see de Campos and Castellano (2007), Flores, Nicholson, Bruskill, Korb, and Mascaro (2011), and Niculescu, Mitchell, and Rao (2006) as examples of use of *a priori* expert knowledge in combination with data to construct a BN. Masegosa and Moral (2013) proposed incorporating expert knowledge after the structure is learned from data, providing guidance on any questionable links identified in the learned structure. Woodberry, Nicholson, Korb, and Pollino (2005) developed a technique for combining elicited expert knowledge and data to parameterize a model. In their discussion of the general development of statistical models, Kjærulff and Madsen (2008) noted the importance of a shift from developing a model designed to replicate the human decision process to developing models to support the human decision process. The recognition of the value of human expertise in model designing addressed by Kjærulff and Madsen is reflected in the popularity of using a hybrid process to construct a BN.

Chapter two of this manuscript proposes such a hybrid approach to model development, in which expert opinion is combined with statistical data to predict retention. This proposed methodology must be considered within the context of over sixty years of discussion of the clinical-statistical controversy, or the argument regarding the inferiority/superiority of clinical/statistical prediction. Meehl's seminal *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (1954/1996) articulated the distinction between statistical/actuarial/formal prediction and clinical/informal/impressionistic prediction, where the "clinical" method of prediction involves an expert human judge relying on informal decision-making procedures (experience, intuition, etc.) and the "statistical" method of prediction involves some formal decision-making rules or formula (statistical regression, actuarial tables, etc.) in order to classify or predict. Noting that predictions using these two methods often differ, Meehl further proposed that statistical prediction is generally at least as or more accurate and less costly than clinical prediction, rendering it the preferable method of prediction. A number of authors further explored this clinical vs. statistical issue (Dawes, 1988; Dawes, Faust, & Meehl, 1989, 1993; Faust, 1991; Goldberg, 1991; Grove & Meehl, 1996; Holt, 1986; Kleinmuntz, 1990; Marchese, 1992; Meehl, 1956, 1957, 1967, 1973, 1986; Sarbin, 1986), and consistently affirmed the superiority of statistical prediction over clinical prediction. Extensive meta-analyses performed by Grove, Zald, Lebow, Snitz, and Nelson (2000) and Ægisdóttir et al. (2006), further confirmed this idea. Both meta-analyses incorporated numerous studies that included comparisons of clinical and statistical prediction of some type of human behavior or medical/psychological diagnosis, and both found statistical prediction to be superior to clinical in most cases. Such conclusions uphold Meehl's original thesis and argue that, when available and possible, statistical prediction should be favored over clinical prediction – not only due to higher likelihood of predictive accuracy, but also because it is generally less costly.

However, Meehl, Grove, and other authors do not completely discount the value of clinical prediction. For example, in the large meta-analysis by Grove et al. (2000), the authors pointed out that clinical prediction was found to be on par with statistical prediction in half of the included studies. The authors attributed inferiority of clinical prediction to commonly cited heuristics, bias, and lack of feedback (see Dawes, Faust, and Meehl, 1986; Garb, 1989; Hogarth, 1987, Kadane and Wlofson, 1998; Tversky and Kahneman, 1974), leaving open the possibility that attempts to alleviate these causes of error could result in improved, and even superior, clinical prediction. Additional authors have addressed the role of clinical or expert judgment within context of its assumed inferiority to statistical prediction. Meehl (1954/1996) recognized “special powers of the clinician,” (p.24) particularly in noting clinicians’ ability to identify or recognize special or unique circumstances that may render statistical predictions inaccurate on a case-by-case basis, but not in a more general, widespread basis. Dana and Thomas (2006) further examined the role of clinician in prediction, presenting examples in which more accurate prediction results from the incorporation of clinician-identified influential factors into statistical predictive models rather than holistic clinician judgment. The authors suggested that a clinician’s expertise and valuable experience and input is best realized through “use of a formal, explicit procedure” (p. 425) – in other words, formal, systematic modeling of clinician decision-making incorporates the best of both worlds in the clinical-statistical controversy. As Dana and Thomas wrote, “Hopefully, the clinical-statistical controversy can move beyond whether we can deny or replace the talents of human judges to determining how we can use the special knowledge of human judges in a more rigorous manner” (p. 425). This study in Chapter Two proposes to do just that – formally and rigorously include the opinions and unique experience of undergraduate student retention experts into a statistical predictive model.

Model evaluation. Assessment of a BN’s functionality is critical regardless of whether the model was built for description, classification, or prediction. BNs designed using structure or parameter learning are often evaluated using measures of fit. The Minimum Description-Length (MDL) metric, described by Lam and Bacchus (1994) and Rissanen (1978), is commonly employed as a measure of fit among learned-structure BNs as it reflects model simplicity against model accuracy. The Bayesian Information Criterion (BIC) is another metric used to evaluate model fit that also considers model parsimony (Kass & Wasserman, 1995).

It is generally recommended that a final BN reflect parsimony in the number of nodes and the size of the CPTs for each node (Chen & Pollino, 2012; Adkison, 2009). Model parsimony is accomplished when there is a balance between the number of nodes and model sensitivity – fewer parent modes simplify the CPTs, but fewer nodes also may result in omitted information. An approach to simplifying models recommended by Olesen et al. (1989) and Neil et al. (2000) is "parent divorcing," in which related parent nodes are combined into one node that effectively describes the influence of the individual nodes thus reducing the number of nodes and subsequent complexity of the associated CPTs. Heckerman & Breese (1994) proposed temporal transformation as an approach to model simplification in which a temporal element is introduced in the analysis and representation of causal relationships, and Kjærulff and Madsen (2008) recommended a technique involving the creation of a new variable that represents structural and functional uncertainty in a model.

Models built using expert knowledge require different approaches to model validation. The experts participating in the elicitation should be asked to provide opinions of the final network's accuracy, and if data are available against which to compare, a BN can be evaluated using measures of predictive accuracy, deviations from expected value, and the extent to which predictions are calibrated (information reward) (Korb & Nicholson, 2010). Pollino, Woodberry, Nicholson, Korb, and Hart (2006) recommended evaluating BNs through sensitivity analyses, in which the magnitude of the effects of changes in a network's structure or parameters are measured. Pitchforth and Mengersen (2013) proposed that, as a BN's performance is a function of the both its structure and parameters, each of these dimensions should be evaluated for validity separately in addition to the performance of the model as a whole. Pitchforth and Mengersen further recommended using psychometric approaches to validity such as those described in Zumbo and Chan (2014) in evaluating an expert-elicited BN.

As Chapter Two of this manuscript will detail the creation and use of a predictive BN, particular attention is paid to the evaluation of a BN's predictive validity. Cowell, Dawid, and Spiegelhalter (1993) proposed that, in addition to overall fit, the predictive accuracy for individual nodes (node fit) and the quality of the modeled relationship between parent and child node (edge fit) should be evaluated. When data are available, comparing predicted values with actual values is the most straightforward way to discern a model's predictive accuracy. In a BN with learned structure, the original data set is typically divided into subsets, where one is used for training/calibration and one is used for testing. In addition to comparing a BN's predictions with observed data and previously collected data, Pollino et al. (2006) called upon experts to review and evaluate a proposed BN. Lalande, Bourguignon, Carlier, and Ducher (2013) evaluated prediction accuracy using Receiver Operating Characteristics curves (ROC) that compare sensitivity against specificity, while Gutierrez, Plant, and Theiler (2011) used modified confusion tables to identify thresholds of acceptable risk of error prediction. In addition to these examples found in the literature, Marcot (2012) provided a review of metrics related to predictive accuracy, including error rates and confusion tables, ROC curves, k-Fold cross-validation, Schwarz' BIC, the true skill statistic, and Cohen's kappa.

It was earlier noted that one of the major benefits of the Bayesian approach, particularly when subjective probabilities and judgments are involved, is the increased transparency and discussion generated. This benefit can be extended into the design and evaluation of BNs. For example, Jakeman, Letcher, & Norton (2006) recommended a holistic evaluation of the model including thorough discussion of the development of the model, the model's sensitivity to changes in structure, parameters, or assumptions, and whether or not the model is actually useful in applied settings. As these authors wrote, "….model accuracy (the traditional modeller's [sic] criterion) is only one of the criteria important in real applications" (p. 612). In conclusion, by evaluating a model's validity, a researcher is not only ensuring that the model actually describes the system of interest but is also perpetuating an ongoing, iterative process of critiquing and improving the model.

Bayesian Networks in Educational and Institutional Research

Chapter Two of this manuscript will describe the development and use of a BN modeling and predicting the first-year to second-year retention of undergraduate students. As this study includes elements of educational research used for an institutional research objective, the following section reviews examples of BNs used in educational and institutional research throughout the literature.

In their discussion of the development of a dynamic tutoring system powered by a BN, Conati, Gertner, and VanLehn (2002) wrote that BNs offer a "unifying framework to manage the uncertainty in student modeling" (p. 372). BNs can be found in a broad array of educational research topics, including psychometrics and item response modeling (Albert, 1992; Desmarais & Pu, 2006; Mislevy, Almond, Yan, & Steinberg, 1999), Evidence-Centered Design assessment (Almond, Mislevy, Steinberg, Yan, & Williamson, 2015; Shute, Hansen, & Almond, 2008), educational psychology (Nussbaum, 2011), and most commonly, in the design and evaluation of intelligent tutoring systems (Bunt & Conati, 2003; Ley, Kump & Albert, 2010; VanLehn, 2008). Xu and Ishitani (2008) and Heckerman (1997) employed Bayesian networks in exploratory modeling of institutional data, and Xu (2012) used Bayesian networks to produce models of female faculty professional experiences.

Of particular interest to institutional researchers is the use of BNs for prediction. Meyer and Xu (2007) developed a BN predicting faculty technology use, and Bekele and Menzel (2005) developed a BN correctly predicting performance in high school math nearly two-thirds of the time. Käser et al. (2013) used a BN predicting students' math knowledge to inform a computerized tutoring system, and Galbraith, Merrill, and Kline (2010) explored the predictive relationship between student evaluations and learning outcomes in college business courses. Sharabiani, Karim, Sharabiani, Atanasov, & Darabi (2014) predicted the end of course grades for students in engineering courses using a BN, and Torabi, Moradi, and Khantaimoori (2012) experimented with a variety of algorithms to build a BN that predicted student performance given teacher attributes. Kotsiantis, Pierrakas, and Pintelas (2004) and Kotsiantis, Patriacheas, and Xenos (2010) explored the capabilities of BNs in predicting performance in distance education courses, ultimately determining that pairing the BN-predicted results with other classification approaches yielded the most accurate results. Lykourentzou, Giannoukos, Nikolopoulos, Mpardis, and Loumos (2009) also addressed distance education, using a BN to model likelihood of course attrition.

A number of authors compared the predictive performance of educational- and institutional research-related BNs to models developed with other techniques and report mixed results: Bukralia, Deokar, Sarnikar, and Hawkes (2012) used the Naïve Bayes' classifier to develop a BN that predicted attrition in online classes less accurately than other methods like artificial neural networks and decision trees. Yukselturk, Ozekes, and Tűrel (2014) reported similar findings when predicting student dropout from an online program. Osmanbegović and Suljić (2012) found that BN outperformed decision trees and neural networks in predicting student success in economics courses, while Taruna and Pandey (2014) reported opposite results for students in engineering courses. In their comparison of BNs and decision trees in predicting general academic performance in terms of GPA, Nghe, Janecek, and Haddawy (2007) found decision trees to be more accurate predictors.

Recall that that one of the two main techniques in designing the structure and assigning the parameters to a BN involves learning from data. Machine learning and data-mining are currently significant ideas in educational and institutional research: Peña-Ayala (2014), Romero and Ventura (2007), and Suhirman, Zain, and Herawan (2014), provided summaries and reviews of recent educational research incorporating data-mining, with Peña-Ayala noting the popularity of BNs as frameworks for educational data-mining. Institutional researchers cited machine-learned BNs as tools for identifying previously unrecognized predictive variables (Antons & Maltz, 2006; Fernandez, Morales, Rodriguez, & Salmerón, 2011; Lykourentzou et al., 2009), and Heckerman (1997) presented a case study in which historical data concerning student demographics and college choices were used to build a BN depicting the causal influences on college plans as a tutorial on the role of data-mining in BNs.

Examples of purely manual construction of BN are rare in the educational and institutional research literature. In their development of a BN used to model learning progressions, West et al. (2010) employed experts' input in addition to other techniques (latent class analysis) into the design of BN's structure, but the authors do not comment on the methodology for specifying the model's parameters. Almond, Mislevy, Steinberg, Yan, & Williamson (2015) described the process of eliciting BN structure and parameters, while noting that models should be updated as data become available. More often, educational and institutional researchers employ a hybrid process involving learning from data as well as expert input. van Duijnhoven (2003) used a hybrid methodology in which expert knowledge and machine learning were applied in the development of a BN modeling student knowledge, and subsequently confirmed the expert-elicited structure and parameters against data-generated models. García, Amandi, Schiaffino, and Campo (2007) designed a BN for identifying learning styles using theory to manually specify the BN’s structure and a combination of Felder’s (1988) learning style definitions and data to assign conditional probabilities. Similarly, Wang and Beck (2013) used a previously developed student skill model to design the structure of a BN, while using data to parameterize the model.

In a discussion of the use of machine learning to develop a model used to predict retention, Delen (2010) highlighted how suited machine learning is for institution-specific settings – an institution’s issues are unique to its population and environment and mining historical data can provide patterns unique to the institution. However, Delen also pointed out that data mining and theory-driven research can be used in tandem to identify important variables and any relationships among them. Although the author did not use a Bayesian approach in modeling, this idea – the idea of leaving no data behind – echoes one of the most influential arguments for a Bayesian approach. Extending this idea to the use of BNs in institutional research, capitalizing on accepted theory and expert input as well as accessibility to historical data seems an ideal approach to developing BNs to address institutional issues.

In conclusion, the incorporation of BNs into educational and institutional research is an approach gaining in popularity and application. While BNs related to intelligent tutoring systems are most popular, they are also found in psychometrics, educational assessment, data modeling, and, most relevant to this study, prediction of outcomes. The literature reveals a number of approaches, although models developed through the use of data mining and machine learning are most common within institutional research. BNs offer an excellent approach to dealing with the uncertainty inherent in educational research and are particularly suited to the narrower field of institutional research, efficiently handling the type of research questions, data, and audiences addressed in higher education. di Pietro, Mugion, Musella, Renzi, & Vicard (2015) advocated for the use BNs in the modeling complexities of higher education, noting that BNs represent a “holistic”, global approach to answering common institutional research questions. In addressing the uncertainty inherent in systems of social science, as well as the complexity and unique nature of institutions, the compatibility of Bayesian approaches with institutional research is clear. A BN handles the uncertainty in student-related data while also offering an intuitive, accessible modeling capability that supports the decision-making and policy-setting processes.

References

Adkison, M.D. (2009). Drawbacks of complex models in frequentist and Bayesian approaches to natural resource management. *Ecological Applications, 19*(1), 198-205.

Ægisdóttir, S., White, M., Spengler, P., Maughermaan, A., Anderson, L., Cook, R., …Rush, J. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341-382.

Aitkenhead, M.J. & Aalders, I.H. (2009). Predicting land cover using GIS, Bayesian and evolutionary algorithm methods. *Journal of Environmental Management, 90,* 236-250.

Akaike, H. (1992). Information theory and an extension of the maximum likelihood principle. In S. Kotz and N.L. Johnson (Eds.), *Breakthrough in statistics* (pp. 193-213). London: Springer-Verlag.

Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K. & Rousseau, J. (2012). *Combining expert opinions in prior elicitation*, *7*(3), 503-532.

Albert, J.H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics, 17,* 251-269.

Almond, R.G., Mislevy, R.J., Steinberg, L.S., Yan, D., & Williamson, D.M. (2015). *Bayesian Networks in Educational Assessment.* New York: Springer.

American Psychological Association. (2013). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Andrews, M. & Baguley, M. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology, 66,* 1-7.

Antons, C.M. & Maltz, E.N. (2006). Expanding the role of institutional research at small private universities: A case study in enrollment management using data mining. *New Directions for Institutional Research, 131,* 69-81.

Aussem, A., de Morais, S.R., & Corbex, M. (2012). Analysis of nasopharyngeal carcinoma risk factors with Bayesian networks. *Artificial Intelligence in Medicine, 54*(1), 53-62.

Axelrad, E.T., Sticha, P.J., Brdiczka, O., & Shen, J. (2013). A Bayesian network model for predicting insider threats. *Proceedings of the 2nd IEEE Security and Privacy Workshops,* 82-89.

Baesens, B., Viaene, S., van den Poel, D., Vanthienen, J. & Dedene, G. (2002). Bayesian neural network learning for repeat purchase modelling in direct marketing. *European Journal of Operations Research*, *138*(1), 191-211.

Bajari, P., & Ye, L. (2003). Deciding between competition and collusion. *The Review of Economics and Statistics*, *85*(4), 971-989.

Bayes, T. (1763). An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of Royal Society of London, 53,* 370-418.

Beach, L. R. & Swenson, R. G. (1966). Intuitive estimation of means. *Psychonomic Science, 5,* 161-162.

Beach, L.R. & Braun, G.P. (1994). Laboratory studies of subjective probability: A status report. In G. Wright and P. Ayton (Eds.). *Subjective Probability.* Chichester: John Wiley.

Bedrick, E.J., Christensen, R., & Johnson, W. (1997). Bayesian binomial regression: Predicting survival at a trauma center. *The American Statistician, 51,* 211-218.

Bekele, R. & Menzel, W. (2005). A Bayesian approach to predict performance of a student (BAPPS): A case with Ethiopian students. In M. Hamza (Ed.), *Proceedings from IASTD International Conference on Artificial Intelligence and Applications*, 189-194.

Berger, J.O (2006). The case for objective Bayesian analysis. *Bayesian Analysis, 1*(3), 385-402.

Berger, J.O. & Bernardo, J. (1992). On the development of the reference prior method. In J.M. Bernardo, J.O. Berger, D.V. Lindley and A.F.M. Smith (Eds.), *Bayesian statistics 4* (pp. 35-60). London: Oxford University Press.

Berger, J.O. & Berry, D.A. (1988). Statistical analysis ane illusion of objectivity. *American Scientist, 76*, 156-165.

Berger, J.O. & Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of *p*-values and evidence (with discussion). *Journal of the American Statistical Association, 82,* 112-139.

Berger, J.O. & Sun, D. (2008). Objective priors for the bivariate normal model. *Annals of Statistics, 36*(2), 963-982.

Berger, J.O. (1994). An overview of robust Bayesian analysis (with discussion). *Test, 3,* 5-124.

Berk, R.A., Western, B. & Weiss, R.E. (1995). Statistical inference for apparent populations. *Sociological Methodology, 25,* 421-458.

Bernardo, J. & Smith, A. (2000). *Bayesian theory*. West Sussex, England: John Wiley & Sons.

Bernardo, J. (2005). Reference analysis. In D.K. Dey and C.R. Rao (Eds.), *Handbook of statistics: Bayesian thinking, modeling and computation* (pp. 17-90). Amsterdam: Elsevier.

Berner, E.S. (2006). *Clinical decision support systems: Theory and practice (health informatics).* New York: Springer.

Blodgett, J.G. & Anderson, R.D. (2000). A Bayesian network model of the consumer complaint process. *Journal of Service Research*, *2* (4), 321-338.

Bogdan, R. C., & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theory and methods* (5th ed.). Boston, MA: Pearson Education.

Bolstad, W. M. (2007). *Introduction to Bayesian statistics* (2nd ed.). Hoboken, NJ: Wiley.

Bonjour, L. (1985). *The structure of empirical knowledge.* Cambridge: Harvard University Press.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein H.R. (2009). *Introduction to meta-analysis.* West Sussex, England: John Wiley & Sons.

Bornstein, B. H. & Zickafoose, D. J. (1999). "I know I do it, I know I saw it": the stability of the confidence-accuracy relationship across domains. *Journal of Experimental Psychology: Applied, 5,* 76-88.

Borsboom, D. & Haig, B.D. (2013). How to practice Bayesian statistics outside the Bayesian church: What philosophy for Bayesian statistical modeling? *British Journal of Mathematical and Statistical Psychology, 66,* 39-44.

Bosworth, K., Gingiss, P.M., Porthoff, S. & Roberts-Gray, C. (1999). A Bayesian model to predict the success of the implementation of education innovations in school-centered pograms. *Evaluation and Program Planning, 22*(1), 1-11.

Bovens, L. & Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Clarendon Press.

Box, G. & Tiao, G. (1973). *Bayesian inference in statistical analysis.* Reading, Mass.: Addison-Wesley.

Buckley, J. (2004). Simple Bayesian inference for qualitative political research. *Political Analysis*, *12*, 386-399.

Bukralia, R., Deokar, A.V., Sarnikar, S., & Hawkes, M. (2012). Using machine learning techniques in student dropout prediction. In H. Burley (Ed.), *Cases on Institutional Research Systems* (pp. 117-131). Hershey, PA: IGI Global.

Bunt, A. & Conatic, C. (2003). Probabilistic student modelling to improve exploratory behaviour. *User Modeling and User-Adapted Interaction, 13*(3), 269-309.

Burgman, M.A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., et al. (2011). Expert states and performance. *PLoS ONE, 6*(7), 1-7.

Burnham, K. & Anderson, D. (2002). *Model selection and inferences: A practical information theoretic approach* (2nd ed.). New York: Springer-Verlag.

Caley, M.J., O'Leary, R.A., Fisher, R., Low-Choy, S., Johnson, S. & Mengersen, K. (2013). What is an expert? A systems perspective on expertise. *Ecology and Evolution, 4*(3), 231-242.

Cano, R., Sordo, C. & Gutiérrez, J.M. (2004). Applications of Bayesian networks is meteorology. In J.A. Gámez, S. Moral and A. Salmerón Cerdan (Eds.), *Advances in Bayesian networks.* New York: Springer.

Carlson, B.W. (1993). The accuracy of future forecasts and past judgments. *Organizational Behavior and Human Decision Processes, 54,* 245-276.

Carver, R.P. (1978). The case against significance testing. *Harvard Educational Review, 48,* 378-399.

Casella, G. & Berger, R.L. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *Journal of the American Statistical Association, 82,* 106-111.

Chaloner, K. & Duncan, G.T. (1987). Some properties of the Dirichlet-multinomial distribution and its use in prior elicitation. *Communications in Statistics-Theory and Methods, 16,* 511-523.

Charniak, E. (1991). Bayesian networks without tears. *AI Magazine, 12*(4), 50-63.

Chen, S.H. & Pollino, C.A. (2012). Good practice in Bayesian network modelling. *Environmental Modelling & Software*, *37*, 134-145.

Cho, I., Park, I., Kim, E., Lee, E., & Bates, D.W. (2013). Using HER data to predict hospital-acquired pressure ulcers: A prospective study of a Bayesian network model. *International Journal of Medical Informatics*, *82*(11), 1059-1067.

Choy, S. L., O'Leary, R., & Mengersen, K. (2009). Elicitation by design in ecology: Using expert opinion to inform priors for Bayesian statistical models. *Ecology*, *90*(1), 265-277.

Cifarelli, M. & Regazzini, E. (1996). de Finetti's contribution to probability and statistics. *Statistical Science, 11*(4), 253-282.

Clark, J.S. (2003). Uncertainty and variability in demography and population growth: A hierarchical approach. *Ecology, 84*(6), 1370-1381.

Clemen, R.T. & Reilly, T. (2001). *Making hard decisions with decision tools.* Pacific Grove, CA: Duxbury Press.

Clemen, R.T. & Winkler, R.L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis, 19*(2), 187-203.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist, 49*, 99-1003.

Conati, C., Gertner, A., & VanLehn, K. (2002). Using Bayesian networks to manage uncertainty in student modeling. *Journal of User Modeling and User-Adapted Interaction, 12*, 371–417.

Cooke, R.M. & Goossens, L.H.J. (2004). Expert judgement elicitation for risk assessments of critical infrastructures. *Journal of Risk Research, 7*(6), 643-656.

Cooke, R.M. (1991). *Experts in uncertainty: Opinion and subjective probability in science.* New York: Oxford University Press.

Cooper, G.F. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence, 42,* 393-405.

Cowell, R.G., Dawid, A.P., & Spiegelhalter, D.J. (1993). Sequential model criticism in probabilistic expert systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15,* 109-129.

Crout, N., Kokkonen, T., Jakeman, A. J., Norton, J. P., Newham, L. T., Anderson, R., ... Whitfield, P. (2008). Good modelling practice. Retrieved from Digital Commons @ University of Nebraska - Lincoln website: http://digitalcommons.unl.edu/usepapapers

Cui, G., Wong, M.L. & Lui, H-K. (2006). Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science, 52*(4), 597-612.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7-29.

Curlette, W. L. (2006). A framework for research studies: Mixed methods through combining Bayesian statistics and qualitative research in individual psychology. *Journal of Individual Psychology, 62*(3), 338-349.

D'Agostini, G. G. (2001). Role and meaning of subjective probability: Some comments on common misconceptions. *AIP Conference Proceedings, 568*(1), 23.

Dagum, P. & Luby, M. (1993). Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence, 60,* 141-153.

Dalkey, N.C. (1969). *The Delphi method: An experimental study of group opinions.* Report No. RM-5888-PR. The Rand Corporation.

Dana, J. & Thomas, R. (2006). In defense of clinical judgment…and mechanical prediction. *Journal of Behavioral Decision Making, 19,* 413-428.

Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks.* Cambridge: Cambridge University Press.

Datta, G.S. & Ghosh, M. (1996). On the invariance of noninformative priors. *The Annals of Statistics, 24*(1), 141-159.

Dawes, R.M. (1988). *Rational choice in an uncertain world.* Chicago, IL: Harcourt Brace Jovanovich.

Dawes, R.M., Faust, D., & Meehl, P.E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668-1674.

Dawes, R.M., Faust, D., & Meehl, P.E. (1993). Statistical prediction versus clinical prediction: Improving what works. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 351-367). Hillsdale, NJ: Erlbaum.

Dawid, A.P. (1982). The well-calibrated Bayesian. *Journal of the American Statistical Association, 77*(379), 605-610.

Dawid, A.P. (1992). Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and Computing, 2,* 25-36.

de Campos, J.G. & Castellano, J.G. (2007). Bayesian network learning algorithms using structural restrictions. *International Journal of Approximate Reasoning, 45*(2), 233-254.

de Finetti, B. (1964). La prévision: ses lois logiques, ses sources subjectives. (H.E. Kyburg, Trans.) In H.E. Kyburg & H.E. Smokler (Eds.), *Studies in subjective probability* (53-118). New York: Wiley. (Original work published 1937).

de Finetti, B. (1972). *Probability, induction, and statistics.* New York: Wiley.

de Finetti, B. (1973). *Theory of Probability, Volumes 1 and 2* as translated from Italian by A. Machi and A. Smith. New York: Wiley.

DeGroot, M.H.. (1970). *Optimal Statistical Decisions.* McGraw Hill.

Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). *Group Techniques for Program Planning.* Glenview, Illinois: Scott Foresman.

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems, 49,* 498-506.

Dennis, B. (1996). Discussion: Should ecologists become Bayesians? *Ecological Applications, 6*(4), 1095-1103.

Denzin, N.K. & Lincoln, Y.S. (Eds.) (2000). *Handbook of qualitative research*. Thousand Oaks, CA: SAGE Publications.

Desmarais, M.C. & Pu, X. (2006). A Bayesian student model without hidden nodes and its comparison with item response theory. *International Journal of Artificial Intelligence in Education*, *15*(4), 291-323.

di Pietro, L., Mugion, R.G., Musella, F., Renzi, M.F. & Vicard, P. (2015). Reconciling internal and external performance in a holistic approach: A Bayesian network model in higher education. *Expert Systems with Applications, 42,* 2691-2702.

Dixon, P. (2003). The *p*-value fallacy and how to avoid it. *Canadian Journal of Experimental Psychology, 57*, 189-202.

Dogan, I. (2012). Analysis of facility location using Bayesian Networks. *Expert Systems with Applications*, *39*, 1092-1104.

Donald, R., Howells, T., Piper, I., Citerio, G., Enblad, P., Gregson, B., et al. (2012). Early warning of ELISIG-defined hypotensive events using a Bayesian artificial neural network. *Acta Neurochirurgica Supplement, 114,* 39-44.

Druzdzel, M. & Simon, H. (1993). Causality in Bayesian belief networks. *Proceedings of ninth conference on uncertainty in artificial intelligence*, 3-11.

Duda, R.O. & Hart, P.E. (1973). *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons.

Edwards, W. (1998). Hailfinder: Tools for and experiences with Bayesian normative modeling. *American Psychologist*, *53*(4), 416-428.

Efron, B. (2003). *Bayesians, frequentists, and physicists*. Keynote address presented at PHYSTAT2003, SLAC. Stanford, CA.

Efron, B. (2013). A 250-year argument: Belief, behavior, and the bootstrap. *Bulletin of the American Mathematical Society, 50*(1), 129-146.

Eisner, E.W. & Peshkin, A. (Eds.), (1990). *Qualitative inquiry in education: The continuing debate.* New York: Teachers College Press.

Erlick, D. E. (1964). Absolute judgement of discrete quantities randomly distributed over time. *Journal of Experimental Psychology*, *67,* 475-482.

Evans, M. & Moshonov, H. (2006). Checking prior-data conflict. *Bayesian Analysis*, *1*(4), 893-914.

Faust, D. (1991). What if we had listened? Present reflections on altered pasts. In W.M. Grove & D. Cicchetti (Eds.), *Thinking clearly about psychology: Vol 1. Matters of public interest* (pp. 185-216). Minneapolis: University of Minnesota Press.

Felder, R. & Silverman, L. (1988). Learning and teaching styles. *Journal of Engineering Education, 78*(7), 674-681.

Fenton, N., Neil, M., & Lagnado, D.A. (2013). A general structure for legal arguments about evidence using Bayesian networks. *Cognitive Science, 37,* 61-102.

Fenton, N., Neil, M., & Marquez, D. (2008). Using Bayesian networks to predict software defects and reliability. *Journal of Risk and Reliability, 222,* 701-712.

Fenz, S. (2012). An otology-based approach for constructing Bayesian networks. *Data & Knowledge Engineering, 73,* 73-88.

Fernández, A., Morales, M., Rodríguez, C., & Salmerón, A. (2011). A system for relevance analysis of performance indicators in higher education using Bayesian networks. *Knowledge Information Systems, 27,* 327-344.

Fidler, F. & Loftus, G.R. (2009). Why figures with error bars should replace p values: Some conceptual arguments and empirical demonstrations. *Journal of Psychology, 217,* 27-37.

Finch, S., Cumming, G., Williams, J., Palmer, L., Griffith, E., Alders, C.,…Goodman, O. (2004). Reform of statistical inference in psychology: the case of memory and cognition. *Behavior Research Methods, Instruments and Computers, 36,* 312-324.

Fisher, R.A. (1925). Applications of "Student's" distribution. *Metron, 5,* 90-104.

Fisher, R.A. (1935). *The design of experiments.* New York: Hafner Press.

Fisher, R.A. (1956). *Statistical methods and scientific inference.* New York: Hafner Press.

Fishhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 422-444). Cambridge: Cambridge University Press.

Flores, M.J., Nicholson, A.E., Brunskill, A., Korb, K., & Mascaro, S. (2011). Incorporating expert knowledge when learning Bayesian network structure: A medical case study. *Artificial Intelligence in Medicine*, *53*(3), 181-204.

French, S. (1985). Group consensus probability distributions: A critical survey (with discussion). In J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (Eds.), *Bayesian Statistics 2* (pp. 183-197). Amsterdam: Elsevier.

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, *29*, 131-163.

Friedman, N., Linial, L., Nachman, I., & Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *Journal of Computational Biology, 7,* 601-620.

Gage, N. L. (1989). The paradigm wars and their aftermath: A "historical" sketch of research on teaching since 1989. *Educational Researcher*, *18,* 4-10.

Galbraith, C.S., Merrill, G.B., & Kline, D.M. (2010). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and Bayesian analysis. *Research in Higher Education*, *53,* 353-374.

Gao, D., Madden, M., Chambers, D., & Lyons, G. (2005). Bayesian ANN classifier for ECG arrhythmia diagnostic system: a comparison study. *Proceedings 2005 IEEE International Joint Conference on Neural Networks, Montreal, 4,* 2383 – 2388.

Gao, F. & Chen, L. (2005). Bayesian or non-Bayesian: A comparison study of item parameter estimation in the three-parameter logistic model. *Applied Measurement in Education, 18(*4), 351-380.

Garb, H.N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105,* 387-396.

García, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education, 49,* 794-808.

Garthwaite, P.H. & Dickey, J.M. (1992). Elicitation of prior distributions for variable selection problems in regression. *Annals of Statistics, 20,* 1697-1719.

Garthwaite, P.H., Al-Awadhi, S.A., Elfadaly, F.G., & Jenkinson, D.J. (2013). Prior distribution elicitation for generalized linear and piecewise-linear models. *Journal of Applied Statistics, 40(1),* 59-75.

Garthwaite, P.H., Kadane, J.B. & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association, 100,* 680-701.

Gelman, A. & Loken, E. (2014). The statistical crisis in science. *American Scientist, 102,* 460-465.

Gelman, A. & Shalizi, C.R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology, 66,* 8-38.

Gelman, A. (2008a). Teaching Bayes to graduate students in political science, sociology, public health, education, economics… *The American Statistician, 62*(3), 202-205.

Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2004). *Bayesian Data Analysis* (2nd ed.). Boca Raton, FL: CRC Press.

Gelman, A., Meng, X., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica, 6*, 733-808.

Genest, C. & Zidek, J.V. (1986). Combining probability distributions. A critique and annotated bibliography. *Statistical Science, 1,* 114-148.

Gershman, S. & Blei, D. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology, 56*, 1-12.

Ghosh, M. (2011). Objective priors: An introduction for frequentists. *Statistical Science, 26*(2), 187-202.

Gigerenzer, G. (1993). The superego, the ego and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp.311-339). Hillsdale, NJ: Erlbaum.

Gigerenzer, G. (1996). A narrow norms and vague heruistics: a reply to Kahneman and Tversky. *Psychological Review, 103*, 592-596.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics, 33,* 587-606.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506-528.

Gill, J. (2009). *Bayesian methods: A social and behavioral sciences approach* (2nd ed.). Boca Raton, FL: CRC Press.

Gill, J., & Walker, L. D. (2005). Elicited priors for Bayesian model specifications in political science research. *Journal of Politics, 67,* 841-872.

Gillies, D. (2000). *Philosophical theories of probability*. London, England: Routledge.

Gneiting, T. & Raftery, A. E. (2007). Strictly proper scoring, rules, prediction and estimation. *Journal of American Statistical Association, 102*, 359-378.

Goldberg, L. R. (1991). Human mind versus regression equation: Five contrasts. In W. M. Grove & D. Cicchetti (Eds.), *Thinking clearly about psychology: Vol. 1. Matters of public interest (pp.* 173-184). Minneapolis: University of Minnesota Press.

Goldstein, M. (2004). Bayes linear analysis. In S. Kotz et al. (Eds.), *Encyclopedia of Statistical Sciences, update volume 3* (pp. 29-34). New York: Wiley.

Goodman, S. N. (1999a). Toward evidence-based medical statistics. 1: The *p* value fallacy. *Annals of Internal Medicine*, *130*, 995-1004.

Goodman, S.N. (1999b). Toward evidence-based medical statistics. 2: The Bayes Factor. *Annals of Internal Medicine*, *130*, 1005-1013).

Gorard, S. (2014). Open dialogue: The widespread abuse of statistics by researchers: What is the problem and what is the ethical way forward? *Psychology of Education Review, 38*(1), 3-32.

Greenes, R.A. (Ed.). (2014). *Clinical decision support: The road to broad adoption* (2nd ed.). London: Elsevier.

Grove, W.M. & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2*(2), 293-323.

Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19-30.

Guba, E. G. (Ed.). (1990). *The paradigm dialog*. Newbury Park, CA: Sage.

Gustafson, D. H., Sainfort, F., Eichler, M., Adams, L., Bisognano, M., & Steudel, H. (2003). Developing and testing a model to predict outcomes of organizational change. *Health Services Research, 38*, 751-776.

Hahn, E. (2006). Re-examining informative prior elicitation through the lens of Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series A (Statistics in Society), 169*(1), 37-48.

Hammitt, J.K. & Zhange, Y. (2013). Combining experts' judgments: Comparison of algorithmic methods using synthetic data. *Risk Analysis, 33*(1), 109-120.

Harlow, L.L., Mulaik, S.A. & Steiger, J.H. (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Hausman, D.M. & Woodward, J. (1999). Independence, invariance and casual Markov condition. *British Journal of Philosophical Sciences, 50,* 521-583.

Heckerman, D. & Breese, J.S. (1994). A new look at causal independence. *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence,* 286-292.

Heckerman, D. & Shachter, R. (1995). A definition and graphical representation of causality. *Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence,* 262-273.

Heckerman, D. (1997). Bayesian networks for data mining. *Data Mining and Knowledge Discovery, 1*, 79-119.

Heckerman, D., Geiger, D., & Chickering, D. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning, 20,* 197-243.

Helsper, E.M. & van der Gaag, L.C. (2002). A case study in ontologies for probabilistic networks. In M. Bramer, F. Coenen and A. Preece (Eds.), *Research and Development in Intelligent Systems XVIII* (pp. 229-242). London: Springer-Verlag.

Hennig, C. (2009). A constructivist view of the statistical quantification of evidence. *Constructivist Foundations*, *5*(1), 39-54.

Henrion, M., Breese, J.S.., & Horvitz, E.J. (1991). Decision analysis and expert systems. *AI Magazine, 12*(4), 64-91.

Higgins, H.M., Huxley, J.N., Wapenaar, W., & Green, M.J. (2014). Quantifying veterinarians' beliefs on disease control and exploring the effect of new evidence: a Bayesian approach. *Journal of Dairy Science*, *97*(6), 3394-3408.

Hoekstra, R., Finch, S., Kiers, H.A., & Johnson, A. (2012). Confidence intervals make a difference: Effects of showing intervals on inferential reasoning. *Educational and Psychological Measurement, 72,* 1039-1052.

Hoekstra, R., Morey, R.D., Rouder, J.N., Wagenmakers, E. (2014). *Robust misinterpretation of confidence intervals, 21*(5), 1157-1164.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, *70*, 271-294.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, *70*, 271-294.

Hogarth, R. M. (1987). *Judgement and Choice* (2nd ed.). Chichester: John Wiley.

Holt, R.R. (1986). Clinical and statistical prediction: A retrospective and would-be integrative perspective. *Journal of Personality Assessment, 50,* 376-386.

Horvitz, E. & Barry, M. (1995). Display of information for time-critical decision making. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 296-305.

Howard, R. & Matheson, J. (1984). Influence diagrams. In R. Howard and J. Matheson (Eds.) *Readings on the Principles and Applications of Decision Analysis* (pp. 721-762). Meade Park, CA: Strategic Decision Group.

Hubbard, R. & Lindsay, R.M. (2008). Why *p* values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology, 18,* 69-88.

Ibrahim, J. & Chen, M. (2000). Power prior distributions for regression models. *Statistical Science, 15*(1), 46-60.

Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), 0696-0701.

Jakeman, A.J., Letcher, R.A. & Norton, J.P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software, 21,* 602-614.

James, A., Choy, S.L. & Mengersen, K. (2010). Elicitator: An expert elicitation tool for regression in ecology. *Environmental Modelling and Software, 25(1),* 129-145.

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., et al. (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science, 302,* 449-453.

Jaynes, E. (1980). Marginalization and prior probabilities. In A. Zellner (Ed.), *Bayesian analysis in econometrics and statistics* (pp. 43-78). New York: Elsevier.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Clarendon Press.

Jenkinson, D. (2005). The elicitation of probabilities: A review of the statistical literature. *Department of Probability and Statistics, University of Sheffield*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.6173&rep=rep1&type=pdf

Jensen, F. & Nielsen, T.D. (2007). *Bayesian networks and decision graphs* (2nd ed.). New York: Springer.

Jensen, F., Lauritzen, S. & Olesen, K. (1990). Bayesian updating in recursive graphical models by local computations. *Computational Statistics Quarterly, 4,* 269-282.

Jiang, X., Xue, D., Brufsky, A., Khan, S., & Neapolitan, R. (2014). A new method for predicting patient survivorship using efficient Bayesian network learning. *Cancer Informatics, 13,* 47-57.

Johansson, T. (2011). Hail the impossible: *p*-values, evidence, and likelihood. *Scandinavian Journal of Psychology, 52*, 113-125.

Johnson, S. R., Tomlinson, G. A., Hawker, G. A., Granton, J. T., & Feldman, B. M. (2010). Methods to elicit beliefs for Bayesian priors: A systematic review. *Journal of Clinical Epidemiology, 63,* 355-369.

Johnson, T. (2013). Item response modeling with sum scores. *Applied Psychological Measurement, 37*(8), 638-652.

Jonsson, A.C. & Allwood, C.M. (2003). Stability and variability in the realism of confidence judgements over time, content domain and gender. *Personality and Individual Differences, 34,* 559-574.

Jorgensen, M., Teigen, K.H. & Molokken, K. (2004). Better sure than safe? Over-confidence in judgement based software development effort prediction intervals. *The Journal of Systems and Software*, *70,* 79-93.

Kadane, J. B., & Wolfson, L. J. (1998). Experiences in elicitation. *The Statistician*, *47*, 3-19.

Kaplan, S. (1992). 'Expert information' versus 'expert opinion.' Another approach to the problem of eliciting/combining/using expert knowledge in PRA. *Reliability Engineering and System Safety, 35,* 61-72.

Käser, T., Busetto, G., Solenthaler, B., Baschera, G., Kohn, J., Kurcian, K….Gross, M. (2013). Modelling and optimizing mathematics learning in children. *International Journal of Artificial Intelligence in Education, 13,* 115-135.

Kass, R.E. & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association, 91*(435), 1343-1370.

Kass, R.E. & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Kass, R.E. & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association, 90,* 928-934.

Kelly, D. & Smith, C. (2011). *Bayesian Inference for Probabilistic Risk Assessment.* London: Springer.

Kim, J.H. & Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference systems. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence,* 190-193.

Kjærulff, U.B. & Madsen, A.L. (2008). *Bayesian networks and influence diagrams: A guide to construction and analysis.* New York: Springer.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin, 107,* 296-310.

Kline, R.B. (2004). *Beyond significance testing: reforming data analysis methods in behavioral research.* Washington, DC: American Psychological Association.

Koller, D. & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques.* Cambridge, MA: The MIT Press.

Korb, K. & Nicholson, A.E. (2010). *Bayesian artificial intelligence* (2nd ed.). London: Chapman & Hall.

Kotsiantis, S., Patriarcheas, K., & Xenos, M. (2010). A combinatorial incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems, 23,* 529-535.

Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2004). Predicting students' performance in distance learning using machine learning techniques. *Applied Artificial Intelligence, 18,* 411-426.

Kruschke, J.K. (2013). Posterior predictive checks can and should be Bayesian: Comment on Gelman and Shalizi 'Philosophy and the practice of Bayesian statistics.' *British Journal of Mathematical and Statistical Psychology, 66,* 45-56.

Kuhnert, P. M., Martin, T. G., Mengersen, K., & Possingham, H. P. (2005). Assessing the impacts of grazing levels on bird density in woodland habitat: A Bayesian approach using expert opinion. *Environmetrics, 16,* 717-747.

Kuhnert, P.M., Martin, T.G. & Griffiths, S.P. (2010). A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters, 13,* 900-914.

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research in interviewing* (2nd ed.). Thousand Oaks, CA: SAGE.

Kyburg, H.E. & Smokler, H.E. (Eds.). (1964). *Studies in subjective probability*. New York: Wiley.

Kynn, M. (2008). The ’heuristics and biases’ bias in expert elicitation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*(1), 239-264.

Lagerström, R., Johnson, P., Höök, D., & König, J. (2009). *Software change project cost estimation - a Bayesian network and a method for expert elicitation*. Paper presented at the Third International Workshop on Software Quality and Maintainability, Kaiserslautern, Germany. Retrieved from http://www.ee.kth.se/php/modules/publications/reports/2009/IR-EE-ICS_2009_015.pdf

Lalande, L., Bourguignon, L., Carlier, C. & Ducher, M. (2013). Bayesian networks: A new method for the modeling of bibliographic knowledge. *Medical and Biological Engineering & Computing, 51*, 657-664.

Lam, W. & Bacchus, F. (1994). Learning Bayesian belief networks – An approach based on the MDL principle. *Computational Intelligence, 10*(3), 269-293.

Landuyt, D., Broekx, S., D’hondt, R., Engelen, G., Aertsens, J., & Goethals, P.L.M. (2013). A review of Bayesian belief networks in ecosystem service modelling. *Environmental Modelling & Software, 46*, 1-11.

Laplace, P.S. (1902). A *philosophical essay on probabilities.* (F.W. Truscott & F.L. Emory, Trans.). London: John Wiley & Sons. (Original work published 1825)

Larrañaga, P., Karshenas, H., Bielza, C., & Santana, R. (2013). A review on evolutionary algorithms in Bayesian network learning and inference tasks. *Information Sciences, 233*, 109-125.

Laru, J., Naykki, P., & Jarvela, S. (2012). Supporting small-group learning using multiple web 2.0 tools: A case study in the higher education context. *Internet and Higher Education, 15*(1), 29-38.

Laskey, K.B. & Mahoney, S.M. (2000). Network engineering for agile belief network models. *IEEE Transactions on Knowledge and Data Engineering, 12*(4), 487-498.

Lau, A.H. & Leong, T.Y. (1999). PROBES: A framework for probability elicitation from experts. *Journal of the American Medical Informatics Association, 1999 Supplement*, 301-305.

Lauritzen, S.L. & Spiegelhalter, D.J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B, 50*(2), 157-224.

Lauritzen, S.L. (1995). The EM algorithm for graphical association models with missing data. *Computational Statistics & Analysis, 19,* 191-201.

Lenk, P. & Orme, B. (2009). The value of informative priors in Bayesian inference with sparse data. *Journal of Marketing Research, 46*(6), 832-845.

Leu, S. & Chang, C. (2013). Bayesian-network-based safety risk assessment for steel construction projects. *Accident Analysis and Prevention, 54,* 122-13.

Lewis, C.I. (1946). *An analysis of knowledge and valuation.* LaSalle, IL: Open Court Press.

Ley, T., Kump, B., & Albert, D. (2010). A methodology for eliciting, modelling and evaluating expert knowledge for an adaptive work-integrated learning system. *International Journal of Human-Computer Studies, 68*(4), 185-208.

Li, B., Han, T. & Kang, F. (2013). Fault diagnosis expert system of semiconductor manufacturing equipment using a Bayesian network. *International Journal of Computer Integrated Manufacturing, 26*(12), 1161-1171.

Li, Y., Zeng, T. & Yu, J. A new approach to Bayesian hypothesis testing. *Journal of Econometrics, 178*(1), 602-12.

Lichtenstein, S., Fischhoff, B., & Phillips, L.D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 306-335). Cambridge, UK: Cambridge University Press.

Lindley, D.V. (1965). *Introduction to probability and statistics from a Bayesian viewpoint.* Cambridge: University Press.

Loeb, J.W. (2003). Hierarchical linear modeling in salary-equity studies. *New Directions for Institutional Research, 117,* 69-96.

Luan, J. & Zhao, C. (2006). Practicing data mining for enrollment management and beyond. *New Directions for Institutional Research, 131*, 1-5.

Lucas, P.J.F., van der Gaag, L.C. & Abu-Hanna, A. (2004). Bayesian networks in biomedicine and health-care. *Artificial Intelligence in Medicine, 30*(3), 201-214.

Luce, R.D. & Suppes, P. (1965). Preference, utility, and subjective probability. In R.D. Luce, R.R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical psychology, III* (pp. 249-410). New York: Wiley.

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education, 53*, 950-965.

Marchese, M.C. (1992). Clinical versus actuarial prediction: A review of the literature. *Perceptual and motor skills, 75,* 583-594.

Marcot, B.G. (2008). Conservation of marbled murrelets in British Columbia. In O. Pourret, P. Naim, & B. Marcot (Eds.), *Bayesian networks: A practical guide to applications* (pp. 127-147). West Sussex, England: Wiley.

Marcot, B.G. (2012). Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling, 230,* 50-62.

Marcot, B.G., Steventon, J.D., Sutherland, G.D. & McCann, R.K. (2006). Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research, 36*(12), 3063-3074.

Margaritis, D. (2003). *Learning Bayesian network model structure from data.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.

Marks, H.M. (1997). *The progress of experiment: Science and therapeutic reform in the United States.* Cambridge, UK: Cambridge University Press.

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. *Conservation Biology, 26*(1), 29-38.

Masegosa, A.R. & Moral, S. (2013). An interactive approach for Bayesian network learning using domain/expert knowledge. *International Journal of Approximate Reasoning, 54,* 1168-1181.

Matheson, J.E. & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science, 22*, 1087-1096.

Matthews, J.R. (1995). Q*uantification and the quest for medical certainty.* Princeton, NJ: Princeton University Press.

May, H. (2006). A multilevel Bayesian item response theory method for scaling socioeconomic status in international studies of education. *Journal of Educational and Behavioral Statistics, 31*(1), 63-79.

McBee, M.T. & Matthews, M.S. (2014). Welcoming quality in non-significance and replication work, but moving beyond the p-value: Announcing new editorial policies for quantitative research in JOAA. *Journal of Advanced Academics, 25*(2), 73-87.

Meehl, P.E. (1956). Symposium on clinical and statistical prediction: The tie that binds. *Journal of Counseling Psychology, 3,* 163-173.

Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology, 4,* 268-273.

Meehl, P. E. (1967). What can the clinician do well? In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 594-599). New York: McGraw-Hill.

Meehl, P. E. (1973). *Psychodiagnosis: Selected papers.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality*

Meehl, P.E. (1996). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Northvale, NJ: Jason Aronson. (Original work published 1954)

Meyer, K.A. & Xu, Y.J. (2007). A Bayesian analysis of the institutional and individual factors influencing faculty technology use. *Internet and Higher Education, 10,* 184-195.

Meyer, M.A. & Booker, J.M. (2001). *Eliciting and analyzing expert judgment: A practical guide*. Philadelphia, PA: ASA-Society of Industrial and Applied Mathematics.

Mislevy, R.J., Almond, R.G., Yan, D., & Steinberg, L.S. (1999). Bayes nets in educational assessment: Where the numbers come from. *Proceedings of the fifteenth conference on uncertainty in artificial intelligence,* 437-446.

Montgomery, J.M. & Nyhan, B. (2010). Bayesian modeling averaging: Theoretical developments and practical applications. *Political Analysis, 18*(2), 245-270.

Moreno, E. & Girón F. J. (2006). On the frequentist and Bayesian approaches to hypothesis testing. *SORT*, *30*(1), 3-28.

Morey, R.D. & Rouder, J.N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods, 16*(4), 406-419.

Morey, R.D., Romeijn, J. & Rouder, J.N. (2013). The humble Bayesian: Model checking from a fully Bayesian perspective. *British Journal of Mathematical and Statistical Psychology*, *66*, 68-75.

Morgan, M.G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *PNAS Proceedings of the National Academy of Sciences in the United States of America, 111*(20), 7176-7184.

Morris, D.E., Oakley, J.E. & Crowe, J.A. (2014). A web-based tool for eliciting probability distributions from experts. *Environmental Modelling & Software, 52*, 1-4.

Moyé, L. A. (2008). Bayesians in clinical trials: Asleep at the switch. *Statistics in Medicine*, *27*, 469-482.

Murphy, A.H. & Winkler, R.L. (1984). Probability forecasting in meteorology. *Journal of the American Statistical Association*, *79*, 489-500.

Murray, J. V., Goldizen, A. W., O'Leary, R. A., McAlpine, C. A., Possingham, H. P., & Choy, S. L. (2009). How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock wallabies. *Journal of Applied Ecology*, *46*, 842-851.

Neapolitan, R.E. (2004). *Learning Bayesian networks.* Upper Saddle River, NJ: Pearson Prentice Hall.

Neapolitan, R.E. (2007). *Probabilistic methods for financial and marketing informatics.* Burlington, MA: Morgan Kaufmann.

Neapolitan, R.E. (2009). *Probabilistic methods for bioinformatics: With and introduction to Bayesian networks.* Burlington, MA: Morgan Kaufmann.

Neil, M., Fenton, N. & Nielsen, L. (2000). Building large-scale Bayesian networks. *The Knowledge Engineering Review, 15*(3), 257-284.

Neyman, J. & Pearson, E. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society, Series A,* 289-337.

Nghe, N.T., Janecek, P., & Haddawy, P. (2007). A comparative analysis of techniques for predicting academic performance. Paper presented at the 37th ASEE/IEEE Frontiers in Education Conference, Milwaukee, WI.

Niculescu, R.S., Mitchell, T.M., & Rao, R.B. (2006). Bayesian network learning with parameter constraints. *Journal of Machine Learning Research, 7*(7), 1357-1383.

Niedermayer, D. (2008). An introduction to Bayesian networks and their contemporary applications. In D.E. Holmes & L. Cain (Eds.), *Innovations in Bayesian networks* (pp. 117-130). Berlin: Springer.

Nussbaum, E.M. (2011). Argumentation, dialogue theory, and probability modeling: Alternative frameworks for argumentation research in education. *Educational Psychologist, 46*(2), 84-106.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., ... Rakow, T. (2006). *Uncertain judgements: eliciting experts' probabilities*. West Sussex, England: John Wiley & Sons, Ltd.

O'Neil, S., Osborn, T., Hulme, M. Lorenzoni, I. & Watkinson, A. (2008). Using expert knowledge to assess uncertainties in future polar bear populations under climate change. *Journal of Applied Ecology, 45,* 1649-1659.

Oakes, M.W. (1986). *Statistical inference: A commentary for the social and behavioral sciences.* Chichester, UK: Wiley.

Oakley, J. E. & O'Hagan, A. (2007). Uncertainty in prior elicitations: A nonparametric approach. *Biometrika, 94*(2), 427-441.

Olesen, K.G., Kjærulff, U., Jensen, F., Jensen, F.V., Falck, B., Andreassen, S., & Anderson, S.K. (1989). A MUNIN network for the median nerve – a case study on loops. *Applied Artificial Intelligence, 3*(2-3), 385-403.

Olmstead, S. (1983). *On representing and solving decision problems.* Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Osmanbegović, E. & Suljić, M. (2012). Data mining approach for predicting student performance. *Journal of Economics and Business, 10*(1), 3-12.

Pearl, J. & Verma, T. (1991). A theory of inferred causation. In J. Allen, R. Fikes & E. Sandewall (Eds.), *Knowledge representation and reasoning: Proceedings of the second international conference* (pp. 441-452). San Francisco: Morgan Kaufmann.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Series in Representation and Reasoning. San Mateo, CA: Morgan Kaufmann Publishers.

Pearl, J. (2000). *Causality: Models, reasoning, and influence.* Cambridge, UK: Cambridge University Press.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications, 41*, 1432-1462.

Phillips, L.D. (1999). Group elicitation of probability distributions: Are many heads better than one? In J. Shanteau, Barbara A. Mellars, and David A. Schum (Eds.), *Decision science and technology: reflections on the contributions of Ward Edwards* (pp. 313-330). Norwell, MA: Kluwer Academic Publishers.

Phillipson, F., Matthijssen, E. & Attema, T. (2014). Bayesian belief networks in business continuity. *Journal of Business Continuity & Emergency Planning, 8*(1), 20-30.

Pitchforth, J. & Mengersen, K. (2013). A proposed validation framework for expert elicited Bayesian networks. *Expert Systems with Applications. 40,* 162-167.

Poirer, D.J. (1988). Frequentist and subjectivist perspectives on the problems of model building in economics. *Journal of Economic Perspectives, 2*, 121-144.

Pollino, C., Woodberry, O., Nicholson, A., Korb, K. & Hart, B. (2006). Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling and Software, 22*, 1140-1152.

Porter, T.M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life.* Princeton, NJ: Princeton University Press.

Pourret, O., Naim, P., & Marcot, B. (2008). *Bayesian networks: A practical guide to applications.* Chichester, West Sussex: John Wiley & Sons.

Prajna, N. V., Krishnan, T., Mascarenhas, J., Rajaraman, R., Srivivasan, M., Raghavan, A., ... Lietman, T. M. (2013). Expert prior elicitation and Bayesian analysis of the Mycotic Ulcer Treatment Trial I. *Investigative Ophthalmology & Visual Science*, *54*, 4167-73.

Press, S. J. & Tanur, J.M. (2001). *The subjectivity of scientists and the Bayesian approach.* New York: John Wiley and Sons.

Press, S. J. (2003). *Subjective and objective Bayesian statistics: Principles, models and applications* (2nd ed.). Hoboken, NJ: John Wiley and Sons.

Press, S.J. (1972). *Applied multivariate analysis: Using Bayesian and frequentist methods of inference.* Melbourne, FL: Krieger.

Price, P.C. (1998). Effects of a relative-frequency elicitation question on likelihood judgment accuracy: The case of external correspondence. *Organizational Behavior and Human Decision Processes*, *76*(3), 277-297.

Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111-163.

Raftery, A.E. (1996). Approximate Bayes factors for accounting for model uncertainty in generalized linear models. *Biometrika*, *183*(2), 251-266.

Raiffa, H. & Schlaifer, R. (1961). *Applied statistical decision theory.* Boston: Harvard University.

Ramsey, F.P. (1931). Truth and probability. In F.P. Ramsey (Ed.), *The foundations of mathematics and other logical essays* (pp. 156-198). London: Kegan Paul.

Ranganathan, S., Spaiser, V., Mann, R.P. & Sumpter, D.J.T. (2014). Bayesian dynamical systems modelling in the social sciences. *PLoS ONE, 9*(1), 1-9.

Reimherr, M., Meng, X., & Nicolae, D.L. (2014). Being an informed Bayesian: Assessing prior informativeness and prior-likelihood conflict. *arVix preprint:* 1406-5958.

Renooij, S. (2001). Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review, 16,* 255-269.

Rezaee, A., Raie, A., Nadi, A. & Ghidary, S.S. (2011). Sensor fault tolerance method using a Bayesian network for robot behavior. *Advanced Robotics, 25*(16), 2039-2064.

Ricketts , C. & Moyeed, R. (2011). Improving progress test score estimation using Bayesian statistics. *Medical Education, 45*, 570-577.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14,* 465-471.

Roback, P.J. & Givens, G.H. (2001). Supra-Bayesian pooling of priors linked by a deterministic simulation model. *Communications in Statistics – Simulation and Computation, 30*(3), 447-476.

Robert, C.P. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation* (2nd Ed.). New York: Springer.

Rocha, G., Loschi, R.H., & Franco, G.C. (2011). Bayesian and classical approaches for hypotheses testing in trisomies. *Communications in Statistics: Simulation & Computation, 40*(3), 307-321.

Romero, C. & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications, 33,* 135-146.

Ronis, D.L. & Yates, J.F. (1987). Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes, 40,* 193-218.

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D. & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225-237.

Royall, R. (2000). On the probability of observing misleading statistical evidence (with discussion). *Journal of the American Statistical Association, 95,* 760-780).

Rubin, D. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics, 12,* (1151-1172).

Rubin, H., & Rubin, R. (2012). *Qualitative Interviewing: The art of hearing data*. Thousand Oaks, CA: SAGE Publications.

Ryan, J.C.H., Mazzuchi, T.A., Ryan, D.J., de la Cruz, J.L. & Cooke, R. (2012). Quantifying information security risks using expert judgment elicitation. *Computers and Operations Research, 39*, 774-784.

Samaniego, F.J. & Renueau, D.M. (1994). Toward a reconciliation of the Bayesian and frequentist approach to point estimation. *Journal of the American Statistical Association, 89*(427), 947-957.

Sandri, M., Berchialla, P., Baldi, I., Gregori, D., & De Blasi, R.A. (2014). Dynamic Bayesian networks to predict sequences of organ failures in patients admitted to ICU. *Journal of Biomedical Informatics, 48,* 106-113.

Sarbin, T.R. (1986). Prediction and clinical inference: Forty years later. *Journal of Personality Assessment, 50,* 362-369.

Saupe, J.L. (1990). *The functions of institutional research* (2nd ed.). Tallahassee, FL: Association for Institutional Research. Available at http://www.airweb.org/EducationAndEvents/Publications/Pages/FunctionsofIR.aspx

Savage, L.J. (1954). *The foundations of statistics.* New York: Wiley.

Savage, L.J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association, 66*(336), 783-801.

Savage, L.J., Hacking, I. & Shimony, A. (1967). A panel discussion of personal probability. *Philosophy of Science, 34*(4), 305-332.

Scheurich, J. J. (1995). A postmodernist critique of research interviewing. *International Journal of Qualitative Studies in Education, 8*(3), 239-252.

Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*, 115-129.

Shachter, R. (1988). Probabilistic inference and influence diagrams. *Operations Research, 36,* 589-604.

Sharabiani, A., Karim, F., Sharabiani A., Atanasov, M., & Darabi, H. (2014). An enhanced Bayesian network model for prediction of students' academic performance in engineering programs. *Proceedings of IEEE Global Engineering Education Conference (EDUCON),* 832-837.

Shenton, W., Hart, B.T. & Chan, T. (2014). A Bayesian network approach to support environmental flow restoration decisions in the Yarra River, Australia. *Environmental Research & Risk Assessment, 28*(1), 57-65.

Shephard, G.G. & Kirkwood, C.W. (1994). Managing the judgmental probabilities elicitation process: A case study of analyst/manager interaction. *IEEE Transactions on Engineering Management, 41*, 414-425.

Shuford, E. H. (1961). Percentage estimation of proportion as a function of element type, exposure type, and task. *Journal of Experimental Psychology, 61*, 430-436.

Shute, V. J., Hansen, E.G., & Almond, R.G. (2008). You can't fatten a hog by weighing it – or can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence in Education, 18*(4), 289-316.

Spetzler, C.S. & Staël von Holstein, C.A.S. (1975). Probability encoding in decision analysis. *Management Science, 22,* 340-358.

Spiegelhalter, D.J. & Rice, K. (2009). Bayesian statistics. *Scholarpedia, 4*(8), 5230. Retrieved from http://www.scholarpedia.org/article/Bayesian_statistics .

Spiegelhalter, D.J., Abrams, K.R. & Myles, J.P. (2004). *Bayesian approaches to clinical trials and health-care evaluation.* New York: John Wiley & Sons.

Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L. & Cowell, R.G. (1993). Bayesian analysis in expert systems. *Statistical Science*, *8*(3), 219-247.

Spirtes, P., Glymour, C. & Scheines, R. (2000). *Causation, prediction and search* (2nd ed.). Cambridge: MIT Press.

Stahlschmidt, S., Tausendteufel, H., & Härdle, W. (2013). Bayesian networks for sex-related homicides: structure learning and prediction. *Journal of Applied Statistics, 40(6),* 1155-1171.

Steck, H. & Tresp, V. (1999). Bayesian belief networks for data mining. *Proceedings of the 2nd Workshop on Data Mining and Data Warehousing als Grundlage moderner entscheiidungunterstuetzender Systeme*, 145-154.

Stevens, J.W. & O'Hagan, A. (2002). Incorporation of genuine prior information in cost-effectiveness analysis of clinical trial data. *International Journal of Technology Assessment in Health Care*, *18*(4), 782-790.

Subbiah, M., Srinivasan, M. R., & Shanthi, S. (2011). Revisiting higher education data analysis: A Bayesian perspective. *International Journal of Science and Technology Education Research*, *1*(2), 32-38.

Suhirman, J., Zain, M. & Herawan, T. (2014). Data mining for education decision support: A review. *International Journal of Educational Technology*, *9*(6), 4-19.

Sun, L. & Shenoy, P. (2007). Using Bayesian networks for bankruptcy predictions. *European Journal of Operational Research*, *180*(2), 738-753.

Swiler, L.P. (2006). *Bayesian methods in engineering design problems.* (Sandia Report SAND2005-3294). Retrieved from Sandia National Laboratories website: https://cfwebprod.sandia.gov/cfdocs/CompResearch/templates/insert/profile.cfm?snl_id=1157

Syversveen, A.R. (1998). Noninformative Bayesian priors. Interpretation and problems with construction and applications. Preprint Statistics 3, Department of Mathematical Science, NTNU, Trondheim. http://-www.math.ntnu.no/preprint/statistics/1998/S3-1998.ps

Taruna, S. & Pandey, M. (2014). An empirical analysis of classification techniques for predicting academic performance. *Advance Computing Conference (IACC), 2014 IEEE International*, 523-528.

Thomson, M.E., Önkal-Atay, Pollock, A.C. & Macaulay, A. (2003). The influence of trend strength on directional probabilistic currency predictions. *International Journal of Forecasting, 19,* 241-256.

Tong, S. & Koller, D. (2000). Active learning for parameter estimation in Bayesian networks. *Advances in Neural Information Processing Systems 13 (NIPS 2000), 13,* 647-653.

Torabi, R., Moradi, P. & Khantaimoori, A.R. (2012). Predict student scores using Bayesian networks. *Procedia – Social and Behavioral Sciences, 46,* 4476-4480.

Tversky, A. & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105-110.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling, 203,* 312-318.

van Duijnhoven, J. (2003). *Knowledge assessment using Bayesian networks: A case study in the domain of algebraic expectation.* Unpublished master's thesis, Utrecht University.

VanLehn, K. (2008). Intelligent tutoring systems for continuous, embedded assessment. In C. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp.113-138). Mahwah: Lawrence Erlbaum.

Venkatesh, J., Cherurveettil, P. & Sivashanmugam, P.M. (2014). Using a prediction model to manage cyber security. *The Scientific World Journal*, 1-6.

Vogelgesang, J., & Scharkow, M. (2009, May). *Is there a place for Bayesian statistics in communications research?* Paper presented at the Vortrag auf der Jahrestagung der DGPuK, Bremen. Retrieved from http://underused.org/docs/vogelgesang_scharkow_dgpuk2009.pdf

Vomlel, J. (2004). Bayesian networks in educational testing. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 12*(1), 83-100.

Wagenmakers, E.J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14*(5), 779-804.

Wagenmakers, E.J., Lee, M.D., Lodewyckx, T. & Iverson, G. (2008). Bayesian versus frequentist inference. In H. Hoijtink, I. Klugkist, & P.A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses in psychology* (pp. 181-207). New York: Springer.

Wainer, H., Wang, X.A., Skorupski, W.P. & Bradlow, E.T. (2005). A Bayesian method for evaluating passing scores: the PPoP curve. *Journal of Educational Measurement, 42,* 271-81.

Walls, L. & Quigley, J. (2001). Building prior distributions to support Bayesian reliability growth modelling using expert judgement. *Reliability Engineering and System Safety, 74,* 117-128.

Wallsten, T.S., Budescu, D.V., Rapoport, A., Zwick, R. & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology, 115*(4), 348-365.

Wang, H., Dash, D. & Druzdel, M.J. (2002). A method for evaluating elicitation schemes for probabilistic models. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, 32,* 38-43.

Wang, Y. & Beck, J. (2013). Class vs. Student in a Bayesian Network Student Model. In H.C. Lane, K. Yacef, J. Mostow & P. Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 151-160). Berlin: Springer-Verlag Berlin Heidelberg.

Wardekker, J.A., van der Sluijs, J.P., Janssen, P.H.M., Kloprogge, P., Petersen, A.C. (2008). Uncertainty communication in environmental assessments: Views from the Dutch science-policy interface. *Environmental Science & Policy, 11*(7), 627-641.

Weatherford, R. (1982). *Philosophical foundations of probability theory*. London: Routledge & Kegan Paul.

West, P., Rutstein, D.W., Mislevy, R.J., Liu, J., Choi, Y., Levy, R., …Behrens, J.T. (2010). *A Bayesian network approach to modeling learning progressions and task performance.* (CRESST Report 776). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

West, R.F. & Stanovich, K.E. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin and Review, 4,* 387-392.

Western, B. (1999). Bayesian analysis for sociologists: An introduction. *Sociological Methods & Research, 28*(1), 7-34.

White, I. R., Carpenter, J., Evans, S., & Schroter, S. (2007). Eliciting and using expert opinions about dropout bias in randomized control trials. *Clinical Trials*, *4*, 125-139.

Winkler, R.L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association, 62,* 776-800.

Woodberry, O., Nicholson, A.E., Korb, K., & Pollino, C. (2005). Parameterising Bayesian networks. In G. Webb & X. Yu (Eds.) *AI 2004: Advances in artificial intelligence. Lecture notes in computer* science (Vol. 339, pp. 711-745). Berlin/Heidelberg: Springer

Wright, G. & Ayton, P. (1992). Judgemental probability forecasting in the immediate and medium term. *Organizational Behavior and Human Decision Processes, 51,* 344-363.

Xu, D.J., Liao, S.S., & Li, Q. (2008). Combining empirical experimentation and modeling techniques: A design research approach for personalized mobile advertising techniques. *Decision Support Systems, 44*(3), 710-724.

Xu, Y.J. & Ishitani, TT. (2008). Bayesian modeling in institutional research: An example of nonlinear classification. *New Directions for Institutional Research, 137,* 83-104.

Xu, Y.J. (2012). The similarities and differences in the work experience of women faculty in traditional and nontraditional disciplines. *New Directions for Institutional Research, 155,* 67-83.

Yang, R. & Berger, J. (1997*). A catalog of noninformative priors.* Institute of Statistics and Decision Sciences Discussion Paper. Durham, NC: Duke University.

Yukselturk, E., Ozekes, S., & Tűrel, Y. (2014). Predicting dropout student: An application of data mining methods in an online education program. *European Journal of Open, Distance and e-Learning, 17*(1), 118-133.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics.* New York: Wiley.

Zhang, L., Wu, X., Skibniewski, M.J.; Zhong, J. & Lu, Y. (2014). Bayesian-network-based safety risk analysis in construction projects. *Reliability Engineering & System Safety, 131*, 29-39.

Zhu, J. Y., & Deshmukh, A. (2003). Application of Bayesian decision networks to life cycle engineering in Green design and manufacturing. *Engineering Applications of Artificial Intelligence*, *16*(2), 91-103.

Zumbo, B.D. & Chan, E.K.H. (Eds.). (2014). *Validity and validation in social behavioral, and health sciences (Social indicators research series).* New York: Springer.

2 DESIGNING A BAYESIAN NETWORK TO PREDICT LIKELIHOOD OF RETENTION OF UNDERGRADUATE STUDENTS

Ensuring that first-year, degree-seeking undergraduate students return for their subsequent academic year ("undergraduate first-to second-year retention") is a high priority issue in higher education. Nationally, 72.9% of all undergraduate students are retained at the same institution from their first to second year. This number is higher for students at private nonprofit institutions (80.3%), slightly lower among public institutions (71.4%), and lowest among for-profit institutions (62.8%) (U.S. Department of Education, 2014). In addition to nationwide calls to consider retention rates as an accountability measure of student progression and institutional success (Carey & Aldeman, 2008; Longden, 2006; Pike & Graunke, 2015), various higher education accreditation agencies and rankings organizations consider retention rates in assessments of institution quality (Wimshurst, Wortley, Bate, & Allard; 2006). Students and other stakeholders are negatively affected by the increased time to graduation often resulting from attrition (Complete College America, 2011), and the costs of recruiting new and replacement students are high (Noel-Levitz, 2013). Retention rates are particularly important among tuition-dependent institutions where even small fluctuations in year-over-year retention result in large impacts in revenue and budget management (Schuh & Gansemer-Topf, 2012).

Due to the importance of retention in enrollment planning and financial management, developing a clear understanding of the factors that influence students' decisions to return for a second year is critical. Through identification of structure, relationships, and interactions in the retention-related data, it is possible to create a statistical model of current and prior student retention that can be used to predict future students' retention decisions. In addition to enhancing institutional knowledge as the process of development requires clear articulation and exploration of aspects of retention, ideal models are easily used and understood by various stakeholders (institution administrators, researchers, academic advisors, admissions counselors) and contain the capacity to be updated as student populations change. Additionally, a model that facilitates prediction of retention aids in the identification of students at risk for attrition, informs institutional intervention and student advising policy, and enables more precise enrollment planning.

Research Questions

This study addresses the following research questions:

1. Based on expert information and data, what are the greatest influences on first-to second-year undergraduate retention at a small, private liberal arts college in the southeast?
2. Using this knowledge from both sources, can a graphical model employing Bayesian Networks be built that adequately predicts retention?

Literature Review

Undergraduate First- to Second-Year Retention

The preeminent literature surrounding retention of U.S. college and university students suggests that retention is influenced by a combination of pre-college student characteristics and students' social and academic experiences once at an institution (Astin, 1993; Bean, 1980, 1985; Cabrera, Nora, & Castaneda, 1993; Pascarella & Terenzini, 1980, 1991; Tinto, 1975, 1988, 1993). These authors cited student integration into and commitment to the educational and institutional environments, faculty-student interaction, and social engagement as key influences on student retention. Tinto's integration framework (1975, 1993), suggesting that students' commitment to and likelihood of graduating from an institution grows as they are socially and academically integrated, has been explored and modified by numerous authors over the years and formed the foundation for the study of undergraduate retention and graduation (Swail, 2004). Astin (1993) focused on the impact of student involvement in college as an influence on retention, and Bean (1980) emphasized the important role of pre-college characteristics such as high school performance and socioeconomic status. Cabrera, Nora, and Castaneda (1993) and Pascarella and Terenzini (1991, 2005) explored the convergence of proposed retention models and theories and began to introduce and investigate different subpopulations of undergraduate students and their unique responses to retention predictors. More recent retention literature indicates a shift towards the inclusion of non-cognitive variables such as motivation, self-efficacy, and academic self-concept into theories of undergraduate retention (Covington, 2000; Demetriou & Schmitz-Sciborski, 2011; Eccles & Wigfield, 2002).

Additional studies discuss and evaluate specific predictors of retention, and a review of the most recent of these reveal that retention predictors tend to fall within the following broad categories: Pre-college student characteristics, pre-college academic preparation, student characteristics in college, and institutional characteristics. Pre-college student characteristics include demographic variables such as race-ethnicity, gender, and socioeconomic and first-generation status, and the literature points to a trend in which students in underrepresented minority groups, student with financial challenges, and students who are the first in their family to attend college are less likely to retain and graduate (The Education Trust, 2004). Pre-college academic preparation is reflected in students' high school GPAs, class ranks, and standardized achievement test scores, and unsurprisingly, are generally positively correlated with retention within the literature (ACT, 2010; Adelman, 1999; Astin & Oseguera, 2005; Lotkowski, Robbins, & Noeth, 2004). A large amount of retention literature addressing specific predictors focuses on student characteristics during college, including financial support, distance from home, social support and engagement, socioeconomic status, academic engagement and participation, and other non-cognitive attributes. These factors are well summarized by Pascarella and Terenzini (2005), and tend to predict retention in the direction that one would intuitively expect. For example, students with less financial support and further distance from home are less likely to retain compared to students with more financial support and attending college closer to home (Bista & Foster, 2011; Titus, 2006), and students with superior study skills and psychosocial attributes tend to retain at higher rates than those without (Robbins et al., 2004). Researchers have also addressed the role of institutional characteristics such as institutional control, selectivity, mission, and size in influencing retention (see Astin & Oseguera, 2012; Pike, 2013; Ryan, 2004; Titus, 2004). The extent to which an institution focuses on retention has also been found to have impact (Howard, 2013; Oseguera & Rhee, 2009; Porter & Swing, 2006). A comprehensive listing of commonly named predictors of retention, along with examples found in the literature, is presented in Appendix A.

As this study focuses on retention within a liberal arts setting, particular attention is paid to findings of the few authors addressing retention specifically within the liberal arts. Nesler (1999) focused on retention at a liberal arts college that offered courses exclusively through distance education to nontraditional students, finding that retention was influenced by a combination of student and environmental characteristics. Howard (2013) explored the influence of first-year programming on student retention at rural liberal arts colleges, finding its impact negligible. Employing analysis of students' social networks, Eckles and Stradley (2011) identified the importance of students' friends on influencing retention at a small liberal arts college, building upon Thomas' (2000) findings of the significance of students' social integration on retention at liberal arts colleges. Finally, Gansemer-Topf, Zhang, Beatty, and Paja (2014) highlighted college transition success, realistic academic expectations, and social integration as important influences on retention at a small, highly selective liberal arts college, while also pointing out that unique campus populations can produce unique retention predictor conclusions. Given smaller liberal arts institutions' unique populations and the aforementioned reliance on enrollment and net tuition revenue, the need for a clear understanding of the forces of retention and the ability to accurately predict the likelihood of retention for individuals or groups of students becomes apparent.

A Bayesian Approach

This study presents the development of a probabilistic network that models and predicts first-to second-year undergraduate retention. The network employs a Bayesian approach, where, through the use of conditional probability rules expressed in Bayes' theorem (Equation 2.1), current knowledge or beliefs about the probability of an event occurring ("prior probability") are updated with new information in order to form a more accurate prediction ("posterior probability"). Bolstad (2007) and Gelman, Carlin, Stern, and Rubin (2004) provided comprehensive introductions to Bayesian statistics and data analysis, while Gill (2009) offered a guide to the approach within the context of social and behavioral sciences. Gigerenzer (2002) illustrated the use of Bayes' theorem with applied examples.

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \tag{2.1}$$

The use of the word "beliefs" in describing the probability updating in Bayesian methodology is important. A Bayesian approach to probability is in direct contrast to a frequentist approach in that Bayesian statistics interprets probability as a degree of belief rather than as the long-term proportion of successful outcomes to number of attempts. This degree of belief represents prior knowledge pertaining to the likelihood of an event, which is then updated with data relevant to this event in order to form a new, or posterior, belief in the probability of the same event occurring. Within the larger field of Bayesian statistics, the use and definition of *a priori* knowledge and beliefs (“priors”) is an ongoing discussion, and the choice and use of prior type is often driven by the weight assigned to and source of the prior knowledge. Uninformative priors are those that provide little additional information or explanatory power (Gill, 2009), and Kass and Wasserman (1996) and Yang and Berger (1997) provided a catalog and review of such priors. Priors that intentionally include knowledge designed to influence posterior probabilities are termed “informative,” and can be derived from a number of sources such as previous studies and results or researcher expertise (Gelman et al., 2004; Gill, 2009). Press (2003) outlined the advantages of informative priors, noting that they capitalize on expert knowledge and present an avenue for analysis when other information (“objective” Bayesian priors, or even a frequentist approach) is unavailable.

It is this use of informative priors that prompts the common criticism that Bayesian statistics employs subjective probability, and thus contradicts the objectivity valued in the modern scientific method. A number of authors have addressed this criticism by pointing out that, regardless of approach, an expectation of pure objectivity is unreasonable. Researchers’ choices regarding research questions, methodologies, and analysis techniques could all be considered subjective choices (Berger, 2006; Gill, 2009; Hennig, 2009; Press & Tanur; 2001; Stevens & O’Hagan, 2002; Weatherford, 1982). Additionally, requirements of coherence and adherence to the laws of probability limit the use of arbitrary and unrealistic informative priors (Bovens & Hartmann, 2003), and, as D’Agostini (2001) describes, the process through which subjective Bayesians consider and account for the conditions affecting their degrees of belief and choice of priors leads to a conclusion that “the coherent subjectivist is far more responsible (and more "objective", in the sense that ordinary parlance gives to this word) than those who blindly use standard 'objective' methods” (p. 25).

Bayesian networks. A goal of this study is the construction of a predictive graphical model that employs Bayesian methodology. A Bayesian network (BN) is a graphical network that, using Bayes’ theorem to calculate conditional and joint probabilities, models the probabilistic relationship between and among variables. A BN consists of two main elements: A directed acyclic graph (DAG) forming the structure of the model, and the independent/dependent relationships between the variables that are quantified by conditional probability distributions (Kjærulff & Madsen, 2008). Each variable included in the network has a finite set of mutually exclusive states, and variables with directed edges pointing towards other variables are considered “parent” nodes of “children” nodes – variables not sharing a directed edge are considered independent of each other. Conditional probability tables (CPTs) are attached to each variable, in which the conditional probabilities of each variable given the state of other variables are presented, and the entire set of probability tables expresses the full model’s parameter set (Equation 2.2). BNs are unique from other graphical models in that, through the use of Bayesian probabilistic inferences, users are provided with a clear representation of independencies, dependencies, and uncertainty. Pearl (1988) and Lauritzen (1996) offered comprehensive introductions to graphical models and BNs, and Pearl and Verma (1991), Druzdzel and Simon (1993), and Heckerman (1997) explored the capability of BNs to illustrate conditional dependence and causal influence. There are a number of techniques for probabilistic inference within BNs, and different algorithms often used are discussed by Dawid (1992), Jensen, Lauritzen, and Olesen (1988), and Heckerman (1997). With the development of computing power, the ability to complete inference in increasingly complex BNs using algorithms has increased (Jensen & Nielsen, 2007).

$$P_a(x_1, \ldots x_n) = \prod_{i=1}^{n} P_a(x_i | \prod x_i) \tag{2.2}$$

Kjærulff and Madsen (2008) highlighted the advantages of Bayesian networks, noting the efficiency in which these networks conduct inference and convey causal relationships, the ease in which the graphical representations can be understood by numerous audiences, and the methodology's firm foundation in decision theory. Speigelhalter, Dawid, Lauritzen, and Cowell (1993) cited the ability of BNs to simultaneously "…be forgiving of limitations in the data but also exploit the accumulated data" (p. 221), and Heckerman (1997) acknowledged the ability of BNs to operate with incomplete data. Additionally, due to the use of Bayesian statistics, a BN can flexibly and efficiently incorporate additional information as it is gathered. BNs are not without their limitations, however. Neidermeyer (2008) pointed out that novel events may threaten the predictive validity of BNs, and cautioned that, even with computing advances, a network with a large number of variables may require unreasonable computing and computational power. Pourret, Naim, and Marcot (2008) considered the requirement that BNs be acyclic to be a limitation as feedback loops are often found in reality. Additionally, poor reliability and quality of prior information included in BNs negatively affects a model's usefulness, although this can be mitigated through adequate model evaluation (Cowell, Dawid, & Spiegelhalter, 1993; Pitchforth & Mengerson, 2013).

Bayesian networks are often developed and used for prediction because BNs can be considered predictive due to the interpretation of the links between variables as causal. For example, if a parent node is a direct cause of a child node, then a change (intervention) in the value of the parent node will change the value of the child mode – the effect can be predicted based on the intervention. Pearl (2000) and Friedman, Linial, Nachman, and Pe'er (2000) provided explanations of causal BNs, and Spirtes, Glymour and Scheines (2000) provided a background of the Causal Markov Assumption, upon which a causal/predictive interpretation of BNs is predicated. In this study, a BN is used to predict the likelihood of first-to second-year undergraduate retention given certain conditions of other variables.

Bayesian network development. Spiegelhalter et al. (1993) named three stages of constructing a Bayesian network: A qualitative stage in which the author defines the relationships among and between variables in terms of conditional independence and develops a graphical model that reflects these relationships, a probabilistic stage in which the author considers the model's joint distribution, and a quantitative stage in which the author assigns values to the underlying CPTs. Approaches to each stage can be manual (theory- and expert-driven) or automatic (data-driven), or even a combination of both. The decision regarding approach to model construction often depends on the field on which the model is based (Chen & Pollino, 2012; Uusitalo, 2007) or the availability of data (Pitchforth & Mengersen, 2013).

Manual construction of a model involves input from experts or previous research. Approaches to manually determining the structure and relationships of a BN range from complex use of idioms and ontology (Fenton, Neil, & Lagnado, 2013; Fenz, 2012) to more simple methods involving identification of each variable and their causal influences/influencers (Kjærulff & Madsen, 2008). Commonly cited authors such as Garthwaite, Kadane, and O'Hagan (2005), Gill (2009), Kadane and Wolfson (1998), and O'Hagan et al. (2006) provided a list of "best practices" related to elicitation of expert knowledge, and review of these works highlights the major stages of a successful elicitation process: Preparation of the researcher/elicitor, selection of expert(s), training of expert(s), confirmed understanding or acceptance of the model for which judgments are being elicited, and the actual elicitation, including assessment and feedback. These practices serve to provide the expert ample opportunity to adequately express her beliefs while also allowing the researcher to gather as much helpful information as possible and verify her own understanding of what the expert is trying to communicate.

The structure and parameters of a BN can also be developed automatically, using machine learning from data. Algorithms that enable learning from data are either score-based, where successive iterations are scored based on data fit, or constraint-based, where *a priori* understandings of independence among variables are incorporated (Margaritis, 2003). Although there are a number of approaches involving these two types of algorithms, many of which are described in Neapolitan (2004) and Kjærulff and Madsen (2008), this study employs Tree Augmented Naïve Bayes' (TAN) learning as described by Friedman, Geiger, and Goldszmidt (1997) as part of the model development process.

Both aforementioned approaches to model development face challenges. Within the manual approach, experts' potential use of heuristics in the formation of probability judgements (see Tversky & Kahneman, 1974) threatens the reliability and accuracy of information elicited. Learning BNs from data has been criticized for over-fitting (Clark, 2003). In response to these difficulties, a hybrid manual/data-driven approach developed. Heckerman, Geiger, and Chickering (1995) first proposed that an expert-generated BN can be subsequently updated and improved upon by observed data, and numerous authors paired expert knowledge with machine learning to build BNs (de Campos & Castellano, 2007; Flores, Nicholson, Bruskill, Korb, & Mascaro, 2011; Masegosa & Moral, 2013; Niculescu, Mitchell, & Rao, 2006; Woodberry, Nicholson, Korb, & Pollino, 2005).

This study engages a hybrid approach to model development, in which formal elicitation of expert opinion is combined with statistical data to predict retention. Note that this hybrid approach to building a predictive model is applicable in larger discussions of clinical and statistical prediction, where the "clinical" method of prediction involves an expert human judge relying on informal decision-making procedures while the "statistical" method of prediction involves some formal decision-making rules or formula (actuarial tables, for example) in order to classify or predict. First discussed by Meehl (1954/1996), the superiority of statistical prediction over clinical prediction within the realms of social sciences, human behavior, and medicine is confirmed by numerous other studies, particularly within terms of accuracy and cost (Ægisdóttir et al., 2006; Dawes, 1988; Dawes, Faust, & Meehl, 1989, 1993; Grove & Meehl, 1996; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Meehl, 1956, 1957, 1967, 1973, 1986). However, these authors did not suggest that clinical, or expert, prediction should simply be ignored. Instead, many authors attributed the overall inferiority of clinical prediction to common heuristics, leaving open the possibility that carefully conducted elicitations designed to minimize bias and error could improve clinical prediction. Dana and Thomas (2006), explored situations in which clinicians' predictions could improve model accuracy, noting that the "use of a formal, explicit procedure" (p. 425) is critical in best eliciting and incorporating clinicians' expertise and valuable experience into models. This study proposes to formally and rigorously include the opinions and unique experience of undergraduate student retention experts into a statistically-designed predictive model.

Implications for Current Study

This study focuses on the hybrid development of a predictive Bayesian network that will be used to model and predict the likelihood of first-to second-year undergraduate retention.

Retention is a topic within the larger realm of institutional research, and a review of the literature reveals that the use of BNs for prediction in institutional research is not rare. See the following as examples of such research: Bekele & Menzel (2005), Galbraith, Merrill, and Kline (2010), Käser et al. (2013), Kotsiantis, Patriacheas, and Xenos (2010), Meyer and Xu (2007), Sharabiani, Karim, Sharabiani, Atanasov, and Darabi (2014), and Torabi, Moradi, and Khantaimoori (2012). However, none of these authors employed a hybrid expert/data approach to model construction, or, when applicable to the research design, provided details concerning the elicitation of expert knowledge. The choice of using a BN to model and predict retention is based on a number of factors – the intuitive nature of a graphical model is appropriate for a range of audiences, the ease in which the model updates when presented with new cases and information, and the ability to incorporate both expert and data-learned information into the model. This hybrid design method incorporates the work of other scholars, unobserved patterns and trends in historical data, and the specific and unique experience and knowledge of campus experts. Such an approach follows the very "Bayesian" idea of leaving no relevant information behind.

Methodology

Data/Population

This study involves the construction of a Bayesian network (BN) that models and predicts the likelihood of a first-year undergraduate student returning for her second academic year at a small, private women's college in the southeast with an annual undergraduate degree-seeking enrollment of approximately 830 students. First-to second-year retention rates from 2009 to 2014 averaged 83%, meaning that, on average, 17% of first-year students have not returned for their second year. Retention rates are calculated using cohorts of students. For example, a student entering as part of the fall 2014 cohort of first-year students is considered retained if she is still enrolled as of an enrollment census date in early fall 2015, and an overall retention rate represents the percentage of cohort students still enrolled in the following fall semester. Student-related quantitative data (admissions, demographic, financial, and academic information) used in the model development process was sourced from institutional databases and includes information on 1,438 degree-seeking first-time students entering the institution from fall 2009 until fall 2014.

Model Development

Development of the BN retention model took place in four major stages:

1. The construction and comparison of two initial BNs, one learned solely by existing data and incorporating variables identified in the literature ("straw man" model), and one designed through the elicitation of expert opinions regarding model structure and important variables to include ("expert-elicited structural" model).
2. The construction of an interim BN incorporating the structural insights provided by the experts and machine-learned parameters for variables on which data were available ("interim" model).
3. The presentation of this interim model to retention experts for review and formal elicitation of prior probabilities on model variables for which no data were available or learned parameters were suspect.
4. The development of a final BN ("final" model), reflecting a hybrid method of construction where the BN's structure and parameters were determined through a combination of expert information and machine learning.

Each iteration of models were evaluated on a number of measures, including predictive validity, internal validity, model complexity, and analysis of sensitivity. The following sections describe these major stages of model development in detail.

Initial data-learned model. Using scholarly literature, the researcher identified variables commonly understood to influence retention. These variables are discussed in detail in the included literature review and are also presented in Appendix A. In addition to compiling quantitative information for the identified variables on which the institution has collected and stored data, the researcher noted literature-identified variables not currently available in institutional databases for future presentation to experts for insight and opinion.

Incorporating the literature-identified variables and compiled data from the 2009 through 2011 cohorts (data from more recent cohorts were reserved for later model iterations), the researcher used Netica 5.2 to develop a “straw man” BN. The BN’s structure was learned through Netica’s Tree Augmented Naïve Bayes (TAN) structure learning, a maximum posterior probability score-based technique that examines correlations and includes more relaxed independence assumptions over Naïve Bayes (Friedman, Geiger, & Goldszmidt, 1997), and parameters were learned using an expectation–maximization (EM) algorithm. It was understood that the resulting model was only to be used as a guide or “first-pass” at developing a retention model, as the researcher recognized a lack of non-cognitive data and institution-specific insight that would ultimately be incorporated through the elicitation of expert information. The “straw man” BN is presented in Figure 2.1.

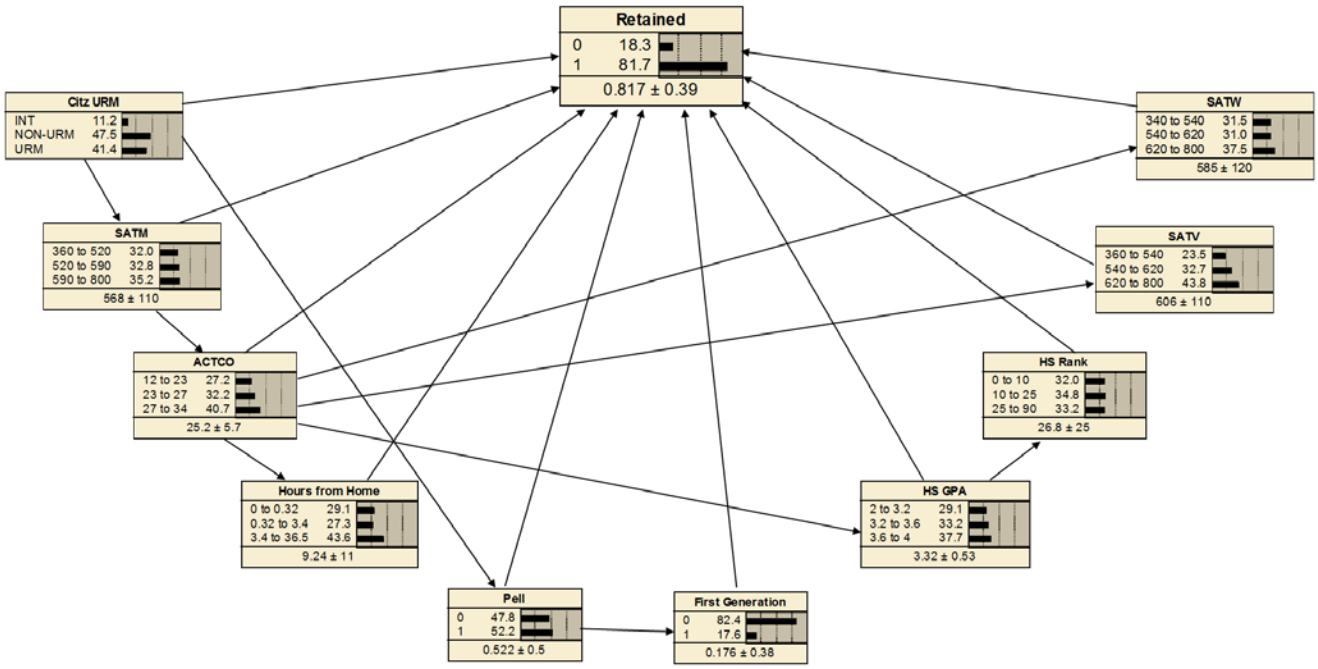

Figure 2.1. Initial data- and structure-learned "Straw-Man" BN.

Network interpretation and use. Without imposing a new student retention scenario and its related properties on the "straw-man" BN, each of the network's nodes show the current-state percentage distribution of all prior cases on each variable. For example, all other conditions held constant, the BN demonstrates 81.7% of first-year students from 2009-2011 are retained to their second year. Just over one-half (52.2%) of students included in the model are eligible to receive Pell grants, less than one-fifth (17.6%) are first-generation students, and 43.6% of students are more than three hours from their home. Netica automatically discretizes continuous variables using existing distributions found in the data – nodes depicting measures of students' academic preparedness (HS GPA, HS Rank, SAT and ACT scores) were thus discretized into three approximately equal bins. Note that all variable classes/categories will ultimately be reviewed and potentially modified by experts.

In addition to demonstrating the current state of retention and each of the variables thought to influence retention, the machine-learned structure of the model reveals that a number of the variables are related to each other in some way through the placement of directed edges (arrows) between different variables. For instance "Citz URM," a node describing whether or not a student is an under-represented minority ("URM"), not an under-represented minority ("NON-URM"), or international student ("INT"), appears to have a relationship with a student's SAT Math score ("SATM") and Pell-eligibility ("Pell"). Unsurprisingly, a student's high school GPA is related to her high school rank percentile ("HS Rank"), and ACT Composite scores ("ACTCO") are related to SAT scores ("SATM," "SATV," and "SATW"). More surprisingly, ACT Composite scores also appear to have a relationship with a student's distance from home ("Hours from Home"), perhaps due to geographic ACT/SAT preferences. The relationships presented in the "straw man" model were ultimately explored in further model iterations, using insights provided by experts.

The predictive application of this type of model is demonstrated through the addition of findings, or "cases." In other words, updating the predictor nodes with the properties of a specific retention case will allow a user to view an updated likelihood of retention. When interpreting the results of the model, it is important to consider that adding a finding on one node tells the network that a new case has been added where information is only known on that one variable – the network will estimate the case's standings on all other variables based on existing data. For example, if interested in the change in likelihood of enrollment based on first-generation status, an update of the "First Generation" node to reflect a student is indeed first-generation decreases the likelihood of retention from 81.7% to 78.7%, *given that all other retention-related variables are held steady*. If a case's status on other variables are known, then these can be entered as well and predicted retention will be updated accordingly using Bayesian inference. This quick updating and dynamic presentation is one of the prime advantages to Netica's BN software. Figures 2.2 and 2.3 show the difference in predicted retention of two different cases where each case's status on one or more of the retention-related variables are known.

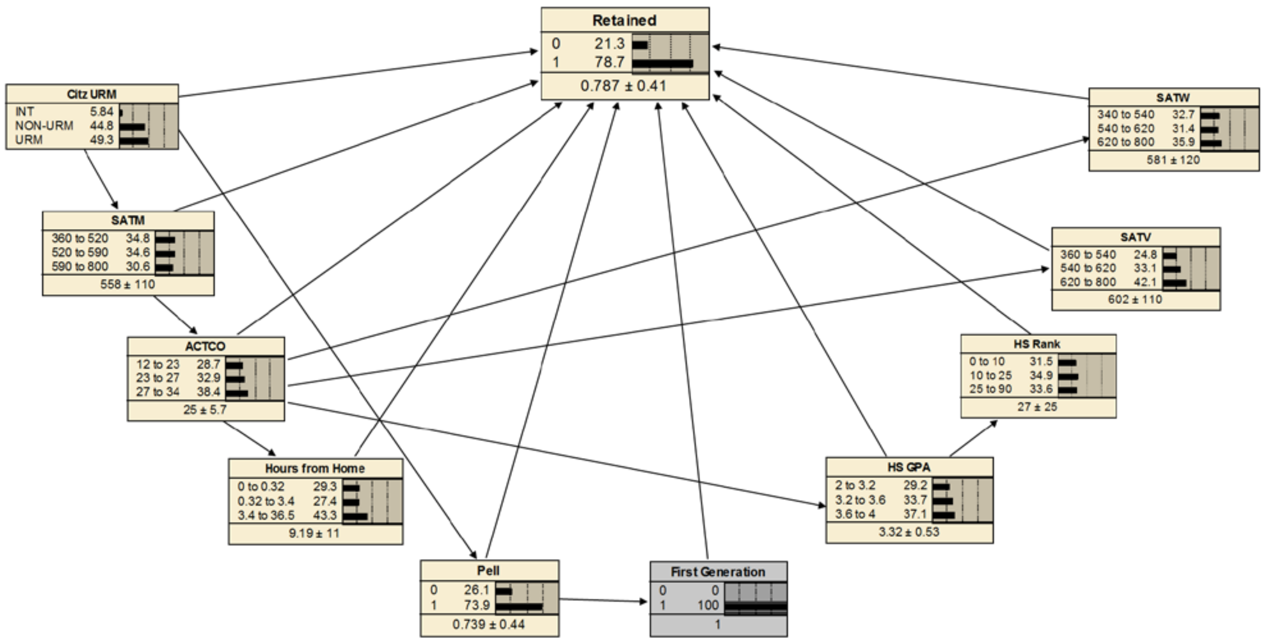

Figure 2.2. "Straw Man" BN Scenario 1. A student is known to be First Generation (coded as "1"). Note that probability of retention decreased from 81.7% to 78.7%.

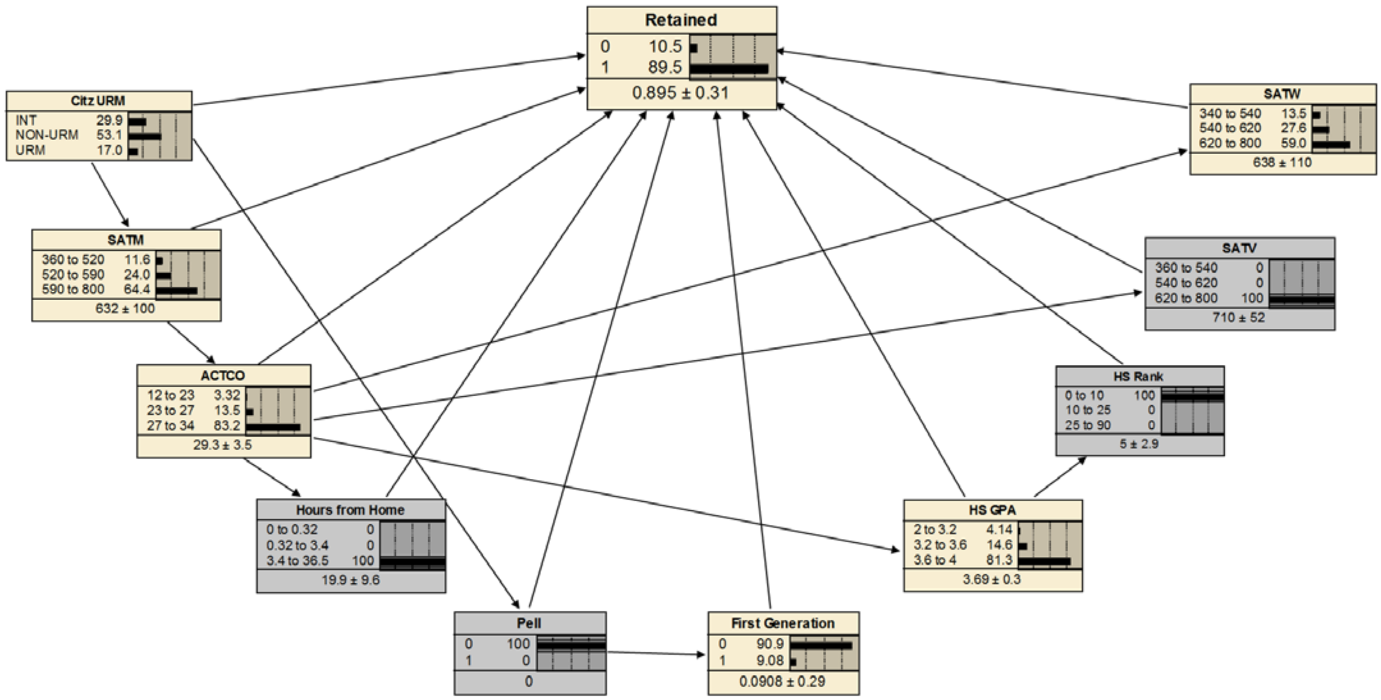

Figure 2.3. "Straw Man" BN Scenario 2. A student is known to be Pell-ineligible (coded "0"), is more than 3.4 hours from home, was ranked within the top 10% of their high school class, and scored at least a 620 on the SAT English component. Note that probability of retention increased from 81.7% to 89.5%.

Model evaluation. As part of the structure and parameter learning process, the data used in the initial "straw man" BN were split into subsets for cross-validation. Cross-validation is often used to evaluate the predictive accuracy of models, and is built upon the premise of partitioning data so that a model can be learned from one data set (the training set) and the resulting derived model's predictive accuracy be evaluated against the remaining data (the testing set) (Geisser, 1975; Stone, 1974). Many approaches to cross-validation appear in the literature, with general advice that the method chosen best represent the research goals and data characteristics while minimizing the trade-off between complexity and performance (Hastie, Tibshirani, & Friedman, 2009; Morrison, Bryant, Terejanu, & Miki, 2013). Given this study's context and objectives, the researcher employed the guidance of other studies incorporating machine-learning and Bayesian networks (see Alqallaf & Gustafson, 2001; Fienen & Plant, 2014; May, Maier, & Dandy, 2010 as examples) and used *k*-fold cross-validation for this model and all future models. *K*-fold cross-validation involves randomly splitting cases into *k* equally-sized partitions, cross-validating each partitioned sample across the remaining partitions, and then averaging predictive performance across all partitions. A main advantage of *k*-fold cross validation is that it allows for the use of as much training data as possible while protecting against model overfit and providing measurements of predictive performance. Additionally, when *k* is greater than two but also not too large, *k*-fold cross-validation at least partially addresses the "bias/variance" dilemma described by Geman, Bienestock, and Doursat (1992), in which minimization of potential bias and prediction error created by an inappropriate data split competes with the minimization of variance that is created by using a number of training sets to estimate a model's parameters. Citing Breiman and Spector (1992) and Kohavi (1995), Hastie et al. (2009) recommended five- to ten-fold cross-validation as a bias/variance dilemma "compromise" (p. 243), and due to a somewhat limited amount of cases (n=729 in cohorts 2009-2011 and n=709 in cohorts 2012-2014), this study employs five-fold cross-validation in the "straw man" BN and all future BNs.

Methods for averaging confusion matrices resulting from each fold as described by Marcot (2012) and Boyce, Vernier, Nielsen, & Schmiegelow (2002) were used to estimate overall model predictive performance using a Confusion Matrix as presented in Table 2.1. In terms of accuracy, the "straw man" model predicted likelihood of retention correctly in 79.7% (570/715) of cases. In terms of misclassification, the model incorrectly predicted retention one-fifth (20.3%) of the time.

Table 2.1
Confusion Matrix and Error Rate – "Straw Man" BN

Predicted		
Retained	Not Retained	Actual
566	19	Retained
126	4	Not Retained

A Receiver Operating Characteristic curve (ROC) is another way to gather information about the predictive capability of the "straw man" BN. Figure 2.4 presents a ROC curve related to the "straw man" BN, where the model's percent of true positive predictions (a measure of sensitivity) is plotted against the percent of false positive predictions (a measure of specificity) at different threshold values. The area between the BN curve and the straight dashed line (a ROC curve representing a completely uninformative model) represents the difference in an uninformative model and the more predictive "straw man" BN at different thresholds. From shape alone, it is clear that the "straw man" BN is more predictively accurate at lower thresholds. Further, the area under the "straw man" BN ROC curve (AUC) is calculated to be .55089, a value that is considered a poor measure of model prediction performance (Hand, 1997).

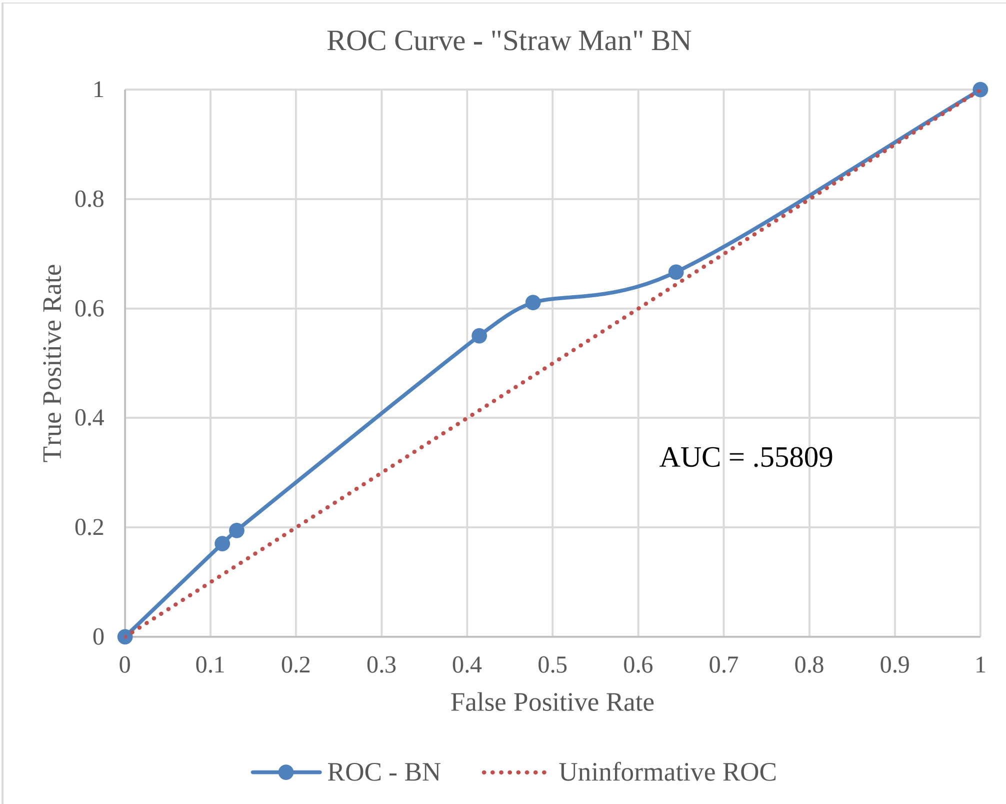

Figure 2.4. “Straw Man” BN ROC Curve. This figure illustrates the model’s predictive accuracy at different thresholds (solid line), as compared to an uninformative model (dashed line), as well as the AUC.

The 20% error rate and the small AUC hint at the fact that there is a large amount of uncertainty due to influences unaccounted for in the network. The sensitivity analysis presented in Table 2.2 supports this suggestion, in that it demonstrates the degree to which variation in retention likelihood is explained by the other included variables (Marcot, Steventon, Sutherland, & McCann, 2006). The Mutual Info column demonstrates the expected decrease in uncertainty (as expressed by entropy) in retention likelihood given a state of another variable included in the model. The most influential variable on likelihood of retention is a student’s SAT Reading score, but even knowing this score will only decrease uncertainty by 3%.

Table 2.2

Sensitivity Analysis of "Retained" to Other Variables Included in "Straw Man" BN

Variable	Mutual Info	Percent	Variance of Beliefs
SATV	.0206	2.99	.0046
SATW	.0179	2.59	.0037
SATM	.0158	2.27	.0033
Citz URM	.0116	1.68	.0025
HS Rank	.0093	1.36	.0018
ACTCO	.0068	.99	.0015
HS GPA	.0038	.56	.0008
Hours from Home	.0028	.40	.0006
First Generation	.0011	.15	.0002
Pell	.0003	.04	.0001

In addition to evaluating a model's predictive performance and sensitivity, a final evaluative measure involves considering model complexity. Complexity can be measured by the number of variables, links, and node states, and is typically used for comparing different models (Marcot, 2012). However, it is also helpful to consider model complexity as part of a holistic evaluation of single models in an effort to examine variable connectivity and dependence. Additionally, complexity is not a necessary condition for reliability or additional insight, so parsimony should receive priority (Jakeman & Hornberger, 1993). The "straw man" BN's complexity measures are summarized in Table 2.3, and indicate a not particularly complex model.

Table 2.3

"Straw Man" BN Model Complexity Metrics

Metric	Count
Nodes	11
Links	19
Node States	30

In summary, although the "straw man" BN can be described as parsimonious based on measures of complexity, evaluating the "straw man" BN in terms of predictive performance and sensitivity reveals a weak model. This was not entirely unexpected for a number of reasons: 1) due to a lack of available non-cognitive quantitative data available, the model incorporated very few of the variables often cited in literature, 2) the model was allowed to be machine-learned with essentially no supervision (e.g. automatic discretization of continuous variable node states), and, 3) the model excludes institution-specific or contextual variables and constraints recommended by retention experts. However, the machine-learned “straw-man” model does provide a starting place for comparison and insight for future model iterations.

Structure elicitation. An important aspect of this research is the combination of machine-learning and expert opinion in order to build a BN that accurately and efficiently predicts retention. The first step in incorporating expert opinion was to query retention experts regarding their beliefs about the structure of a graphical network that predicts retention, which was then compared to the machine-learned structure of the “straw man” BN in the creation of an interim model. The following section outlines this process of structure elicitation from the experts, and incorporates the recommendations of Fenz (2012) to emphasize consistent operating definitions and O’Hagan et al. (2006) in the design of the structure elicitation protocol.

Experts. The experts participating in the session have a combined twenty-five years of experience at the institution working directly with students. Expert A is an Assistant Dean of the College and Director of Academic Advising who has been with the institution for fifteen years, and Expert B is an Associate Director of Academic Advising who has been at the institution for ten years. Both individuals have extensive personal experience with the reasons first-year students leave the institution, as well as domain knowledge regarding industry-wide retention issues.

Variable elicitation. As a first step in elicitation of the expert-elicited structural BN, the researcher presented the experts with a list of variables identified in the literature as related to undergraduate retention. The experts were also provided with proposed operating definitions and variable classes. An example of what the experts were asked to review is provided in Appendix B. After providing insight on the inclusion or exclusion of certain variables based on relevance to the institution, the experts offered revisions to operating definitions and the groupings (variable classes) of states of variables. The experts also proposed new variables that, in the experts' experience through working with and counseling students at risk of attrition, are important predictors of retention that were not identified in the literature.

Table 2.4 summarizes the list of variables agreed upon by the experts. In their discussion of race/ethnicity or first-generation status as predictors of retention, the experts noted that while there is nothing about race/ethnicity or fist-generation status alone that influences retention alone, it is highly related to stereotype threat. For example, the experts agreed that women of color are more likely to succumb to stereotype threat, and are therefore less likely to seek and access academic support and tutoring. This idea is also supported in the literature – see Steele (1997) and Aronson & Steele (2005) as examples. The experts also proposed two variables related to academic support as influences of retention, while emphasizing that institutional emphasis on academic support (as measured by spending or resource allocation to academic support programming) is separate and distinct from whether or not a student actually accessed such academic support programming. Institutional emphasis on increasing retention was also noted as an influence on retention, and the experts identified a shift in institutional focus on retention occurring in the year 2011 as evidenced by the hiring of personnel charged with addressing retention issues. Experts noted that students' study skills (time management, academic discipline, etc.) and course attendance patterns, while potentially related to each other, should also be included as exclusive variables in the retention model. See Allen, Robbins, Casillas, and Oh (2008), Kennett and Reed (2009), and Seo (2012) for discussions of these variables and their influence on retention in the literature. Note that the experts first recommended that the study skills and attendance variables should be used in place of other commonly employed academic preparedness variables (standardized test scores, high school academic records), opining that the institution's students are typically prepared intellectually and academically, but may lack the confidence or study skills necessary for academic success. After extensive discussion, the experts agreed that a composite variable representing academic preparation should be employed – the "Academic Rank" variable in Table 2.4 is such a composite variable and incorporates high school quality, high school academic record, and standardized test scores. In terms of social support and its influence on retention, the experts cited Eckles' and Stradley's (2011) analysis of social networks on student retention and affirmed the relevance of the study's conclusions that the negative feelings of one student act as a contagion towards others while students' positive feelings can act as an inoculation against negativity, albeit in a weaker capacity. The experts discussed the need for including variables addressing students' financial need, noting that financial issues and a lack of financial aid literacy are more likely to influence retention when other factors, such as academic struggle, are present as well. Additionally, the experts argued that Pell-eligibility, a commonly cited influence on retention, is not an appropriate proxy for students with financial issues at the institution as one-half of the institution's students typically receive Pell. Anticipating that students who may struggle academically and have high levels of unmet financial need have very little "wiggle room" for financial or academic shortfalls, the experts concluded a financial variable should be included that addresses students with high unmet need, while also encompassing academic risk, an overall need profile, and evidence of financial aid literacy. In a finding contradictory to much of the literature, the experts expressed an opinion that students with homes closer to the institution (particularly within the institution's metro area) are less likely to retain. Additionally, noting a commonly expressed reason for leaving during student exit interviews, the experts also recommended a variable capturing mismatch between a student's first-year academic advisor and the student's expressed program of study interest.

Table 2.4
List of Variables Proposed and Approved by Experts in Structure Elicitation (Elicitation Session I)

Variable	Operating Definition	Variable Classes	Weight 1=Highest 10=Lowest	Quantitative Data Availability/Measure
Study Skills	Whether or not a student demonstrates good study skills – successful time management, minimal procrastination, adequate note-taking and review, course engagement, etc.	Developed UnderDeveloped	1	Data proxy: Hours reported spent studying on institutional survey (2011 and later cohorts only)
Financial Need & Risk Profile	Encompasses students' academic preparedness, high financial need, unmet need, and understanding of financial aid literacy.	High (High Need/High Risk Profile) Low (Low Need/Low Risk Profile)	2	Data available.
Social Support	The extent to which a student is exposed to negative or positive attitudes towards retention within their social network.	Positive social support Negative social support	3	Data proxy: First-year housing placement within a dorm floor with unusually high first- to second-year attrition.

Academic Support – Student Access	Whether or not a student accessed academic support programming.	Accessed Not Accessed	4	Data Not Available.
Academic Support – Institutional	Whether or not an institution provides more than a nominal amount of academic support programming.	Very Present Less Present	5	Data Proxy: Receipt of grant funding development of academic resource center in 2011.
Attendance Patterns	Whether or not a student consistently attends scheduled course meetings.	Consistent - attends > 80% of course meetings Inconsistent - attends < 80% of course meetings	6	Data not available.
Academic Rank	Composite index value to represent student's academic preparedness. Incorporates high school quality, curriculum quality, high school academic record (GPA and rank) and standardized test scores.	Lowest Low Medium High Highest	7	Data available.
Race/Ethnicity	Self-reported, federally-defined race/ethnicity categories.	Underrepresented minority International Other	8	Data available.

Institutional Focus on Retention	Institution-wide implementation of programs, services, and resources designed to address retention.	Very Present Less Present	9	Data proxy: "Less Present" if student enrolled before 2011. "Very Present" if student enrolled in 2011 or after.
Distance from Home	Distance, in hours, of institution from student's permanent home address	Within one hour More than one hour	10	Data available.
Advisor/Major Mismatch	Assignment of academic advisor belonging to an academic department outside the student's expressed program of student interest.	Advisor/Major match Advisor/Major mismatch	11	Data available.

Structure elicitation. In preparation for drafting a basic structural model of a retention BN, experts were next asked to discuss the relationships between and among the chosen list of variables. Experts were first asked to weight each of the variables in terms of influence on retention, beginning with strongest and weakest and moving inward from there (see Kjærulff & Madsen, 2008). These ranks are included in Table 2.4, and were elicited in an effort to introduce the idea of causal influence to the experts and ultimately be incorporated in the specification of model parameters and CPTs in development of a final model (Netica allows for the inclusion of uncertainty using a special case file format). Using the weightings as reference, the researcher prompted the experts to discuss which variables could be considered direct causes of retention, and which variables actually influence other variables and should be considered indirect causes of retention. The discussions of causal influence were included to provide insight as to the model's appropriate structure and variables' relationships to each other. The researcher also facilitated discussion of conditional independence, asking experts to consider if a student's state on one variable reveals a large amount of information on how this same student might appear in another variable. Examination of variable dependencies were included in order to aid the researcher and experts in identifying redundant variables, assure conditional independence among included variables, and ultimately contribute to model parsimony. As the representation of uncertainty is a unique feature of BNs, experts were continuously encouraged to express their uncertainty in any of these discussions. Any instances of high uncertainty were ultimately included in the specification and evaluation of the final model.

The literature consistently includes training experts as a best practice in expert elicitation, and in order to provide context and demonstrate the basic components and function of a predictive BN, the researcher presented the experts with a BN that predicts the likelihood of coronary artery disease based on a number of symptoms (Figure 2.5). The training BN was used to provide examples of parent/child nodes, leaf nodes, and conditional independence. Experts were encouraged to consider the variables influencing retention similarly to the symptoms or conditions modeled to influence risk of coronary artery disease, and to note the direct and indirect relationships between symptoms and disease. Additionally, the researcher demonstrated the dynamic updating capability of a Netica BN, providing an example of the utility and ease in which a BN can be used to predict retention given certain conditions or new cases.

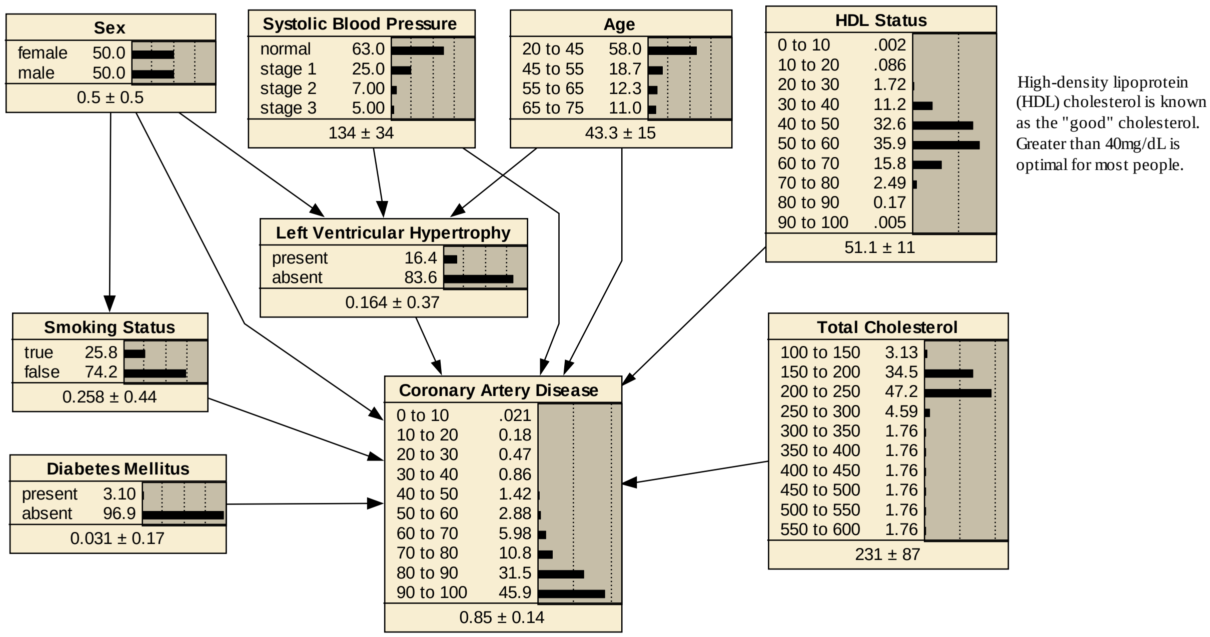

Figure 2.5. Training BN: Coronary Artery Disease Risk Estimator, *Norsys Software Corp. Bayes Net Library*, http://www.norsys.com/netlibrary/index.htm . Copyright 2004 by Assessment Technologies, Inc.

Finally, the experts were asked to draft a basic structural model of retention incorporating the variables and the relationships between and among the variables and retention. The researcher reminded the experts of earlier discussions and decisions regarding causal influence, conditional independence, and weight in the formation of the structure, and allowed the experts to collaborate in the design although they ultimately drafted their own versions. Throughout the drafting process, the researcher verbally articulated scenarios depicted in the structure in order to ensure that the structure accurately represented the experts' judgments. A composite of the experts' structure designs is provided in Figure 2.6.

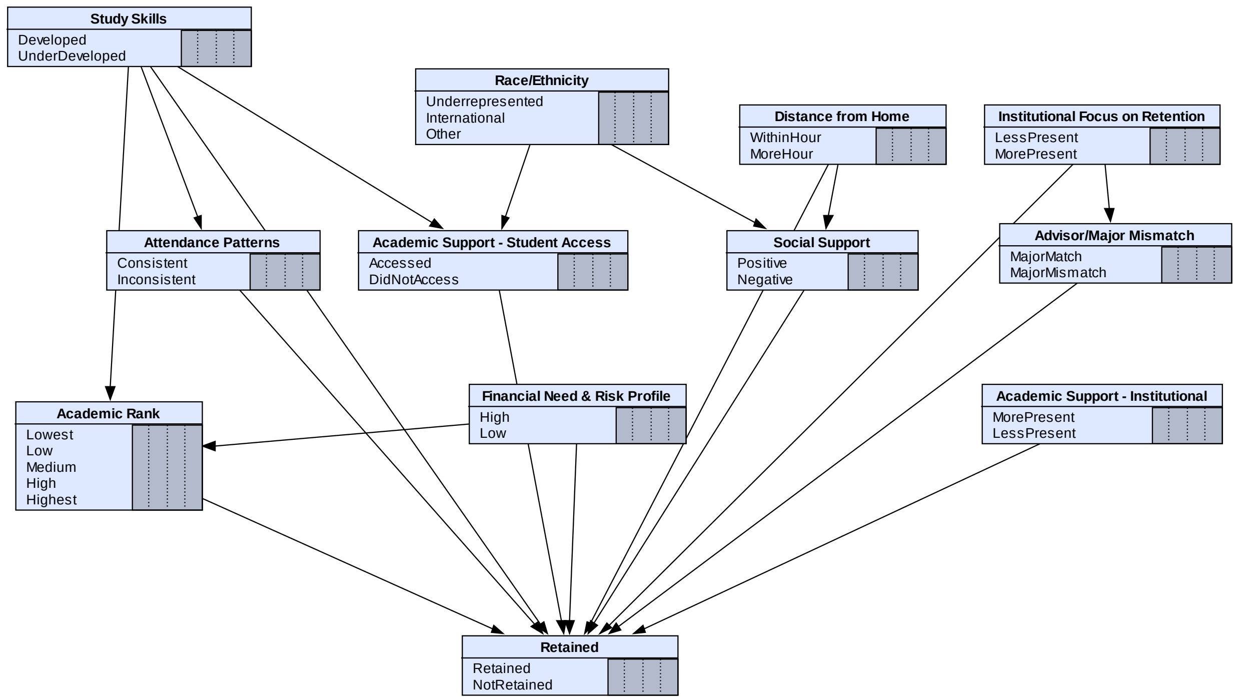

Figure 2.6. Expert-elicited structural model.

In summary, the expert-elicited structural BN includes twelve nodes. All but one ("Race/Ethnicity") are portrayed as direct influences on retention, although five ("Study Skills," "Race/Ethnicity", "Distance from Home," "Institutional Focus on Retention," and "Financial Need & Risk Profile") are also thought to be influences on other included variables. Table 2.5 summarizes the model's complexity, which is very similar to the "straw man" BN.

Table 2.5
Expert Elicited Structural Model Complexity Metrics

Metric	Count
Nodes	12
Links	18
Node States	28

Development of interim model. Development of an interim model designed to incorporate the insights gathered from the original "straw man" data- and structure-learned model as well as the expert-elicited model involved the comparison and examination of these models' structures, variables, and data-learned parameters. Ultimately, this second model was designed as an interim stage in preparation of creation of a final model that includes a hybrid of expert-elicited and data-learned parameters and structure.

Comparison of "straw man" and expert-elicited structural model. Cursory comparison of the data- and structure-learned "straw man" model and the expert-elicited structural model revealed few commonalities in included nodes. For example, the experts dismissed the importance of including distinct nodes for generally accepted proxies for academic preparedness (high school academic records and standardized test scores) in a model, opining that while the institution's students are academically and intellectually prepared, academic success and retention at the institution is much more dependent on students' study skills. However, the experts did endorse the inclusion of a composite variable that incorporates a number of indicators of academic preparedness ("Academic Rank"), noting that, in addition to having a direct influence on retention, it is also influenced by study skills. Examination of the links learned in the "straw man" model support the inclusion of the composite "Academic Rank" variable in a second model– many of the academic preparedness variables were learned to be related to each other in the "straw man" model and the combination of them into one variable creates a more parsimonious model. Expert discussion of demographic variables included in the "straw man" model revealed experts' uncertainty of any direct role played by race/ethnicity on retention while highlighting their belief that race/ethnicity is involved in other important causes of retention. Hence, this variable was still included in a second model but was not shown to have a direct influence on retention. Additionally, the "straw man" model included a variable depicting students' socioeconomic status as determined by eligibility for Pell grants, and indicated that socioeconomic status is closely linked with other demographic variables. The experts' opinion that the role of finances is much more complicated than simple socioeconomic status and ultimate suggestion of a composite financial risk/need variable addresses the confounding and redundant influence of demographics on retention expressed in the “straw man” model. Finally, as evaluation of the “straw man” model revealed a relatively weak predictive model with unaccounted for influences, all other variables suggested by the experts were included in the interim model under the expectation that the experts’ familiarity with institution-specific factors is superior to the literature-identified variables on which quantitative data were available that were used to build the “straw man” model.

Parameterizing interim model. Using Netica, the researcher prepared a second, interim BN that incorporated the structural insights provided by experts and was parameterized by available existing data (2009-2011 first-year cohort data). Two of the variables included did not have associated existing data (“Attendance Patterns” and “Academic Support-Student Access”) and other variables included numerous missing data. For example, data indicating whether or not a student exhibits study skills (as measured by responses on an institutional questionnaire) were only available for students in the 2011 cohort. In order to address these incomplete data, Netica allows parameter learning via an expectation-maximization (EM) algorithm where maximum *a posteriori* parameter estimates are computed using maximization of the expected log-likelihoods of parameters after a number of iterations (see Dempster, Laird, & Rubin, 1977). In other words, the EM algorithm is used to maximize the probability of data given the BN’s structure and CPTs. The resulting model reflected the current state of each of the variables for which data were available and showed the two unknown variables (“Attendance Patterns” and “Academic Support – Student Access”) as having uniform prior probability distributions, or equally likely states. The model also included the data-learned CPTs for each variable when data were available. As part of design of a final model, the parameters and associated CPTs were ultimately presented to the experts in a second elicitation session for review, critique, and elicitation of the unknown parameters.

Development of final model. Development of the final model included the following stages: Expert review of the interim model, formal elicitation of unknown parameters, parameterization of the final model, and evaluation of final model performance.

Expert review of interim model. Expert review of the interim model began with evaluation of the included variables' operating definitions. Noting that the two institutional variables ("Academic Support-Institutional" and "Institutional Focus on Retention") are irrelevant for any post-2011 cohorts as both variables served as indicators of a shift in institutional priority to increasing retention beginning in 2011 and continuing forward, the experts recommended exclusion of these two variables from any final model. The experts also discussed the quantification of "consistent" and "inconsistent" attendance patterns, confirming that attending 80% or more course meetings is appropriately considered "consistent." Additionally, the experts considered the "Academic Support-Student Access" variable, focusing on what activities should be considered "academic support programming" and the frequency at which student access of such programming would begin to influence retention likelihood.

The researcher next introduced a comparison of the variable influence rankings established by the experts in the first elicitation session against the rankings suggested by sensitivity analysis of the data-learned parameters in the second/interim model. The only major difference in the relative importance suggested by the experts and reflected in the data-learned parameters of the interim model was the influence of "Academic Rank," with the data-learned parameters suggesting retention was more sensitive to academic preparedness than the experts anticipated. Discussion of the ranks was included in an effort to gage experts' uncertainty with specific variables that would be incorporated into the final model. For example, a conflict between expert-understood rank and data-learned rank could be addressed and mitigated by using Netica's uncertain case file format in development of the final model. The final list of included variables and their related operating definitions, ranks, variable classes and data sources are summarized in Table 2.6.

Table 2.6
List of Variables Operationalized and Approved by Experts in Model Review and Parameter Elicitation Session (Elicitation Session II)

Variable	Operating Definition	Variable Classes	Weight 1=Highest 10=Lowest	Quantitative Data Availability/Measure
Study Skills	Whether or not a student demonstrates good study skills – successful time management, minimal procrastination, adequate note-taking and review, course engagement, etc.	Developed UnderDeveloped	1	Expert estimations.
Academic Rank	Composite index value to represent student's academic preparedness. Incorporates high school quality, curriculum quality, high school academic record (GPA and rank) and standardized test scores.	Lowest Low Medium High Highest	2	Data available.

Social Support	The extent to which a student is exposed to negative or positive attitudes towards retention within their social network.	Positive social support Negative social support	3	Data proxy: First-year housing placement within a dorm floor with unusually high first- to second-year attrition.
Financial Need & Risk Profile	Encompasses students' academic preparedness, high financial need, unmet need, and understanding of financial aid literacy.	High (High Need/High Risk Profile) Low (Low Need/Low Risk Profile)	4	Data available.
Academic Support	Whether or not a student accessed academic support programming. Academic support programming includes student-initiated meetings with instructors or attendance at academic resource center programming.	None Low – Student attended at least one and less than five self-initiated meetings with instructor or other academic support programming. High – Student attended five or more self-initiated meeting with instructor or academic support programming.	5	Expert estimation.
Attendance Patterns	Whether or not a student consistently attends scheduled course meetings.	Consistent - attends > 80% of course meetings Inconsistent - attends < 80% of course meetings	6	Expert estimation.

Distance from Home	Distance, in hours, of institution from student's permanent home address	Within one hour More than one hour	7	Data available.
Advisor/Major Mismatch	Assignment of academic advisor belonging to an academic department outside the student's expressed program of student interest.	Advisor/Major match Advisor/Major mismatch	8	Data available.
Race/Ethnicity	Self-reported, federally-defined race/ethnicity categories.	Underrepresented minority International Other	N/A (no direct influence on retention)	Data available.

Formal elicitation of conditional probabilities. As noted earlier, two of the variables included in the interim model did not have associated available quantitative data ("Attendance Patterns" and "Academic Support") and one variable deemed particularly important by previous models and experts ("Study Skills") contained substantial missing data. Consequently, the parameters associated with these variables (their conditional probabilities) were elicited from the two experts using formal and rigorous methodology designed to accurately represent the experts' knowledge and minimize bias. Elicitation techniques followed the best-practice guidelines outlined in Kadane and Wolfson (2008) and O'Hagan et al. (2006), and utilized materials and software from an expert elicitation framework (SHELF: the Sheffield Elicitation Framework version 2.0) designed by Oakley and O'Hagan (2010). Kadane and Wolfson described a successful elicitation as one in which the researcher assures the process is "as easy as possible for subject-matter experts to tell us what they believe, in probabilistic terms, while reducing how much they need to know about probability theory to do so" (p. 4), and this idea was used as a guiding principle throughout the session with the experts.

Prior to the session, the experts were provided with pre-elicitation materials describing the purpose and objectives of the upcoming elicitation session, as well as a basic probability review and summary of common causes of bias. The pre-elicitation materials also emphasized the important role of uncertainty throughout the process. The session began with expert review of the interim model and its included variables and related operating definitions and rankings (see earlier discussion), but the majority of the session focused on the elicitation of probabilities. In order to set the stage for the session's discussions and ensure experts' proper understandings of basic probability theory, the researcher first posed a practice elicitation where a known probability was elicited (the retention rate for the 2014 cohort) and the experts were asked to describe what that known probability meant in terms of how many students stay, how many leave, etc. The expert also posed a known retention scenario involving conditional probability (the likelihood of retention given Pell-eligibility) to gauge and prompt discussion of experts' understanding of conditional probabilities. Based on expert responses and explanations, the researcher determined the experts were prepared to provide estimations of conditional probabilities for the three variables in question.

In accordance with the SHELF materials and guidance, experts were first asked to estimate the extreme lower and upper bounds of an overall retention rate, and move inward to more likely rates from there. Oakley and O'Hagan (2010) posited that this technique mitigates experts' overconfidence and encourages experts to consider models outside of what they're most familiar. In addition to these extreme lower and upper bounds of overall retention rates, the experts provided “medium”, “high,” and “low” estimations of an overall retention rate on which they would base their judgements throughout the rest of the session. As with all estimations during the session, the researcher consistently provided feedback (“…given Z, you’re suggesting that out of X students, Y wouldn’t return the following fall semester…”) and encouraged experts to express their level of confidence and uncertainty in their conclusions. Additionally, the researcher remained mindful of incoherence of judgements, and was prepared to ask experts to account for any such incoherence.

For each variable on which elicitation took place, experts were first asked to estimate the current state of each variable given no other information. For example, as no data were available regarding how many students attend courses consistently, experts were first asked to estimate the percentage of students consistently attending courses. This information was entered into a pre-designed spreadsheet containing conditional probability tables and would ultimately be used to simulate the underlying data distributions used in development of the final model. Experts were next asked to provide judgements regarding the conditional probabilities for each variable in question. This topic served as the most intensive in the session, as the number of conditional probabilities required grew as the relationships between variables grew. For example, based on the interim model, a full CPT for “Attendance Patterns” involves “Study Skills” and “Retention,” while a full CPT for “Academic Support” involves “Race/Ethnicity,” “Study Skills,” and “Retention.” Finally, experts were asked to provide predictions of the probability of retention given a number of hypothetical variable scenarios, and their responses were later used as a measure of the final model’s internal validity.

Expert discussion during the conditional probability elicitations revealed a number of insights related to model. For example, the experts expressed a large amount of uncertainty regarding the actual amount of students accessing academic support, but were much more confident in estimating the role of accessing academic support on retention. The experts also openly expressed difficulty avoiding availability bias, especially in terms of allowing particularly memorable student retention scenarios to overpower more typical situations. Additionally, the experts identified certain probabilistic conditions that, while puzzling to an outside viewer, are specifically relevant and unique to the institution. For example, given the experts' conclusion that students' study skills exact a heavier influence on likelihood of retention than students' attendance patterns, once would expect a pattern of probability in which, no matter the level of attendance, a student with developed study skills is more likely to retain. However, the experts provided a probability distribution in which students with developed study skills who inconsistently attend class are less likely to retain than students with underdeveloped study skills who do consistently attend (see Table 2.7). The experts described this as a situation unique to the institution, noting that inconsistent course attendance of a student with developed study skills is an indicator of a larger, more significant problem that will ultimately lessen the likelihood of retention. It is this type of expert-identified situation that speaks to the value of including expert opinion in model development. If data surrounding these variables were available, this type of pattern might be viewed as an anomaly without the experts' insights. Without data or the experts' opinions, this particular situation may never have been recognized or represented in a model. See Appendix C for more detail concerning probability elicitation techniques and protocol.

Table 2.7

Elicited Probability of "Retention" Given "Study Skills" and "Attendance Patterns:" Example of Unanticipated Probability Patterns

Study Skills	Attendance Patterns	p(Retained)	p(Not Retained)
Developed	Consistent	86%	14%
UnderDeveloped	Consistent	83%	17%
Developed	**Inconsistent**	**79%**	**21%**
UnderDeveloped	Inconsistent	77%	23%

Parameter estimation. After elicitation and collection of conditional probabilities on the variables for which no or very little quantitative data were available, the researcher employed a hybrid approach to parameterization of the final BN's nodes. All variables and structure recommended by the experts were maintained in the final model as all of the most influential variables included in the original "straw man" model were also somehow represented in the expert-generated design (e.g., standardized test scores are incorporated into "Academic Rank" variable). Further, the variables recommended by the experts were also found to have theoretical underpinnings in the larger retention literature. Heretofore unused 2012-2014 cohort data were used to parameterize seven of the ten nodes for which data were available. Following the guidance of Woodberry, Nicholson, Korb & Pollino (2005) and Pollino, Woodberry, Nicholson, Korb, & Hart (2007), data were simulated through Netica to mirror the conditional probabilities elicited from the experts for the unknown variables and then manually input into the CPTs using Netica. The final BN incorporating expert-designed structure and a hybrid data-learned/expert-learned approach to parameterization is shown in Figure 2.7. Note that this figure only shows the current state of each variable – the dynamic and predictive nature of the model is viewed through Netica when adding new case information.

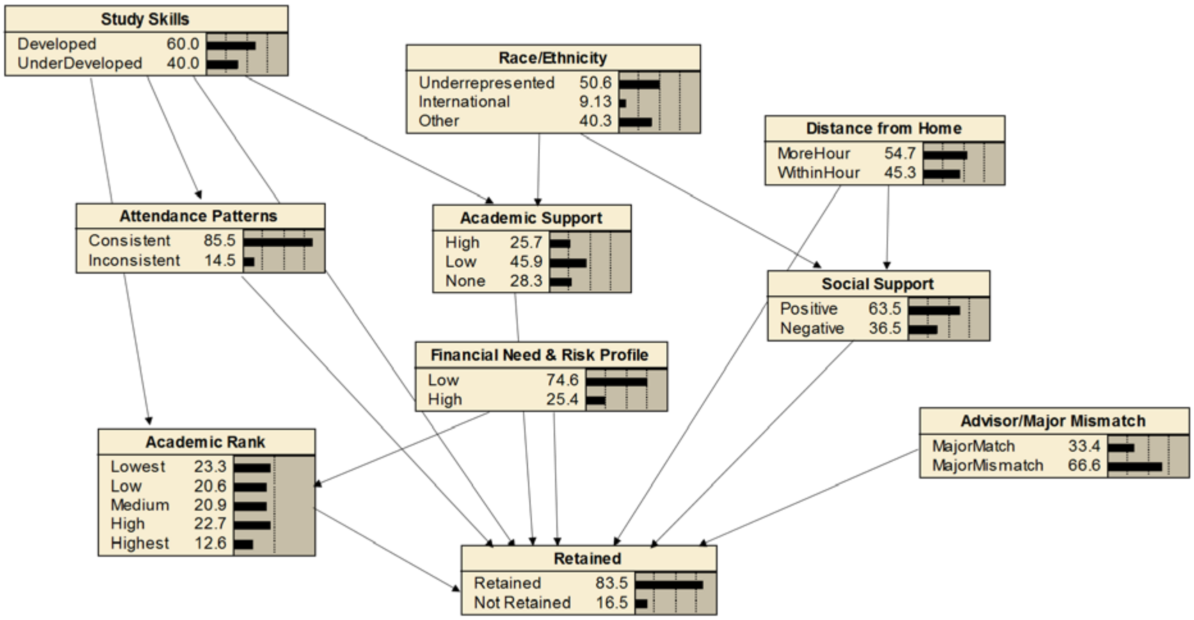

Figure 2.7. Final Retention BN. Incorporates hybrid expert- and data-learned construction.

For example, Figure 2.8 demonstrates the effect on predicted retention for a hypothetical student with low levels of social and academic integration. The student's degree of social integration is reflected by her negative standing on "Social Support" and her academic integration is reflected by inconsistent "Attendance Patterns" and no access of "Academic Support." Viewed within the context of Tinto's model of retention (the degree to which a student is integrated into an institution's social and academic framework is positively related to her likelihood of retention) (Tinto, 1975), this student reflects a high risk of attrition. This is corroborated by the BN, as predicted retention drops from 83.5% to 74.4%. Table 2.8 summarizes the model's retention prediction for a number of scenarios within Tinto's framework. The final BN suggests that the included population adheres to Tinto's model in that higher degrees of academic and social integration lead to higher likelihood of retention. Additionally, reviewing the final BN's performance in the context of a commonly accepted and cited theory of undergraduate retention illustrates the relevance of the model.

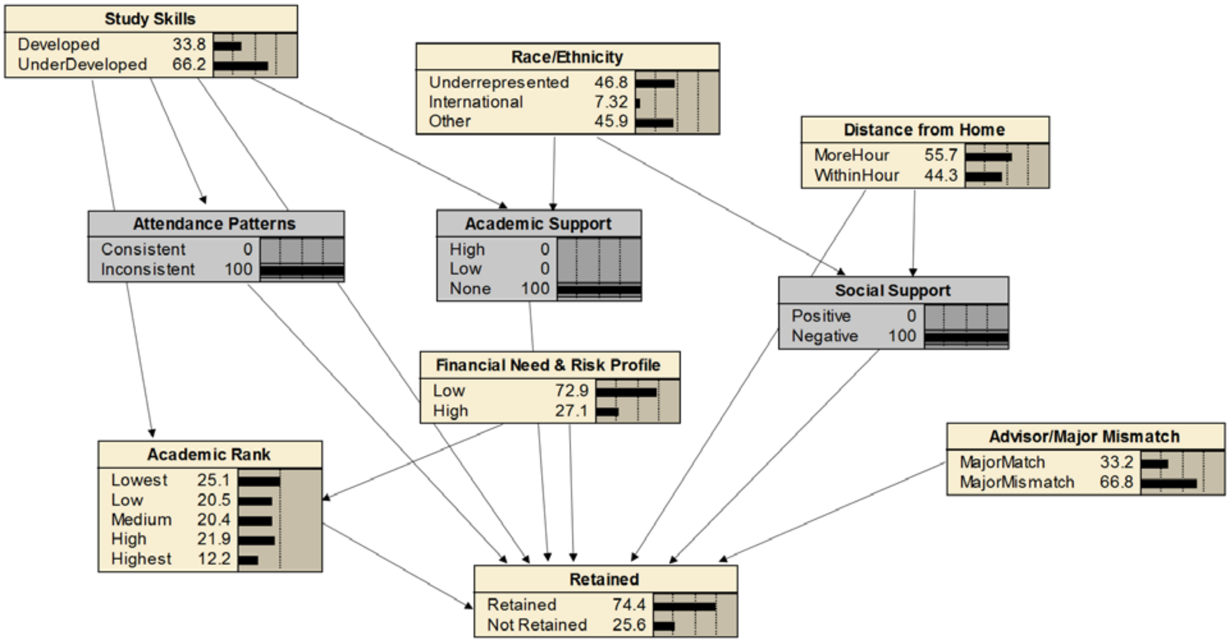

Figure 2.8. Final Retention BN, Tinto high attrition risk. Hypothetical student demonstrates low academic integration (inconsistent course attendance and no access to academic support) and low social integration (negative social support). Under Tinto's model of retention, this student represents a high risk of attrition. Note that probability of retention has decreased from 83.5% to 74.4%.

Table 2.8

Retention Effects of Variables Related to Tinto's Model

Academic Integration		Social Integration		
Attendance Patterns	Academic Support	Social Support	p(Not Retained)	p(Retained)
Consistent	High	Positive	11.0%	89.0%
Inconsistent	High	Positive	16.8%	83.2%
Consistent	High	Negative	13.3%	86.7%
Inconsistent	High	Negative	20.1%	79.9%
Consistent	Low	Positive	15.6%	84.4%
Inconsistent	Low	Positive	25.3%	74.7%
Consistent	Low	Negative	18.8%	81.2%
Inconsistent	Low	Negative	29.6%	70.4%
Consistent	None	Positive	14.9%	85.1%
Inconsistent	None	Positive	21.7%	78.3%
Consistent	None	Negative	17.9%	82.1%
Inconsistent	None	Negative	25.6%	74.4%

Results

As with the original “straw man” model, 5-fold cross-validation was employed in order to evaluate the predictive accuracy of the final model. The confusion matrix for the model is presented in Table 2.9. Over the five folds, the final model predicted likelihood of retention correctly in 83.5% (592/709) of cases. The model misclassified cases 16.5% of the time. This measure of overall model performance suggests a more accurate model than the original “straw man” network. However, a closer look at some of the additional measures of performance that can be calculated using the confusion matrix (Table 2.10) reveals that the apparent success of the model is tempered by other factors. For example, while the model’s sensitivity, or true positive rate, is high at 98% (when a student is actually retained, the model predicts retention 98% of the time), the model’s specificity or true negative rate (when a student is not actually retained and the model predicts attrition) suffers at merely 5% (6/116). Similar issues with specificity and false positive rates were present in the original “straw man” model, and are most likely due to the highly imbalanced class distribution within the dataset. Within this population, the vast majority of students typically retain – it is a “rare” event that students do not retain and such imbalanced datasets lead to model overfit and overstated predictive accuracy (Chawla, 2010). While this doesn’t negate the usefulness of a model, it does require the consideration of other evaluative measures.

Table 2.9
Confusion Matrix and Error Rate – Final BN

Predicted		
Retained	Not Retained	Actual
586	7	Retained
110	6	Not Retained

Table 2.10

Common Performance Metrics – Final BN and Original "Straw Man" BN

Metric	Calculation	Final BN	"straw man" BN
Accuracy	(TP + TN)/All Cases	83.5%	79.7%
Misclassification	1-Accuracy	16.5%	20.3%
Sensitivity/Recall/True Positive Rate	TP/(TP+FN)	98.8%	96.8%
Specificity/True Negative Rate	TN/(FP+TN)	5.2%	3.1%
Fall-out/False Positive Rate	FP/(FP+TN) or 1-Specificity	94.8%	96.9%
Precision/Positive Prediction Value	TP/(TP+FP)	84.2%	81.8%
F-value (combined measure of precision and recall using harmonic mean)[a]	$\frac{1+\beta^2 * recall * precision}{\beta^2 * recall + precision}$	90.9%	88.6%

Note. TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative
[a]For F-value, see Buckland & Gey (1994), β represents the importance of precision relative to recall and is typically set to 1.

One such measure is the ROC curve, as it presents transparent information about the model's performance in predicting the minority class and is not dependent on class distributions (Kotsiantis, Kanellopoulos, Pintelas, 2006). Figure 2.9 presents a ROC curve showing the final BN's tradeoff between sensitivity and specificity. While the AUC (.62446) is slightly higher than that found in the "straw man" model, it is still low enough to indicate that the model's sensitivity, or ability to correctly predict retention, is only slightly larger than the model's *inability* to correctly predict attrition. In other words, the ROC shows that the model's true positive rates are similar to its false positive rates, suggesting that the model is not necessarily discriminating between retention and non-retention and that it's simply mirroring the high prevalence of retention within the underlying data.

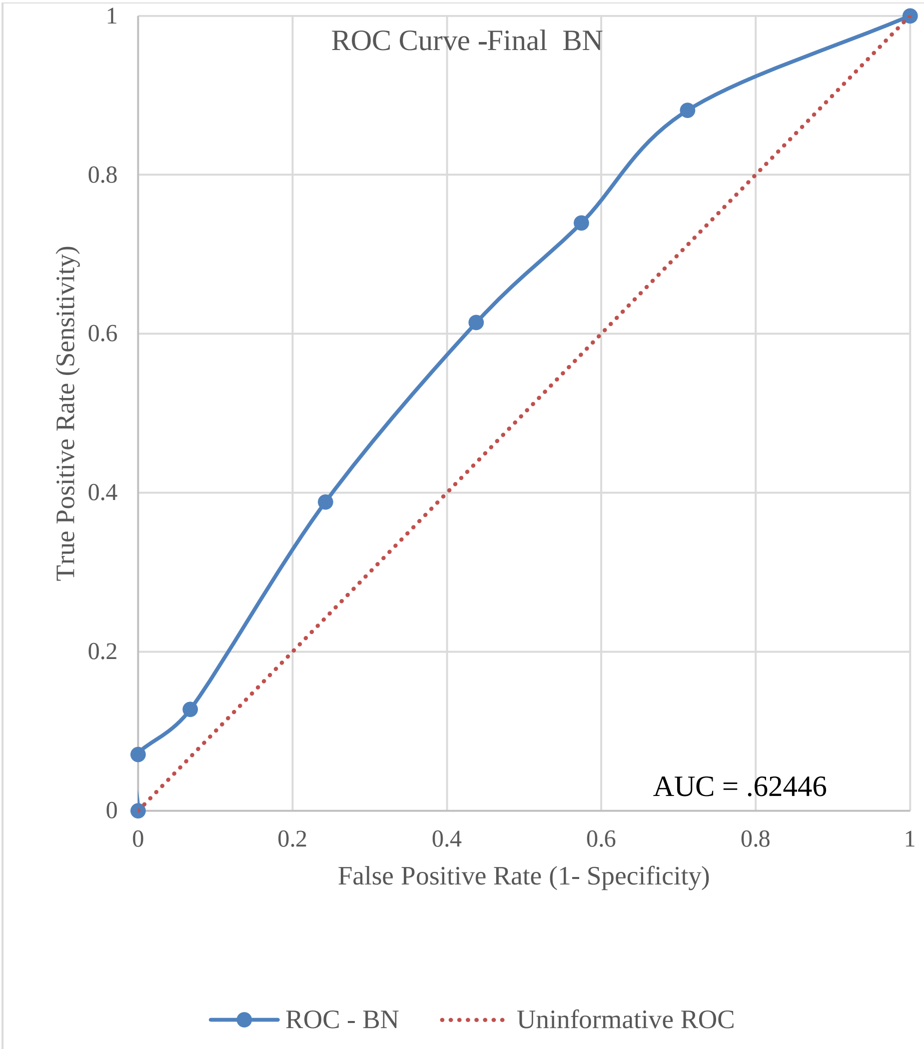

Figure 2.9. Final BN ROC Curve. This figure illustrates the model's relationship between sensitivity and specificity (solid line) at different thresholds as compared to an uninformative model (dashed line).

This is further confirmed with examination of the model's Cohen's kappa coefficient (*k*), a calculation that incorporates the possibility of chance effects into the measure of agreement between a model's prediction and actuality (Rosenfield & Fitzpatrick-Lins, 1986). Cohen's kappa coefficient is described Equation 2.3, where *P(a)* is the proportion of correctly predicted cases and *P(e)* represents the hypothetical probability of a chance agreement.

$$k = \frac{P(a) - P(e)}{1 - P(e)} \tag{2.3}$$

The final model's kappa coefficient equals .062, indicating that classification agreement is most likely due to chance as there is only 6.2% agreement above what is expected by chance alone. In other words, there is little difference between the model's correct prediction and what might be predicted due to chance alone. Again, this is most likely the result of imbalanced data with a high prevalence of retention.

An analysis of sensitivity of the "Retained" node to the other predictive nodes (Table 2.11) reveals a substantial amount of uncertainty still unexplained by the final BN. While the two most influential variables ("Academic Rank" and "Financial Need & Risk Profile") do explain away a slightly greater amount of the uncertainty of likelihood of retention (as expressed in the Mutual Info column) than found in the "straw man" model, it is notable that the total amount of retention uncertainty explained in the final BN (8.56%) is less than that in the original "straw man" model (13.04%). However, as explained earlier, the particularly influential variables included in the "straw man" model (SAT scores) are incorporated into the composite "Academic Rank" variable in the final BN, suggesting that they are still represented in a parsimonious model while also allowing room for other variables deemed important by the experts.

Table 2.11
Sensitivity Analysis of "Retained" to Other Variables Included in Final BN

Variable	Mutual Info	Percent	Variance of Beliefs
Academic Rank	.0207	3.21	.0043
Financial Need & Risk Profile	.0180	2.80	.0037
Study Skills	.0059	.91	.0011
Attendance Patterns	.0041	.63	.0008
Academic Support	.0028	.44	.0005
Distance from Home	.0023	.36	.0004
Social Support	.0011	.17	.0002
Advisor Match/Mismatch	.0003	.04	.0000
Race/Ethnicity	.0000		

Examining Table 2.11 also reveals consensus and conflict with the final influence ratings assigned to the variables by the experts. The model confirms the importance of "Academic Rank" and "Study Skills" in predicting retention, but places other variables deemed important by the experts ("Social Support," "Academic Support") lower in the scale of influence. Note that "Race/Ethnicity" was purposely not included as a direct influence on retention, and this is represented accordingly with zero mutual information. A look at the sensitivity of each individual node with other nodes (see Appendix D) is also helpful in evaluating the experts' recommended structure. For example, "Academic Rank" is highly sensitive to "Financial Need & Risk Profile," suggesting that the assumption of conditional independence among these two nodes may need to be further investigated. In another example, while the experts opined that students' race/ethnicity influenced both "Academic Support" and "Social Support," sensitivity findings of "Race/Ethnicity" to other nodes find that it is merely related to "Academic Support."

Recommendations to review a model's sensitivity to changes in informative priors are also included in the literature (Gill, 2009; Reimherr, Meng, & Nicolae, 2014). Recall that three of the nodes included in the final model are comprised of expert-elicited probabilities that serve as the informative prior ("Study Skills," "Attendance Patterns," and "Academic Support"). In order to evaluate the impact of these prior choices, each of these three nodes were set to uniform probability distributions (indicating equally likely states) and the resulting models were compared to the final BN in terms of inference. Unsurprisingly, the expert-elicited priors do heavily influence outcomes predicted by the model. For example, imposing the same Tinto high attrition scenario as described in Figure 2.8 on a model in which all of the expert-elicited priors are considered uniform results in predicted probability of retention of 82% compared to 74%

when using the informative priors. Conceding the importance of these priors on conclusions that can be drawn from the final model further highlights the importance of careful elicitation.

Table 2.12 summarizes the final model's complexity measures. The final model is slightly less complex than the original "straw man" model, and should be considered parsimonious and not overly complex. This simplicity is advantageous for explaining and demonstrating the model's use to various audiences.

Table 2.12
Final BN Model Complexity Metrics

Metric	Count
Nodes	10
Links	15
Node States	25

A final evaluative measure of the final model surrounds examination of its internal validity as measured by whether or not the model performs as expected by the experts. The incorporation of expert information in the development of the final BN's structure as well as select parameters requires that some evaluation of the model's capacity to adequately reflect the experts' expectations take place. During the second elicitation session experts were asked to hypothesize the probability of retention given a number of scenarios. The scenarios chosen for testing were done so based on discussions during both elicitation sessions and focused on expert indications of important and influential variables or situations unique to retention at the institution. Comparison of the experts' predicted retention rates with those resulting from imposition of the case scenarios on the final model reveals varied results and is summarized in Table 2.13. Large differences between the experts' and final BN predictions appear to be a result to two main factors: Systematic under-prediction by the experts and disagreement between the variables deemed most influential by the experts and the model. For example, the experts considered "Financial Need & Risk Profile" to be only mildly influential on retention, while the model found it to be a top predictor. The difference between expert and model prediction in scenario four is due to this disagreement – the model recognizes the large impact of a high "Financial Need & Risk Profile" on retention, while the experts discounted this influence. Similarly, differences in the predictions for scenario six reflects the experts' expectation that negative "Social Support" heavily influences retention and the model's estimation that any impact of negative "Social Support" is greatly outweighed by high "Academic Rank" and developed "Study Skills."

Table 2.13
Comparison of Final BN Predictions and Expert-Elicited Predictions of Retention Given Different Scenarios

Scenario: What is the probability of retention, given:	Expert Prediction	Final BN Prediction
1. Within an hour, strong academic background, developed study skills	85%	89%
2. More than an hour, strong academic background, developed study skills	88%	92%
3. Underdeveloped study skills, moderate academic background, consistent class attendance and positive social support	84%	85.5%
4. Underdeveloped study skills, low academic rank, high financial need/risk profile	79%	74.5%
5. Developed study skills, high academic rank, high financial need/risk profile	85%	57%[a]
6. Developed study skills, high academic rank, within an hour, low social support	80%	90%
7. Underdeveloped study skills, low academic rank, within an hour, low social support	79%	80%

[a]Unreliable result due to very few examples of this scenario found in underlying or simulated data

In summary, cursory measures of predictive accuracy reveal a fairly strong final model. However, closer examination of other evaluative measures indicate challenges due to imbalanced training data with a strong prevalence of retention over attrition. The model is particularly weak in classifying the "rare" cases of attrition, with a high rate of false positive classifications of retention. In terms of sensitivity, very few of the included variables explain an adequate amount of uncertainty in the model, but those that are most influential were also weighted heavily by the experts. Issues of the model's internal validity in terms of agreement with experts' predictions center on these same disagreements in expert and model weightings, although some predictions were very close.

Conclusion

This study was designed to explore two main ideas: Using the literature, data, and expert information to identify important causal influences on retention, and employing a hybrid data- and expert-learned approach to constructing a Bayesian network that adequately predicts retention. While the predictive power of the final BN created using a combination of expert information and data suffers from imbalanced training data, the employment of experts in the identification, discussion, and quantification of influential retention variables can be considered successful. Regardless of the research outcomes, both stages of the research revealed that the process of incorporating expert information into designing models adds a level of insight and institutional knowledge that might otherwise be unrecognized.

Identification of Retention Variables

The development of a Bayesian network incorporating both expert knowledge and prior data allows for an individualized model that is specific to the institution, reflecting its student population, culture, and other characteristics. As Robbins et al. (2004) explained, relying solely on research literature to guide the choice of variables is limiting as "the research literature ranges across many psychological and educational content domains, which dampens efforts at integrating or evaluating the empirical literature…" (p. 262). Combining data and expert knowledge introduces insights that might otherwise be unrecognized or unacknowledged within the presence of only one of these sources of information. The experts identified patterns that were unique to the institutions (e.g. greater distance from home correlates with higher retention, students with highly developed study skills and inconsistent course attendance trigger more significant retention red flags than students with less developed study skills), and were able to parse among the numerous retention prediction variables found in the literature to suggest a simple and parsimonious model structure. While data-mining procedures may have ultimately identified these unexpected patterns, preemptory knowledge of such institution-specific events allows for more directed and efficient modeling and evaluation.

In future iterations or replications of this research, it is recommended that the elicitation facilitator be very familiar with the research topic and its coverage within the wider scholarly literature. Extensive background knowledge on the part of the facilitator helps maintain focus during discussions with experts, and allows everyone involved in the discussion to speak the same topical language. It is equally important to spend ample time on collaboratively developing and finalizing explicit operating definitions for the identified variables. Recognition of the importance of this might have resulted in fewer changes in the variables' definitions between elicitation sessions one and two and redirected valuable discussion time in elicitation session two from finalizing operating definitions to actual performance of the model (see the change in "Academic Support" as an example). Finally, it is highly recommended that, given experts' and researchers' limited time and resources, all elicitation sessions be accompanied by pre-elicitation materials outlining the sessions' expectations, goals, and even a timed agenda. Keeping the session on track and focused is critical in producing usable and relevant information within an often limited timeline.

Construction and Performance of BNs

The process of model construction revealed a number of insights. In terms of expert participation, discussions during both elicitation sessions confirmed a high amount of consensus and agreement among the two experts. For example, comparison of the individually drafted structures that were ultimately transformed into the expert-elicited structural model exposed striking agreement between the experts' understandings of the causal influences and relationships between retention variables. Similar consensus was found during the probability elicitations, particularly in terms of expressions of uncertainty and concerns over bias. While expert solidarity does not necessarily assure expert accuracy or precision, it does provide reassurance of consistent opinion.

As addressed earlier, the final BN's performance suffers in terms of identifying and predicting non-retention. This is particularly troubling considering that it is this specific group of students, those at risk of not returning, in whom model stakeholders are most interested. This finding highlights the importance of in-depth model evaluation outside of simple predictive accuracy. Perhaps due to the capability of BNs to handle scarce data and uncertainty, thus making evaluation and validation more difficult, many studies involving BNs simply do not include quantitative model evaluation (Aguilera, Fernández, Fernández, Rumi, & Salmerón, 2011). Without evaluation of the additional metrics described in Table 2.9, the false positive classification issues resulting from imbalanced dataset would have gone unnoticed and an inadequate model would be adopted by stakeholders for use in decision-making and intervention.

Finally, model construction confirmed the importance of carefully considering variable operating definitions and the assignment of underlying data to less-than-certainly defined variables. For example, analysis of node sensitivity (Table 2.10) revealed conflict between data-learned and expert-anticipated importance of certain variables like "Social Support." This finding does not unilaterally negate the experts' opinions that positive or negative social support plays an important role in influencing retention. Rather, it could simply mean that the data used to parameterize this variable (whether or not a student lived on a dorm floor that had unusually high attrition) was not particularly descriptive of or relevant to what experts' perceive as social support. Again, this finding emphasizes the importance of comprehensive model evaluation, including analysis of sensitivity and unexpected results.

Use of a formalized elicitation framework like SHELF proved invaluable to accomplishing the goals of the probability elicitation session within a limited timeframe. Additionally, use of a spreadsheet pre-populated with formulas that could be used to quickly demonstrate the conditional probabilistic impact of elicited distributions on the variables pleased the experts and allowed them additional opportunities for feedback and revision. As a major goal of any elicitation of expert information is to assure that the opinions of experts are communicated and received clearly and accurately, it is recommended that preparation of such materials and tools be repeated in any future iterations or replications of this research.

Importance of Process

While the final BN includes major limitations concerning its current usefulness for predicting retention or acting as an early intervention tool, it is important to recognize that the process of model development can be considered as important as the final model itself. Those involved in shaping the decisions and policies related to a predicted behavior are required to formally discuss and articulate influential factors. These formal discussions result in a deeper understanding of the problem at hand, allowing decision-makers to set future priorities for resource allocation, data collection, and additional study. In other words, the process of reviewing the literature and shaping a model, especially in the first elicitation session, revealed a large number of variables that the experts think the institution should be tracking. For example, the experts were confident that whether or not a student accessed academic support programming affects likelihood of retention. However, the formal structural and probability elicitations required them to trouble this anecdotal and somewhat vague conclusion and operationally define and quantify student access to academic support programming and its implications. Further, the elicitation revealed a lack of quantitative data surrounding an activity that the experts feel is an important factor in predicting retention. Given their expressed uncertainty about the prevalence and definition of the "Academic Support" variable, it became clear that collection of quantitative data related to student access to academic support programming and retention should be initiated in order to support or negate the expert intuition. Even variables for which data were available included concerns of whether or not they were the "correct" data for describing a condition (see earlier discussion of "Social Support"). Any and all formal consideration and discussion of these types serves to only increase the knowledge-base and awareness of retention issues, thus setting the stage for even better model construction and utilization.

Limitations and Future Study

A major advantage of the use of BNs is their capacity to adapt to new information. As cohorts mature and new data are gathered, these data can be added as new cases from which the model will learn. Parameters can easily be updated given new insights from data or additional experts, and the model will reflect different predictions given a changing student body and other new information. Given emphasis on non-cognitive influences found in the most recent retention literature, particular attention will be paid to inclusion of these types of variables. Future versions of the model will be re-presented to experts for review and evaluation, as well as elicitation of any structural and probabilistic shifts that may arise. Maintaining an open feedback loop with the experts through model creation, evaluation and refinement is critical, and will be employed going forward. As recommended by Pollino, Woodberry, Nicholson, Korb, & Hart (2007), an unaffiliated third party reviewer may be brought in to review and evaluate any future expert-elicited CPTs.

However, as attrition is already a “rare” event, and may grow even rarer as additional focus and resources are allocated towards increasing retention, the problems created by such imbalanced data will not be alleviated. In order to create a more adequate predictive BN, it is necessary to explore alternatives to traditional training/testing data sets for parameterizing the variables for which data are available. Future proposed study involves the investigation of sampling strategies that are designed to mitigate the influences of imbalanced data. A number of authors review and recommend strategies for handling imbalanced data sets within the context of classification, ranging from simple over/under-sampling to more complex algorithmic approaches (Chawla, 2010; Kotsiantis, Kannellopoulos, Pintelas, 2006; Weiss & Provost, 2003), as well as quantification of the costs of misclassification (Japkowicz & Stephen, 2002; Monard & Batista, 2002). Comparison of these approaches based on model performance metrics like those summarized in Table 2.9 are planned for future study.

Additionally, future study will include more model comparison in general. As more data become available, a comparison of the performance, complexity, and sensitivity of strictly data-learned, strictly expert-informed, and even hybrid constructed models would offer additional insights into the understanding and reliability of experts, the quality of retention-related data, and the interaction between both these sources. Metrics of model comparison include Bayesian

Information Criterion (BIC) (Schwarz, 1978) and Minimum Description Length (MDL) (Rissanen, 1996).

A further option for future study involves contrasting BN performance and advantages against those of logistic regression and other forms of discriminant analysis. While logistic regression is often used as a technique for prediction and classification, the inclusion of experts without extensive experience in statistical methodology, the transparent depiction of uncertainty, and the user-friendly graphical and dynamic representation of variable relationships in a BN called for the exploration of its usefulness in this research. Depending on the type and nature of the predicted variable, the performance of BNs versus logistic regression is mixed (Ducher et al., 2013; Schmeits & Kok, 2010). However, simply the act of comparing performance and other aspects of predictive BNs to logistic regression models offers an opportunity for insights related to overfitting, missing data, and variable importance that can be helpful in selecting and defending model choice (Roos, Wettig, Grünwald, Myllmäki, & Tirri, 2005; Tu, 1996).

While the limitations introduced by imbalanced and unavailable training/testing data are important and encourage further study, it is important to consider that a major contribution of this study lies within the lessons learned through the process of combining expert and quantitative data. Initiating and maintaining formal elicitation practices that reinforce focus and discipline during sessions, allowing experts to quickly view their judgements' implications, and prioritizing the development of clear and detailed operating definitions are recommendations generalizable to other studies formally incorporating expert judgements of any subject-matter. Additionally, this research stresses importance of complete model evaluation in any context as a critical step, the exclusion of which could lead to seriously flawed conclusions. Most importantly, the flexibility and usefulness of a Bayesian networks in the incorporation of unique and valuable expert judgement, often minimized or ignored, is highlighted. While data-mining has its advantages and place within predictive modeling, BNs allow for the combination of both sources while still transparently accounting for uncertainty in a format that is easily understood and employed by multiple audiences.

References

ACT (2010). *The Condition of College and Career Readiness 2010.* Iowa City, IA.

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment.* Washington, DC: Office of Education Research and Improvements, U.S. Department of Education.

Ægisdóttir, S., White, M., Spengler, P., Maughermaan, A., Anderson, L., Cook, R., …Rush, J. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, *34*, 341-382.

Aguilera, P.A., Fernández, R., Fernández, A., Rumi, R., & Salmerón, A. (2011). Bayesian networks in environmental modelling. *Environmental Modeling Software, 26*(12), 1376–1388.

Allen, J., Robbins, S.B., Casillas, A., & Oh, I. (2008). Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness. *Research in Higher Education*, *49*(7), 647-664.

Alqallaf, F. & Gustafson, P. (2001). On cross validation of Bayesian models. *The Canadian Journal of Statistics, 29*(2), 333-340

Aronson, J. & Steele, C.M. (2005). Stereotypes and the fragility of academic competence, motivation, and self-concept. In A.J. Elliot & C.S. Dweck (Eds.), *Handbook of competence and motivation*, pp. 436-456). New York: Guilford.

Astin, A.W. (1993). *What matters in college: Four critical years revisited.* San Francisco: Jossey – Bass.

Astin, A.W. & Oseguera, L. (2005). Pre-college and institutional influences on degree attainment. In Seidman (Ed.), *College student retention: formula for student success* (245-276). Westport, CT: Prager Publishers.

Astin, A.W. & Oseguera, L. (2012). Pre-college institutional influences on degree attainment. In A. Seidman (Ed.), *College student retention: Formula for student success* (2nd ed, pp. 119-146). Lanham, MD: Rowan & Littlefield.

Bean, J.P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education, 12,* 155–187.

Bean, J.P. (1985). Interaction effects based on class level in an explanatory model of college student dropout syndrome, *American Educational Research Journal, 22*(1), 35–64.

Bean, J.P. & Metzner, B.S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research, 55*(4) 485-540.

Beck, H. P. & Davidson, W. D. (2001). Establishing an early warning system: Predicting low grades in college students from survey of academic orientation **scores**. *Research in Higher Education, 42*(6), 709-723.

Bekele, R. & Menzel, W. (2005). A Bayesian approach to predict performance of a student (BAPPS): A case with Ethiopian students. In M. Hamza (Ed.), *Proceedings from IASTD International Conference on Artificial Intelligence and Applications,* 189-194.

Berger, J. B. (2001–2002). Understanding the organizational nature of student persistence: Empirically based recommendations for practice. *Journal of College Student Retention, 3*(1), 3–21.

Berger, J.O (2006). The case for objective Bayesian analysis. *Bayesian Analysis, 1*(3), 385-402.

Berger, J. B. & Milem, J. F. (2000). Organizational behavior in higher education and student outcomes. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 15, pp. 268–338). New York: Agathon Press.

Bista, K. & Foster, C. (2011). Issues of international student retention in American higher education. *The International Journal of Research and Review, 7*(2), 1-10.

Bolstad, W. M. (2007). *Introduction to Bayesian statistics* (2nd ed.). Hoboken, NJ: Wiley.

Bonous-Hammarth, M. (2000). Value congruence and organizational climates for undergraduate persistence. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 15, pp. 339-370). New York: Agathon Press.

Boulter, L. (2002). Self-concept as a predictor of college freshman academic adjustment. *College Student Journal*, *36*, 234-246.

Bovens, L. & Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Clarendon Press.

Bowen, W.G., Chingos, M.M., & McPherson, M.S. (2009). *Crossing the finish line: Completing college at America's public universities.* Princeton, NJ: Princeton University Press.

Boyce, M.S., Vernier, P.R., Nielsen, S.E., & Schmiegelow, R.K.A. (2002). Evaluating resource selection functions. *Ecological Modelling, 157*, 281-300.

Breiman, L. & Spector, P. (1992). Submodel selection and evaluation in regression: the X-random case. *International Statistical Review*, *60*, 291-319.

Buckland, M. & Gey, F. (1994). The relationship between recall and precision. *Journal of the American Society for Information Science*, *45*(1), 12-19.

Cabrera, A., Nora, A. & Castaneda, M.B. (1993). College persistence: Structural equations modelling test of an integrated model of student retention, *Journal of Higher Education, 64*(2), 125–139.

Carey, K. & Aldeman, C. (2008). *Ready to assemble: A model state higher education accountability system*. Washington DC: Education Sector.

Chawla, N.V. (2010). Data mining for imbalanced datasets: An overview. In O. Maimon and L. Rokach (Eds.) *Data mining and knowledge discovery handbook* (pp. 875-886). New York: Springer.

Chemers, M.M., Hu, L., & Garcia, B.F. (2001). Academic self-efficacy and first year college student performance and adjustment. *Journal of Educational Psychology*, *93*, 55-64.

Chen, S.H. & Pollino, C.A. (2012). Good practice in Bayesian network modelling. *Environmental Modelling & Software*, *37*, 134-145.

Choy, S.P. (2001). *Students whose parents did not go to college: Postsecondary access, persistence, and attainment* (NCES 2001-126). Washington, DC: US Department of Education, National Center for Education Statistics.

Choy, S. L., O'Leary, R., & Mengersen, K. (2009). Elicitation by design in ecology: Using expert opinion to inform priors for Bayesian statistical models. *Ecology*, *90*(1), 265-277.

Clark, J.S. (2003). Uncertainty and variability in demography and population growth: A hierarchical approach. *Ecology*, *84*(6), 1370-1381.

Clemen, R.T. & Winkler, R.L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, *19*(2), 187-203.

Complete College America. (2011). Time is the enemy: The surprising truth about why today's college students aren't graduating…and what needs to change. Washington, DC. Retrieved from *www.completecollege.org/docs/Time_Is_the_Enemy.pdf*

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, *51*, 171–200.

Cowell, R.G., Dawid, A.P., & Spiegelhalter, D.J. (1993). Sequential model criticism in probabilistic expert systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15,* 109-129.

D'Agostini, G. G. (2001). Role and meaning of subjective probability: Some comments on common misconceptions. *AIP Conference Proceedings*, *568*(1), 23-30.

Dana, J. & Thomas, R. (2006). In defense of clinical judgment…and mechanical prediction. *Journal of Behavioral Decision Making, 19,* 413-428.

Dawes, R.M. (1988). *Rational choice in an uncertain world.* Chicago, IL: Harcourt Brace Jovanovich.

Dawes, R.M., Faust, D., & Meehl, P.E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668-1674.

Dawes, R.M., Faust, D., & Meehl, P.E. (1993). Statistical prediction versus clinical prediction: Improving what works. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 351-367). Hillsdale, NJ: Erlbaum.

Dawid, A.P. (1992). Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and Computing, 2,* 25-36.

de Campos, J.G. & Castellano, J.G. (2007). Bayesian network learning algorithms using structural restrictions. *International Journal of Approximate Reasoning, 45*(2), 233-254.

DeBerard, M.S., Spielmans, G.I., & Julka, D.L. (2004). Predictors of academic achievement and retention among college freshmen: A longitudinal study. *College Student Journal, 38*(1), 66-80.

Demetriou, C. & Schmitz-Sciborski, A. (2011). Integration, motivation, strengths and optimism: Retention theories past, present and future. In R. Hayes (Ed.), *Proceedings of the 7th*

National Symposium on Student Retention, 2011, Charleston. (pp. 300-312). Norman, OK: The University of Oklahoma.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the em algrorithm. *Journal of the Royal Statistical Society: Series B., 39*(1), 1-38.

Druzdzel, M. & Simon, H. (1993). Causality in Bayesian belief networks. *Proceedings of ninth conference on uncertainty in artificial intelligence,* 3-11.

Ducher, M., Kalbacher, E., Cornbarnous, F., Finaz de Vilaine, McGregor, B., Fouque….Fauvel, J. (2013). Comparison of a Bayesian network with a logistic regression model to forecase IgA nephropathy. *BioMed Research International, 2013*, 1-7.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53,* 109–132.

Eckles, J.E. & Stradley, E.G. (2011). A social network analysis of student retention using archival data. *Social Psychology of Education: An International Journal*, *15*(2), 165-180.

Education Trust. (2004). *A matter of degrees: Improving graduation rates in four-year colleges and universities.* Washington, DC: The Education Trust.

Fenton, N., Neil, M., & Lagnado, D.A. (2013). A general structure for legal arguments about evidence using Bayesian networks. *Cognitive Science*, *37,* 61-102.

Fenz, S. (2012). An otology-based approach for constructing Bayesian networks. *Data & Knowledge Engineering, 73,* 73-88.

Fienen, M.N. & Plant, N.G. (2014). A cross-validation package driving Netica with python. *Environmental Modelling & Software, 63,* 14-23.

Flores, M.J., Nicholson, A.E., Brunskill, A., Korb, K., & Mascaro, S. (2011). Incorporating expert knowledge when learning Bayesian network structure: A medical case study. *Artificial Intelligence in Medicine, 53*(3), 181-204.

Fredricks, J., Blumenfield, P., & Paris, A. (2004). School engagement: Potential of the concept: State of the evidence? *Review of Education Research,* 74, 59-119.

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, *29*, 131-163.

Friedman, N., Linial, L., Nachman, I., & Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *Journal of Computational Biology, 7,* 601-620.

Galbraith, C.S., Merrill, G.B., & Kline, D.M. (2010). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and Bayesian analysis. *Research in Higher Education, 53,* 353-374.

Gansmer-Topf, A.M. & Schuh, J.H. (2006). Institutional selectivity and institutional expenditures: Examining organizational factors that contribute to retention and graduation. *Research in Higher Education, 47,* 613-642.

Gansemer-Topf, A.M., Zhang, Y., Beatty, C.C., & Paja, S. (2014). Examining factors influencing attrition at a small, private, selective liberal arts college. *Journal of Student Affairs Research and Practice, 51*(3), 270-285.

Garthwaite, P.H., Kadane, J.B. & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association, 100,* 680-701.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association, 70,* 320-328.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian Data Analysis* (2nd ed.). Boca Raton, FL: CRC Press.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and bias/variance dilemma. *Neural Computation, 4,* 1-58.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

Gill, J. (2009). *Bayesian methods: A social and behavioral sciences approach* (2nd ed.). Boca Raton, FL: CRC Press.

Grove, W.M. & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2*(2), 293-323.

Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19-30.

Hamilton, S., Pollino, C. & Jakeman, A. (2015). Habitat suitability modelling of rare species using Bayesian networks: Model evaluation under limited data. *Ecological Modelling, 299*, 64-78.

Harrington, P. & Fogg, N, (2009). *White Paper: College Access and Retention of Career and Technical Education Graduates.* Cambridge, MA: Rennie Center for Education Research and Policy.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2nd ed.). New York, NY: Springer-Verlag.

Hausmann, L.R.M., Schofield, J.W., & Woods, R.L. (2007). Sense of belonging as a predictor of intentions to persist among African American and White first-year college students. *Research in Higher Education, 48*(7), 803-839.

Heckerman, D. (1997). Bayesian networks for data mining. *Data Mining and Knowledge Discovery, 1*, 79-119.

Heckerman, D., Geiger, D., & Chickering, D. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning, 20,* 197-243.

Hennig, C. (2009). A constructivist view of the statistical quantification of evidence. *Constructivist Foundations*, *5*(1), 39-54.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, *70*, 271-294.

Hosch, B. J. (2008). *Institutional and student characteristics that predict graduation and retention rates.* Paper presented at the annual meeting of the North East Association for Institutional Research, Providence, RI.

Howard, J.S. (2013). Student retention and first-year programs: A comparison of students in liberal arts colleges in the mountain south. *Electronic Theses and Dissertations*, Paper 2270. Retrieved from: http://dc.etsu.etd/2270.

Hurtado, S. & Carter, D.F. (1997). Effects of college transition and perceptions of the campus racial climate on Latino college students' sense of belonging. *Sociology of Education, 70,* 324-345.

Jakeman, A.J. & Hornberger, G.M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research, 29*(8), 2637–2649.

Jakeman, A.J., Letcher, R.A., & Norton, J.P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling and Software*, *21*, 602-614.

Japkowicz, N. & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis, 6*(5), 429-449.

Jensen, F. & Nielsen, T.D. (2007). *Bayesian networks and decision graphs* (2nd ed.). New York: Springer.

Jensen, F., Lauritzen, S. & Olesen, K. (1990). Bayesian updating in recursive graphical models by local computations. *Computational Statistics Quarterly, 4,* 269-282.

Kadane, J. B. & Wolfson, L. J. (1998). Experiences in elicitation. *The Statistician*, *47*, 3-19.

Käser, T., Busetto, G., Solenthaler, B., Baschera, G., Kohn, J., Kurcian, K….Gross, M. (2013). Modelling and optimizing mathematics learning in children. *International Journal of Artificial Intelligence in Education*, *13,* 115-135.

Kass, R.E. & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association, 91*(435), 1343-1370.

Kennett, D.J. & Reed, M.J. (2009). Factors influencing academic success and retention following a 1st year post-secondary success course. *Educational Research and Evaluation: An International Journal on Theory and Practice, 15*(2), 153-166.

Kiser, A. & Price, L. (2008). The persistence of college students from their freshman to sophomore year. *Journal of College Student Retention: Research, Theory & Practice, 9*(4), 421-436.

Kjærulff, U.B. & Madsen, A.L. (2008). *Bayesian networks and influence diagrams: A guide to construction and analysis.* New York: Springer.

Kotsiantis, S., Kanellopoulos, D., Pintelas, P. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering, 30*, 25-36.

Kotsiantis, S., Patriarcheas, K., & Xenos, M. (2010). A combinatorial incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems, 23,* 529-535.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *14th International Joint Conference on Artificial Intelligence (IJCAI),* 20-25 August 1995, Montreal, Quebec, Canada.

Kuh, G. D. (2001–2002). Organizational culture and student persistence: Prospects and puzzles. *Journal of College Student Retention, 3*(1), 23–39.

Kuhnert, P.M., Martin, T.G., & Griffiths, S.P. (2010). A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters, 13,* 900-914.

Lauritzen, S.L. (1996). *Graphical models.* Oxford: Clarendon Press.

Longden, B. (2006). An institutional response to changing student expectations and their impact on retention rates. *Journal of Higher Education Policy and Management, 28,* 173-187.

Longwell-Grice, R., & Longwell-Grice, H. (2008). Testing Tinto: How do retention theories work for first-generation, working class students? *Journal of College Student Retention: Research, Theory & Practice*, *9*(4), 407–420.

Lotkowski, V., Robbins, S., & Noeth, R. (2004). *The role of academic and non-academic factors in improving college retention.* Iowa, IA: ACT Policy Report.

Makuakane-Drechsel, T. & Hagedorn, I. (2000). Correlates of retention among Asian Pacific Americans in community colleges: The case for Hawaiian students. *Community College Journal of Research and Practice, 24*, 639-655.

Marcot, B. (2012). Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling, 230,* 50-62.

Margaritis, D. (2003). *Learning Bayesian network model structure from data.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.

Masegosa, A.R. & Moral, S. (2013). An interactive approach for Bayesian network learning using domain/expert knowledge. *International Journal of Approximate Reasoning, 54,* 1168-1181.

May, R.J., Maier, H.R., & Dandy, G.C. (2010). Data splitting for artificial neural networks using SOM-based stratified sampling. *Neural Networks, 23*, 283-294.

Meehl, P.E. (1956). Symposium on clinical and statistical pred iction: The tie that binds. *Journal of Counseling Psychology, 3*, 163-173.

Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology, 4,* 268-273.

Meehl, P. E. (1967). What can the clinician do well? In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 594-599). New York: McGraw-Hill.

Meehl, P. E. (1973). *Psychodiagnosis: Selected papers.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality*

Meehl, P.E. (1996). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Northvale, NJ: Jason Aronson. (Original work published 1954)

Meyer, K.A. & Xu, Y.J. (2007). A Bayesian analysis of the institutional and individual factors influencing faculty technology use. *Internet and Higher Education, 10,* 184-195.

Monard, M.C. & Batista, G.E. (2002). Learning with skewed class distributions. *Advances in Logic, Artificial Intelligence, and Robotics: LAPTEC 2002, 85,* 173-181.

Morrison, R.E., Bryant, C.M., Terejanu, G., Prudhomme, S., & Miki, K. (2013). Data partition methodology for validation of predictive models. *Computers and Mathematics with Applications, 66,* 2114-2125.

Neapolitan, R.E. (2004). *Learning Bayesian networks.* Upper Saddle River, NJ: Pearson Prentice Hall.

Nesler, M.S. (1999). *Factors associated with retention in a distance-based liberal arts program.* Paper presented at the 26th North East Association for Institutional Research Conference, Newport, RI.

Niculescu, R.S., Mitchell, T.M., & Rao, R.B. (2006). Bayesian network learning with parameter constraints. *Journal of Machine Learning Research, 7*(7), 1357-1383.

Niedermayer, D. (2008). An introduction to Bayesian networks and their contemporary applications. In D.E. Holmes & L. Cain (Eds.), *Innovations in Bayesian networks* (pp. 117-130). Berlin: Springer.

Noel-Levitz. (2013). *2013 Cost of recruiting an undergraduate student: Benchmarks for four-year and two-year institutions.* Coralville, IA: Author. Retrieved from: www.noellevitz.com/BenchmarkReports.

Oakley J. E. & O'Hagan, A. (2010). SHELF: the Sheffield Elicitation Framework (version 2.0), School of Mathematics and Statistics, University of Sheffield, UK. (http://tonyohagan.co.uk/shelf)

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., ... Rakow, T. (2006). *Uncertain judgements: eliciting experts' probabilities*. West Sussex, England: John Wiley & Sons, Ltd.

Oseguera, L. & Rhee, B.S. (2009). The influence of institutional retention climates on student persistence to degree completion: A multilevel approach. *Research in Higher Education, 50,* 546-569.

Osherson, D., Lane, D., Hartley, P., & Batsell, R.R. (2001). Coherent probability from incoherent judgment. *Journal of Experimental Psychology: Applied*, *7*(1), 3-12.

Pascarella, E.T. & Terenzini, P.T. (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model, *Journal of Higher Education, 51*, 60–75.

Pascarella, E.T. & Terenzini, P.T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco: Jossey-Bass.

Pascarella, E.T. & Terenzini P.T. (2005). *How college affects students: A third decade of research* (Vol 2). San Francisco, CA: Jossey-Bass.

Pearl, J. & Verma, T. (1991). A theory of inferred causation. In J. Allen, R. Fikes & E. Sandewall (Eds.), *Knowledge representation and reasoning: Proceedings of the second international conference* (pp. 441-452). San Francisco: Morgan Kaufmann.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,* Series in Representation and Reasoning. San Mateo, CA: Morgan Kaufmann Publishers.

Pearl, J. (2000). *Causality: Models, reasoning, and influence.* Cambridge, UK: Cambridge University Press.

Pike, G.R. (2013). NSSE benchmarks and institutional outcomes: A note on the importance of considering the intended uses of a measure in validity studies. *Research in Higher Education, 54,* 149-170.

Pike, G.R. & Graunke, S.S. (2015). Examining the effects of institutional and cohort characteristics on retention rates. *Research in Higher Education, 56,* 146-165.

Pike, G.R., Hansen, M.J., & Childress, J.E. (2014). The influences of students' pre-college characteristics, high school experiences, college expectations, and initial characteristics on degree attainment. *Journal of College and Student Retention, 16,* 1-23.

Pitchforth, J. & Mengersen, K. (2013). A proposed validation framework for expert elicited Bayesian networks. *Expert Systems with Applications. 40,* 162-167.

Pollino, C., Woodberry, O., Nicholson, A., Korb, K., & Hart, B. (2007). Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling & Software, 22,* 1140-1152.

Porter, S.R. & Swing, R.L. (2006). Understanding how first-year seminars affect persistence. *Research in Higher Education, 47*(1), 89-109.

Pourret, O., Naim, P., & Marcot, B. (2008). *Bayesian networks: A practical guide to applications.* Chichester, West Sussex: John Wiley & Sons.

Press, S. J. & Tanur, J.M. (2001). *The subjectivity of scientists and the Bayesian approach.* New York: John Wiley and Sons.

Press, S. J. (2003). *Subjective and objective Bayesian statistics: Principles, models and applications* (2nd ed.). Hoboken, NJ: John Wiley and Sons.

Reason, R. (2003). Student variables that predict retention: recent research and new developments. *National Association of Student Personnel Administrators Journal, 20*(4), 172-191.

Reimherr, M., Meng, X., & Nicolae, D.L. (2014). Being an informed Bayesian: Assessing prior informativeness and prior-likelihood conflict. *arVix preprint:* 1406-5958.

Rissanen, J. (1996). Fisher information and stochastic capacity. *IEEE Transactions on Information Theory, 42*(1), 40-47.

Robbins, S.B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261-288.

Roos, T., Wettig, H., Grünwald, P., Myllmäki, P., & Tirri, H. (2005). On discriminative Bayesian network classifiers and logistic regression. *Machine Learning, 59,* 267-296.

Roska, J., Davis, J., Jaggars, S., Zeidenbert, M. & Cho, S-W. (2009). *Strategies for promoting gatekeeper course success among students needing remediation: Research report for the Virginia Community College System*. New York: Community College Research Center: Teachers College, Columbia University.

Rosenfield, G.H. & Fitzpatrick-Lins, K. (1986). A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing, 52*(2), 223-227.

Ryan, J. F. (2004). The relationship between institutional expenditures and degree attainment at baccalaureate colleges. *Research in Higher Education, 45*, 97–113.

Schmeits, M. J. & Kok, K. J. (2010). A comparison between raw ensemble output, (modified) Bayesian model averaging, and extended logistic regression using ECMWF ensemble precipitation reforecasts. *Monthly Weather Review, 138*(11), 4199-4211.

Schuh, J.H. & Gansemer-Topf, A.M. (2012). Finances and retention: Trends and potential implications. In A. Seidman (Ed.), *College student retention: Formula for student success.* New York, NY: Rowman & Littlefield.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461-464.

Scott, M., Bailey, T., & Kienzl, G. (2006). Relative success? Determinants of college graduation rates in public and private colleges in the U.S. *Research in Higher Education, 47*, 249-279.

Scott-Clayton, J. & Rodriguez, O. (2015). Development, discouragement, or diversion? New evidence on the effects of college remediation policy. *Education Finance and Policy, 10*(1), 4-45.

Seidman, A. (2012). Taking action: A retention formula and model for student success. In A. Seidman (Ed.), *College student retention* (2nd ed., pp. 267-284). New York, NY: Rowman & Littlefield.

Seo, E.H. (2012). Cramming, active procrastination, and academic achievement. *Social Behavior and Personality, 40*(8), 1330-1340.

Sharabiani, A., Karim, F., Sharabiani A., Atanasov, M., & Darabi, H. (2014). An enhanced Bayesian network model for prediction of students' academic performance in engineering programs. *Proceedings of IEEE Global Engineering Education Conference (EDUCON),* 832-837.

Soria, K. M. & Stebleton, M. J. (2012). First-generation students—academic engagement and retention. *Teaching In Higher Education*, *17*(6), 673–685.

Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L. & Cowell, R.G. (1993). Bayesian analysis in expert systems. *Statistical Science, 8*(3), 219-247.

Spirtes, P., Glymour, C. & Scheines, R. (2000). *Causation, prediction and search* (2nd ed.). Cambridge: MIT Press.

Steele, C.M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613-629.

Stevens, J.W. & O'Hagan, A. (2002). Incorporation of genuine prior information in cost-effectiveness analysis of clinical trial data. *International Journal of Technology Assessment in Health Care*, *18*(4), 782-790.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Ryal Statistical Society, 36*(2), 111-147.

Swail, W.S. (2004). *The art of student retention: A handbook for practitioners and administrators.* Austin, TX: Educational Policy Institute.

Thayer, P.B. (2000). Retention of students from first-generation and low-income backgrounds. *Opportunity Outlook, May,* 2-9.

Thomas, S.L. (2000). Ties that bind: A social network approach to understanding student integration and persistence. *The Journal of Higher Education*, *71*(5), 591-615.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89–125.

Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. *Journal of Higher Education*, *59*(4), 438–455.

Tinto, V. (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition* (2nd ed.). Chicago/London: University of Chicago Press.

Titus, M. A. (2004). An examination of the influence of institutional context on student persistence at 4-year colleges and universities: A multilevel approach. *Research in Higher Education, 45*(7), 673–699.

Titus, M.A. (2006). Understanding the influence of the financial context of institutions on student persistence at four-year colleges and universities. *Journal of Higher Education, 77*(2), 353-375.

Torabi, R., Moradi, P., & Khantaimoori, A.R. (2012). Predict student scores using Bayesian networks. *Procedia – Social and Behavioral Sciences, 46,* 4476-4480.

Tross, S.A., Harper, J.P., Osher, L.W., & Kneidinger, L.M. (2000). Not just the usual cast of characteristics: Using personality to predict college performance and retention. *Journal of College Student Development, 41,* 323-334.

Tu, J.V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for prediction of medical outcomes. *Journal of Clinical Epidemiology, 49*(11), 1225-1231.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS). (2014). Table 326.30: Retention of first-time degree-seeking undergraduates at degree-granting postsecondary institutions, by attendance status, level and control of institution, and percentage of applications accepted: 2006 to 2013. In U.S. Department of Education, National Center for Education Statistics (Ed.), *Digest of Education Statistics* (2014 ed.). Retrieved from http://nces.ed.gov/programs/digest/d14/tables/dt14_326.30.asp

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling, 203,* 312-318.

Waugh, G., Micceri, T., & Takalkar, P. (1994). *Using ethnicity, SAT/ACT scores, and high school GPA to predict retention and graduation rates.* Paper presented at the 1994 Florida Association for Institutional Research Conference, Orlando, FL.

Weatherford, R. (1982). *Philosophical foundations of probability theory*. London: Routledge & Kegan Paul.

Webber, D.A. & Ehrenberg, R.G. (2010). Do expenditures other than instruction affect graduation and persistence rates in American higher education? *Economics of Education Review, 29*(6), 947-958.

Weiss, G., & Provost, F. (2003). Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research, 19,* 315-354.

Wilcox, P., Winn, S. & Fyvie-Gauld, M. (2005). 'It was nothing to do with the university, it was just the people': The role of social support in the first-year experience of higher education. *Studies in Higher Education, 30*(6), 707-722.

Wimshurst, K., Wortley, R., Bates, R., & Allard, T. (2006). The impact of institutional factors on student academic results: Implications for 'quality' in universities. *Higher Education Research and Development, 25*(2), 131-145.

Winston, G.C. & Zimmerman, D.J. (2004). Peer effects in higher education. In C.M. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 395-421). Chicago: University of Chicago Press.

Woodberry, O., Nicholson, A.E., Korb, K., & Pollino, C. (2005). Parameterising Bayesian networks. In G. Webb & X. Yu (Eds.) *AI 2004: Advances in artificial intelligence. Lecture notes in computer* science (Vol. 339, pp. 711-745). Berlin/Heidelberg: Springer.

Yang, R. & Berger, J. (1997*). A catalog of noninformative priors.* Institute of Statistics and Decision Sciences Discussion Paper. Durham, NC: Duke University.

APPENDICES

Appendix A

Table A1
Variables Commonly Cited as Influences on Retention and Examples in the Literature

Broad Category	Predictors	Literature Examples
Student Characteristics (Pre-College)	Race-Ethnicity	Scott , Bailey, & Kienzl, 2006 Webber & Ehrenberg, 2010 Pike, Hansen, & Childress, 2014
	Gender	Pike, 2013 Scott , Bailey, & Kienzl, 2006 Webber & Ehrenberg, 2010
	Socioeconomic Status/Pell Eligibility	Hosch, 2008 Webber & Ehrenberg, 2010 Pike, Hansen, & Childress, 2014
	First Generation	Thayer, 2000 Longwell-Grice & Longwell-Grice, 2008 Soria & Stebleton, 2012
Pre-College Academic Records	High School GPA	Waugh, Micceri, & Takalkar, 1994 Adelman, 1999 Fredricks, Blumenfield, & Paris, 2004 ACT, 2010
	Standardized Achievement Test Scores	Kahn & Nauta, 2001 Reason, 2003 Lotkowski, Robbins, & Noeth, 2004 Astin & Oseguera, 2005
	Class Rank	Adelman, 1999
Student Characteristics (College)	Financial Support/Ability to Pay	Titus, 2006 Astin & Oseguera
	Socioeconomic Status/Pell Eligibility	Hosch, 2008 Webber & Ehrenberg, 2010
	Remediation/Remedial Courses	Roska, Davis, Jaggars, Zeidenberg, & Cho, 2009 Scott-Clayton & Rodriguez, 2015
	Academic Engagement	Beck & Davidson, 2001
		Bowen, Chingos, & McPherson, 2009
	Self-Efficacy/Personality	Boulter, 2002 Chemers, Hu, & Garcia, 2001 Tross, Harper, Osher, & Kneidinger, 2000
	Study Skills	Robbins, Lauver, Le, Davis, Langley, & Carlstrom, 2004
	Attendance Patterns	Harrington & Fogg, 2009

	First-Year Academic Performance	Makuakane-Drechsel & Hagedorn, 2000 Kiser & Price, 2008
	Commuter/Boarder	Ryan, 2004 Scott , Bailey, & Kienzl, 2006 Hosch, 2008
	Distance from Home	Bista & Foster, 2011
	Full-Time/Part-Time	Bean & Metzner, 1985 Reason, 2003
	Social Support	DeBerard, Spielmans, & Julka, 2004 Wilcox, Winn, & Fyvie-Gauld, 2005 Eckles & Stradley, 2011
	Social Involvement/Engagement	Hurtado & Carter, 1997 Winston & Zimmerman, 2004 Hausmann, Schofield, & Woods, 2007
Institutional Characteristics	Institutional Control	Ryan, 2004 Astin & Oseguera, 2012 Pike, 2013
	Institutional Selectivity	Gansmer-Topf & Schuh, 2006 Astin & Oseguera, 2012
	Academic Support	Ryan, 2004 Oseguera & Rhee, 2009 Pike, 2013
	First-Year Programs	Porter & Swing, 2006 Howard, 2013
	Institutional Mission	Titus, 2004 Astin & Oseguera, 2012 Seidman, 2012 Pike & Graunke, 2015
	Institution Size	Ryan, 2004 Pike, 2013
	Organizational Behavior	Berger, 2001-2002 Kuh, 2001-2002
	Focus on Retention	Bonous-Hammarth, 2000 Longden, 2006 Oseguera & Rhee, 2009
	Campus Climate	Berger & Milem, 2000 Titus, 2004 Oseguera & Rhee, 2009

Appendix B

Structure Elicitation Session Worksheet

Variable	General Variable Category	Proposed Operating Definition	Variable Classes	Causal Influence (Retention)	Causal Influence (Other Variables)	Conditional Independence	Variable Weight
				Which variables can be considered direct causes of retention/attrition? Which variables are considered indirect causes of retention/attrition as they actually influence other included variables?		Does a student’s state on variable X give you a lot of information about how they might be represented on (non-output) variable Y?	Please rank the included variables in terms of ultimate influence on retention – identify the strongest and weakest and move inward from there.
Gender	Student Characteristics (Pre-College)	Self-identified gender	Male/Female				
Socioeconomic Status/Pell Eligibility	Student Characteristics (Pre-College)	Pell eligibility acts as a proxy for socioeconomic status. Students eligible for Federal Pell Grants are considered low income/high financial need.	Pell eligible/Not Pell eligible				

Race-Ethnicity	Student Characteristics (Pre-College)	Federally-defined race/ethnicity categories, as reported by student off Common Application.	IPEDS Race/Ethnicity Categories			
First Generation	Student Characteristics (Pre-College)	Self-reported and defined in Common Application as an individual both of whose parents did not complete a baccalaureate degree, or, in the case of an individual who regularly resided with and received support from only one parent, an individual whose only such parent did not complete a baccalaureate degree.	First Generation/ Not First Generation			
High School GPA	Academic Preparation (Pre-College)	High school GPA used to determine admission eligibility - this is calculated from a student's high school transcript, but only using certain courses of interest to admitting institution.	0.0-.99 1.0-1.99 2.0-2.99 3.0-4.0			

Standardized Achievement Test Scores	Academic Preparation (Pre-College)	ACT or SAT scores typically used for admissions decisions. SAT Scores consist of Critical Reading, Math, and Writing Components. ACT Composite score calculated from scores on English, Math, Reading, and Science tests.	TBD			
Class Rank	Academic Preparation (Pre-College)	Rank among high school class - typically presented as percentile.	TBD			
Financial Support/Ability to Pay	Student Characteristics (College)	Percent of FAFSA-determined need met by institution	0-50% 51-75% 76-100%			
Socioeconomic Status/Pell Eligibility	Student Characteristics (College)	Pell eligibility acts as a proxy for socioeconomic status. Students eligible for Federal Pell Grants are considered low income/high financial need.	Pell eligible/Not Pell eligible			
Remediation/ Remedial Courses	Student Characteristics (College)	Elementary courses required as a prerequisite to college-level coursework.	Participation/Non-Participation			

Academic Engagement	Student Characteristics (College)	Frequency of: student interaction with faculty, insightful, co-curricular contribution to class discussions, synthesis of coursework.	TBD			
Study Skills	Student Characteristics (College)	Activities necessary to organize and complete school work tasks, and prepare for and take tests, including time management, test taking skills, using information resources, taking notes in class and interacting with faculty.	TBD			
Attendance Patterns	Student Characteristics (College)	Whether or not a student consistently attends scheduled classes.	Yes = 75% of classes?			
First-Year Academic Performance	Student Characteristics (College)	Cumulative GPA at the end of first academic year. *Or first academic semester?*	0.0-.99 1.0-1.99 2.0-2.99 3.0-4.0			
Commuter/ Boarder	Student Characteristics (College)	Whether or not a student resides in campus housing during first year.	Yes/No			

Distance from Home	Student Characteristics (College)	Distance from students main family/caregivers.	0-1 hours More than 1 hour, less than 2 More than 2 hours, less than 4 More than 4 hours			
Full-Time/Part-Time	Student Characteristics (College)	Full time enrollment indicated by twelve or more hours during both first-year semesters.	Full-time/Part-time			
Social Support	Student Characteristics (College)	Integration with, and emotional support received by, friends in the first year of college.	TBD			
Social Involvement/ Engagement	Student Characteristics (College)	Integration into and participation in social activities (college-wide and residential) that fosters a sense of belonging.	TBD			
Institutional Control	Institutional Characteristics	Whether or not an institution is privately controlled or part of a public system.	Private/Public			
Institutional Selectivity	Institutional Characteristics	The level at which applying students are typically admitted. More selective institutions have lower admission rates and higher pre-	50-65% 66-75% 76-100%			

		college academic admissions requirements.				
Academic Support	Institutional Characteristics	The extent to which an institution provides and promotes academic support to first-year students.	TBD			
First-Year Programs	Institutional Characteristics	Whether or not an institution provides programming aimed at first-year students, and the extent to which these are supported, promoted and sustained.	Yes/No			
Institutional Mission	Institutional Characteristics	The educational mission of the institution, as measured by persistence goals. The persistence goals of a community college are very different than those of a four-year institution.	Focus on Persistence/No Focus on Persistence			
Focus on Retention	Institutional Characteristics	The extent to which an institution has prioritized retention, as measured by staffing and programming geared	Prioritized Retention/Not Prioritized Retention			

		towards increasing retention and graduation rates.				
Campus Climate	Institutional Characteristics	Culture in which students feel valued and empowered by campus peers, faculty, and administration	Positive Campus Climate/ Negative Campus Climate			

Appendix C

Examples of Elicitation Session II Materials & Technique

Appendix C.1 - Pre-Elicitation Materials (Provided One Week Prior to Elicitation Session II)

Elicitation Date/Time: Thursday, December 17, 2015, 9:00 – 11:00 am (Jenn's Office)

Participants: Jennifer Cannady, Machamma Quinichett, Corey Dunn (facilitator)

Objectives:

1. Review and critique interim model incorporating proposed structure and variables from first elicitation session
2. Obtain probability distributions that represent experts' experience and uncertainty about specific variables related to retention (see highlighted variables in attached list of variables included in model)
3. Review resulting model specifications

Elicitation of Probabilities Notes:

- You will *not* be asked to provide single estimates of probabilities
- You will be asked to discuss plausible ranges of probabilities for each uncertain variable, and whether or not some values are more likely than others
- Uncertainty is part of the process – feel free to express your uncertainty

Common Causes of Bias to Avoid:

- Availability – easier-to-recall occurrence may incorrectly be deemed more important or likely.

- Example: Notable exceptions, students to whom experts were more familiar, etc.

- Representativeness - similarity doesn't necessarily mean events are probabilistically related. The conjunction of two events can't be more probable than either event separately.
  - Example: expert suggesting that the likelihood that a student has unmet need and is retained is greater than the likelihood that a student has unmet need.
  - Particularly relevant to the elicitation of conditional probabilities
- Adjustment & Anchoring – experts may calculate probability based on an initial value
- Overconfidence
- Hindsight Bias – experts who have seen sample data may let it influence their opinion

Appendix C.2 - Probability Review Provided to Experts (Provided One Week Prior to Elicitation Session II)

Probability – measure of the likelihood of a random phenomenon or chance behavior. Describes the long-term proportion with which a certain outcome will occur in situations with short-term uncertainty.

- Probabilities are numbers between zero and one – the closer it is to one, the more likely the event is to occur.
- Computing Probability Using the Classical Method:
  If an experiment has *n* equally likely outcomes and if the number of ways that an event *E* can occur is *m*, then the probability of *E*, *P*(*E*), is

$$P(E) = \frac{\text{number of ways that } E \text{ can occur}}{\text{number of possible outcomes}} = \frac{m}{n}$$

Subjective probability of an outcome is probability obtained on the basis of personal judgment.
Experts will be asked to provide estimates of subjective probability.

- Independence – two events E and F are independent if the occurrence of event *E* in a probability experiment does not affect the probability of event *F*.
  - Example – obtaining heads on first coin toss has no effect on the likelihood of obtaining heads on second toss.
- Dependence – two events are dependent if the occurrence of event *E* in a probability experiment affects the probability of event *F*.
  - Example – the likelihood of higher career earnings is related to education level.

Conditional Probability – the probability that event F occurs, given that the event E has occurred

P(*F*|*E*) – the probability of event *F* given event *E*

- Very important concept in our retention model

Important Rules of Probability

- The probability of any event must be between 0 and 1, inclusive. If we let *E* denote any event, then $0 \leq P(E) \leq 1$.
- The sum of the probabilities of all outcomes must equal 1.
- If *E* and *F* are independent events, then *P (E* and *F) = P(E) * P(F)* **(Multiplication Rule)**
- If E and F are any two events, then

$$P(F|E) = \frac{P(E \text{ and } F)}{P(E)}$$

- The probability of event *F* occurring, given the occurrence of event *E*, is found by dividing the probability of *E* and *F* by the probability of *E*. **(Conditional Probability Rule)**
- Two events E and F are independent if *P(E|F)* = P*(E)* or, equivalently, if P(*F*|*E)* = P*(F).* **(Conditional Independence)**

Appendix C.3 - Probability Elicitation Protocol – “Attendance Patterns” Example

In our earlier session, you indicated that Attendance Patterns are related to Retention, but are also affected by a student's Study Skills. As of yet, we don't have any data or data proxy to quantify students' attendance patterns.

First, let's talk about what you think about students’ attendance patterns, given no other information. We’ve defined attendance of more than 80% of courses as "Consistent." Let’s talk about how attendance is related to retention - given a student consistently attends courses, what is the likelihood that they'll retain (given no other knowledge at this point)? What is your level of confidence in this?

We also talked about how attendance patterns are influenced by study skills - now I'm going to ask you to think about attendance patterns as the effect and Study Skills as the cause. Given a student has Developed Study Skills, what is the probability that they will consistently attend class? What is your confidence in this estimate? Do you think that this influence is strong or weak?

Now, we need to pull this whole line of influence together - let's talk about the probability of retention given SS and AP…

Questions:

Given a student has Developed SS and Consistent AP, what is an estimate of a high/low/medium probability they'll retain?

Given a student has UnderDeveloped SS and Consistent AP, what is an estimate of a high/low/medium probability they’ll retain?

Given a student has Developed SS and Inconsistent AP, what is an estimate of a high/low/medium probability they’ll retain?

Given a student has UnderDeveloped SS and Inconsistent AP, what is an estimate of a high/low/medium probability they'll retain?

Remember your conclusions about the strong/weak influences on Retention - if you think that developed study skills are more influential on retention than consistent attendance patterns (as you indicated in initial session), make sure that the probabilities reflect this.

Appendix D

Table D1
Final BN Node Sensitivities

	Node	Mutual Info	Percent	Variance of Beliefs
Sensitivity of "Retained" to a finding at another node:	Academic Rank	.02072	3.21	.0042552
	Financial Need & Risk Profile	.01804	2.8	.0037342
	Study Skills	.00585	.907	.0011329
	Attendance Patterns	.00405	.628	.0008366
	Academic Support	.00284	.44	.0005185
	Distance from Home	.00234	.363	.0004477
	Social Support	.00112	.174	.0002162
	Advisor/Major Mismatch	.00028	.043	.0000524
	Race/Ethnicity	.00000	0	.0000000
Sensitivity of "Study Skills" to a finding at another node:	Attendance Patterns	.01431	1.47	.0048639
	Academic Support	.01427	1.47	.0047703
	Retained	.00585	.602	.0019775
	Academic Rank	.00018	.0189	.0000612
	Financial Need & Risk Profile	.00016	.0166	.0000537
	Distance from Home	.00002	.00199	.0000064
	Social Support	.00001	.000959	.0000031
	Advisor/Major Mismatch	.00000	.000234	.0000008
	Race/Ethnicity	.00000	0	.0000000
Sensitivity of "Race/Ethnicity"	Academic Support	.01741	1.3	.0045111

to a finding at another node:	Social Support	.00000	0	.0000000
	Attendance Patterns	.00000	0	.0000000
	Distance from Home	.00000	0	.0000000
	Advisor/Major Mismatch	.00000	0	.0000000
	Academic Rank	.00000	0	.0000000
	Study Skills	.00000	0	.0000000
	Financial Need & Risk Profile	.00000	0	.0000000
	Retained	.00000	0	.0000000
Sensitivity of "Distance from Home" to a finding at another node:	Retained	.00234	.235	.0008069
	Academic Rank	.00007	.00731	.0000250
	Social Support	.00007	.00679	.0000232
	Financial Need & Risk Profile	.00006	.00642	.0000219
	Study Skills	.00002	.00195	.0000066
	Attendance Patterns	.00001	.00144	.0000049
	Academic Support	.00001	.000892	.0000030
	Advisor/Major Mismatch	.00000	.00000	.0000003
	Race/Ethnicity	.00000	0	.000000
Sensitivity of "Attendance Patterns" to a finding at another node:	Study Skills	.01431	2.4	.0025088
	Retained	.00405	.679	.0007532
	Academic Support	.00025	.0415	.0000429
	Academic Rank	.00013	.0226	.0000233
	Financial Need & Risk Profile	.00012	.0198	.0000205

	Distance from Home	.00001	.0024	.0000025
	Social Support	.00001	.00115	.0000012
	Advisor/Major Mismatch	.00000	.000288	.0000003
	Race/Ethnicity	.00000	0	.000000
Sensitivity of “Academic Support” to a finding at another node:	Race/Ethnicity	.01741	1.13	.0031333
	Study Skills	.01427	.93	.0025795
	Retained	.00284	.185	.0003825
	Attendance Patterns	.00025	.0161	.0000402
	Academic Rank	.00008	.0055	.0000122
	Financial Need & Risk Profile	.00007	.00483	.00000107
	Distance from Home	.00001	.000578	.0000013
	Social Support	.00000	.000274	.0000006
	Advisor/Major Mismatch	.00000	.000000	.0000002
Sensitivity of “Social Support” to a finding at another node:	Retained	.00112	.118	.0003647
	Distance from Home	.00007	.00713	.0000217
	Academic Rank	.00004	.0037	.0000113
	Financial Need & Risk Profile	.00003	.00325	.0000099
	Study Skills	.00001	.000996	.0000030
	Attendance Patterns	.00001	.000728	.0000022
	Academic Support	.0000	.000448	.0000014
	Advisor/Major Mismatch	.0000	.000000	.0000001
	Race/Ethnicity	.0000	0	.0000000

Sensitivity of "Academic Rank" to a finding at another node:	Financial Need & Risk Profile	.32327	14.1	.0214356
	Retained	.02072	.904	.0021506
	Study Skills	.00018	.00801	.0000130
	Attendance Patterns	.00013	.00587	.0000098
	Academic Support	.00008	.00369	.0000058
	Distance from Home	.00007	.00317	.0000051
	Social Support	.00004	.00153	.0000025
	Advisor/Major Mismatch	.00001	.000373	.0000006
	Race/Ethnicity	.0000	0	.0000000
Sensitivity of "Financial Need & Risk Profile" to a finding at another node:	Academic Rank	.32327	39.6	.0734777
	Retained	.01804	2.21	.0051425
	Study Skills	.00016	.0197	.0000424
	Attendance Patterns	.00012	.0145	.0000313
	Academic Support	.00007	.00908	.0000194
	Distance from Home	.00006	.0078	.0000167
	Social Support	.00003	.00376	.0000081
	Advisor/Major Mismatch	.00001	.000917	.0000020
	Race/Ethnicity	.0000	0	.0000000
Sensitivity of "Advisor/Major Mismatch" to a finding at another node:	Retained	.00028	.0302	.0000849
	Academic Rank	.00001	.000926	.0000026
	Financial Need & Risk Profile	.00001	.000815	.0000023
	Study Skills	.00000	.000251	.0000007
	Attendance Patterns	.00000	.000185	.0000005

	Academic Support	.00000	.000116	.0000003
	Distance from Home	.00000	.000000	.0000003
	Social Support	.00000	.000000	.0000001
	Race/Ethnicity	.00000	0	.0000000