

**DESIGNING DATA-DRIVEN VIRTUAL PATIENTS FOR
HEALTH SCIENCES EDUCATION**

by

Dmitriy Babichenko

Masters of Information Sciences (MSIS), University of
Pittsburgh School of Information Sciences, 2011

Submitted to the Graduate Faculty of
the School of Computing and Information in partial
fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH
SCHOOL OF COMPUTING AND INFORMATION

This dissertation was presented

by

Dmitriy Babichenko

It was defended on

September 21, 2018

and approved by

Marek Druzzel, PhD, Professor and Advisor, School of Computing and Information

Yu-Ru Lin, PhD, Associate Professor, School of Computing and Information

Michael Lewis, PhD, Professor, School of Computing and Information

James B. McGee, MD, Assistant Dean for Medical Education Technology, Director,

Laboratory for Educational Technology, School of Medicine

Dissertation Director: Marek Druzzel, PhD, Professor and Advisor, School of

Computing and Information

DESIGNING DATA-DRIVEN VIRTUAL PATIENTS FOR HEALTH SCIENCES EDUCATION

Dmitriy Babichenko, PhD

University of Pittsburgh, 2018

Electronic virtual patients (VPs) are interactive screen-based computer simulations of real-life clinical scenarios that are widely used for the purposes of health sciences education. Advances in computational modeling and availability of large patient cohort datasets from Electronic Medical Record (EMR) systems have created an opportunity to develop a new type of VPs, where cases are based on and simulate real patient clinical treatment processes and outcomes.

Traditional VP cases are static narrative representations of clinical scenarios that are presented to health sciences students in order to teach a clinical topic of interest. This research investigates the feasibility of authoring and presenting virtual patient cases that leverage Bayesian network (BN) models learned from EMR data to present clinical scenarios and control outcomes of learners' decisions within the context of a presented VP. Because the underlying models are based on real patient data, each decision made by a learner would affect the probability of each outcome occurring in the same way as with real patients.

Additionally, this dissertation explores the challenges related to using BN models in the context of VP case authoring and presentation, and experimentally compares a VP case based on a BN model to one created using a traditional narrative-branched VP system across multiple categories, including meeting learning objectives, accuracy in depicting the chosen clinical scenario, introducing/reinforcing relevant clinical skills,

providing formative feedback, scenario realism, and learner engagement.

Furthermore, this work investigates the extent to which the use of annotated BN models in VP cases facilitates modifying an existing VP case by allowing case authors to manipulate the underlying model in such a way that the modified VP case meets alternate learning objectives.

Last, but not least, this research provides practical and methodological contributions to the body of work in the areas of health sciences education, problem-based learning, and clinical simulation design and evaluation. More specifically, this work (1) defines criteria and guidelines for designing VP cases based on BN models, (2) identifies and describes shortcomings and challenges associated with different BN modeling approaches for different types of clinical scenarios, (3) and presents a framework for evaluating VP cases.

TABLE OF CONTENTS

PREFACE	xvii
1.0 INTRODUCTION	1
1.1 Motivation	4
1.2 Aims of My Dissertation	6
1.3 Research Questions	6
1.4 Findings	7
1.4.1 Research Question 1 Findings Summary	7
1.4.2 Research Question 2 Findings Summary	7
1.5 Contributions	8
1.6 Dissertation Outline	10
2.0 SIMULATION AND TUTORING SYSTEMS IN MEDICAL EDUCATION	11
2.1 Serious Games in Medical Education	11
2.2 Simulation Types	13
2.3 Virtual Patient Simulators	15
2.4 Virtual Patient Case Authoring	16
2.5 Generating VP Cases From Clinical Knowledge Bases	21
3.0 BAYESIAN NETWORK MODELS	22
3.1 Bayes' Theorem	22
3.2 Bayesian Networks	24
3.3 Bayesian Network Models	25
3.3.1 Naive Bayes	26

3.3.2	Augmented Naive Bayes	27
3.4	Constraint-Based Search	28
3.5	Dynamic Bayesian Network Models	32
3.6	Proposed Use of Bayesian Network Models in Virtual Patient Cases	34
4.0	MODEL PATIENT SYSTEM IMPLEMENTATION AND USER INTERFACE	
	DETAILS	38
4.1	System Architecture Summary	43
4.2	Learning BN Models From The Data	45
4.3	Interacting with BN Models	46
4.4	Non-destructive Model Updating	47
4.5	Database	49
4.6	User Interface	50
4.6.1	Case Authoring	50
4.6.2	BN Model Annotation	51
4.6.3	Case Introduction	56
4.6.4	Additional Case Content	56
4.6.5	Case Metrics	57
4.6.6	Case Rules	58
4.6.6.1	Rule Triggers	60
4.6.6.0	Rule Outcomes	62
4.6.7	VP Case Player	62
5.0	VIRTUAL PATIENT CASES	67
5.1	VP Cases	67
5.2	Patient Cohort Data Set	68
5.3	Virtual Patient Case 1: Managing Sepsis	69
5.3.1	Managing Sepsis: vpSim Case	69
5.3.2	Sepsis BN Model	71
5.3.3	Managing Sepsis: ModelPatient Case	75

5.4	Virtual Patient Case 2 – The Management of Diabetic Ketoacidosis	79
5.4.1	Managing DKA: vpSim Case	79
5.4.2	Diabetic Ketoacidosis Model	83
6.0	STUDY DESIGN, ANALYSIS, AND FINDINGS	86
6.1	Research Question 1 (RQ1)	87
6.1.1	Subject Selection Criteria and Demographic Information	88
6.1.2	RQ1 Study Design	90
6.1.3	Analysis and Findings	92
6.1.3.1	Analysis of Free-Text Responses (Affinity Diagramming)	96
6.1.3.2	Common Themes From Text Responses About ModelPatient	98
6.1.3.3	Common Themes From Text Responses About vpSim	99
6.2	Research Question 2 (RQ2)	100
6.2.1	Subject Selection Criteria and Demographic Information	100
6.2.2	RQ2 Study Design	102
6.2.3	Analysis and Findings	105
7.0	DISCUSSION AND FUTURE WORK	110
7.1	Discussion	110
7.1.1	Reflections on Research Question 1	110
7.1.2	Reflections on Research Question 2	116
7.2	Case Authoring Lessons Learned: Feasibility Criteria / Guidelines for Creating VP Cases Based on Bayesian Network Models	118
7.3	Limitations	123
7.4	Future Work	126
7.4.1	Alternate Modeling Approaches	126
7.4.1.1	Representing Temporal VP Cases with Multiple Chained BN Models	126

7.4.1.2	Representing Temporal VP Cases with Hidden Markov Models	130
7.4.1.3	Representing Temporal VP Cases with Interval Temporal Bayesian Networks	131
7.4.2	Automatically Generating Association Rules	133
7.4.3	Automatically Generating Feedback	134
7.4.4	Future Studies	135
	APPENDIX A. PATIENT COHORT DATA DESCRIPTION	137
A.1	Original Patient Cohort Description	137
A.1.1	Inclusion criteria	137
A.1.2	Additional variables	138
A.2	Sepsis Patient Cohort	138
A.3	Dataset Features	139
A.4	Diabetic Ketoacidosis Patient Cohort Data	144
	APPENDIX B. NORMAL VALUE RANGES FOR VITALS AND LABORATORY TESTS	149
	APPENDIX C. MODEL PATIENT CASE PRESENTATION AND AUTHORIZING EVALUATION QUESTIONNAIRES	152
C.1	Questionnaire: ModelPatient Case Presentation Evaluation	152
C.2	Questionnaire: Case Content and Learning Objectives Comparison Between the “Managing Sepsis” VP Case Presented in vpSim and ModelPatient	154
C.3	Questionnaire: ModelPatient Case Authoring Evaluation	155
	APPENDIX D. SAMPLE VP CASE COMPLETION REPORTS	158
D.1	Sample vpSim Completion Report	158
	APPENDIX E. SUBJECT RECRUITMENT MATERIALS	162
E.1	Recruitment Email for Research Question 1 Experiment	162
E.2	Recruitment Email for Research Question 2 Experiment	163

APPENDIX F. PYTHON SCRIPT USED TO SELECT RESEARCH QUESTION

1 SUBJECTS INTO TWO GROUPS 165

APPENDIX G. RESEARCH QUESTION 1 STUDY LOG 166

APPENDIX H. RESEARCH QUESTION 2 STUDY LOG 170

BIBLIOGRAPHY 172

LIST OF TABLES

1	Prior probability distribution table for the “Family History of Thyroid Cancer” node	25
2	Conditional probability distribution table for the “Tall Cell Variant (TCV)” node	25
3	VP cases reviewed for authoring features design	39
4	VP case features common to vpSim, DecisionSim, OpenLabyrinth, and Laerdal SimMan	40
5	Author-defined rule triggers	60
6	Author-defined rule outcomes	61
7	Accuracy of Predicting Discharge Location	73
8	Sepsis Model Confusion Matrix	73
9	DKA VP treatment timeslices presented in the vpSim version of the case	81
10	Subjects’ Clinical Specialties	89
11	Item-Total Statistics (Cronbach’s Alpha if Item Deleted)	92
12	Wilcoxon Signed Ranks Test Results	94
13	Summary Statistics — Case Presentation and Learning Objectives Comparison	96
14	Subjects’ Clinical Specialties	102
15	“Managing Sepsis” VP Case Primary Learning Objectives	103
16	“Managing Sepsis” VP Case Alternate Learning Objectives	103
17	Summary Statistics — Authoring System Difficulty Rating	105

18	Summary Statistics - How Well Does the Modified Case Reflect Alternate Learning Objectives	106
19	Medical problems commonly encountered by the University of Pittsburgh 3 rd and 4 th -year medical students during clinical rotations + common medical problems encountered by the Henry Ford Health System residents [1].	119
20	“Allen’s 13 atomic temporal relations to represent temporal relationships between two events X and Y” [2, 3]	132
21	Sepsis Dataset — Demographic Information	139
22	Sepsis Dataset — Comorbidities	140
23	Sepsis Dataset — Administered (Bedside) Medications	141
24	Sepsis Dataset — Laboratory Tests	142
25	Sepsis Dataset — Vital Signs	143
26	Sepsis Dataset — Outcomes	143
27	DKA Dataset — Demographic Information	144
28	DKA Dataset — Comorbidities	145
29	DKA Dataset — Administered (Bedside) Medications	146
30	DKA Dataset — Laboratory Tests — Initial Values at the Time of Hospital Admission	147
31	DKA Dataset — Initial Vital Signs at the Time of Hospital Admission	148
32	DKA Dataset — Outcomes	148
33	Normal Value Ranges for Vitals	149
34	Normal Value Ranges for Laboratory Tests	149
35	Virtual Patient Case Evaluation Questionnaire 1	153
36	Case Content and Learning Objectives Comparison Questionnaire	154
37	Virtual Patient Case Authoring Evaluation	155
38	Virtual Patient Alternate Objective(s) Case Evaluation	156
39	Sample vpSim Completion Report	158
40	Sample ModelPatient Completion Report	160

41	Research Question 1 Study Log	166
42	Research Question 2 Study Log	170

LIST OF FIGURES

1	Blank narrative template in vpSim	3
2	Blank state template in Laerdal Learning Application (LLEAP) SimDesigner	3
3	Blank narrative template in OpenLabyrinth	4
4	Decision structure of a diabetic ketoacidosis virtual patient case.	5
5	Initial nodes (choices) of the DKA case designed in vpSim	17
6	A “Patient History” node from the DKA case designed in vpSim in the edit mode (being edited by a case author)	18
7	Example of rules (logic) authoring in vpSim	19
8	Subset of the decision structure of a diabetic ketoacidosis virtual patient case.	20
9	An example of a Bayesian network modeling the relationships among Thyroid Cancer, TCV, Radiation Exposure, and Family History of Thyroid Cancer.	24
10	A Naive Bayes model for the example classification problem.	27
11	A TAN model for the augmented Naive Bayes example problem.	28
12	A causal graph modeling the relationships among Smoking (S), Yellow Fingers (F), Cancer (C), and Asbestos Exposure (A).	29
13	The initial structure assumed by the PC algorithm.	30
14	The causal structure reflecting conditional independencies.	30
15	Example of final causal structure after the PC algorithm orients arcs.	31
16	Example structure of a standard BN model	33

17	Example structure of a 2-slice dynamic BN model	33
18	A model for patients with acid reflux learned from the data. Bars in the nodes CYP2C19 and GERD show marginal probability distributions for those nodes.	34
19	Probability distribution over GERD symptoms severity given Ibuprofen. . .	35
20	Probability distribution over GERD symptoms severity given PPI.	36
21	Probability distribution over GERD symptoms severity given PPI and CYP2C19.	36
22	System architecture diagram	44
24	Relationship between a patient’s smoking status and use of opiates with default state values.	46
25	Relationship between a patient’s smoking status and use of opiates after change in smoking status from the default to “Never Smoked.”	46
26	Relationship between a patient’s smoking status and use of opiates after change in smoking status from the default to “Current Smoker.”	46
27	Adobe Lightroom “rendering engine in its simplest form: The source image and your processing settings (in the form of rendering metadata) are processed in the rendering software and a finished image is created.” [4] .	48
28	“Create / Edit VP case” screen.	51
29	BN model editing and annotating tool.	52
30	State and annotation editor for a single BN node.	53
31	Virtual patient case introduction.	56
32	VP Case Player Content Editor.	57
33	Custom case metrics editor.	58
34	Rule editor.	59
35	VP Case Player — Learner View — Patient Demographics Screen.	63
36	VP Case Player — Learner View — Prescribed Drugs.	64
37	VP Case Player - Learner View - New Information Notifications.	64
38	VP Case Player - Learner View - New Information Notifications.	65

39	vpSim – Sepsis Case Narrative – Patient Introduction.	70
40	vpSim – Sepsis Case Quiz.	71
41	Sepsis BN model graph	72
42	ROC curve for “Discharge Location” state (AUC=0.816)	73
43	ROC curve for “Managed Care” state (AUC=0.662)	74
44	ROC curve for “Self Care” state (AUC=0.622)	74
45	ModelPatient — Sepsis Case Clinical Notes.	76
46	ModelPatient — Sepsis Case Labs.	76
47	ModelPatient — Learner prescribing normal saline bolus hydration treatment.	77
48	ModelPatient — Changes in the model’s probability distributions result in new information begin presented to the learner. Sections of the “EMR” that contain new information are denoted with red notification icons.	77
49	ModelPatient — The system displays changes to six vital signs based on the changes in the states’ probabilities of the corresponding BN model nodes following learner’s selection of “Normal Saline” hydration treatment (Figure 47)	78
50	vpSim DKA Case – The system displays changes vital signs and laboratory values based on learner’s choices on prior vpSim screen(s).	81
51	vpSim DKA Case – Fluids administration tutorial	83
52	DKA model structure	85
53	RQ1 Study Flowchart.	91
54	Partial BN model of sepsis trained with PC algorithm (Chapter 3, Section 3.4).	122
55	Partial conditional probability table (CPT) of the “discharge location” node from a BN model of sepsis trained with PC algorithm (Chapter 3, Section 3.4).	123

56 Initial parameters and response variables — first “sub-model” in a chain of BN network models. 127

57 Second “sub-model” in a chain of BN network models representing the “Saline Bolus” event. 127

58 Two “sub-model” chained together. Dashed arrow lines represent setting the initial states of “Anion Gap 1” and “Glucose 1” to the outcome probabilities of the corresponding states of the previous model in the chain. . . . 128

59 Multiple time slices represented as discrete BN models. 129

60 DKA treatment progression modeled with HMM. “S Init” represents initial vital signs and laboratory test results at ICU admission. S1 - S5 are latent states representing a patients’ state given a treatment. Death, Hospitalization, and Discharge are absorbing states. 131

61 A graphical representation of temporal dependencies among temporal entities Hydration (H), Vasopressor (V), and Systolic BP (SBP). In this example, the temporal relationships are “H occurs before V,” “H occurs before SBP,” and “V occurs before SBP or overlaps SBP.” 133

62 BN model representation (graphical explanation) in ModelPatient case authoring tool. 135

PREFACE

I would like to thank my adviser, Dr. Marek Druzdzel for providing help, guidance, and invaluable advice with my academic pursuits. Also, my committee members, Dr. Yuru Lin, Dr. Michael Lewis, and Dr. J.B. McGee for all their help, insight, feedback, and guidance. Dr. David Binion, Dr. Claudia Ramos Rivers, Dr. Benjamin Click and Dr. Neal Benedict for including me in their research projects, giving me access to data, and being patient with my lack of medical background. Dr. Gary Tabas and Dr. Neil Benedict for helping with modeling virtual patient cases as Bayesian network models. Dr. Lorin Grieve, Dr. Ravi Patel, and Dr. Randall Smith for support, insight, and feedback on clinical and educational aspects of my research. Dr. Kayla Booth and Dr. Robert Perkoski for giving me constant feedback on my research methods. Kelly Shaffer, Rachel Kelly, Dr. Robert Perkoski, Dr. Kayla Booth, Dr. Lorin Grieve and Dr. Colleen Culley for spending countless hours helping me with text analysis and with rating interview text responses.

I would also like to thank all the doctors, nurses, pharmacists, simulation experts, instructional designers, and educators who took hours from their incredibly busy schedules to participate in the studies associated with my work – I hope that I can repay your kindness one of these days.

I am incredibly grateful to my mother, Irina, my grandmother, Olga, and my baby sister Olga (not to be confused with my grandmother Olga) for their constant love and support.

Most importantly, I never would have been able to go on this journey without love and support of my amazing wife Irina and my beautiful daughters, Daniella and

Sophia. During all of the most difficult moments of my work, studies, and research you never let me feel sorry for myself and never allowed me to doubt my abilities.

Thank you!

1.0 INTRODUCTION

Problem-based learning (PBL) has gained wide acceptance in many educational domains, including medicine, business, engineering, and computer science. PBL originated in medical education in the 1950s [5] to address issues with application of foundational and theoretical knowledge to practical problems. Educators presented students with complex patient case histories to engage learners and to encourage them to apply knowledge learned in classroom to real-world problems [6].

One of the most important aspects of PBL is that if implemented correctly, it can drive the development of critical skills, such as presentation, communication, writing, and teamwork [7]. Furthermore, PBL provides a direct gateway for applying theory and knowledge presented in traditional didactic lectures to real-world problems. Another outcome of PBL is the positive effect on student engagement. Because PBL comprises of activities with immediate feedback and tangible outcomes, and because it focuses on practical problems with stakeholder engagement and realistic assessments, students find themselves more involved and engaged than they would be with more traditional homework assignments or in-class projects [7]. A systematic study of students' perceptions of skills that they have developed during PBL projects showed that students do indeed find PBL more engaging than traditional pedagogical methods [8].

Patient simulation is a type of PBL that has been a well-established part of the medical education paradigm since the 1950s. Clinical simulations allow for safe and supportive experiences for healthcare students to practice problem solving, diagnostic

skills and critical procedures without endangering real patients. Traditionally, medical schools have been employing standardized patients (actors trained to portray a patient in a medical situation) to simulate various clinical situations and to teach students bedside manner, diagnostic skills, and critical thinking. Traditional simulations, however, are logistically time-consuming and require considerable space to perform adequately. Advances in technology made it possible to create software-based virtual patient (VP) systems [9]. These VP simulations could be used at a lower logistical cost and have been shown to enhance knowledge and skill retention as well as teamwork, critical thinking, and satisfaction in learning [10, 11]. Software-based VPs have become widely adopted in all fields of healthcare education because of these strengths. Currently, the majority of software-based VPs are created within the institution that employs them. This internal development of VPs is driven by the lack of robust commercial alternative options. On account of this, VP programs are considered to be time and resource intensive to produce [12]. Additionally, the amount of time a faculty member must devote to learning the system, building the cases, and deploying them for students to use is prohibitively large. This has led to VPs being under-utilized in most facilities [13].

The current platforms for VP cases are essentially finite state machines that use blank templates to allow VP case authors to define and describe individual states (Figures 1, 2, 3). This means that all the information must be added by the author, increasing the burden of time on the creator and limiting the pool of authors to subject-matter experts. Due to this burden on faculty, the VP cases that are made gravitate towards linear experiences (a learner simply navigates through the teaching case by going from screen to screen in a linear fashion) that fail to capture the nuances of actually caring for a patient.

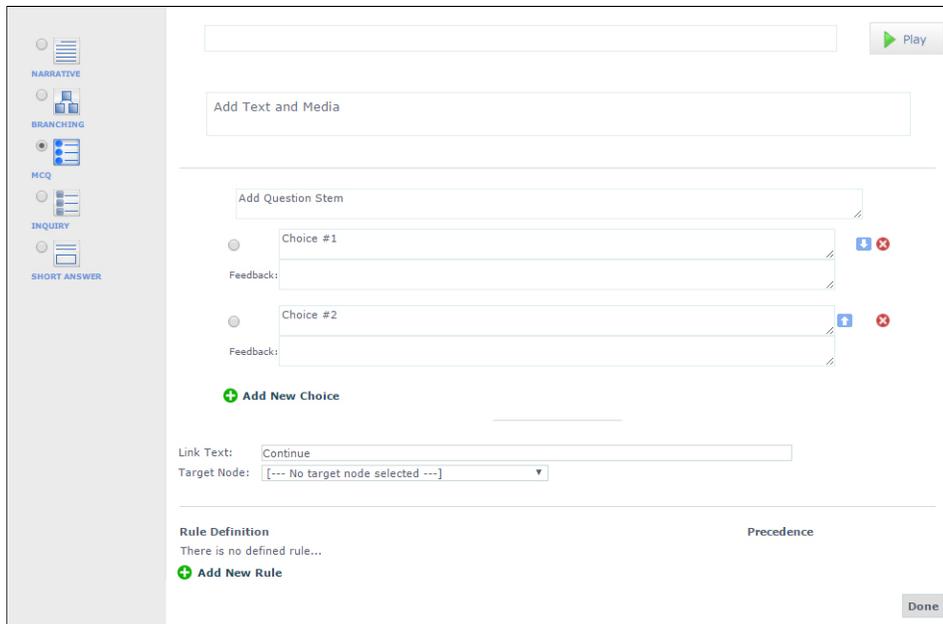


Figure 1: Blank narrative template in vpSim

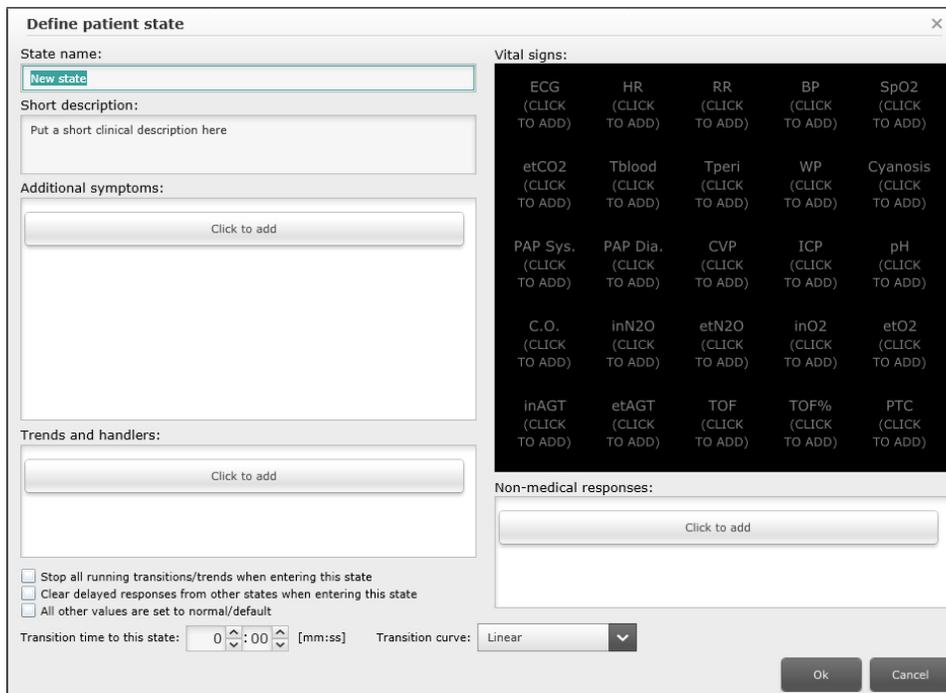


Figure 2: Blank state template in Laerdal Learning Application (LLEAP) SimDesigner

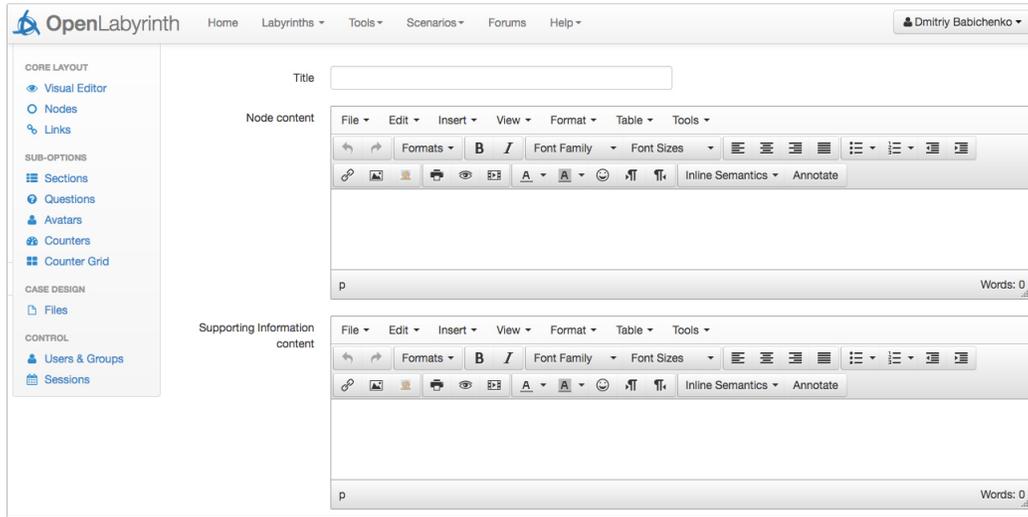


Figure 3: Blank narrative template in OpenLabyrinth

These current barriers present an opportunity for a VP platform to leverage the plethora of available electronic medical records (EMR) data sets and machine learning techniques to benefit educators and learners.

1.1 MOTIVATION

One of the most commonly discussed problems with current VP authoring systems is the difficulty of managing case complexity [14, 15, 16]. To manually create a complex teaching case that addresses the majority of common outcomes, common treatments, and diagnostic mistakes, as well as offers just-in-time learning opportunities to students often results in a visual equivalent of “spaghetti code.” Figure 4 shows a small subset of the decision structure of a diabetic ketoacidosis virtual pa-

tient case¹ [17] created in vpSim.² This particular case has a total 92 nodes (screens accessible by learners) and 67 decision points (screens where learners have to make decisions). Maintaining such a case, adding or removing pathways or decision options is a very time-consuming task indeed. Furthermore, existing VP systems provide low fidelity and limited re-playability, making the learning experience less engaging and ultimately less immersive [18].

The clear benefit of simulation in health sciences education combined with the need for realistic dynamic learner experiences and time-constraints of educators presents a need for a data- and model-driven VP platform.

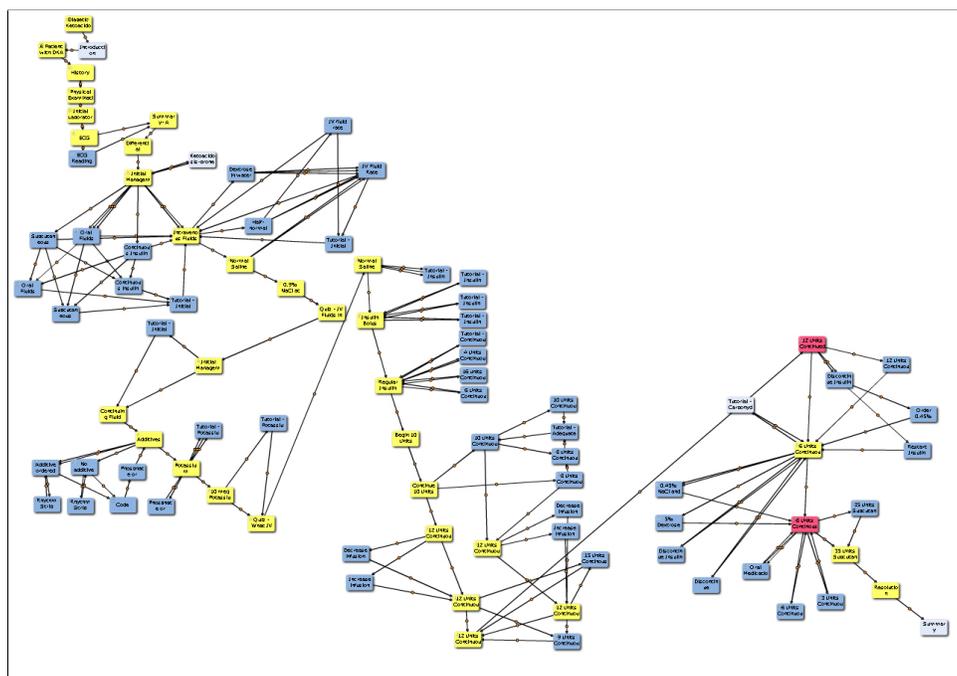


Figure 4: Decision structure of a diabetic ketoacidosis virtual patient case.

¹Permission is granted for one-time non-exclusive, non-transferable worldwide re-use and modification of the case, “Diabetic Ketoacidosis,” for research purposes, as long as you credit Annals Virtual Patients as the original source of the material as follows. Adapted from Tabas G, Kosytkowski M, Harsha Rao H (2015). Diabetic Ketoacidosis [Online patient case]. Annals Virtual Patients. Philadelphia, PA: American College of Physicians. Copyright 2016 American College of Physicians. Used with permission.

²vpSim: <http://vpsim.pitt.edu>

1.2 AIMS OF MY DISSERTATION

This research investigates the feasibility of authoring and presenting VP cases that leverage Bayesian network (BN) models learned from EMR data to present clinical scenarios and control outcomes of learners' decisions within the context of a presented VP. Because the underlying models are based on real patient data, each decision made by a learner would affect the probability of each outcome occurring in the same way as with real patients.

Furthermore, this dissertation explores the challenges related to using BN models in the context of VP case authoring and presentation, and experimentally compares a VP case based on a BN model to one created using a traditional narrative-branched VP system on multiple categories, including meeting learning objectives, accuracy in depicting the chosen topic, representing clinical variations, introducing/reinforcing relevant clinical skills, providing formative feedback, scenario realism, and learner engagement (Chapter 6).

Last, but not least, this work investigates the extent to which the the use of annotated BN models in VP cases facilitates modifying an existing VP case by allowing case authors to manipulate the underlying model in such a way that the VP case meets alternate learning objectives.

More specifically, my work strives to address the following questions:

1.3 RESEARCH QUESTIONS

- **Research Question 1:** To what extent is it possible to create a virtual patient case using a VP system based on a Bayesian network model that is comparable to a case created using a branched-narrative VP system? (Refer to Appendix C, Section C.1 for comparison categories.)
- **Research Question 2:** To what extent does the proposed system facilitate modi-

ifying an existing VP by allowing case authors to change the model's states in such a way that the VP case meets alternate learning objectives?

1.4 FINDINGS

1.4.1 Research Question 1 Findings Summary

Branched-narrative case presented in vpSim was rated significantly higher than a VP case based on a Bayesian network model on the following criteria:

- Reflecting the learning objectives
- Introducing / reinforcing clinical skills, attitudes, and behaviors relevant to sepsis treatment
- Providing formative feedback for choices / outcomes
- Being more effective in teaching the subject matter to novice clinical practitioners

The VP case based on a Bayesian network model was rated significantly higher than the branched-narrative version on:

- Representing clinical variations associated with sepsis
- Scenario realism
- Learner Engagement

Neither version of the evaluated VP case showed a significant difference in ratings on how accurately they depicted the chosen topic (both versions of the evaluated VP case presented a sepsis scenario).

1.4.2 Research Question 2 Findings Summary

While it is possible to modify an existing Bayesian network model-based VP case by manipulating the model's states and the associated rules, many other parts of the case need to be modified in order to address alternate learning objectives. Almost

unanimously case authoring expert subjects indicated that in order to meet alternate learning objectives, case narrative, patient demographic information, prior medical history, and comorbidities would also have to be updated.

1.5 CONTRIBUTIONS

This research provides practical and methodological contributions to the body of work in the areas of health sciences education, problem-based learning, and clinical simulation design and evaluation. More specifically, this work (1) defines criteria and guidelines for designing VP cases based on BN models, (2) identifies and describes shortcomings and challenges associated with different BN modeling approaches for different types of clinical scenarios, (3) and presents a framework for evaluating VP cases.

1. **Defines criteria and guidelines for designing VP cases based on BN models.**

Using empirical evidence from RQ1 and RQ2 studies, I identified and catalogued criteria, guidelines, and constraints for designing VP cases based on BN models. Even though this work's primary aim is to investigate the feasibility of authoring and presenting VP cases that leverage BN models learned from EMR data, the methods, criteria, and guidelines presented and explained in Chapter 7, Section 7.2 can be generalized to other intelligent tutoring systems (ITS) that present scenario-based cases. For example, methods described in this work could be applied to develop crisis management models derived from historical crime data for FBI training [19], or to control the behaviors of non-player characters (NPC) in serious games used for cyber security training [20].

2. **Identifies and describes shortcomings and challenges associated with different BN modeling approaches for various types of clinical scenarios.** This work helped identify a number of shortcomings and challenges associated with ap-

plying different types of BN models to various clinical scenarios. The challenges I experienced in creating BN models representative of complex treatment scenarios such as diabetic ketoacidosis (DKA) (Chapter 5, Section 5.4.2) indicated that – at least in its current iteration – a VP system based on BN models is more suitable to present clinical scenarios for acute or emergent clinical conditions.

With the help from the University of Pittsburgh School of Medicine and School of Pharmacy clinical faculty, I identified a list of clinical conditions / scenarios that are more suitable for implementation with the methods described in this work.

Moreover, shortcomings of the methods described in this work provided insight into approaches that could work better in future model-based VP systems, as well as in computational modeling endeavors associated with modeling complex time-dependent clinical treatment scenarios (Chapter 7.4, Section 7.4.1).

Last, but not least, these findings can inform and guide selection of modeling approaches when designing decision support systems (DSS). Machine learning approaches are often employed in the design of general clinical DSS [21], as well as in development of disease-state-specific computational models [22, 23]. It is possible that a practitioner attempting to design a model for a complex clinical condition such as DKA will encounter problems similar to the ones described in this work — it is my hope that my “lessons learned” will help others to avoid or to overcome the same difficulties.

3. **Presents a framework for evaluating VP cases.** In order to experimentally compare a VP case based on a BN model to one created using a traditional narrative-branched VP system (Chapter 6, Section 6.1.2), I developed a VP case evaluation framework based on (1) features and uses of high-fidelity medical simulations that lead to effective learning as proposed by Issenberg, et. al. [24], (2) Posel’s dissertation work on validating criterion-referenced guidelines for virtual patient case authoring [25], and (3) the Bateman et. al. exploratory study on virtual patient design [26]. The VP case evaluation questionnaire was validated as part of the RQ1

study ((Chapter 6, Section 6.1.3) and can be used for evaluating and rating VP cases, PBL problems, and clinical simulations for use in health sciences curricula. Moreover, this framework could be generalized to evaluation of educational materials outside of health sciences. For example, as case- and problem-based learning are widely used in educational programs everywhere from middle schools [27] to law schools [28], the framework described in this dissertation could be used by educational experts to evaluate the perceived quality of such learning experiences.

1.6 DISSERTATION OUTLINE

The remainder of this dissertation is structured as follows. Chapter 2 covers background information on virtual patients, case-based learning, and simulation in medical education. Chapter 3 provides background information on Bayesian network models, including Bayesian classifiers, constraint-based learning (causal discovery), and dynamic Bayesian network models. Chapter 4 describes ModelPatient’s system architecture, including underlying technologies, frameworks, programming languages, and user interfaces. Chapter 5 present VP cases, related data, and underlying BN models that were used to evaluate the feasibility of the system. Chapter 6 addresses the research questions, subject inclusion criteria, experimental designs, analysis, and findings. Finally, Chapter 7 covers limitations, discussion, and future work.

2.0 SIMULATION AND TUTORING SYSTEMS IN MEDICAL EDUCATION

2.1 SERIOUS GAMES IN MEDICAL EDUCATION

Simulation technology has been used for nearly a century to train pilots, soldiers, and astronauts. The military has been the primary investor in the development of simulation training technology throughout most of its history, but the gaming industry has recently stepped forward as the driving force toward the advancement of this technology [9]. The industry has shifted some of its focus away from entertainment, and “serious games” now represent a \$20 billion market within the \$100 billion per year digital gaming industry, and have become more significant within the global education and training market, a market estimated in the trillions of dollars [29].

A game-based approach to learning is being applied to many different areas of training and curriculum. Introduction of the Serious Games Initiative and the Games for Health Project has encouraged the development of serious games for training medical professionals [30].

The expression “serious games” describes games that implement advanced video graphics and computer technology for the purposes of developing games and simulations for education and training, rather than mere entertainment [31]. A simulation is defined as a representation of a real-world system in which simulation activity is flexible and variable according to rules and strategies particular to that system. Simulation activity can take on aspects of reality in which learning may transfer while maintaining a low cost of error, thus protecting the user from the severity of real-

world mistakes by containing the consequences within the simulation [32].

The potential of serious games as platforms for training has captured the attention of training experts and educators for three reasons [33]. First, instructional models have shifted toward a learner-centered model that emphasizes a more active learner role in which information is applied rather than recalled. A second reason underlying the interest of training professionals in serious games is the empirical evidence suggesting that games can be implemented as instructional tools for enhancing learning and understanding of complex subject-matter. The literature has begun to establish connections between instructional strategies, motivational processes, and learning outcomes to better guide research and development within this domain. The third reason supporting efforts in the development of serious games application is the intensity of engagement that can be invoked by games.

Changes in the delivery of healthcare engender major shifts in the methods of medical education [34]. Clinical skills training is directly influenced by the pressures of managed care and limited financial resources. Consequently, physicians in training have reduced educational time in managed care settings and have fewer opportunities to assess patients with a wide variety of diseases and physical findings. Such problems result in a reduction in the quality of healthcare providers' bedside skills, and this transfers to a decline in the ability to provide cost-effective, high-quality healthcare [35]. Simulation systems have been recognized as valuable tools that support the resolution of such problems influencing training and assessment within medical education [36, 37].

Serious games allow students to follow individualized learning paths, engage in experiences equivalent to those of their peers, and engage in practice without risk to the patient [38, 39]. These simulations offer some extent of compensation for the limited experience that students receive within managed care settings, and can expose physicians in training to rare but critical cases that are essential to their development. The student and instructor can discuss clinical decisions about the condition of a patient,

explore the psychosocial aspects of the case, and raise bioethical issues, without the need of having access to a real patient [40].

The Accreditation Council for Graduate Medical Education (ACGME)¹ has classified six domains of clinical medical competence: patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice [41]. Serious games should require users to demonstrate competence within these domains at four levels (Miller, 1990): knows (knowledge) — recall of facts, principles, and theories; knows how (competence) — ability to solve problems and describe procedures; shows how (performance) — demonstration of skills in a controlled setting; and does (action) — behavior in real practice.

Serious games have the ability to implement learner-specific conditions, provide standardized experiences for all examinees, and include outcome measures that yield reliable data [34]. This provides a platform for evaluating the first three levels of assessment. Leveraging games to engage the learner in deliberate practice can also lead to more successful outcomes at the fourth level of assessment [34].

Despite many advantages to using serious games, two major disadvantages are that these systems can be resource intensive and difficult to manage. Developing low-cost and low-time-investment systems that support easy authoring, management, and presentation of simulated patient cases without the assistance of technical specialists is a significant concern that must be addressed early on in the design process [10, 42].

2.2 SIMULATION TYPES

Health sciences simulation training falls into five broad categories:

- *Standardized patients* - actors hired by medical schools to play scripted patient scenarios to teach medical students patient interaction, clinical diagnostic, and

¹The Accreditation Council for Graduate Medical Education. <http://www.acgme.org/>

critical decision-making skills.

- *Mannequin simulators* - companies such as Laerdal² produce high-fidelity “human-looking” simulators that are capable of simulating human physiological processes and responses to external stimuli. An additional benefit of mannequin simulators is that they allow health sciences students to practice invasive procedures such as central line placement or emergency intubation without endangering a real patient.
- *Specialized procedure simulators / task trainers* - companies such as Surgical Science³ produce special-purpose devices that simulate invasive procedures by combining physical representations of surgical tools (e.g., laparoscopic surgery controls) with haptic feedback mechanism and some form of virtual reality (VR) representation of human organs.
- In-game scenarios - leveraging existing entertainment gaming platforms, such as Second Life⁴ and World Of Warcraft⁵ to create interactive clinical scenarios [43].
- *Electronic Virtual patients (VP)* - interactive screen-based “computer simulations of real-life clinical scenarios for the purpose of medical training, education, or assessment” [44]. Electronic Virtual Patient term is also synonymous with **Virtual Patient Case** or **Virtual Patient Simulation**.
 - **Linear VP cases** are classified into two categories: **linear-passive** and **linear-interactive**. Linear-passive cases “progress in one direction without options” - in other words, learners can move from one screen to the next without affecting or influencing the progression and the outcomes of a case. Linear-interactive cases allow learners some degree of exploration, but learners still cannot make choices or selection that would affect the outcome of a case [12].
 - **Branched-narrative VP cases** present learners with a challenge, offer a series of choices where each choice result in a consequence. “Use of this model

²Laerdal. <http://www.laerdal.com/>

³Surgical Science. <http://www.surgical-science.com/>

⁴Second Life. <http://secondlife.com/>

⁵World of Warcraft. <https://worldofwarcraft.com/en-us/>

allows learner input to directly affect the outcomes of the virtual patient where appropriate recommendations will improve the simulated patient's condition, whereas suboptimal recommendations will worsen it [45]."

2.3 VIRTUAL PATIENT SIMULATORS

Traditional virtual patients have characteristics that are reflective of different pedagogical frameworks [12]. For example, cases may be characterized as static or dynamic. Static cases train students in the process of asking relevant questions, ordering relevant tests, and developing a treatment plan based on a patient's medical condition. Cases characterized as dynamic progress over time and simulate the pressure of decision-making.

Since 1990, Karolinska Institute (KI) R&D group in e-Learning and simulation at the Department of Learning, Information, Management, and Ethics (LIME) have developed over 13 different patient case simulation systems [10]. One of their most advanced systems was the Interactive Simulated Patient (ISP). With this interactive web-based system, authors develop patient cases by combining real and fictive patient-history questions, answers, laboratory data, and physiological examination data [42]. A series of tests proved that ISP was a very realistic tool that is profoundly beneficial for clinical and preclinical learning. The Web-based Simulation of Patients (Web-SP) project followed with the aim to incorporate a flexible platform to be used in any health care curriculum that appropriately involves patient cases. It also sought to provide a solution to several user and development issues that arise in non-web-based applications such as the inconvenience of maintenance for the software [3]. The web-based virtual patient system, vpSim was developed in 2009 [38]. It was designed as an active way to enhance student learning of advanced therapeutic concepts. Virtual patient sessions are built using a branched-outcome decision-making model and a problem-based learning practicum, both effective applications of the aforementioned

criteria which fostered effective learning for students [46].

2.4 VIRTUAL PATIENT CASE AUTHORIZING

All existing VP systems require educators to manually create teaching cases. Regardless of whether the training system in question employs a linear or a branched model approach, educators must identify learning objectives, find a relevant medical case that fits those objectives, describe and present that case to the learners.

The advantages of this approach to case design is that educators can rely on their knowledge, experience, and research literature to tailor the learning experience to specific learning objectives. Furthermore, manual case creation gives authors full control of every detail of the teaching case. Even if there is no real patient on whom one can base a case, authors can make up details. For example, if an educator wanted to teach students about managing diabetes in cancer patients undergoing chemotherapy, they would not have to search through medical records to find a patient who provides a perfect fit for all learning objectives. The authors could either find relevant cases in the literature [47, 48], or create a list of symptoms based on the authors' own expertise in the subject matter. But then, will the case be realistic enough?

Another advantage of manual case creation (provided that the case is a text-based VP simulation) is that it is relatively easy for educators to modify in order to introduce new possible outcomes, decision points, symptoms, etc.

On the opposite side of the coin, many VP cases are linear-passive — they present information in a very linear fashion without options [12]. To add interaction and to teach critical decision-making skills many medical schools employ branched cases, such as the ones that can be designed using vpSim,⁶ an online VP simulator developed at the University of Pittsburgh School of Medicine. Such cases allow students to experience the consequences of each decision — the cases will take learners down

⁶vpSim: Online virtual player simulator. <http://vpsim.pitt.edu>

different (and often incorrect) paths based on every choice made within the case.

Figure 5 shows the starting nodes of the diabetic ketoacidosis (DKA) VP case designed in vpSim and described in Chapter 1. Each node shown in Figure 5 represents a decision point within a case, and connections between the nodes represent possible pathways through a case, from the introduction to an outcome.

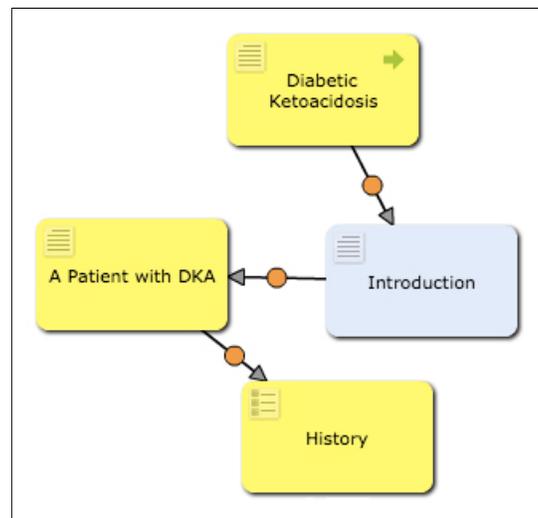


Figure 5: Initial nodes (choices) of the DKA case designed in vpSim

When designing a vpSim case, authors must carefully think about how different types of learning and assessment can be represented by different node types. Currently, vpSim supports 5 different node types (Figure 6):

1. *Narrative*: presents a story (narrative) about a patient, clinical condition, aspect of a case, or just-in-time learning
2. *Branching*: a decision point within a case. When learners make a choice within a branching node, they effectively choose a treatment pathway that might lead them to one of the possible outcomes.
3. *MCQ*: a multiple choice question with one correct answer. This type of node acts as a gate - a learner must answer a question correctly in order to continue. This

type of node is generally used for assessment or for just-in-time learning. The multiple choice question in this type of node is represented by HTML radio button controls; only one choice is possible.

4. *Inquiry*: a multiple choice question with multiple correct answers. This type of node is generally used as a checklist. For example, an inquiry node might be used when a learner needs to select appropriate questions to ask a patient during the initial encounter. The multiple choice question in this type of node is represented by HTML checkbox controls that allow for multiple selections.
5. *Short Answer*: a question that requires an open-ended text-based response from the learner. Answers to such questions are evaluated by the instructor upon VP case completion.

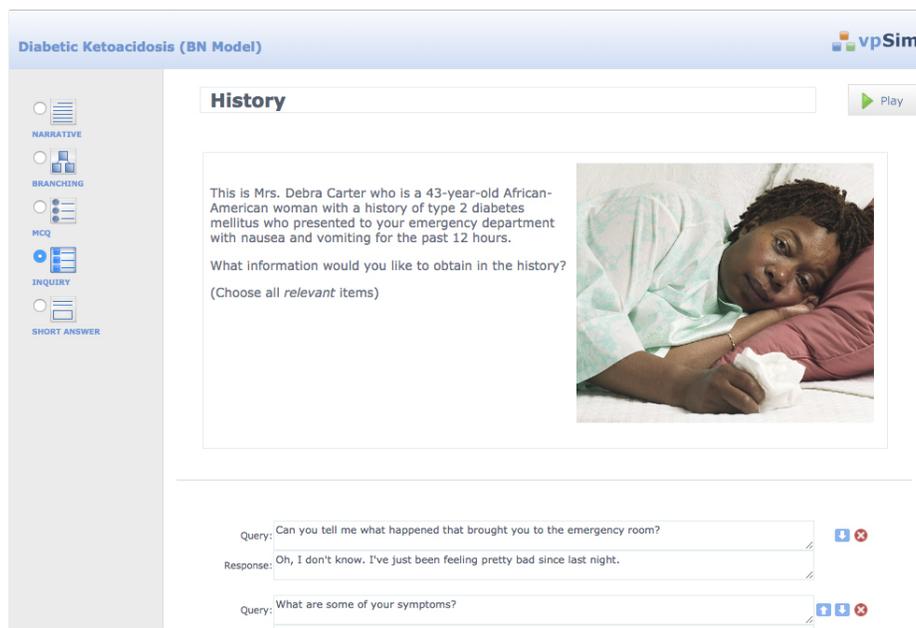


Figure 6: A “Patient History” node from the DKA case designed in vpSim in the edit mode (being edited by a case author)

Recently, *Clinical Tools, Inc*⁷ began developing branched cases similar to the one

⁷Clinical Tools, Inc. <http://www.clinicaltools.com/>

created in vpSim using an open-source storytelling software Twine.⁸ Twine offers an editing environment similar to that of vpSim that allows to extend stories with variables, conditional logic, images, CSS, and JavaScript.

One advantage that Twine has over vpSim and other proprietary VP simulator platforms is that authors can modify their VP cases' presentation with simple CSS changes. Twine's ability to export cases to HTML, CSS, and JavaScript package files allows reasonably easy deployment to mobile platform via PhoneGap⁹ or Xamarin.¹⁰

vpSim, on the other hand, offers more robust simulation logic programming via rules and variables (called counters) (Figure 7).

The screenshot shows a dialog box for creating a rule. It has a title "Choose rule type:". Underneath, there are two radio button options: "A previously visited node" (which is unselected) and "A counter/timer value" (which is selected). Below the radio buttons, there is an "IF" section with a dropdown menu set to "Real Time", followed by the word "is", another dropdown menu set to "more than ...", and a text input field containing the number "50". Below the "IF" section is a "THEN" section with a checked checkbox and a dropdown menu set to "Go to" with "Increase infusion" selected. At the bottom of the dialog, there is an unchecked checkbox labeled "Keep active after triggered (not recommended)". In the bottom right corner, there are two buttons: "Create New Rule" and "Cancel".

Figure 7: Example of rules (logic) authoring in vpSim

While branched cases have a number of advantages compared to their linear counterparts [16], the downside of such cases is that they are time and resource intensive to produce [12]. The amount of time a faculty member must devote to learning the VP system, building the cases, and deploying them for students to use is prohibitively large. This has led to VPs being underutilized in most facilities [13].

Another significant issue with manually created cases comes from the difficulty in managing case complexity. To manually create a complex teaching case that addresses

⁸Twine. <https://twinery.org/>

⁹PhoneGap. <http://phonegap.com/>

¹⁰Xamarin. <https://www.xamarin.com/>

the majority of common outcomes, common treatment and diagnostic mistakes, as well as offers just-in-time learning opportunities to students often results in a visual equivalent of “spaghetti code.” Figure 8 shows a small subset of the decision structure of a diabetic ketoacidosis virtual patient case [17] created in vpSim. This particular case has a total 92 nodes (screens accessible by learners) and 67 decision points (screens where learners have to make decisions). Maintaining such a case, adding or removing pathways or decision options, is a very time-consuming task indeed.

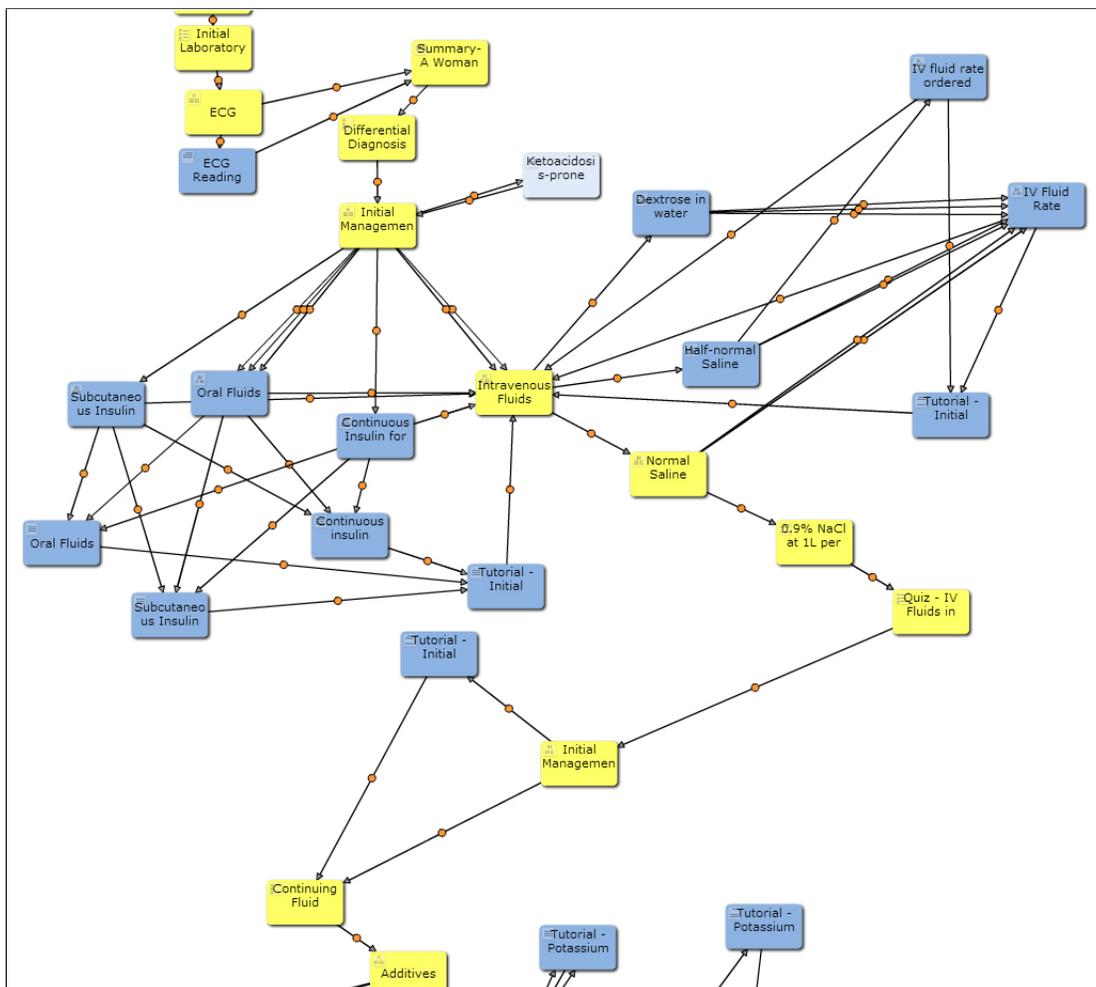


Figure 8: Subset of the decision structure of a diabetic ketoacidosis virtual patient case.

Last, but not least, a major issue with manually-created teaching cases is replaya-

bility [49]. Even with complex branching educational cases, most VPs are essentially state machines with a limited number of states and a limited number of choices to change these states. Once learners go through each possible branch of a VP case, they can no longer learn anything new by playing through the same simulation.

2.5 GENERATING VP CASES FROM CLINICAL KNOWLEDGE BASES

Early work on automating VP case creation concentrated on generation of cases from medical knowledge bases and decision support systems (DSS) such as Internist-1 [50], HELP [51], RECONSIDER [52], DXplain [53], and MYCIN/NEOMYCIN [54], to name a few.

The idea of developing a tutoring system from a clinical DSS knowledge base was first described in 1975 by Shortliffe and Davis as a question-answer program based on a semantic network of disease knowledge extracted from MYCIN [55]. In 1986 Clancey build on Shortliffe's and Davis' work with a tutoring system called GUIDON. GUIDON was based on a list of approximately 200 reasoning rules extracted from a generalized version of MYCIN [56, 57]. In their 1989 work, Chin and Cooper describe KBSimulator — a case-based tutorial system that combines Internist-1's rules with a BN model to calculate the probability of each possible diagnosis given the symptoms presented to the learner [58]. The most recent work by Carberry et. al. (1996) in the area of VP case generation described a system based on TraumAID knowledge base [59] and used a set of reasoning rules similar to those outlined by Clancey in 1986 [60].

This research builds on the idea presented by Chin and Cooper in their 1989 work and evaluates the feasibility of using BN models learned from EMR data rather than from specialized clinical knowledge bases to author and present VP cases.

3.0 BAYESIAN NETWORK MODELS

3.1 BAYES' THEOREM

The foundation behind all Bayesian statistics is Bayes' theorem, which, in turn, is based on the idea of conditional probabilities. For example, according to the National Institute of Health (NIH) cancer.gov website, based on 2009-2013 cases and deaths and adjusted for age, thyroid cancer rate in the United States was 13.9 per 100,000 men and women per year.¹ That means that an average US citizen has a 0.00014 probability of developing thyroid cancer in his or her lifetime. These are pretty good odds, unless... you are not an average citizen. What if you have been exposed to radiation? Then the probability of you developing a thyroid cancer $P(\text{thyroid cancer})$ becomes conditioned on radiation exposure $P(\text{thyroid cancer} \mid \text{radiation exposure})$. What if you add a few more variables into the mix, something along the lines of genetic predisposition to an aggressive tall cell variant (TCV) of papillary thyroid carcinoma (PTC) [61]? What if you have family history of thyroid cancer?

Bayes' theorem allows us to calculate the probability of an event based on conditions that might be related to this event. Given the aforementioned scenario, we can calculate the probability of an individual developing thyroid cancer in his or her lifetime if we know the related conditions, history of radiation exposure, probabilities of TCV, and relevant family history.

¹SEER Stat Fact Sheets: Thyroid Cancer. <http://seer.cancer.gov/statfacts/html/thyro.html>

Mathematically, Bayes' theorem is stated as the following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

where:

- P(A) and P(B) are the probabilities of events A and B occurring respectively
- P(A | B) is the conditional probability of an event A occurring given that the event B has occurred
- P(B | A) is the conditional probability of an event B occurring given that the event A has occurred
- P(B) \neq 0 - the probability of an event B cannot be 0 (in other words, an event B cannot be impossible)

Applying Bayes' theorem to the aforementioned thyroid cancer example, we can calculate the probability of developing thyroid cancer given one's exposure to radiation.

- TC: Thyroid cancer
- RE: Radiation exposure

$$P(TC|RE) = \frac{P(RE|TC)P(TC)}{P(RE)},$$

where:

- P(TC) is the probability of the hypothesis before we see the relevant data, called the **prior probability**
- P(TC | RE) is the probability of the hypothesis after we see the data
- P(RE | TC) is the probability of the data under the hypothesis
- P(RE) is the probability of the data under any hypothesis, called the **normalizing constant** [62]

3.2 BAYESIAN NETWORKS

A Bayesian network (BN) model is an acyclic directed graph that consists of nodes that represent domain variables and arcs that represent probabilistic (possibly causal) influences among connected nodes. The qualitative part of a Bayesian network model is the graph itself, showing a pattern of dependencies among variables in a data set. The quantitative part of encodes the joint probability distribution over these variables [63].

As Thomas Bayes showed in his famous billiard ball experiment, if we update our initial beliefs with objective new information, we get a new and improved belief [64]. In Bayesian networks, this updating of initial beliefs essentially translates into updating the probability of a hypothesis (e.g., thyroid cancer) given new evidence (e.g., radiation exposure).

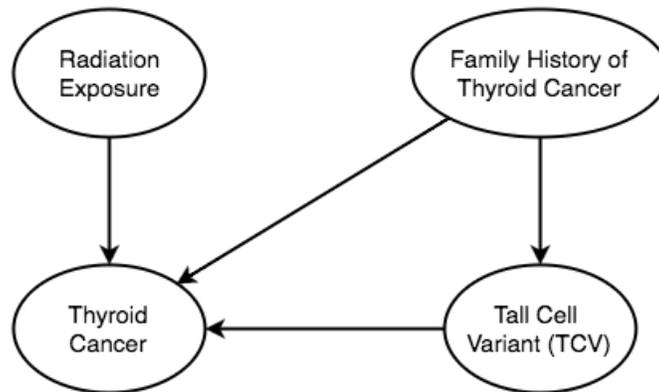


Figure 9: An example of a Bayesian network modeling the relationships among Thyroid Cancer, TCV, Radiation Exposure, and Family History of Thyroid Cancer.

States of nodes that do not have predecessors (prior knowledge nodes) are defined by prior probability distribution tables. Given the example shown in Figure 9, the prior probability distribution table for the “Family History of Thyroid Cancer” node would like like Table 1, where No/Yes states indicate absence or presence of family history of thyroid cancer with respective probabilities of 0.7 and 0.3.

Table 1: Prior probability distribution table for the “Family History of Thyroid Cancer” node

No	0.7
Yes	0.3

States of nodes that do have predecessors are defined by conditional probability distribution tables (CPT). The example in Table 2 shows conditional probability distributions for all possible combination of states in the “Tall Cell Variant (TCV)” node:

- $P(\text{TCV} \mid \text{Family History of Thyroid Cancer})$
- $P(\text{TCV} \mid \text{No Family History of Thyroid Cancer})$
- $P(\text{No TCV} \mid \text{Family History of Thyroid Cancer})$
- $P(\text{No TCV} \mid \text{No Family History of Thyroid Cancer})$

Table 2: Conditional probability distribution table for the “Tall Cell Variant (TCV)” node

Family History of Thyroid Cancer	No	Yes
No	0.7	0.4
Yes	0.3	0.6

Bayesian networks are widely used in decision support systems (DSS), including clinical DSS [65, 66], business analytics [67], and even forecasting the results of presidential elections [68].

3.3 BAYESIAN NETWORK MODELS

Given a data set containing observations (features) about a specific data point (class variable) whose category association is known, it is possible to identify a category to

which a new observation belongs. In machine learning, this process is called *classification*. Machine learning algorithms that perform the classification process based on models learned from training data are called *classifiers*.

For example, let us consider a data set containing data about patients diagnosed with lung cancer. Let us also assume that this data set contains four propositional variables:

- Exposure to asbestos (A): denotes whether or not a patient had been exposed to asbestos prior to the lung cancer diagnoses
- Yellow fingers (F): a denotes presence or absence of yellow fingers
- Smoking (S): indicates whether or not a patient is a smoker
- Lung cancer (C): denotes presence or absence of lung cancer.

Now let us assume that given the existing data we need to build a model that uses values from each instance of A, F, and S to predict the value (or the category) of C.

The following sections describe different classification algorithms that learn the structure and the parameters of a model from an existing (training) data set where the category of each instance of the class variable (in our example the class variable is C - lung cancer) is known. Then, leveraging the model learned from the training data set, the algorithms can attempt to predict (classify) an unknown instance, for example a new patient who has yellow fingers and is a heavy smoker, but has not yet been diagnosed with lung cancer.

3.3.1 Naive Bayes

Naive Bayes is a well-known Bayesian classifier that tends to outperform more sophisticated classifiers, especially in datasets where the variables are not strongly correlated [69]. Naive Bayes has several advantages over many other classifiers, including the PC algorithm. Because it assumes conditional independence among the features given the class in a data set, its structure is given a priori. Therefore, constructing a Bayesian network using the Naive Bayes classifier requires no structure learning

procedure. Furthermore, because of the assumed independence, the classification process is very efficient [70].

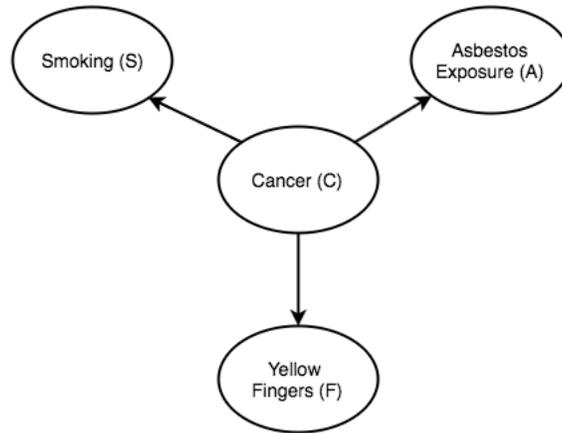


Figure 10: A Naive Bayes model for the example classification problem.

3.3.2 Augmented Naive Bayes

In real life we cannot assume that variables in a data set are independent. Augmented Naive Bayes classifiers include the Tree Augmented Naive Bayes (TAN), BN Augmented Naive Bayes (BAN), and General Bayesian Network (GBN) classification algorithms. What differentiates augmented classifiers from the simple Naive Bayes is that their classification algorithms add a tree structure to a Naive Bayes model by connecting the most dependent attributes with directed arcs [70]. This results in a more accurate modeling of the joint probability distributions and improves the predictive accuracy of the model.

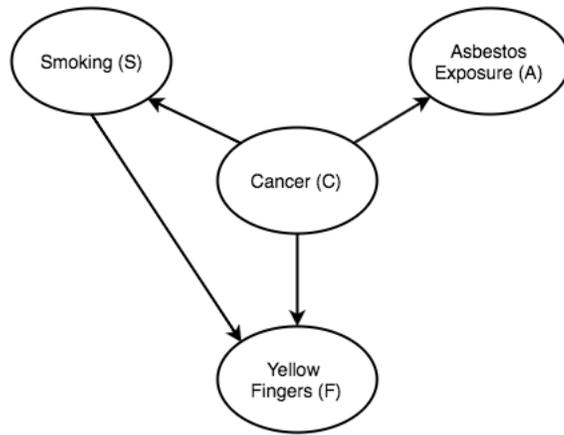


Figure 11: A TAN model for the augmented Naive Bayes example problem.

3.4 CONSTRAINT-BASED SEARCH

Causal discovery procedures focus on creating a model of the joint probability distribution over the variables in a dataset. A critical element in such a model is an explicit representation of independences among variables. The starting point is the intuitive idea that causal connections result in dependencies, whereas lack of causal connections shows as independence between variables in a data set generated by the system. Causal discovery procedures focus on testing for independence and inferring from the observed independencies the causal structure that has most likely generated the data. Independence and dependence can be conditional on other variables. For example, yellow fingers and cancer are correlated, but this correlation disappears when we condition on smoking (that is, when we look at smokers and non-smokers separately). Figure 12 shows the causal graph modeling this example.

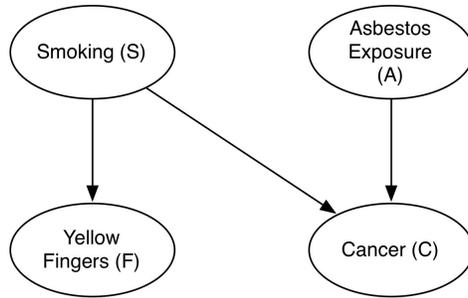


Figure 12: A causal graph modeling the relationships among Smoking (S), Yellow Fingers (F), Cancer (C), and Asbestos Exposure (A).

Conditioning can also lead to dependencies. This happens when we condition on common effects. For example, smoking and exposure to asbestos may be independent in the general population. However, they are very likely negatively correlated among cancer patients.

The PC algorithm [71] is a prominent member of the constraint-based search class of algorithms. Constraint-based search is a causal discovery procedure that tests each pair of variables in a data set for conditional independence and then produces a set of directed graphs representing the discovered dependencies. Conditional independence tests are classical statistical tests with a predefined significance level α . When causal discovery algorithms begin to analyze a data set, they assume that all variables are dependent on each other. For the example described above, the initial causal structure would look like the diagram in Figure 13.

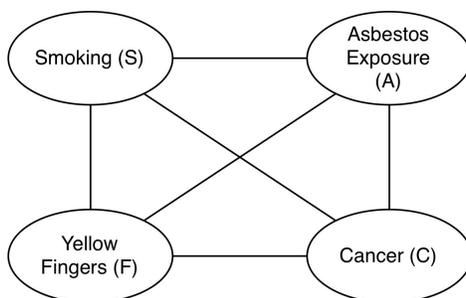


Figure 13: The initial structure assumed by the PC algorithm.

If two variables are found to be independent, the algorithm drops the connecting edge. For example, observation of independence between *Smoking (S)* and *Asbestos Exposure (A)* leads to removal of the edge between *S* and *A*. Observation of independence between *Asbestos Exposure (A)* and *Yellow Fingers (F)* allows for removal of the edge between *A* and *F*.

Next, the algorithm checks each variable in each pair to verify whether they are independent conditional on other variables in the data set. If the procedure finds such a dependency, it removes the corresponding edge from the resulting graph (Figure 14). For example, observing that *Yellow Fingers (F)* are independent of *Cancer (C)* given *Smoking (S)* allows us to remove the edge between *F* and *C*.

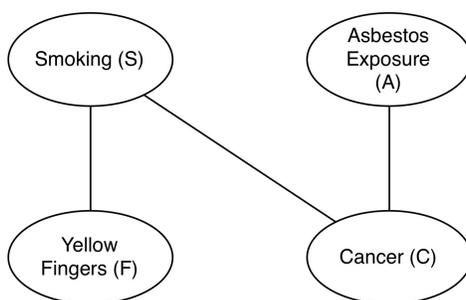


Figure 14: The causal structure reflecting conditional independencies.

Finally, the algorithm tries to orient each arc. For every triple of variables *A*, *B*, and

C, such that A is connected to B and B is connected to C, but A is not directly connected to C, the algorithm tests whether A and C are dependent conditionally on B. Spirtes et al. [72] prove that if A and C are dependent conditionally on B, the arcs from A and C must be directed toward B. In our example, *Smoking* and *Asbestos Exposure* are independent until we condition them on *Cancer*. The resulting graph will have three types of arcs: (1) directed arcs, where the algorithm found correct orientation, (2) undirected arcs, where the algorithm was unable to determine an orientation, (3) bidirectional arcs, where the algorithm was unable to determine an orientation, but inferred a possible hidden common cause between two variables, which is possible under certain circumstances.

The output of this causal discovery procedure is a class of acyclic directed graphs, where each node represents a variable in a data set and each directed edge (arrow) represents a direct dependence between two variables.

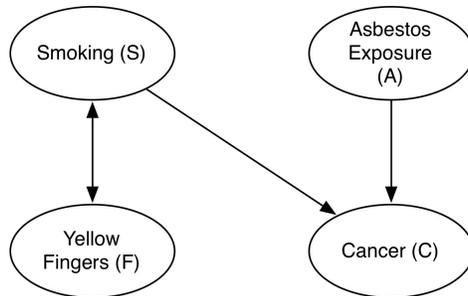


Figure 15: Example of final causal structure after the PC algorithm orients arcs.

Note that the arc between the variables *Smoking* and *Yellow Fingers* in (Figure 15) is bidirectional. However, we may add our experiential knowledge that smoking and yellow fingers have a temporal relationship. In order to develop yellow fingers, a person must smoke first. If we are willing to assume that smoking comes before yellow fingers and there is no hidden common cause of both, we can reorient the direction of the arc to demonstrate that *Smoking* is a likely underlying cause of *Yellow Fingers*.

Each test of independence in the algorithm is a classical statistical test performed at some level of significance α . A high value of α makes rejection of the independence hypothesis of two variables more likely. When the independence hypothesis cannot be rejected, the connection between two variables is removed. The result is that lower significance levels lead to sparser graphs (graphs with fewer connections), and higher significance levels lead to denser graphs. It is a good practice to repeat the search for a wide spectrum of α values. Graphs retrieved for high values of alpha are denser but lack of an arc is a robust finding: it was not possible to reject independence even though the value of α was high. Conversely, these arcs that remained at a very low level of α will serve as good indicators of possible causal relationships between connected variables: even though α was very low, independence was rejected.

Additional details of causal discovery procedures are available in Spirtes et al. [72].

3.5 DYNAMIC BAYESIAN NETWORK MODELS

Dynamic BN models extend standard BNs by allowing us to model time series or sequences of events [73, 74]. For example, let us consider a scenario where a diabetic ketoacidosis patient needs to be treated with hydration and insulin over a period of time. In a standard Bayesian network models, we could envision this scenario as the probability of some outcome (O) (hopefully an improvement in the patient's condition) given insulin I and hydration H . We can express this relationship as $P(O|I \cup H)$, or as a graphical model shown in Figure 16.

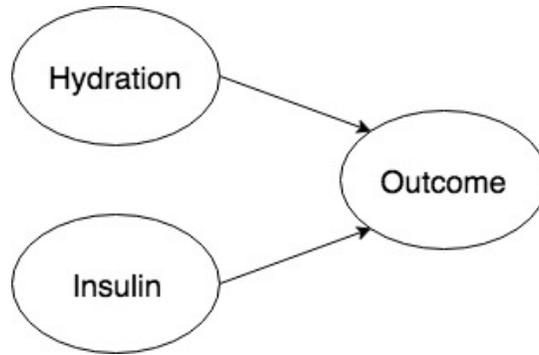


Figure 16: Example structure of a standard BN model

However, in a more likely scenario, a DKA patient would have to receive hydration and insulin multiple times over the course of the treatment and every administration of hydration and insulin will affect the outcome probabilities. To model such a scenario, we need to model the interactions between the two treatments and the outcome as multiple time slices. In other words, we want to be able to determine $P(O_1|I_1 \cup H_1)$ at time point 1, then $P(O_2|I_2 \cup H_2 \cup O_1)$ at time point 2, and so on. If we were to model this scenario as a 2-slice dynamic BN, the graphical model may look like Figure 17.

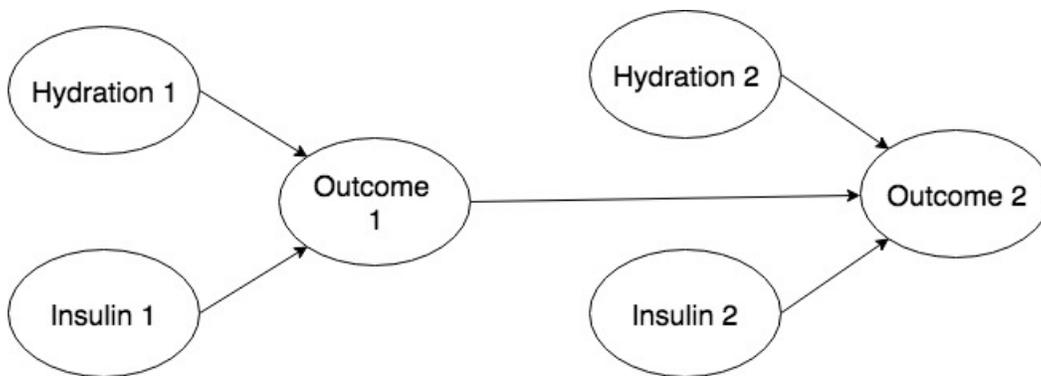


Figure 17: Example structure of a 2-slice dynamic BN model

Dynamic BN models are commonly used in clinical decision support systems where a treatment scenario has a temporal component (e.g. a treatment needs to be repeated

multiple times over a period of time). Examples of usage of such models include dynamic models of sepsis treatment [22], personalized medicine applications [75], and ICU organ failure decision support systems [76]. Non-clinical applications include speech processing [77], gene sequencing [78], face recognition [79], and sociology research [80].

3.6 PROPOSED USE OF BAYESIAN NETWORK MODELS IN VIRTUAL PATIENT CASES

Let us consider the following hypothetical VP simulation scenario. A 38-year-old male patient presents with symptoms of acid reflux and stomach ache. The patient’s medical record indicates past history of proton pump inhibitor (PPI) and Ibuprofen 800 prescriptions. Which medication would offer the best treatment option?

This example case is based on a simple BN model learned from a small data set of patients who presented with acid reflux symptoms. (Figure 18).

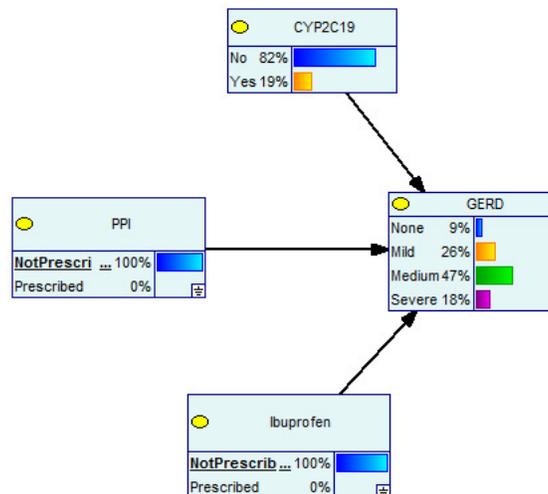


Figure 18: A model for patients with acid reflux learned from the data. Bars in the nodes CYP2C19 and GERD show marginal probability distributions for those nodes.

This model contains four nodes:

1. GERD: Severity of gastroesophageal reflux disease.
2. PPI: Whether or not a PPI medication (such as over-the-counter Omeprazole) was prescribed. PPI medications reduce the symptoms of acid reflux.
3. Ibuprofen: Whether or not Ibuprofen was prescribed.
4. CYP2C19: Whether a patient has a rare variant of the CYP2C19 gene that makes people rapid metabolizers of PPIs. In other words, rapid metabolizers do not benefit from PPI treatment for GERD.

If the learner chooses to prescribe Ibuprofen, the system uses SMILE API to update the state of “Ibuprofen” node to “Prescribed.” SMILE, in turn, updates the probability distributions over the BN model (probability of GERD symptoms severity given Ibuprofen). The learner’s action changes the outcome of the case – now the VP has a higher probability of having severe GERD symptoms (Figure 19).

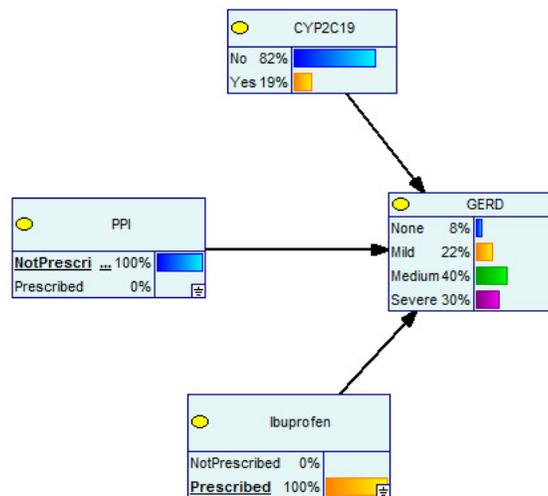


Figure 19: Probability distribution over GERD symptoms severity given Ibuprofen.

If the learner chooses to prescribe a PPI medication, the VP’s symptoms will improve (Figure 20).

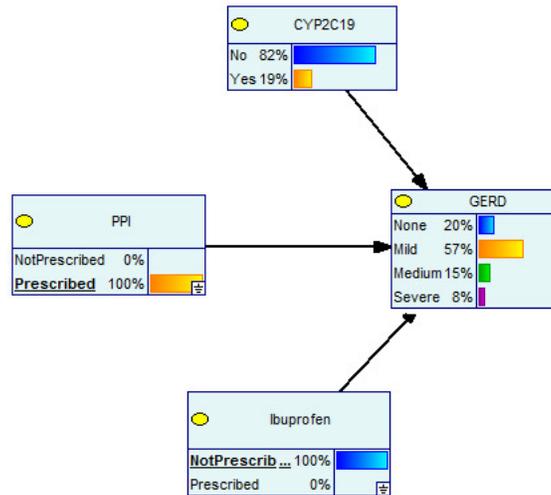


Figure 20: Probability distribution over GERD symptoms severity given PPI.

However, if the learner fails to take into account presence of the CYP2C19 gene variant, the PPI prescription will have virtually no effect, because rapid metabolizers do not benefit from PPI treatment for GERD (Figure 21).

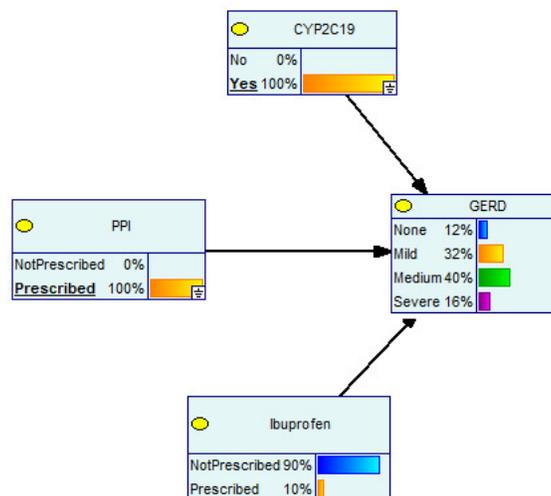


Figure 21: Probability distribution over GERD symptoms severity given PPI and CYP2C19.

Each decision that the learner makes while interacting with a VP results in immediate update of the probability distributions over the model, reflecting the VP's state of health in real time and providing immediate feedback to the learner's actions.

4.0 MODELPATIENT SYSTEM IMPLEMENTATION AND USER INTERFACE DETAILS

I implemented some of the ModelPatient’s features as part of the preliminary work described in *Designing the Model Patient: Data-Driven Virtual Patients in Medical Education* [81]. In validating ModelPatient with health sciences educators and medical simulation experts, I have discovered a number of shortcomings, both from educational and system engineering and design perspectives.

The design decisions for the current version of ModelPatient were largely based on outcomes of a series of iterative semi-structured interviews with eight VP case authoring experts¹, the MedBiquitous Virtual Patient Specifications [82], and an extensive literature review of publications on the subjects of virtual patient design, serious game design, intelligent learning system design, and best practices in instructional design of clinical simulations. Furthermore, I carefully reviewed several existing VP cases and simulation scenarios designed using vpSim [46, 45], DecisionSim [83, 84, 85], OpenLabyrinth [86, 87, 88], and Laerdal SimMan [89, 90, 91] case authoring tools and compiled a list of features that were common to all four systems.

More specifically, I reviewed the authoring side of the following VP cases:

¹VP case authoring expertise criteria are defined in Chapter 6, Section 6.2.1

Table 3: VP cases reviewed for authoring features design

System	Case/Scenario Title	Developed At
vpSim	Managing Sepsis	University of Pittsburgh School of Pharmacy
DecisionSim	Management of Diabetic Ketoacidosis (DKA)	American College of Physicians (ACP)
SimMan	Lost a/w in PACU	WISER Simulation Cen- ter, University of Pitts- burgh
SimMan	Laparoscopy	WISER Simulation Cen- ter, University of Pitts- burgh
OpenLabyrinth	OpenLabyrinth v3 cases — Robin: A Dermatology Case	University of Calgary

In addition to the features presented in Table 4, every interviewed case author requested the ability to add elements of randomness and unpredictability to VP cases. The primary rationale behind this feature request was guided by the premise that in real life sometimes patients do not respond to treatment protocols in a predictable and consistent fashion and that adding unpredictable outcomes will increase the variability in outcomes and presumably make the case more interesting / challenging to advanced learners. Moreover, a number of studies and pedagogical frameworks show benefit of learning to deal with unpredictable situations early on in ones clinical education career [92, 93, 94]. A system feature that allows for a degree of uncertainty and unpredictability of outcomes can be an important teaching tool in the virtual patient toolbox (Subsection 4.6.1).

Table 4: VP case features common to vpSim, DecisionSim, OpenLabyrinth, and Laerdal SimMan

Feature	vpSim	Decision Sim	Open Labyrinth	Laerdal SimMan
Introductory narra- tive and instructions	X	X	X	
Feedback (text)	X	X	X	
Feedback (changing vitals)				X
Distractors (incor- rect decision options or incorrect treat- ment pathways)	X	X	X	X
Case metrics (coun- ters)	X	X	X	X
Rules	X	X	X	X

Even though all features listed in Table 4 will be discussed and explained in detail further in this chapter, it is important to understand the motivation behind some of the design decisions. The descriptions of the system architecture, user interfaces, and interactions with BN models provided in this chapter are designed to help the reader gain a better insight into VP case design and study design decisions described in subsequent chapters (Chapters 5 and 6), as well as into possible future directions (Chapter 7).

Introductory Narrative and Case Instructions

While not every VP case reviewed as part of ModelPatient’s system design contained elements of narrative, most of them started by providing learners with instructions and by giving learners some background information on the patient. The initial designs of ModelPatient’s case authoring user interface (UI) was based entirely on the premise that the learner will only interact with the data and metadata contained in the case’s underlying BN model [81]. Evaluating existing VP systems and the related body of literature provided invaluable insight into necessary elements of VP case design and VP case authoring. Providing guidance and instructions to the learner prior to allowing them to delve into a virtual patient case is a commonly used instructional design technique, as well as a recommended feature of designing case-based scenarios [95, 96, 15, 97]. Moreover, even though patient history information can be presented in ModelPatient’s EMR-like interface, most case authors interviewed during the requirements collection phase requested the ability to present VP history as part of a cases’ instructions in a linear narrative fashion at the start of the case (Subsection 4.6.3).

Feedback Feedback is an integral part of every VP case and every VP system reviewed as part of this study. Moreover, feedback mechanisms are defined in the MedBiquitous Virtual Patient Specifications [82] and are strongly advocated for in many VP-related studies and publications [98, 97, 99, 100]. ModelPatient combines text-based feedback mechanisms available in vpSim, DecisionSim, and OpenLabyrinth

with a feedback mechanism based on changing vital signs and laboratory test results (Laerdal SimMan). Text-based feedback in ModelPatient can be deployed via custom rules (Subsection 4.6.6), while feedback based on changing vital signs and laboratory values is automatically triggered every time probability distributions change in the underlying BN model (Subsections 4.3 and 4.6.2).

Distractors

Every vpCase reviewed as part of the initial design study presented learners with distractors — incorrect decision choices or incorrect treatment options deliberately added to the case in order to increase complexity. In their cross-disciplinary review of literature on developing teaching cases Kim, et. al. state that “[teaching] cases should include both pertinent information (positive and negative) and unnecessary information to simulate the real challenge of data collection and synthesis [101].” Moreover, Posel, et. al, Grunwald, et. al., and Kim, et. al. advocate for the use of distractors in clinical scenarios to increase case complexity in order to “help learners experience complex and unpredictable decision-making [97, 102, 101].” Details describing how ModelPatient implements distractors are presented in Subsections 4.6.4 and 4.6.6.

Metrics

The idea of tracking author-defined learner performance metrics in VP cases originated in entertainment games [103, 104, 105]. For example, first-person shooter (FPS) games routinely keep track of and display real-time values for each player hit points, lives, weapons, and ammunition. The use of these metrics allows players (in our case learners) to self-evaluate their performance and progress. For educators, these metrics allow for quantitative evaluation of learners’ performance, as well as for evaluation of the quality of the VP case itself. Medbiquitous VP standard [44, 106] calls for use of counters (metrics) and rules for all VP systems, and most branched-narrative VP systems such as OpenLabyrinth, vpSim, and DecisionSim implement them to some degree. vpSim and DecisionSim case authoring systems provide VP authors with pre-defined metrics (score, number of steps, real time, virtual time, health), as well as with

the ability to define custom metrics. Based on feedback collected from case authoring experts during initial user interviews, ModelPatient was designed to allow authors to define custom metrics without providing predefined ones (Subsection 4.6.5).

Rules

Rule-based systems have been extensively used in entertainment games from Pac-Man [107] to Super Mario Brothers [108]) to control behaviors of non-player characters (NPCs), as well as in clinical decision support systems [109, 51, 52, 53, 54, 110], and in existing virtual patient systems [111]. User studies with the original version of ModelPatient [81] showed that a BN model alone does not provide enough flexibility, complexity, and variability to a VP case. All interviewed case authoring experts unanimously requested the ability to augment a BN model-based case with custom-defined rules. Details of rules implementation, as well as specific examples and use cases are discussed in Subsection 4.6.6.

4.1 SYSTEM ARCHITECTURE SUMMARY

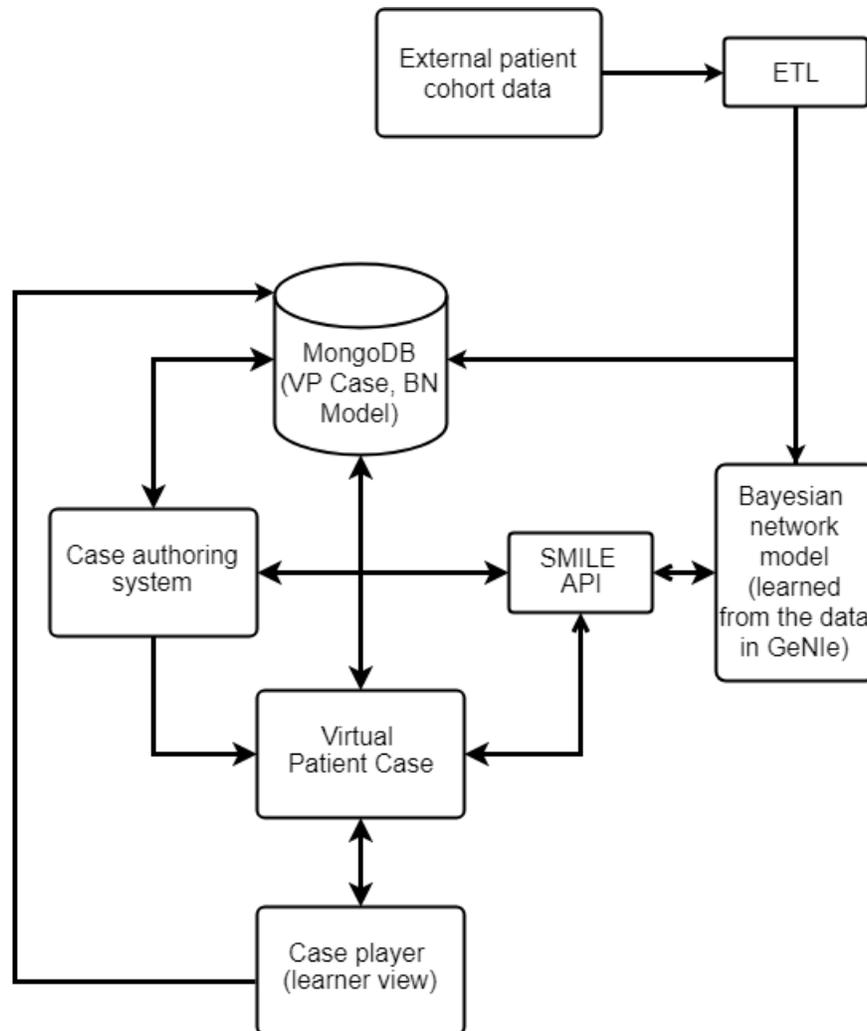


Figure 22: System architecture diagram

The core of the system is implemented as an ASP.NET MVC web application, using the C# programming language. The system also integrates Knockout.js JavaScript library² to facilitate client-side data binding. BN model visualization and annotation tool is implemented using D3js document visualization library.³ All BN model, VP case, user, and event log data is stored in a MongoDB database.⁴

²Knockout.js Javascript Library: <http://knockoutjs.com/>

³D3js: <https://d3js.org/>

⁴MongoDB: <https://www.mongodb.com>

4.2 LEARNING BN MODELS FROM THE DATA

For the current version of the system, the models are learned using GeNIe, a Bayesian modeling environment developed at the University of Pittsburgh School of Information Sciences Decision Systems Laboratory.⁵ Figure ?? shows a model learned from a data set of 2,300 IBD patients [112] using the PC algorithm [71]. This model was originally created for a study on identifying high-cost IBD patients [113] and used to validate the original version of ModelPatient [81].

ModelPatient is designed to work with such models to allow learners to manipulate the model's states by making choices using a user interface (UI) that had been modeled after an Electronic Medical Record (EMR) system. Because each state within such models is represented by conditional probability tables (CPT), if a learner changes one state, the model will recalculate the simulated clinical outcome probabilities of all descendant nodes, thus changing outcomes of the VP.

For example, consider a small "submodel" of the larger IBD treatment cost model. Figure 24 shows two nodes from the overall model from Figure ?. The relationship between these two nodes from the larger model shows that IBD patients who do not smoke have lower probability of managing pain via prescribed opiates.

If we change the state of the "Smoking Status" node, the probability of opiate use goes up significantly (Figure 25). Now, let us imagine this scenario as part of a VP simulator system. A learner is presented with a teaching case. As part of collecting patient history, a learner might identify that the patient is a heavy smoker who also has an IBD. By updating patient history options in the system, the learner essentially changes states of corresponding nodes in the model. The model recalculates probabilities of states within connected nodes.

⁵GeNIe, a Bayesian modeling environment: <http://www.bayesfusion.com/>



Figure 24: Relationship between a patient’s smoking status and use of opiates with default state values.



Figure 25: Relationship between a patient’s smoking status and use of opiates after change in smoking status from the default to “Never Smoked.”



Figure 26: Relationship between a patient’s smoking status and use of opiates after change in smoking status from the default to “Current Smoker.”

Because the model was originally learned from real patient data it will respond to learner’s choices in a manner similar to that of a real patient.

4.3 INTERACTING WITH BN MODELS

ModelPatient uses SMILE (Structural Modeling, Inference, and Learning Engine) [114], a library of C++ classes implementing graphical decision-theoretic methods, such as

Bayesian networks and influence diagrams to interact with author-created BN models. Both the BN model annotation tool (Section 4.6, Subsection 4.6.2) and VP case player (Section 4.6, Subsection 4.6.7) rely on SMILE API to update BN models' states and conditional probabilities.

4.4 NON-DESTRUCTIVE MODEL UPDATING

GeNIe stores its BN models in XML files with .xdsl extensions. SMILE API allows other applications to modify the models by adding/chaging nodes, states, and edges.⁶ These changes can happen either “in-memory” and only be available while the program is running, or they can be saved back into the .xdsl model file.

One of the key ideas behind ModelPatient is to use BN models as reusable templates. A single GeNIe model learned from the data can potentially serve as a starting templates for multiple VP cases. Once authors select a model for a VP case, set the initial states and annotate it accordingly, the annotated model could also be used as a template for different VP case scenarios. Finally, learners can modify the model's states when they select diagnostic and treatment options for the VP using the case player EMR interface. Managing these three “versions” of for any given BN model became a major design and software architecture challenge.

In order to preserve the initial model unchanged, to keep track and to selectively apply any author- and learner-driven changes, I borrowed the idea of non-destructive imaging (NDI) editing implemented by Adobe Systems Incorporated in their Lightroom and Photoshop products [4]. The concept of NDI originated from film photography where the original image was represented by a film negative and variations in the actual print were achieved by purposeful variations in the print process. The advent of digital photography brought new editing challenges – while a photographer may improve a photograph by adding contrast or altering color hues, the changes essen-

⁶SMILE API and XDSL XML schema definitions: <https://www.bayesfusion.com/smile-engine>

tially destroy the original information captured in the image file. Once an image file is saved and the photo editing software is closed, the file is essentially altered forever without any way to bring back the original digital information. To address this issue, Lightroom and Photoshop systems implement NDI - any changes made to the original raw image are not applied to the image itself but rather stored in an XML metadata file and are applied to the image in-memory for preview purposes or when exported for printing (Figure 27).

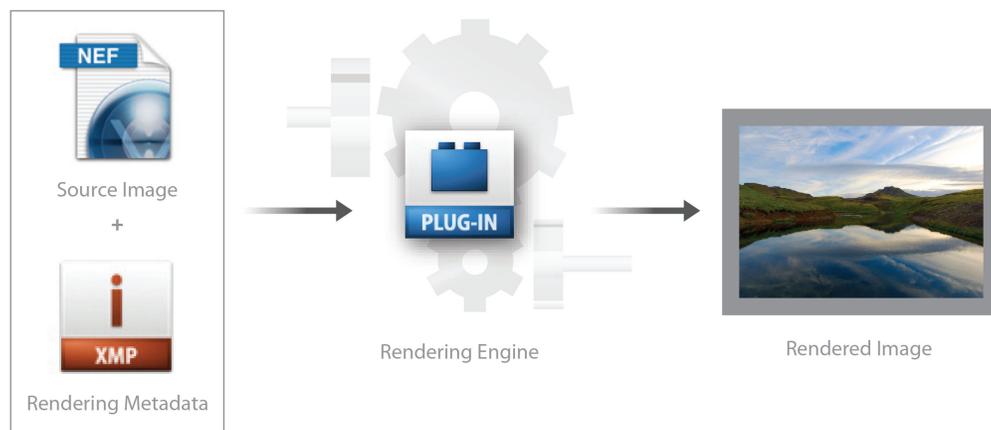


Figure 27: Adobe Lightroom “rendering engine in its simplest form: The source image and your processing settings (in the form of rendering metadata) are processed in the rendering software and a finished image is created.” [4]

I designed the BN model editing logic in a similar fashion. Once associated with a VP case, each model has two associated MongoDB documents - one document contains data about initial model states and metadata set by a VP case author. The second document (one document per learner who attempts the corresponding VP case) is identical in structure to the author BN model metadata document and contains data about states selected by each learner or triggered by a rule while a learner was attempting to complete the VP case. The author-created metadata is applied to a BN model at runtime when the author opens the model for editing. The learner-created metadata is applied to a BN model at runtime when the learner is interacting with the

VP case player / EMR view.

For VP case authors, this approach allows reuse of a single BN model for multiple cases and creates an “undo” buffer where authors can always revert to the original model or deselect a node state in case of a mistake.

For learners, this approach may improve case re-playability as learners can go through the same case multiple times, make different choices, and presumably observe different results. Furthermore, metadata documents recorded in MongoDB can be used to track learners’ choices for feedback and/or debriefing.

4.5 DATABASE

The original version of this system used MySQL Community Edition database management system (DBMS) to store data for the BN model, VP cases, simulation events, user information, and usage logs. However, as I began working with larger models, I came across a number of performance issues, most of them stemming from the number of SQL joins required to map the data from multiple relations to a single complex C# object that represents a BN model. Introduction of NHibernate,⁷ an open source object-relational mapper (ORM) for the .NET framework, helped solve some of the performance issues related to rendering graphical representations of a model, but the system still ground to a halt when users tried to change states on nodes with 6 or more children.

The current version of ModelPatient implements MongoDB - an open source database that, unlike table- and row-based schemas of relational databases, uses a document-oriented data model. The data model is built on an architecture of collections and documents, where documents comprise of key-value pair sets, and collections contains sets of documents [115]. This document-based architecture allows for one-to-one mapping between MongoDB documents and application objects [116], eliminates the need

⁷NHibernate: <http://nhibernate.info/>

for ORM frameworks such as NHibernate, and reduces run-time overhead [117]. Furthermore, MongoDB’s use of a Binary Javascript Object Notation (BSON) document storage and data interchange format allows direct asynchronous passing of objects between the client and the server using jQuery AJAX without having to implement custom object serialization / de-serialization logic.

4.6 USER INTERFACE

4.6.1 Case Authoring

Authors can begin designing a case by providing basic VP case background information, selecting a desired model from an existing list of available models, and specifying learning objectives (Figure 28).

One notable feature of this screen is the ability to specify how ModelPatient will respond to learners’ choices. By default, for each learner’s decision, the system will respond by selecting the most likely outcomes (states) from the model’s conditional probability tables. Case authors have the option to override this default behavior and have the system randomly select an outcome from available model states. It is worth noting that the outcome selection is not completely random, but is rather based on distributions of “states” in the original dataset.

Consider a VP case where an outcome of a learner’s choice has three possible states — (1) systolic blood pressure drops, (2) systolic blood pressure stays unchanged, and (3) systolic blood pressure increases. If in the original data 70% of patients responded to treatment with a decrease in systolic blood pressure, the state of “blood pressure drops” has a higher likelihood of being selected than the other two states.

Case Details

Case title:

Description:

In this simulation, you will be presented with a patient case followed by recommendations that need to be made. Choosing the most correct recommendation from a list of possibilities will not only be more beneficial to your virtual patient, but will also help your score.

 Your mission, therefore, is to keep our virtual patient alive and well as...

Name of the attending/supervisor:

Select patient treatment model:

Learner choices result in:

- The most statistically likely outcome
- Outcome randomly selected from the data

Learning Objectives

Learning Objectives:

Evaluate patient's risk for septic shock

Understand steps and guidelines for emergent sepsis treatment

Select appropriate empirical drug therapy

Figure 28: “Create / Edit VP case” screen.

4.6.2 BN Model Annotation

The next step requires annotating the BN model selected on the previous screen. The initial visualization of the BN model is based on GeNIe user interface, where nodes are represented by rectangles and edges are represented by directed arcs (arrows). Each node’s rectangle contains a list of possible states and their corresponding probabilities as learned from the data by the PC algorithm. Authors can specify initial conditions for the VP (patient history) by clicking on states of interest in relevant nodes (Figure 29).

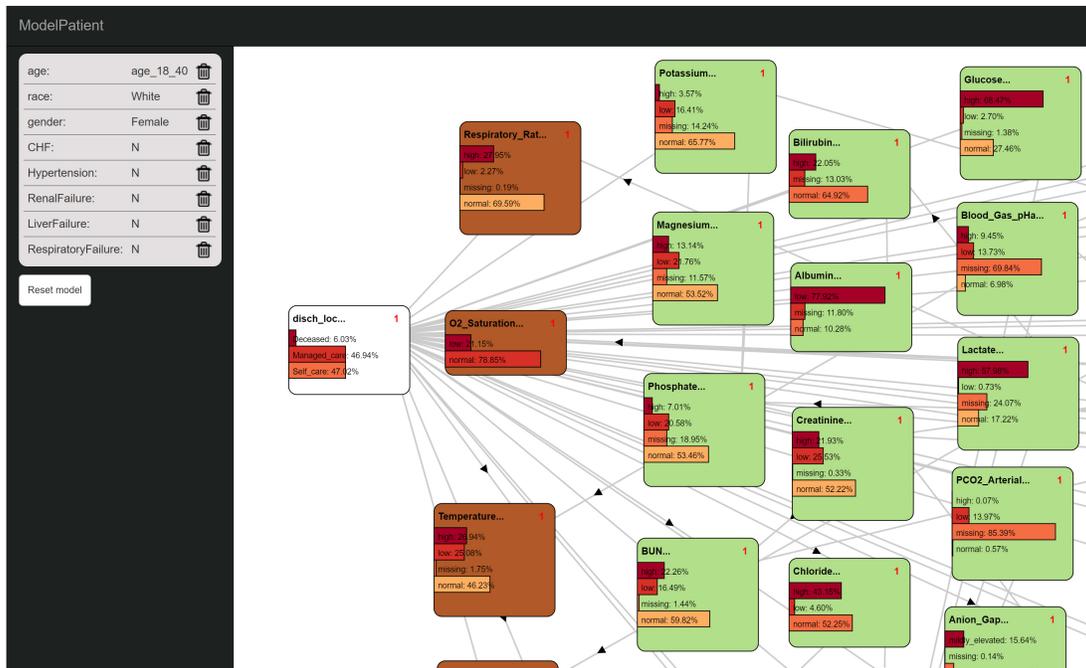


Figure 29: BN model editing and annotating tool.

One of the cases described in Chapter 5 Section 5.3 is a VP case designed to teach pharmacy students about managing sepsis. The author might set the following initial parameters:

- Sex: female
- Race: white
- Age: 18-40

The author would set the states in sex, race, and age nodes by clicking on the state of interest. ModelPatient uses SMILE API (Section 4.3) to update conditional probability tables in the BN model based on authors initial selections. States (values) for the nodes that were set as initial parameters appear in the VP player (learners' EMR view) under a section that corresponds to the node's category. For example, if the author set a state in a "Renal Failure" node as an initial parameter, and the "Renal Failure" node is in the "Diagnoses" category, then the value will appear as prior history in the "Diagnoses" tab of the VP player.

Authors can double-click on a node to add / edit the node's metadata. In the node details editor (Figure 30), authors can add or change the following metadata:

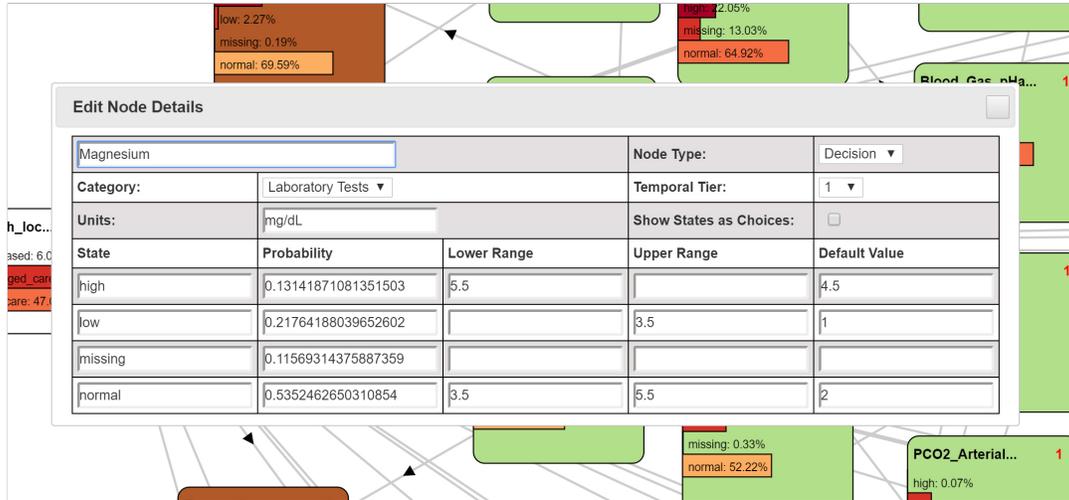


Figure 30: State and annotation editor for a single BN node.

- Rename node: change the node's label to something more suitable for a specific case presentation
- Change node category: by default, all nodes are imported as "uncategorized." Authors can classify each node into one of the categories listed below. Each category corresponds to a section in learners' EMR user interface (Subsection 4.6.7). During the initial requirements collection case authors identified these categories as a minimum necessary subset of the tabs available in Epic EMR. Moreover, most of these categories are defined in the Virtual Patient Data (VPD) component of the MedBiquitous Virtual Patient (MVP) standard [44].
 - Allergy: the node represents a particular allergy and the states represent severity levels or frequencies of occurrence
 - Demographic: the node represents some demographic information about the VP: gender, age, income, etc.
 - Diagnosis: the node represents a specific diagnosis (e.g., hypertension)

- Encounter: the node represents an instance of contact between the patient and a physician. For example, the node may represent phone calls between the patient and a clinical practice and the states represent frequencies of phone calls.
 - Laboratory Test: the node represents laboratory test results
 - Medication: the node represents medication prescriptions or drug requests and the states represent dosages, frequencies, or durations
 - Note: the node represents a note about the patient. This type of node is usually selected as an “initial state,” or “patient history” node and its selected state appears as a text note in the “Notes” section of the learners’ EMR user interface.
 - Procedure: the node represents surgical or diagnostic procedures.
 - Radiology: the node represents results of radiological tests (e.g., X-ray or CT scan)
 - Social History: the node represents social history variables (e.g., smoking history)
 - Vitals: the node represents vital signs (e.g., blood pressure, heart rate)
 - Uncategorized: default state of a node. Uncategorized nodes will not appear in the learners’ EMR view.
- Node type: ModelPatient’s node types are partially based on activity node types defined in the MedBiquitous Virtual Patient Specifications [106, 82]. More specifically, ModelPatient’s nodes fall into one of the three types described below:
 - Decision: this node’s states will appear as choices / decisions in the VP player. For example, if a “Lab” node is set as a decision node, learners will be able to “order” that lab.
 - Initial: this node becomes part of the prior patient history that ModelPatient displays for the learner at the beginning of a VP case.
 - Outcome: nodes designated as “outcomes” are displayed as responses to learner’s

decision. For example, if a learner chooses to prescribe “Normal saline,” the state of “Normal saline” node will change to “yes” and all nodes designated as “outcomes” that affected by this change (nodes where states’ probabilities change) will be shown to the learner as lab results, vitals, or clinical notes.

- Temporal Tier: allows authors to specify the order in which information is presented to the learner, as well as the order in which the model(s) respond to learners’ choices. For example, in a sepsis case an author may not wish for the learner to prescribe hydration before prescribing a vasopressor. In that case, the “Hydration” node in the BN model would be set to temporal tier one, and the “Vasopressor” node would be set to temporal tier two.
- Change / add metadata for each state:
 - State name: Name / label of a state
 - Lower and upper range boundaries: since states are represented by discrete categories rather than by continuous values, authors can specify a numeric range for each category. This is useful in cases where states represent laboratory values. For example, in a node representing systolic blood pressure, the author can specify four numeric ranges to signify specific health conditions: (1) less than 120: normal blood pressure, (2) 120-139: prehypertension, (3) 140-159: stage 1 hypertension, (4) 160 or higher: stage 2 hypertension
 - Default value: if this value is specified, it will appear in the learners’ EMR view if the corresponding state is selected or triggered by a rule. For example, instead of showing a patient’s age as “age_below_30” as seen in the corresponding state’s label (Figure 30), authors may chose to associate that particular state with 30 years old for the purposes of presenting a VP case.
 - Feedback: if the state is selected by a learner or triggered by a rule, the feedback text will appear in the “Notes” section of the EMR as a notification.
- Set node status as “Patient History.” Information contained in nodes marked as “Patient History” will be visible to learners at the start of a VP case. By default, all

demographic node are marked as “Patient History.”

4.6.3 Case Introduction

Case introduction interface allows authors to develop narrative / background information about the VP case or the patient (Figure 31), provide learners with instructions, or administer a multiple-choice pretest.

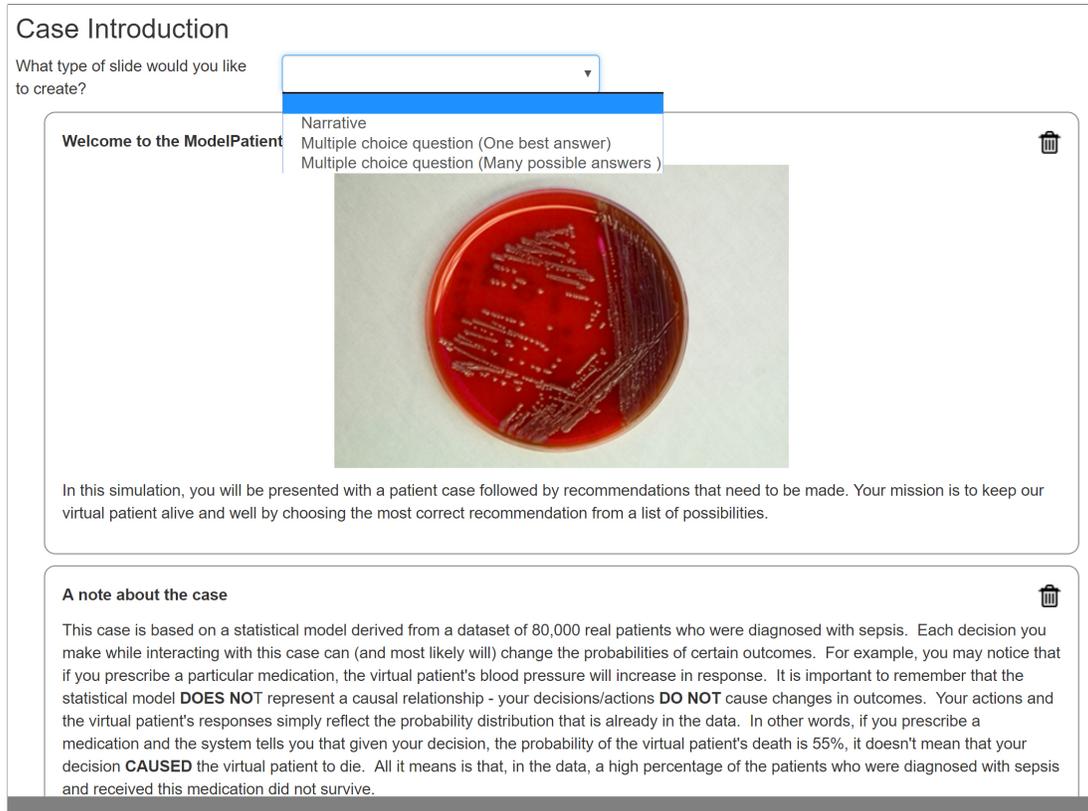


Figure 31: Virtual patient case introduction.

4.6.4 Additional Case Content

Authors also have the option to augment a VP case with additional content. The content can be presented to learners as patient history data, or be available as choices along with states from the underlying BN model. These additional choices can either

act as distractors that increase complexity of the case [97], or they can be configured to trigger case rules along with custom metrics and changes to BN node states.

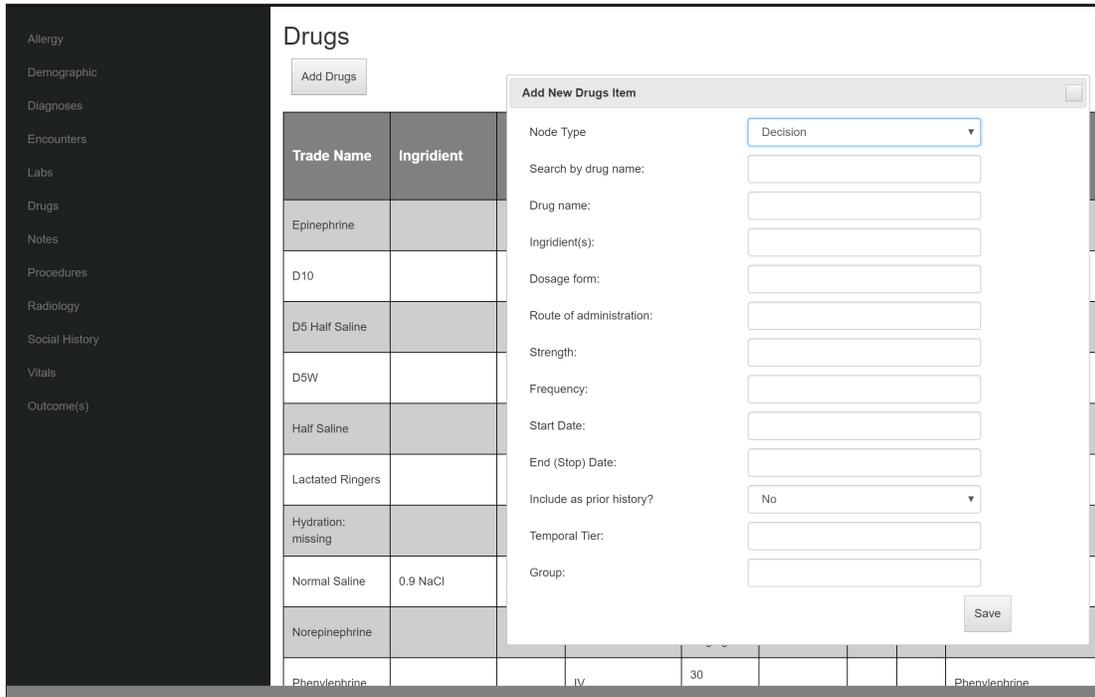


Figure 32: VP Case Player Content Editor.

4.6.5 Case Metrics

ModelPatient allows VP case authors to define any combination of custom variables that will be tracked throughout the VP case and will change based on learners' actions and on author-defined rules. Figure 33 shows a bare-bones user interface that allows authors to specify the following information for each metric:

- Metric name: desired name / label for a variable
- Lower range and upper range boundaries: the boundaries beyond which the value of the variable cannot change (for example, an author may choose to define a "Score" variable whose value must be between 0 and 100)

- Initial Value: initial / starting value for the variable - the variable will be initialized to this value at the start of the VP case
- Step: if the variable is auto-incremented (or auto-decremented), this value specifies by how much the variable should change every time.

Manage Case Metrics

Metric Name:

Lower Range:

Upper Range:

Initial Value:

Step:

Metric Name	Lower Range	Upper Range	Initial Value	Step	Delete
Score	0	100	100	-1	
Health	0	100	100	-1	

Figure 33: Custom case metrics editor.

4.6.6 Case Rules

Rule builder (Figure 34) allows authors to specify *IF – THEN* rules that affect the flow and the possible outcomes of the learners’ experiences with the VP case.

Rule Builder

If

Model outcome
 Metric value
 Distractor value
 Time Elapsed

Health less than 50

Then

Change model outcome
 Change metric value
 Provide feedback
 Change distractor value
 Show choices for next temporal tier

Your patient is not doing well

IF 'Patient meets criteria for single organ dysfunction' IS SELECTED THEN provide feedback: [Correct! The patient's Glasgow coma scale = 7 representing significant Central Nervous System dysfunction]	🗑️
IF 'Patient meets criteria for sepsis' IS SELECTED THEN provide feedback: [Incorrect! Must have at least 2 points on SOFA organ dysfunction scale. She currently has only one]	🗑️
IF 'Patient meets criteria for septic shock' IS SELECTED THEN provide feedback: [Incorrect. Septic shock is defined as sepsis plus hypotension refractory to fluids and pressors.]	🗑️
IF 'Time elapsed' = 30 THEN provide feedback: [Hannah has been transferred to MICU and intubated (mechanical ventilation) for airway protection since she is unresponsive. A central venous catheter (CVC) has been inserted. Her blood pressure which was relatively normal on admission (MAP 75) is now rapidly decreasing due to septic shock: MAP 60, HR 88 CVP 4, SVR 550, SvO2 66%]	🗑️
IF 'Norepinephrine' = Norepinephrine THEN provide feedback: [Nice! Norepinephrine is a first line vasopressor in septic shock.]	🗑️
IF 'Dobutamine Hydrochloride' IS SELECTED THEN provide feedback: [Dobutamine is an inotrope!]	🗑️
IF 'Epinephrine' = Epinephrine THEN provide feedback: [Epinephrine is a second/third line option. No need to go to epi just yet.]	🗑️
IF 'Phenylephrine' = Phenylephrine THEN provide feedback: [Phenylephrine is not a first line option - not potent enough.]	🗑️

Figure 34: Rule editor.

The current version of ModelPatient’s rule builder supports four types of rules, and each type of rule supports five possible outcome changes. I implemented these particular rule types based on the rules available in vpSim at the time of this writing, as well on the case authors’ responses collected during the preliminary user studies.

Authors can create a rule that will be triggered based either on a change in the underlying BN model, a change in a tracked metric value, a change in a distractor value,⁸ or a timer-driven event).

⁸Distractors are deliberately incorrect (or almost correct) choices added to the case to increase VP complexity

Table 5: Author-defined rule triggers

Trigger	Description
Model outcome	<p><i>If [Total Cholesterol] = [high] Then [do something...]</i></p> <p>In this rule [Total Cholesterol] represents the name of a node in the BN model, and [high] represents that current state of that node. In other words, this rule will be triggered when learners' actions cause a change in [Total Cholesterol] node to update the node's state to [high].</p>
Metric	<p><i>If [Health] < [50] Then [do something...]</i></p> <p>In this rule [Health] is an author-defined variable that represents the VP's health score and changes throughout the case based on learner actions. This rule will be triggered when learners' actions change the value of [Health] to some number below 50.</p>
Distractor	<p><i>If [Meningitis] is selected Then [do something...]</i></p> <p>In this rule [Meningitis] is an author-added distractor - in this example, it is a diagnosis that did not exist in the BN model or in the original data set. This rule will be triggered when a learner select "Meningitis" diagnosis from the list of diagnoses available in the VP case.</p>
Time elapsed	<p><i>If time elapsed is 30 seconds Then [do something...]</i></p> <p>Timer-based rules are based on the time elapsed (in seconds) since a learner entered the VP case.</p>

4.6.6.1 Rule Triggers

Table 6: Author-defined rule outcomes

Outcome	Description
Model outcome	<p><i>If [something triggers this rule...] Then [Stabilized Glucose] = [high]</i></p> <p>In this rule [Stabilized Glucose] represents the name of a node in the BN model, and [high] represents that current state of that node. In other words, when this rule is triggered by a condition defined in the <i>If</i> clause, the rule will update the state of [Stabilized Glucose] node to [high].</p>
Metric	<p><i>If [something triggers this rule...] Then [Score] - [10]</i></p> <p>When this rule is triggered by a condition defined in the <i>If</i> clause, the rule will subtract 10 points from the value of the author-defined [Score] metric.</p>
Feedback	<p><i>If [something triggers this rule...] Then feedback = [Author-defined feedback]</i></p> <p>When this rule is triggered by a condition defined in the <i>If</i> clause, the system will show learners author-provided text feedback. The feedback will appear in the “Notes” section of the EMR view as a “consultant note.”</p>
Distractor	<p><i>If [something triggers this rule...] Then [MAP] = 80</i></p> <p>When this rule is triggered by a condition defined in the <i>If</i> clause, the system will set the value of MAP (Mean Arterial Pressure) distractor to 80.</p>

Temporal tier	<p><i>If [something triggers this rule...]</i> Show choices for next temporal tier</p> <p>Temporal tiers allow authors to specify the order in which information is presented to the learner, as well as the order in which the model respond to learners' choices. When this rule is triggered by a condition defined in the <i>If</i> clause, the system will make the choices from the subsequent temporal tier available to the learner.</p>
---------------	---

4.6.6.2 Rule Outcomes

4.6.7 VP Case Player

The case player (the part of the systems that learners interact with) is designed to resemble a patient chart of an EMR (Figure 32).

Nodes and states from the underlying BN model, as well as author-generated distractors that are flagged as "Initial", are displayed to the learner at the start of the VP case (Figure 35). The BN model editor (Section 4.6.2 allows authors to categorize model's nodes as "Demographic," "Diagnoses," "Labs," etc... The information associated with BN model nodes is displayed in sections of the VP player that correspond to each node's category. For example, state(s) from a node categorized as "Laboratory Test" will be displayed in the "Labs" section of the VP player.

ModelPatient EMR

Score

Health

Allergy

Demographic

Diagnoses

Encounters

Labs

Drugs

Notes 1

Procedures

Radiology

Social History

Vitals 5

Outcome(s)

Demographic

			Cancel	Added By
<input type="radio"/>	Age	23		Dr. Alexander Fleming
<input type="radio"/>	Gender	Female		Dr. Alexander Fleming
<input type="radio"/>	Race	White		Dr. Alexander Fleming
<input type="radio"/>	Height	5'7"		Dr. Alexander Fleming
<input type="radio"/>	Weight	58 kg		Dr. Alexander Fleming
<input type="radio"/>	Photo			Dr. Alexander Fleming
<input type="radio"/>	Name	Hanna		Dr. Alexander Fleming

Figure 35: VP Case Player — Learner View — Patient Demographics Screen.

Any selections that the learner makes while interacting with the VP either update beliefs in the model and change outcome probabilities, or trigger author-created rules. In the example shown in Figure 36, the learner would be able to only select drugs that have corresponding nodes categorized as “Drugs” in the model, or that have been added as distractors by the case author (Subsection 4.6.4).

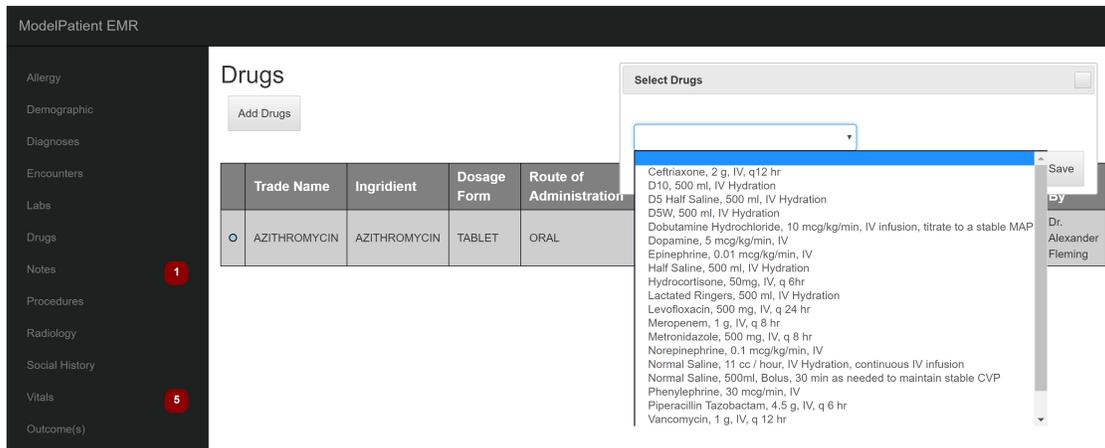


Figure 36: VP Case Player — Learner View — Prescribed Drugs.

When learners make decisions that result in changes in BN model states or trigger a feedback-generating rules, the VP player displays a red notification icon next to the section where the new information appears (Figure 37). This functionality is designed to imitate EMR notifications that clinicians receive when new consultant notes, lab results, and information from other providers become available.

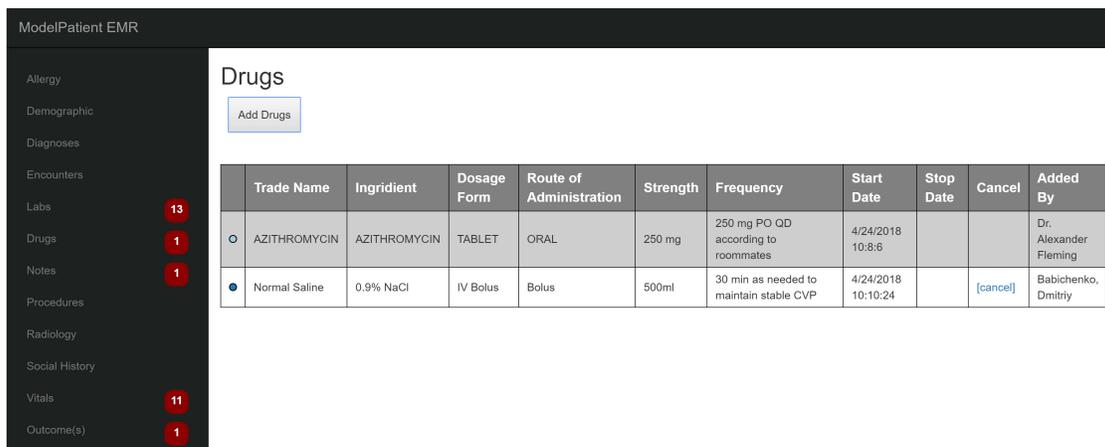


Figure 37: VP Case Player - Learner View - New Information Notifications.

Last, but not least, the “Outcomes” section shows how the learner’s choices had affected the probabilities in the model’s node categorized by the case author as “Out-

come.” Figure 38 shows how outcome probabilities changed for a learner in the “Managing Sepsis” case based on two sequential drug therapy choices. In this example, after the initial normal saline bolus hydration treatment, the probability of death was 7%, the probability of managed care was 50%, and the probability of discharge from the hospital was 44%. These probabilities essentially represent to $P(\text{Outcome} \mid \text{“Normal Saline Bolus”})$. However, when learner prescribes Vasopressin, the outcome probabilities change to 21% for death, 41% for managed care, and 38% for discharge from the hospital. These probability distribution reflect $P(\text{Outcome} \mid \text{“Normal Saline Bolus” and “Vasopressin”})$

Date	Note Text	Outcome	Cancel	Added By
4/24/2018 10:12:8	Based on the outcomes of your decisions, the probability of death is 7.000000000000001%, the probability of managed care (long-term hospitalization) is 50%, the probability of discharge from the hospital is 44%	death	[cancel]	Babichenko, Dmitry
4/24/2018 10:12:16	Based on the outcomes of your decisions, the probability of death is 21%, the probability of managed care (long-term hospitalization) is 41%, the probability of discharge from the hospital is 38%	death	[cancel]	Babichenko, Dmitry

Figure 38: VP Case Player - Learner View - New Information Notifications.

This section of the VP player received the most criticism from authoring experts and learners alike and required an additional explanation displayed at the beginning of the case:

This case is based on a statistical model derived from a dataset of real patients who were diagnosed with sepsis. Each decision you make while interacting with this case can (and most likely will) change the probabilities of certain outcomes. For example, you may notice that if you prescribe a particular medication, the virtual patient’s blood pressure will increase in response. It is important to remember that the statistical model DOES NOT represent a causal relationship - your decisions/actions DO NOT cause changes in outcomes. Your actions and the virtual patient’s responses simply

reflect the probability distribution that is already in the data. In other words, if you prescribe a medication and the system tells you that given your decision, the probability of the virtual patient's death is 55%, it doesn't mean that your decision CAUSED the virtual patient to die. All it means is that, in the data, a high percentage of the patients who were diagnosed with sepsis and received this medication did not survive.

Even with the aforementioned explanation, "Outcomes" section caused confusion with users. Specific comments and suggestions will be discussed further in Chapter 6 (Virtual Patient Case Evaluation) and Chapter 7 (Discussion and Future Work).

5.0 VIRTUAL PATIENT CASES

5.1 VP CASES

In order to assess the feasibility of BN model-based VP cases, as well as to address research questions 1 and 2 presented in the introduction to this dissertation (Chapter 1), I identified two existing VP cases that have been extensively used in teaching and training curricula of accredited health sciences institutions. Both of these cases have been used in academic research and have been described in published literature.

The “Managing Sepsis” case was originally created in vpSim and used in the University of Pittsburgh School of Pharmacy curriculum [118, 119, 120].

The “The Management of Diabetic Ketoacidosis (DKA)” case was created in DecisionSim and used for continuing education by the American College of Physicians (ACP) [121, 122].

The “Managing Sepsis” case was used in this work with permission of its original author, Dr. Neal Benedict, PharmD, Associate Professor, Pharmacy and Therapeutics, University of Pittsburgh School of Pharmacy. The “Managing DKA” case was used with permission from the ACP.¹

Another important reason for selecting these two cases is that they present two very different disease states that require very different approaches to treatment. The “Sepsis” case teaches drug therapies required in emergency treatment of an acute

¹Adapted from Tabas G, Korytkowski M, Harsha Rao H (2015). *Diabetic Ketoacidosis [Online patient case]*. Annals Virtual Patients. Philadelphia, PA: American College of Physicians. Retrieved online April 28, 2016 from <http://vp.acponline.org/virtualpatients/> Copyright (c) 2016 American College of Physicians. Used with permission.

condition, while the “DKA” case presents a multi-step treatment scenario that occurs over the course of multiple hours and requires a more temporal approach. The specifics of both cases are discussed in sections 5.3 and 5.4 respectively.

5.2 PATIENT COHORT DATA SET

The data used in creation of both cases, as well as in learning the parameters of the Sepsis BN model (Section 5.3.2) and the DKA BN model (Section 5.4.2) was originally obtained by Dr. Neal Benedict, PharmD for a research study titled “Comparison of Normal Saline and Plasma-Lyte for Fluid Resuscitation in Septic Shock,” IRB protocol #PRO14100151. The original purpose of this study was “to compare a retrospective cohort of ICU patients receiving normal saline for fluid resuscitation in septic shock with a retrospective cohort of patients receiving Plasma-Lyte for fluid resuscitation in septic shock.”² This data set contains de-identified patient history of 71,202 patients admitted to the emergency rooms (ER) or hospitals of the UPMC hospital system between 2001 and 2014. The data was retrieved from UPMC’s Medical Archival Systems (MARS) data archive by an honest broker. The data and the patient cohort selection criteria are described in Appendix A.

Since both sepsis treatment guidelines [123] and DKA treatment guidelines [124] recommend hydration as a critical line of treatment, I was able to extract a cohort of patients who were admitted to UPMC hospitals with a primary, secondary, or tertiary diagnosis of sepsis and DKA respectively. Inclusion details of the original data set and of the sepsis and DKA patient sub-cohorts are available in Appendix A.

The use of the “Comparison of Normal Saline and Plasma-Lyte for Fluid Resuscitation in Septic Shock” study data in this work has been approved by the University of Pittsburgh Institutional Review Board (IRB protocol #PRO17080163, *Designing Data-*

²Quoted from the IRB protocol for “Comparison of Normal Saline and Plasma-Lyte for Fluid Resuscitation in Septic Shock” study, IRB protocol #PRO14100151

5.3 VIRTUAL PATIENT CASE 1: MANAGING SEPSIS

5.3.1 Managing Sepsis: vpSim Case

“Sepsis of any type is one of the most time-sensitive non-trauma emergencies a health team is likely to face. The guidelines from the *Surviving Sepsis* campaign emphasize the importance of rapid identification of sepsis and quick deployment of therapies. Practice in this area is crucial for the pharmacy student to be comfortable enough to react swiftly and accurately when faced with a real sepsis case when in practice. Virtual patients provide a safe environment to practice these skills.”³

A vpSim version of the Sepsis case has been used for problem-based learning (PBL) in University of Pittsburgh School of Pharmacy and University of the Sciences College of Pharmacy courses [46, 119] since 2010. This case’s original author, Dr. Neal Benedict, PharmD, Associate Professor, University of Pittsburgh School of Pharmacy has provided a tremendous amount of help and insight in designing a ModelPatient version of the vpSim case and in developing a BN model representative of sepsis treatment.

The “Managing Sepsis” case presents learners with a 23-year-old female college student who was found unresponsive by her roommates and brought to the emergency department by an ambulance. “The roommates report that the patient went to the University health clinic for a sore throat about a week ago and was given Azithromycin for strep throat. For the last two days she had complained of a headache and in the last 24 hours, a stiff neck and photophobia.”⁴ (Figure 39).

³From an interview with Dr. Lorin Grieve, PharmD, Instructor, University of Pittsburgh School of Pharmacy

⁴From the vpSim “Managing Sepsis” case introduction

PATIENT CASE: Hannah Foore

Hannah is a 23 year-old female college student at the University of Pittsburgh. She was brought to the ED by ambulance this morning after she was found unresponsive to voice by her roommates. The roommates report that Hannah went to the University health clinic for a sore throat about a week ago and was given azithromycin for strep throat. For the last two days she had complained of a headache and in the last 24 hours, a stiff neck and photophobia.



PMH:

Unknown (none according to roommates)

SHX:

Unknown

FHX:

Unknown

Vitals:

Temp 39.9° C, HR 80, mean arterial BP (MAP) 75, RR 8

Ht 5'7"; Wt 58 kg

Allergies:

Unknown

Medications:

Only azithromycin 250 mg PO QD according to roommates

Imaging/Scans:

EKG: Normal sinus rhythm

CXR: clear

Figure 39: vpSim – Sepsis Case Narrative – Patient Introduction.

The learners are also presented with a set of initial vital signs and laboratory test results and are asked to evaluate, diagnose, and select appropriate drug treatments for the patient. The primary learning objectives for this case are (1) to evaluate patient's risk factors for septic shock, (2) to understand steps and guidelines for emergent sepsis treatment, and (3) to select appropriate empirical drug treatment therapy.

The VP case proceeds to guide the learner through a series of choices where the learner must make decisions on hydration, vasopressor, and antibiotic therapies needed to stabilize the virtual patient. The vpSim version of the case [118, 38] provides learners with options to correct their mistakes and offers real-time feedback, just-in-time learning opportunities [125, 126], and assessments in the form of multiple choice quiz questions (Figure 40).

Early antibiotics

Cultures are drawn; a lumbar puncture is performed and a meningitis regimen of IV ceftriaxone, vancomycin, and dexamethasone has been started.

Blood and CSF are sent to the laboratory.



In what time frame should antibiotics be administered for patients with severe sepsis or septic shock?

- Within the 1st hour from presentation to the hospital
- ✓ Within the 1st hour from severe sepsis or septic shock diagnosis
Correct! Evidence has shown a linear increase in the risk of mortality for each hour delay in antibiotic administration from diagnosis of severe sepsis or septic shock.
- Within the 1st hour from presentation to the ICU
- Within the 1st hour from patient symptoms of severe sepsis or septic shock

Figure 40: vpSim – Sepsis Case Quiz.

5.3.2 Sepsis BN Model

The ModelPatient version of the “Sepsis” case is based on a BN model learned from a cohort of 3,991 patients extracted from the master dataset described in Section 5.2. The inclusion criteria for the sepsis cohort and the resulting dataset’s features are described in Appendix A, Section A.2.

Initially five BN models were trained from the Sepsis cohort dataset. Four of the models were static BN models trained using TAN (Chapter 3, Section 3.3.2) and PC (Chapter 3, Section 3.4) at significance levels $\alpha = 0.5$, $\alpha = 0.1$, and $\alpha = 0.01$. The fifth model was a 5-timeslice dynamic BN (Chapter 3, Section 3.5).

Out of the five BN models the TAN model (Figure 41) performed best at predicting discharge location (what happens to the patient as a result of learners’ decisions)

(Table 8, 7 and Figures 44, 43, 42). The dynamic BN had the lowest overall accuracy of the five (0.433); after the 3rd timeslice most nodes' states exhibited uniform probability distributions.

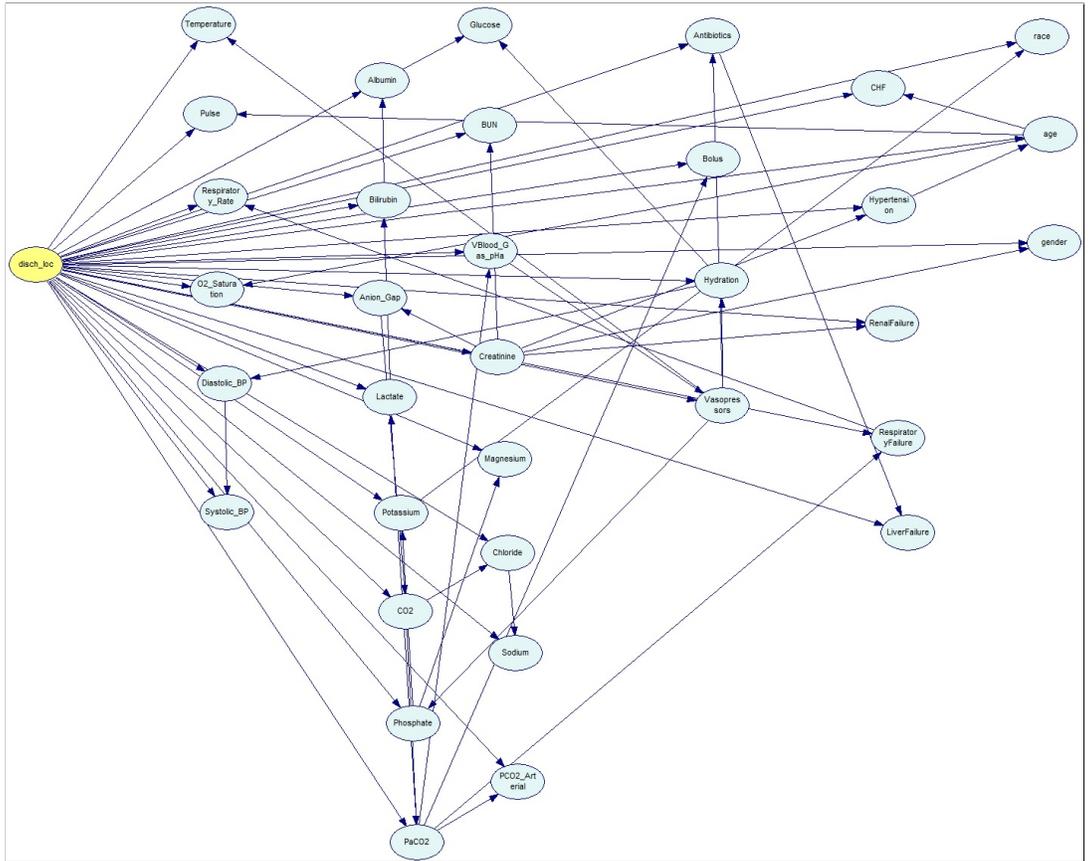


Figure 41: Sepsis BN model graph

Table 7: Accuracy of Predicting Discharge Location

Discharge location	0.614999	2329/3787
Deceased	0.692057	636/919
Managed Care	0.733015	1381/1884
Self Care	0.317073	312/984

Table 8: Sepsis Model Confusion Matrix

	Deceased	Managed Care	Self Care
Deceased	636	240	43
Managed Care	245	1381	258
Self Care	158	514	312

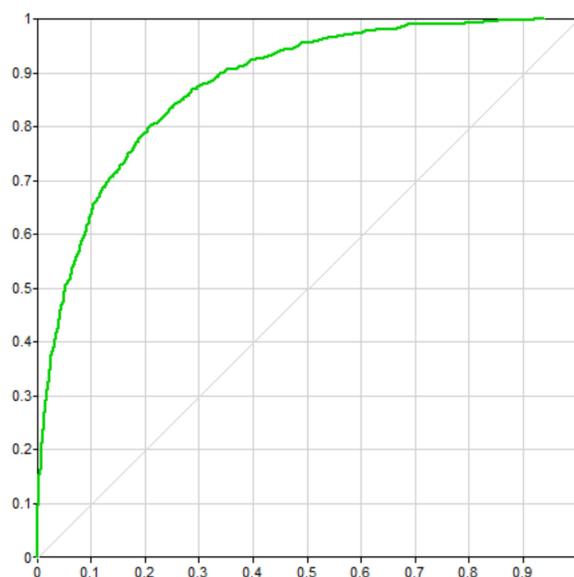


Figure 42: ROC curve for “Discharge Location” state (AUC=0.816)

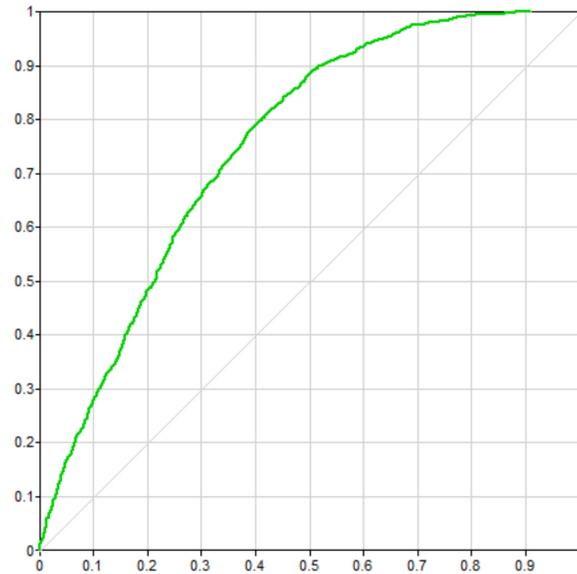


Figure 43: ROC curve for “Managed Care” state (AUC=0.662)

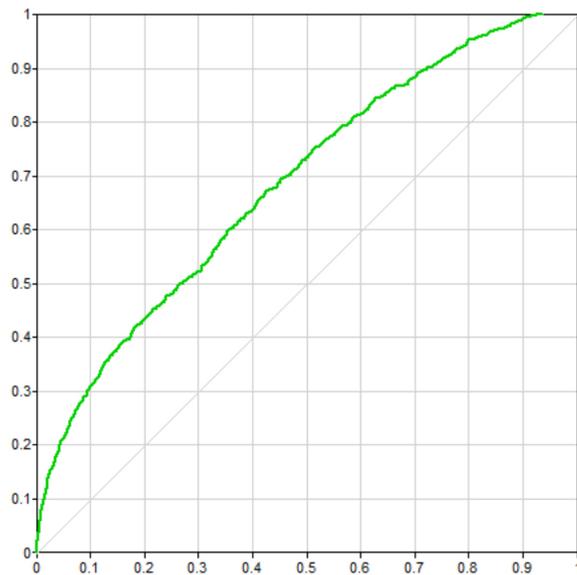


Figure 44: ROC curve for “Self Care” state (AUC=0.622)

All five models were evaluated by two clinicians to determine if the models accurately represented responses to hydration, vasopressor, and antibiotic treatments in septic patients. In this particular evaluation we relied on sepsis criteria that are defined as “bacterial infection plus two or more of the following systemic inflammatory

response syndrome (SIRS) conditions: (1) heart rate > 90 beats per minute, (2) respiratory rate > 20 per minute, or PaCO₂ < 32 mm Hg, (3) temperature < 96.8 F or temperature > 100.4 F, and (4) white blood cell count > 12,000/mm³ or < 4,000/mm³ or >10% immature bands [127, 128, 129].” Note that since all patients in the data set were already diagnosed with sepsis based on the ICD-9 codes used to select the cohort, I made the assumption that the bacterial infection criteria has already been met for all patients. It is important to note that the aforementioned sepsis selection criteria have been replaced by the sequential organ failure assessment, or SOFA score, “based primarily on laboratory results.” [130] Since the data set did not contain features related to organ failures (except for renal and respiratory failure) and did not have all the labs required for calculating a full SOFA score, we relied on the “quick SOFA,” or qSOFA score for additional model validation. The qSOFA score includes only two vital signs (systolic blood pressure ≤ 100 mm Hg and respiratory rate ≥ 22) [131]; changes to these vital signs could be examined at any point in the ModelPatient VP case and while observing the changes to the model’s states.

Both clinicians selected the TAN model as the most representative of a sepsis patient treatment scenario presented in the vpSim “Managing Sepsis” case and of the sepsis criteria defined above.

5.3.3 Managing Sepsis: ModelPatient Case

The ModelPatient version of the “Sepsis” case presents the same initial information as the vpSim version; the main difference in the presentation is that ModelPatient offers learners a view similar to that of an Electronic Medical Record (EMR) system. Learners have to explore the patient chart without any guidance from the system and make decisions based on available information rather than go through a series of multiple-choice decision points (Chapter 4, Section 4.6.7). Figures 45 and 46 show how clinical notes and laboratory test results are presented to the learner.

ModelPatient EMR	
Date	Note
4/25/2018 9:8:11	 <p>Hannah is a 23 year-old female college student at the University of Pittsburgh. She was brought to the ED by ambulance this morning after she was found unresponsive to voice by her roommates. The roommates report that Hannah went to the University health clinic for a sore throat about a week ago and was given azithromycin for strep throat. For the last two days she had complained of a headache and in the last 24 hours, a stiff neck and photophobia.</p>
4/25/2018 9:8:11	<p>PMH: Unknown (none according to roommates)</p> <p>SHX: Unknown</p> <p>FHX: Unknown</p>
4/25/2018 9:8:11	<p>Physical Exam:</p> <ul style="list-style-type: none"> • Neuro: withdrawals to painful stimuli; makes incomprehensible sounds; does not open eyes to painful stimuli (Glasgow Coma Scale score 7) • Resp: shallow breathing. Recent intubation (mechanical ventilation) in ICU for airway protection • CV: weak peripheral pulses • Skin: warm, diaphoretic
4/25/2018 9:8:11	Glasgow Coma Scale = 7

Figure 45: ModelPatient — Sepsis Case Clinical Notes.

ModelPatient EMR								
Laboratory Tests								
Add Laboratory Tests								
OrderDate	LabName	Result	Unit(s)	Low	High	Flag	Cancel	Added By
4/25/2018 9:1:59	Sodium	high	mEq/L					Dr. Alexander Fleming
4/25/2018 9:1:59	CO2	high	mEq/L					Dr. Alexander Fleming
4/25/2018 9:1:59	Sodium	136	mEq/L			Normal		Dr. Alexander Fleming
4/25/2018 9:1:59	Potassium	3.6	mEq/L			Normal		Dr. Alexander Fleming
4/25/2018 9:1:59	Chloride	100	mEq/L			Normal		Dr. Alexander Fleming
4/25/2018 9:1:59	CO2	29	mEq/L			High		Dr. Alexander Fleming
4/25/2018 9:1:59	BUN	9	mg/dL			Normal		Dr. Alexander Fleming
4/25/2018 9:1:59	Creatinine (Scr)	0.5				Low		Dr. Alexander Fleming
4/25/2018 9:1:59	Glucose	102				High		Dr. Alexander Fleming
4/25/2018 9:1:59	WBC	19				High		Dr. Alexander Fleming
4/25/2018 9:1:59	Hemoglobin (Hgb)	12	g/dL			Normal		Dr. Alexander Fleming
4/25/2018 9:1:59	Hematocrit (Hct)	36.4				Normal		Dr. Alexander Fleming
4/25/2018 9:1:59	Platelet count (Plt)	325	K/uL			Normal		Dr. Alexander Fleming
4/25/2018 9:1:59	Prothrombin Time (Pt)	13	seconds			Normal		Dr. Alexander Fleming
4/25/2018 9:1:59	INR	1				Low		Dr. Alexander Fleming
4/25/2018 9:1:59	aPTT	24	seconds			Low		Dr. Alexander Fleming
4/25/2018 9:1:59	AST	6	units per liter			Low		Dr. Alexander Fleming
4/25/2018 9:1:59	ALT	15	units per liter			Normal		Dr. Alexander Fleming
4/25/2018 9:1:59	Lumbar puncture	CSF results						Dr. Alexander Fleming

Figure 46: ModelPatient — Sepsis Case Labs.

This VP case is based on the Sepsis BN model described in the previous subsection

(5.3.2). All responses to learner decisions are driven by changes in probability distributions across the underlying model (described in Chapter 4, Sections 4.3, 4.6.7).

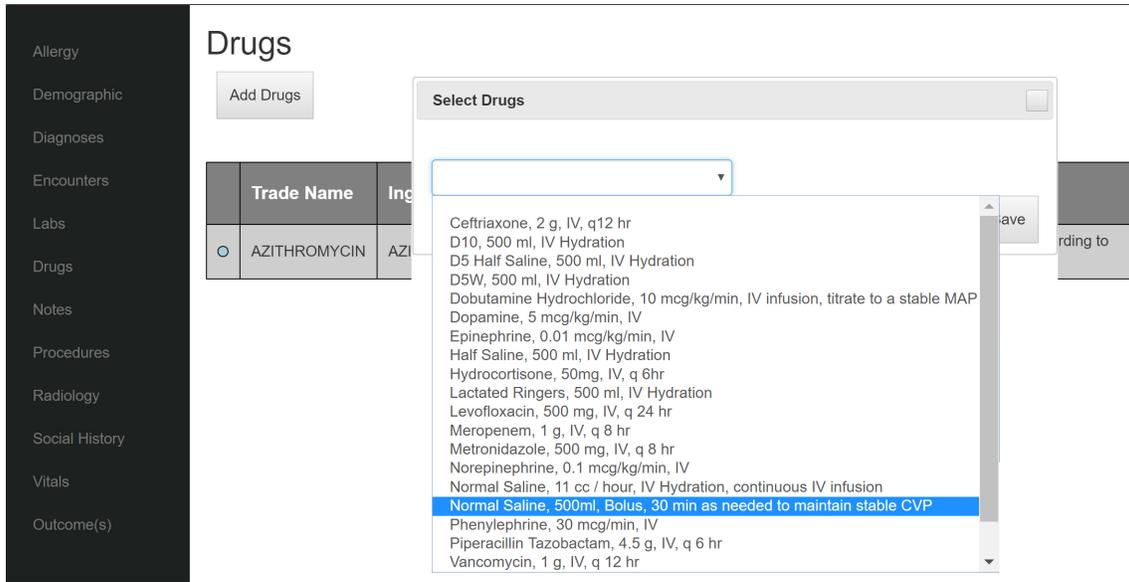


Figure 47: ModelPatient — Learner prescribing normal saline bolus hydration treatment.

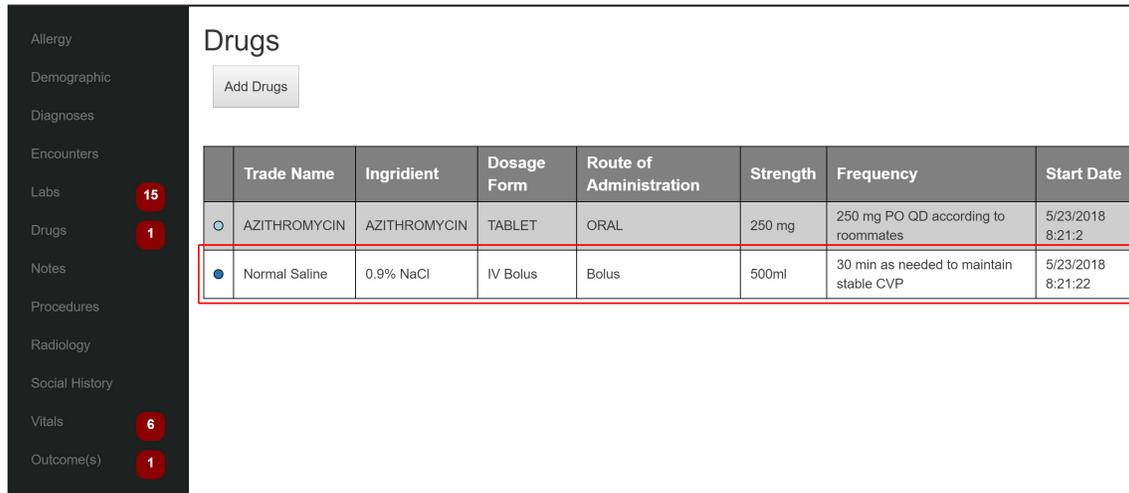


Figure 48: ModelPatient — Changes in the model’s probability distributions result in new information begin presented to the learner. Sections of the “EMR” that contain new information are denoted with red notification icons.

	Visit Date	Vital Name/Category	Value	Cancel	Added By
<input type="radio"/>	5/23/2018 8:22:9	Temperature	high		Dr. Alexander Fleming
<input type="radio"/>	5/23/2018 8:22:9	Temperature	39.9 C		Dr. Alexander Fleming
<input type="radio"/>	5/23/2018 8:22:9	Heart Rate	80		Dr. Alexander Fleming
<input type="radio"/>	5/23/2018 8:22:9	Mean arterial BP (MAP)	75		Dr. Alexander Fleming
<input type="radio"/>	5/23/2018 8:22:9	Respiratory Rate	8		Dr. Alexander Fleming
<input checked="" type="checkbox"/>	5/23/2018 8:22:15	Diastolic BP	74.95	[cancel]	Babichenko, Dmitriy
<input checked="" type="checkbox"/>	5/23/2018 8:22:15	Systolic BP	119.6	[cancel]	Babichenko, Dmitriy
<input checked="" type="checkbox"/>	5/23/2018 8:22:15	O2 Saturation	98.08	[cancel]	Babichenko, Dmitriy
<input checked="" type="checkbox"/>	5/23/2018 8:22:15	Pulse	80.18	[cancel]	Babichenko, Dmitriy
<input checked="" type="checkbox"/>	5/23/2018 8:22:15	Respiratory Rate	18.03	[cancel]	Babichenko, Dmitriy
<input checked="" type="checkbox"/>	5/23/2018 8:22:15	Temperature	37.46	[cancel]	Babichenko, Dmitriy

Figure 49: ModelPatient — The system displays changes to six vital signs based on the changes in the states’ probabilities of the corresponding BN model nodes following learner’s selection of “Normal Saline” hydration treatment (Figure 47)

To stay consistent with the vpSim “Managing Sepsis” case, the ModelPatient version only allows learners to select diagnosis and prescribe drug treatments — ordering procedures and labs was disabled for the purposes of this study. Figures 47, 48, and 49 show how choices and outcomes are presented to the learner.

Another important distinction between the vpSim and the ModelPatient versions of the “Sepsis” case is that vpSim provides learners with a clear stopping point — the patient is either “cured,” or the learner has been “kicked off the team” for making too many poor decisions. ModelPatient does not have an explicit stopping point — it is up to the learner to examine the changes in laboratory results and vital signs and recognize when to stop the treatment(s).

5.4 VIRTUAL PATIENT CASE 2 – THE MANAGEMENT OF DIABETIC KETOACIDOSIS

“Diabetic ketoacidosis (DKA) is a metabolic disorder seen as a complication of diabetes mellitus and is associated with significant morbidity and mortality. Management requires complex decision-making and improper treatment increases the risk for prolonged metabolic dysregulation as well as recurrence. General internists, emergency medicine physicians, and endocrinologists play a role in managing DKA. However, only a minority of health sciences students will encounter a patient with DKA during clinical experiential learning rotations because DKA accounts for [a very small fraction] of admissions to the hospital, necessitating an increased educational role for computer simulation. DKA as a topic lends itself to teaching clinical decision making because management requires the integration of several forms of clinical information for optimal outcomes.”⁵

At the time of this writing a vpSim / DecisionSim version of this case is currently available through the American College of Physicians’ (ACP)⁶ Annals Virtual Patients interactive online program.⁷

One of this case’s original authors, Dr. Gary Tabas, MD, Professor of Medicine, Director, Ambulatory Education, UPMC Shadyside work with me to identify patient cohort inclusion criteria, build a BN model representative of DKA treatment, and develop a VP case using ModelPatient’s authoring system.

5.4.1 Managing DKA: vpSim Case

The DKA case presents learners with partial patient history of Mrs. Debra Carter, a 43-year-old African-American woman with a history of type 2 diabetes mellitus who presented to the emergency department with nausea and vomiting.

⁵DKA description was provided by Dr. Gary Tabas, MD, Professor of Medicine, Director, Ambulatory Education, UPMC Shadyside

⁶American College of Physicians (ACP): <https://www.acponline.org/>

⁷ACP Annals Virtual Patients: <http://vp.acponline.org/virtualpatients>

The DKA case learning objectives are (1) to diagnose DKA, (2) to manage fluids, electrolytes, and insulin, and (3) to transition the patient from continuous IV insulin to subcutaneous insulin.

The case guides the learner through obtaining patient information (history, physical exam, diagnostic tests) through multiple choice / multiple answers questions and allows learners to make clinical decisions through multiple choice / one correct answer questions. Each choice/decisions directly affects the patient's clinical status (medications, procedures, consultation) - for each decision vpSim presents learner with a table of laboratory test result changes over time (Figure 50). It is important to note that the data displayed on each of the vpSim's screens is manually entered by the case author(s).

Continuous Insulin for One Hour

OK - Your order is implemented.

One hour later: Respirations are deep and labored (Kussmaul respirations). Her mouth and axilla are dry. Lungs are clear and heart exam reveals tachycardia. The abdomen is mildly diffusely tender. There is no peripheral edema.

BP:

110/50 mmHg (supine)

90/palp mmHg (sitting)

Resp: 18/min

Pulse: 115

Lab Values	New	Original (if changed)
Glu	602 mg/dL	736 mg/dL
P _{NA}	126 mEq/L	128 mEq/L
HCO ₃	8 mg/dL	
Cl	100 mg/dL	102
Anion gap	18	18

What would you like to prescribe at this point

- ▶ Intravenous fluids
- ▶ Continuous insulin infusion
- ▶ Subcutaneous insulin
- ▶ Oral fluids

Figure 50: vpSim DKA Case – The system displays changes vital signs and laboratory values based on learner’s choices on prior vpSim screen(s).

In order to successfully treat the virtual patient the learner must navigate through twelve treatment timeslices (Table 9).

Table 9: DKA VP treatment timeslices presented in the vpSim version of the case

Timeslice	Description	Time Since Last Treatment

1	Initial Management	0 hours
2	Continuous insulin infusion and intravenous fluids (0.9% Normal Saline [NaCl], 1L per hour)	1 hour
3	Continuing Fluid Choices (0.9% Normal Saline [NaCl], 1L per hour)	1 hour
4	Additives (Potassium, 10 meq, 500 cc per hour)	1 hour
5	Regular Insulin Bolus	1 hour
6	Continuous intravenous insulin infusion, 10 units per hour	1 hour
7	Continuous intravenous insulin infusion, 10 units per hour	2 hours
8	Continuous intravenous insulin infusion, 12 units per hour	1 hour
9	0.45% NaCl and insulin infusion rate at 12 units per hour	2 hours
10	0.45% NaCl and insulin infusion rate at 12 units per hour	1 hour
11	5% dextrose in 0.45% NaCl at 6 units per hour	1 hour
12	Insulin glargine (Lantus) or insulin detemir (Levemir) subcutaneously at 35 units	2 hours

This case's vpSim implementation allows learners to mediate poor decisions by going back to the last "correct" treatment choice and provides just-in-time learning opportunities through tutorials.

Tutorial - Initial Management of DKA

Fluid administration is the first step the initial management of patients with diabetic ketoacidosis, and the **intravenous route** is recommended.

Isotonic saline (0.9% NaCl) is the recommended initial fluid used to replace fluid deficits (up to 6L) in DKA.

The amount of fluid retained in the intravascular space by giving a litre (1000 cc) of 0.9% NaCl, 0.45% NaCl or dextrose in water is, respectively, 300-400 ml, about 200 ml and only about 100 ml.

Therefore, the initial rate of 0.9% NaCl administration should be 1L/hr.

Lactated Ringer`s solution is similar to saline at the correctly prescribed rate.

Optimal fluid therapy in DKA:

1. Expands intravascular and interstitial volume
2. Restores renal perfusion
3. Reduces blood sugar
4. Decreases resistance to low-dose insulin
5. Decreases levels of counter-regulatory hormones

1 L of NSS is recommended to be given IV in the first 30-60 minutes.

► Return to Intravenous Fluids

Figure 51: vpSim DKA Case – Fluids administration tutorial

Unlike the “Managing Sepsis” case (Section 5.3), the DKA vpSim case does not allow the learner to fail and resolves with the patient’s blood sugar controlled on 35 units of subcutaneous insulin glargine. At the end of the case the learner is presented with a summary of learning in the form of DKA treatment guidelines as they relate to the individual learning objectives.

5.4.2 Diabetic Ketoacidosis Model

Creating a BN model of DKA turned out to be a challenging task, as diabetic ketoacidosis requires a complex treatment scenario. Over the course of 6 months I have worked with Dr. Gary Tabas, one of the original authors of the vpSim DKA case, to design a BN model of DKA and a corresponding virtual patient case. The original vpSim case guides learners through 12 time-based treatment steps (Section 5.4). We attempted several approaches to building a DKA model:

- Manually create the model’s structure for each timeslice based on the features (treatments, lab values, vital signs) listed in the vpSim case and learn the model’s parameters from the data using GeNIe and SMILE (Chapter 4, Section 4.3).
- Create a static BN model by learning both the structure and the parameters from the DKA patient cohort dataset using the PC algorithm (Chapter 3, Section 3.4).
- Create a dynamic BN model by learning both the structure and the parameters from the DKA patient cohort dataset. For the dynamic BN we attempted creating 3-, 5-, and 7-timeslice models.

The DKA patient cohort dataset (Appendix A, Section A.4 contains data for a little over 3,000 DKA patients. For some of these patients there are as few of 3 time slices of treatments and responses, and one patient has as many as 57 time slices. To complicate things further, the treatments in each time slice are different. For example, in the first time slice the patient may be given a normal saline bolus, in the second time slice the treatment is switched to continuous hydration, in the third time slice the learner needs to select an additive to the hydration solution (e.g. potassium). Presence of these very different treatments in each time slice makes it difficult to model the DKA disease state using a dynamic Bayesian network model.

Furthermore, the feature space is so large (there are many variables required in administering and evaluating results of a treatment) that in a BN model with 5 static time slices the distributions become mostly uniform after the 3rd timeslice, making the model useless. A DKA virtual patient case based on a 3-timeslice model was immediately rejected by an expert as too simplistic and not representative of DKA treatment protocols.

In the end, we used a 3-timeslice model (Figure 52), but heavily augmented it with expert-defined rules in order to make the case comparable to the one in vpSim. That being said, with a rule-augmented case the model is responsible only for about 20% of the treatment scenario – the rest of the case is governed by a total of 195 rules.

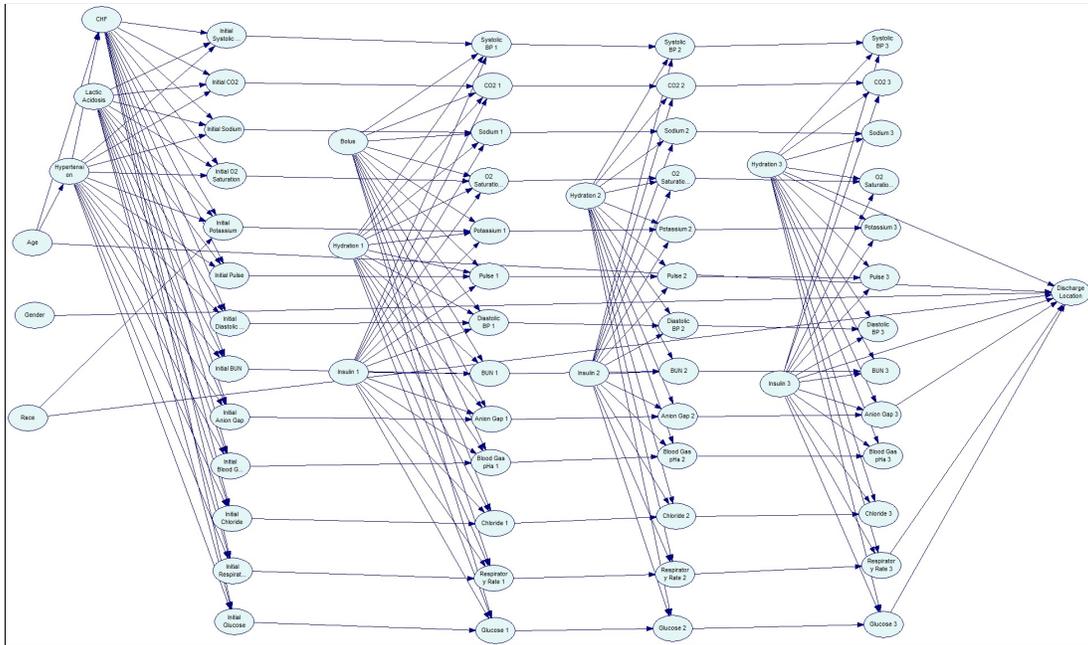


Figure 52: DKA model structure

Because the DKA BN model was responsible for such a small percentage of the resulting VP case, I did not use the “Managing DKA” case in the experiments described in Chapters 6. That being said, the difficulties discovered while attempting to develop the DKA BN model and the “Managing DKA” VP case provided ideas and insights into how such issues could be addressed in future iterations of ModelPatient (discussed in Chapter 7).

6.0 STUDY DESIGN, ANALYSIS, AND FINDINGS

This chapter is divided into two main sections, one for each of the research questions. These sections are further subdivided into the following subsections: (1) subject selection criteria and subject demographic information, (2) experimental design, and (3) analysis and findings.

Studies associated with research questions 1 and 2 (RQ1 and RQ2) have been approved by the University of Pittsburgh Institutional Review Board (Study title: “*Designing Data-Driven Virtual Patients for Health Sciences Education*,” IRB Protocol #PRO17080163).

Subject recruitment email verbiage and consent forms are available in Appendix [E](#). Questionnaires used in both RQ1 and RQ2 studies are available in Appendix [C](#).

There was no overlap between subjects who participated in RQ1 and RQ2 studies. In other words, subjects from RQ1 study did not participate in RQ2 study, and vice versa.

Both studies were conducted at a variety of locations, including hospital lounges, hospital cafeterias, coffee shops, private offices, University of Pittsburgh libraries, and through videoconferencing software (Skype).

Subjects were not compensated for their participation in either study.

Both studies were conducted using the “Managing Sepsis” VP case. I selected the “Managing Sepsis” case as the baseline for this study because its vpSim version has been extensively used in teaching at the University of Pittsburgh School of Pharmacy. Moreover, several academic research studies based on the “Managing Sepsis” case

have been conducted and described in published literature [118, 119, 120].

6.1 RESEARCH QUESTION 1 (RQ1)

To what extent is it possible to create a virtual patient case using a VP system based on a Bayesian network model that is comparable to a case created using a branched-narrative VP system?

The purpose of this study was to benchmark the clinical and educational content and presentation of a VP case based on a BN model (ModelPatient) to the content and presentation of an established branched-narrative VP case (vpSim) in order to determine whether or not a VP case based on a BN model can be an effective teaching tool. This study relied on the following nine criteria to compare the clinical and educational content and the presentation of the two aforementioned versions of the “Managing Sepsis” VP case:

1. How well did the learning experience reflect the stated learning objectives? (Q1)
2. How accurately did the case depict the chosen topic? (Q2)
3. How well did the case represent clinical variations associated with the chosen topic? (Q3)
4. How well did the case introduce / reinforce relevant clinical skills? (Q4)
5. How well did the case introduce / reinforce behaviors and attitudes relevant to sepsis treatment? (Q5)
6. How well did the case provide formative feedback for choices / outcomes? (Q6)
7. How realistic was the case? (Q7)
8. Based on the presentation method of the virtual patient case, how engaging was the case? (Q8)
9. Based on the presentation method of the virtual patient case, how effective was the case in teaching the subject matter to novice clinical practitioners? (Q9)

The comparison and evaluation criteria for the VP case clinical and educational content and presentation were based on and derived from:

1. Features and uses of high-fidelity medical simulations that lead to effective learning proposed by Issenberg, et. al. [24]
2. Posel's dissertation work on validating criterion-referenced guidelines for virtual patient case authoring [25]
3. Bateman et. al. exploratory study on virtual patient design [26]

6.1.1 Subject Selection Criteria and Demographic Information

In order to address RQ1, the VP cases used in this study needed to be evaluated from clinical and educational perspectives by subjects with both clinical and educational background / expertise. For that reason participating subjects were selected based on meeting **BOTH** of the following inclusion criteria:

- The subject is actively practicing in a clinical field and is familiar with either sepsis or diabetic ketoacidosis treatment guidelines.
- The subject is actively teaching in an accredited health sciences institution (medical, pharmacy, nursing, dental, public health, epidemiology), or is a practicing clinician who performs preceptor duties with medical, pharmacy, or nursing students, and is familiar with either sepsis or diabetic ketoacidosis treatment guidelines.

A total of 40 subjects were recruited for the study. Data collected from the first two subjects were removed from the analysis because based on their feedback the Model-Patient version of the "Managing Sepsis" VP case had to be updated with a clarifying explanation detailing how the outcome probabilities change and how that affects feedback and results of treatment choices. Adding a clarifying explanation to the VP case effectively changed its content; in order to ensure that all subjects complete the same VP case, the first two subjects' responses were removed from the final analysis. Data collected from one more subject was removed from the analysis because they rated

all Likert questions as “4” without actually reading the questions and did not provide any text responses. From the remaining cohort, 33 subjects completed the “Managing Sepsis” case and four completed the “Managing DKA” case. Due to the shortcomings of the “Managing DKA” case outlined in Chapter 5, Section 5.4, only the data collected from the 33 subjects who completed the “Managing Sepsis” case were used in the analysis and evaluation.

Selected subjects represented a wide variety of clinical specialties (Table 10) and years of experience (mean of 13 years of experience with a standard deviation of 7.34 and median of 11).

Table 10: Subjects’ Clinical Specialties

Specialty	Count
Emergency medicine	2
Otolaryngology (ENT)	1
Gastroenterology	1
General Internal Medicine	5
Maxillofacial surgery	1
Neurology	1
Nursing	6
Oncology	1
Orthopedic surgery	1
Palliative Care	1
Pediatrics	3
Pharmacy	4
Public health	3
Pulmonology	3

6.1.2 RQ1 Study Design

The subjects were randomly selected into two groups (Group A and Group B) using a Python script (Appendix F). Each subject was asked to complete one VP case in both vpSim and ModelPatient. Participants in Group A (n = 17) completed their assigned cases in vpSim first, followed by the same case in ModelPatient. Participants in Group B (n = 16) completed their assigned cases in ModelPatient first, followed by the same case in vpSim. This helped reduce the priming bias based on which system was used first [132].

Subjects completed the vpSim version of the “Managing Sepsis” case using a web-based vpSim¹ instance hosted by the Laboratory for Educational Technologies (LET)² at the University of Pittsburgh School of Medicine.

Subjects completed the ModelPatient version of the “Managing Sepsis” case deployed as a localhost web application on a laptop.

After completing each version of the case, each participant completed a web-based Qualtrics³ questionnaire (Appendix C, Section C.1) on the same laptop that they used for completing VP cases.

The questionnaire was divided into three sections: (1) demographic information, (2) VP case content and presentation evaluation based on the nine criteria described in Table 35, and (3) open-ended questions asking subjects to provide free-text feedback on strengths and weaknesses of each version of the “Managing Sepsis” case and each VP system.

After completing BOTH versions of the case (one in vpSim and one in ModelPatient), subjects completed an additional 2-question Qualtrics Likert-scale questionnaire comparing educational content and the degree to which the learning objectives were met between the two versions of the “Managing Sepsis” case (Appendix C, Ta-

¹vpSim: <https://vpsim.pitt.edu>

²Laboratory for Educational Technology at the University of Pittsburgh School of Medicine. <https://let.pitt.edu/>

³Online Survey Service (Qualtrics): <http://technology.pitt.edu/services/qualtrics-survey-service>

ble 36).

Flowchart shown in Figure 53 summarizes the steps and the flow of the RQ1 study.

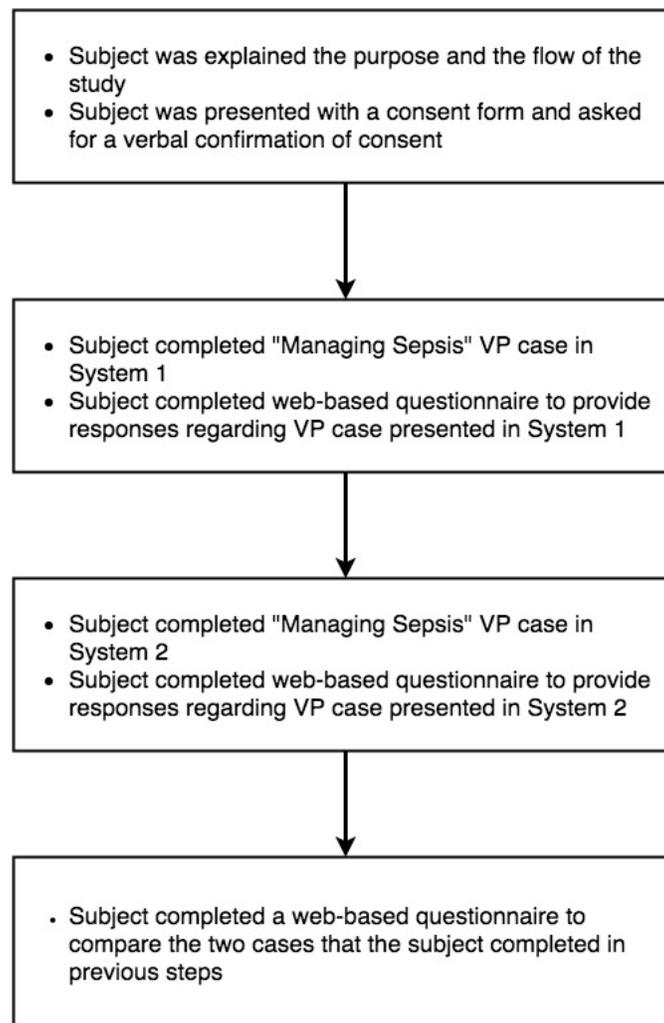


Figure 53: RQ1 Study Flowchart.

Screen captures of subjects' interactions with both systems were recorded using Open Broadcaster Software (OBS) Studio.⁴ Since this study did not involve interview questions or capturing "think-out-loud" observations, audio was not recorded as part of this experiment.

Subjects spent an average of 57 minutes on the on RQ1 study (median of 53 minutes, standard deviation of 14 minutes). Subjects spent an average of 8 minutes 34

⁴Open Broadcaster Software (OBS) Studio. <https://obsproject.com/>

seconds in vpSim (median of 7 minutes 11 seconds, standard deviation of 6 minutes 8 seconds) and 11 min 30 seconds in ModelPatient (median of 10 min 28 seconds, standard deviation: 3 min 49 seconds) (Appendix G).

Both vpSim and ModelPatient record users’ decisions, time between decisions, each visited screen and time spent on each screen in user activity logs. While not directly relevant to this study, this data was captured and stored for future analysis. Examples of user activity logs for both vpSim and ModelPatient are available in Appendix D. Moreover, backups of ModelPatient’s MongoDB database were exported and saved as JSON files using the MongoDB export utility⁵ after each case completion.

6.1.3 Analysis and Findings

Likert scale questions in the RQ1 questionnaire (Appendix C, Section C.1) were evaluated for internal consistency and reliability using Cronbach’s alpha. Overall Cronbach’s alpha for the entire questionnaire was 0.837, indicating a high level of internal consistency [133]. Removing responses to individual questions did not significantly affect the overall reliability (Table 11).

Table 11: Item-Total Statistics (Cronbach’s Alpha if Item Deleted)

Question	Cronbach’s Alpha if Item Deleted	
	vpSim	ModelPatient
Q1: How well did the learning experience reflect the stated learning objectives?	0.838	0.821
Q2: How accurately did the case depict the chosen topic?	0.842	0.834

⁵MongoDB Export Utility: <https://docs.mongodb.com/manual/reference/program/mongoexport/>

Q3: How well did the case represent clinical variations associated with the chosen topic?	0.843	0.819
Q4: How well did the case introduce / reinforce relevant clinical skills?	0.824	0.817
Q5: How well did the case introduce / reinforce behaviors and attitudes relevant to [<i>DKA or sepsis</i>] treatment?	0.824	0.818
Q6: How well did the case provide formative feedback for choices / outcomes?	0.841	0.839
Q7: How realistic was the case?	0.836	0.815
Q8: Based on the presentation method of the virtual patient case, how engaged were you?	0.833	0.819
Q9: Based on the presentation method of the virtual patient case, how effective do you think this case is in teaching the subject matter to novice clinical practitioners?	0.833	0.830

I used the Wilcoxon rank-sum test to identify if there is a statistically significant difference in the per-question responses (responses about vpSim were compared to responses about ModelPatient) given the effect size $d = 0.5$, significance level $\alpha = 0.05$, and power $\beta = 0.80$ (80%) (Table 12).

Table 12: Wilcoxon Signed Ranks Test Results

Question	Mean		St.Dev.		Z	Sig (α)
	vpSim	MP	vpSim	MP		
Q1: How well did the learning experience reflect the stated learning objectives?	4.64	3.79	0.489	0.696	-4.242	0.001
Q2: How accurately did the case depict the chosen topic?	4.15	4.21	0.508	0.650	-0.462	0.644
Q3: How well did the case represent clinical variations associated with the chosen topic?	3.24	4.39	0.561	0.864	-3.999	0.001
Q4: How well did the case introduce / reinforce relevant clinical skills?	4.64	3.48	0.653	0.834	-4.573	0.001
Q5: How well did the case introduce / reinforce behaviors and attitudes relevant to [DKA or sepsis] treatment?	4.55	3.30	0.666	0.770	-4.557	0.001
Q6: How well did the case provide formative feedback for choices / outcomes?	4.73	3.30	0.626	0.728	-4.608	0.001

Q7: How realistic was the case?	3.42	4.45	0.663	0.869	-3.912	0.001
Q8: Based on the presentation method of the virtual patient case, how engaged were you?	4.27	4.61	0.674	0.788	-2.338	0.019
Q9: Based on the presentation method of the virtual patient case, how effective do you think this case is in teaching the subject matter to novice clinical practitioners?	4.82	3.70	0.584	1.159	-3.880	0.001

The Wilcoxon rank-sum test results (Table 12) indicate that a branched-narrative case presented in vpSim was rated significantly higher than its ModelPatient counterpart on reflecting the learning objectives (Q1), introducing / reinforcing clinical skills, attitudes, and behaviors relevant to sepsis treatment (Q4, Q5), providing formative feedback for choices / outcomes (Q6), and being more effective in teaching the subject matter to novice clinical practitioners (Q9).

ModelPatient case was rated significantly higher than the vpSim version on representing clinical variations associated with sepsis (Q3), on realism (Q7) and on engagement (Q8).

Both versions of the “Managing Sepsis” case showed no significant difference in ratings on how accurately they depicted the chosen topic (Q2).

Response means for the Likert scale questions presented in the educational content and learning objectives survey (Table 13) indicate that subjects found the two systems “somewhat comparable”, with a mean rating of 4 on the scale of 1 to 5, with

1 being “Not comparable at all” and 5 being “Extremely Comparable”.

Table 13: Summary Statistics — Case Presentation and Learning Objectives Comparison

Question	Mean	St.Dev	Median
In terms of educational content, how comparable is the case presented in ModelPatient to the case presented in vpSim?	4.030	0.684	4
In terms of meeting the learning objectives, how comparable is the case presented in ModelPatient to the case presented in vpSim?	3.970	0.770	4

6.1.3.1 Analysis of Free-Text Responses (Affinity Diagramming) In order to identify common themes from the participants’ responses about strengths and weaknesses of the VP systems and the VP cases presented in this experiment, answers to questions 2 - 5 from the RQ1 questionnaire were analyzed using the affinity diagram method. “Affinity diagramming is a method of qualitative analysis wherein common themes are identified from [qualitative] data, such as statements from a free response interview” [134, 135]. Affinity diagrams are commonly used in qualitative analysis across multiple domains and disciplines, including identifying customer requirements [136], creating business and marketing models [137], and software systems usability analysis [138, 139].

Since responses to most questions contained statements about multiple aspects of each version of the “Managing Sepsis” case and of the two VP systems used in this experiment, I separated each response into a series of atomic statements. For example, one subject responded to the question “Describe the best/strongest features of this case?” with “I think this case is a much better fit for novice learners. The other system (ModelPatient) presented a case much more suited to residents or practicing

physicians. The feedback and guidance in this case were excellent. This case is better designed to both test someone's knowledge and to teach." I separated this response into the following statements:

- "This case is a much better fit for novice learners."
- "The other system (ModelPatient) presented a case much more suited to residents or practicing physicians."
- "The feedback and guidance in this case were excellent."
- "This case is better designed to both test someone's knowledge and to teach."

Seven independent raters reviewed text response statements associated with questions 2 - 5 from the RQ1 questionnaire to identify common themes. All raters received the same affinity diagramming instructions at the same time and analyzed the text data while sitting in the same room. Three of the raters were faculty from the University of Pittsburgh School of Computing and Information, two of the raters were faculty from the University of Pittsburgh School of Pharmacy, and two of the raters were staff from the University of Pittsburgh School of Computing and Information. Two of the raters had prior experience with designing virtual patient cases.

Each rater identified and labeled the common themes independently of each other. After responses were grouped according to independently identified common themes, the labels for the common themes were reviewed and reconciled as a group. For example, six raters identified "Feedback" as a common theme, while one of the raters labeled the same theme as "System responses." Both labels carried the same meaning and were reconciled under a single label of "Feedback."

While there is no consensus in the literature on a threshold for a satisfactory level of agreement among raters, there are some suggestions that inter-rater agreement of 70% or higher is acceptable [140]. Common themes were defined as similar statements about reviewed VP cases and the respective systems made by at least 25% of the experiment subjects (9 subjects out of 33) [141] and recognized as "common" by at least 70% of the reviewers (5 reviewers out of 7).

6.1.3.2 Common Themes From Text Responses About ModelPatient In describing the best/strongest features of the “Managing Sepsis” case presented in ModelPatient, 67% of subjects commented that they liked that the results are reflective of learner decisions and that *“feedback to learners [is] provided in terms of immediate changes in lab and vital values in response to learner decisions similar to the way a real patient would respond to treatments.”* 33% of subjects felt that the EMR-like case presentation *“teaches students how to read patient charts and select appropriate treatment.”*

In response to the question *“What features of this case did you struggle with the most?”* 30% of subjects commented that the case is not suitable for novice learners because it does not provide *“enough formative feedback for incorrect decisions”*, does not *“enforce [...] temporal treatment order,”* and does not provide *“just-in-time learning opportunities.”* 27% of subjects reported that they did not like the open-ended nature of the case because *“there are no clear stopping criteria.”* 42% found it difficult to interpret outcome probabilities (hospitalization, discharge, death) presented in the “Outcomes” tab of the VP player, stating that *“[it] either needs a better explanation, or a different way to show learner the results / outcomes.”* 36% of subjects expressed frustration with *“not being able to order additional labs and procedures.”*

In response to the question *“Describe the best/strongest features of this VP system?”* 48% of subjects commented on ModelPatient’s realism in portraying VP treatment scenarios — *“cases are built on real patient data and respond to treatments like real patients would.”* 42% liked that virtual patient cases are presented as a patient chart and that the learner-facing user interface looks like an electronic medical records (EMR) system.

In describing ModelPatient’s features that could use improvement, 27% of respondents said that while the current user interface resembles an EMR, it needs to look even more like Epic or Powerchart. These suggestions included statements that *“labs and vitals results should show newest results first”* and *“system should show changes*

in lab values over time either as a side-by-side display or a trend graph” and “drug ordering screen should not be a single dropdown with a list of medications.” 51% of subjects felt that ModelPatient needs a better formative feedback mechanism — “changing vitals and lab values and popup messages are not enough for a novice learner.” Last, but not least, 27% mentioned that ModelPatient needs to have an ability to allow learners to recover from incorrect decisions — “learners should be able to make a wrong decision, observe the effect, receive formative feedback, and ‘cancel’ an incorrect order.”

6.1.3.3 Common Themes From Text Responses About vpSim While describing the best/strongest features of the vpSim “Managing Sepsis” case, 42% of subjects found the case to be well-suited for novices *“because of the guided nature of the case.”* 70% of subjects commented on the *“excellent feedback for both correct and incorrect decisions,”* and 36% liked the ability to *“go back and review previous choices.”* 27% of participants indicated that just-in-time learning opportunities and build-in assessments in the form of mini-quizzes were the vpSim “Managing Sepsis” case’s strongest features. As one of the subject’s commented, *“This case was a nice guided experience with feedback and just in time learning opportunities, which is what you want to see in a case for novice clinical practitioners.”*

In response to *“What features of this case did you struggle with the most?,”* 64% indicated that the case needs to have more decisions points and branches — *“some of the decision options and multiple choice questions have too few choices and make the correct decision too obvious.”*

Only one common theme emerged from the responses to *“Describe the best/strongest features of this VP system?,”* with 73% of subjects listing “ease of use” as vpSim’s strongest feature — *“Very simple to use, [...], no ambiguity about what needs to be done to make choices / decisions.”*

Finally, 30% of subjects noted that when it comes to vpSim’s features that could use improvement, the system is not interactive / engaging enough and needs *“more*

dialogue / Q&A with the patient” and the ability to “ask open-ended questions [to] collect [patient] history.” 27% commented that vpSim needs to be able to support more choices — “the system needs to be able to generate more choices to make cases more challenging, to drive actual critical decision-making rather than answering multiple choice questions.”

6.2 RESEARCH QUESTION 2 (RQ2)

To what extent does the proposed system facilitate modifying an existing VP by allowing case authors to change the model’s states in such a way that the VP case meets alternate learning objectives?

As mentioned several times in the “*Introduction*” and “*Simulation and Tutoring Systems in Medical Education*” chapters of this dissertation (Chapters 1 and 2), some of the barriers to authoring VP cases with current VP authoring systems include the difficulty of managing case complexity and the burden that authoring/editing VP cases puts on already time-constrained content experts [14, 15, 16]. This study’s primary aim is to assess the degree to which the the use of annotated BN models in VP cases facilitates modifying an existing VP case by allowing case authors to manipulate the underlying model in such a way that the VP case meets alternate learning objectives.

6.2.1 Subject Selection Criteria and Demographic Information

Participating subjects were selected based on meeting **BOTH** of the following inclusion criteria in order to be classified as VP authoring experts:

- The subject authored at least 2 VP cases that have been used in a curriculum of a medical, nursing, dental, or pharmacy school, or a simulation center
- The subject is an author or a co-author of at least one publication that is related to simulation authoring, evaluation, or use of simulation in health science curricula

A total of 27 subjects were contacted via email (Appendix E) about participation in this study. All subjects were identified based on either prior research collaborations with the author of this dissertation, a Google Scholar publication search, or a direct introduction through their work with virtual patient case design for DecisionSim,⁶ a company that provides a simulation-based technology platform for developing and deploying virtual patient cases.

Five subjects participated in preliminary interviews that informed some of the design features of ModelPatient's authoring system; 14 subjects agreed to participate in the study related to RQ2.

Out of the 14 subjects who agreed to participate in the RQ2 study, one subject dropped out due to scheduling issues. Two subjects realized in the middle of their respective interviews that they do not meet all of the inclusion criteria — their data were excluded from the analysis. One subject was removed from the cohort due to multiple technical issues that occurred while trying to conduct the experiment using Microsoft Skype for Business for videoconferencing and screen sharing.

It is important to note that the intention of the RQ2 study is to extrapolate common themes related to VP case authoring rather than generalizing the outcomes to a population. In other words, this is a preliminary qualitative study designed to elicit feedback from VP case authoring experts. Other published studies that relied on qualitative analysis of experts' responses suggest that the number of expert subjects used in the RQ2 study ($n = 10$) is sufficient for preliminary qualitative analysis [142, 143].

The 10 participating subjects represented a wide variety of clinical specialties (Table 14), years of clinical experience (mean of 16.7 years of experience with a standard deviation of 10.92 and median of 16.5), years of virtual patient case design experience (mean of 8.1 with a standard deviation of 3.31 and median of 6.5), and the number of designed and deployed virtual patient cases (mean of 15.5 with a standard deviation of 8.1 and median of 17.5). The majority of subjects (7 out of 10) had prior experience designing VP cases with vpSim. The remaining three had experience creating

⁶Kynectiv/DecisionSIM. <https://www.kynectiv.com/>

VP cases using a combination of Twine,⁷ Unity,⁸ DecisionSim, or Laerdal SimManager/SimDesigner⁹ case authoring tools.

Two of the subjects reported that they did not have enough knowledge of sepsis management and treatment guidelines and did not answer the questions that were specifically related to sepsis learning objectives (Appendix C, Table 38).

Table 14: Subjects' Clinical Specialties

Subject's Specialty	Count
Pharmacy	3
Instructional Design and Emergency Medicine	1
Anesthesiology	1
Thoracic Surgery and Patient Safety	1
Palliative Care	1
Transplant and Nephrology	1
Clinical Simulation	1
Oral and Maxillofacial Pathology	1

6.2.2 RQ2 Study Design

Each subject was presented with a 15-minute demonstration of ModelPatient, followed by a 5-minute overview of the sepsis model (Chapter 5, Section 5.3.2). The demonstration included all aspects of the VP case authoring system and the VP player (Chapter 4), as well as an opportunity to experiment with the model's states and observe the changes in probability distributions.

⁷Twine, an open-source tool for telling interactive, nonlinear stories. <http://twinery.org/>

⁸Unity3D: <https://unity3d.com/>

⁹Laerdal SimManager / SimDesigner: <https://www.laerdal.com/us/products/simulation-training/manage-assess-debrief/simcenter/simmanager/>

After the completion of the demonstration and the overview, each subject was presented with a set of primary learning objectives (Table 15) and was asked to review and interact with the “Managing Sepsis” VP case. The primary learning objectives presented to this study’s subjects were defined by the case’s original author, Dr. Neal Benedict, PharmD.

Table 15: “Managing Sepsis” VP Case Primary Learning Objectives

	Learning Objective
1	Understand steps and guidelines for emergent sepsis treatment
2	Select appropriate empirical drug therapy

After having an opportunity to “prescribe” several drug treatments and observe the feedback in the form of changes in laboratory test and vitals values, the subjects were presented with alternate learning objectives (Table 16) and guided through modifying the “Managing Sepsis” VP case to fit the new objectives.

Table 16: “Managing Sepsis” VP Case Alternate Learning Objectives

	Alternate Learning Objective
1	Understand steps and guidelines for emergent sepsis treatment in a patient with acute kidney injury / renal failure
2	Select appropriate empirical drug therapy for a patient with acute kidney injury / renal failure

These alternate learning objectives were selected for the following clinical and educational reasons (reasoning provided by Dr. Neal Benedict, PharmD, Associate Professor, Pharmacy and Therapeutics, University of Pittsburgh School of Pharmacy,

and Dr. Lorin Grieve, PharmD, Instructor, Pharmacy and Therapeutics, University of Pittsburgh School of Pharmacy):

- **Fluid administration:** In patients with severe renal failure, not yet on dialysis, learners would have to reduce the volume of fluid that is administered. Since renal failure patients are not able to clear large volumes of fluid (due to the kidney injury), learners will need to be cautious with the amount of fluid administered so not to put them in a fluid overload state.
- **Antibiotic dosing:** Nearly all the antibiotics administered in sepsis and septic shock would require dose adjustments based on renal function.
- **Response to therapy:** Learners should closely monitor kidney function and specifically urine output during sepsis and septic shock as markers of therapy responsiveness. Renal failure clouds the learners ability to use this indicator to gauge fluid and/or vasopressor response.

The subjects were asked to ONLY modify states for the nodes and rules relevant to the new learning objectives. For example, in the “Managing Sepsis” VP case, subjects were be asked to only change states of the nodes and rules associated with the renal failure comorbidity (e.g., presence/absence of renal failure, related labs and drugs). The three subjects that were interviewed using Microsoft Skype for Business (with screen sharing) could not directly interact with the case, so instead they guided the interviewer through their desired actions (i.e., “Please set the state of the ‘Renal Failure’ node to ‘yes’.”, or “Create a new rule that is triggered by selecting [X].”).

Once the subjects modified and reviewed their respective cases, they were asked to complete the ModelPatient case authoring evaluation questionnaire (Appendix C, Section C.3). Subjects provided responses to demographic, clinical experience, prior VP case authoring experience, and ModelPatient authoring system rating questions using a paper questionnaire form. Subjects also provided verbal responses to the open-ended interview questions regarding the authoring system’s usability, strengths, weaknesses, and ability to modify the provided VP case to meet alternate learning

objectives.

The mean duration of the RQ2 study was 1 hour 17 minutes (median of 1 hour 18 minutes, standard deviation of 23 minutes) [H](#)).

Screen captures of subjects' interactions with ModelPatient were recorded using Open Broadcaster Software (OBS) Studio. Audio was also captured as part of this experiment.

Audio recording failed in two of the interviews due to a technical issue with the laptop used for the experiment. One subject asked for the audio not to be recorded.

Audio recordings from the experiments were transcribed using Temi - an automated online audio transcription service.¹⁰

Each experiment took approximately 1 hour and 20 minutes to complete.

6.2.3 Analysis and Findings

Responses to the two Likert scale questions from the ModelPatient VP case authoring questionnaire ([Appendix C](#), [Section C.3](#)) were evaluated for internal consistency and reliability using Cronbach's alpha.

Responses to interview questions were separated into atomic statements and analyzed by seven independent raters using the affinity diagram method [[134](#)] (Described in [Section 6.1.3](#), Subsection [6.1.3.1](#)).

Overall Cronbach's alpha for questions regarding the ModelPatient's authoring system usability ([Table 37](#)) is 0.753, and 0.954 for questions regarding the alternate learning objectives ([Table 38](#)). Even with the small sample size of responses and a small number of questions, both values indicate a high level of internal consistency for their respective questions [[133](#)].

Table 17: Summary Statistics — Authoring System Difficulty Rating

¹⁰Temi - an automated online audio transcription service. <https://www.temi.com/>

Question	Mean	St.Dev	Median	n
How difficult or easy was it to update the model's states?	4.30	1.25	5	10
How difficult or easy was it to observe the effects of changes that you made to the model's states?	4.50	0.97	5	10
Overall, how difficult or easy was it to understand the use of models in a VP case?	4.40	0.70	4	10
How difficult or easy was it to modify the rules?	4.00	0.94	4	10
How difficult or easy was it to observe the effect(s) of changes to the rules	3.80	1.32	4	10
Overall, how difficult or easy was it to use the VP case authoring system?	3.60	0.52	4	10

Question	Mean	St.Dev	Median	n
How well did the case reflect the alternate learning objectives?	4.22	0.83	4	8
How accurately did the case depict the clinical scenario associated with the chosen topic?	4.37	1.06	5	8
How well did the case introduce / reinforce relevant clinical skills?	4.22	0.97	4	8
How realistic was the case?	4.25	1.16	5	8

Table 18: Summary Statistics - How Well Does the Modified Case Reflect Alternate Learning Objectives

In response to “*Compared to other VP authoring systems that you have used in the past, how easy or difficult was it for you to modify this VP case?*” question, most subjects found modifying a VP case in ModelPatient as having “*about the same level of difficulty*” (4 subjects out of 10) or being “*somewhat easier*” (4 subjects out of 10) than case authoring tool(s) that they previously used (Question 4 on the ModelPatient case authoring evaluation questionnaire, Appendix C, Section C.3).

Most subjects’ response ratings to the questions related to ModelPatient’s authoring system usability (Appendix C, Section C.3, Table 37) indicated that when modifying an existing ModelPatient case it was easy to update the model’s states (mean = 4.3, median = 5), easy to observe the effects of changes made to the model’s states on the VP case (mean = 4.5, median = 5), somewhat easy to understand the use of models in VP cases (mean = 4.4, median = 4), and somewhat easy to modify the rules (mean = 4, median = 4). Subjects also indicated that it was neither easy nor difficult to observe the effects of changes to the rules (mean = 3.8, median = 4) and neither easy nor difficult to use the ModelPatient VP authoring system (mean = 3.6, median = 4).

Subjects with clinical expertise reported that the modified case reflected the alternate learning objectives reasonably well (mean rating = 4.22, median = 4), accurately depicted the clinical scenario associated with sepsis (mean rating = 4.37, median = 5), reinforced relevant clinical skills (mean rating = 4.22, median = 4), and was realistic (mean rating = 4.25, median = 4).

After reviewing the sepsis BN model and modifying the case to reflect alternate learning objectives described in Section 6.2.2, authors commented that given the existing model, the ModelPatient VP case could be modified to address learning objectives such as worsening sepsis (8/10 subjects), treatment of sepsis with comorbidities present in the model, such as congestive heart failure (CHF), hypertension, or acute renal disease (7/10 subjects), and time-to-treatment (5/10 subjects).

Most subjects commented that simply modifying the model’s states does not make the resulting VP case address alternate learning objectives. In order to address the

alternate learning objectives described in Section 6.2.2, case authors also suggested that they would at the very least need to modify patient demographics and prior medical history (10/10 subjects) and rules (8/10 subjects). As one subject commented, “[...] *the demographics should be completely different. If you’re trying to create an older patient, pretty much any of these extra conditions are not going to be seen in a 23 year old college student. You’re not going to have congestive heart failure and hypertension, or renal, liver failure, and respiratory failure.*” Another subject indicated that “*for the learning objective of renal failure, [we] would need to add dialysis needs, complications, anemia, phosphate levels, and selection of dialysis*”.

In response to “*Describe the best/strongest features of this VP system’s case authoring*” question, seven subjects commented that they liked the system’s flexibility — “*almost any part of the VP case, including the underlying model, can be augmented with additional information, metrics, and rules to make the case more robust*”. Five subjects indicated that they liked the “*ability to relatively quickly generate realistic VP cases from a library of [BN] models.*” Five subjects suggested that very robust rule authoring system is also one of the strongest features of ModelPatient’s authoring system. Four subjects commented that the real strength of the system is “*that the case is generated from real data and has additional labs and comorbidities that are representative of the patient population*”.

When analyzing interview responses to the “*Describe features of this VP system’s case authoring that you struggled with the most?*,” two out of seven raters reported that they could not find common themes. Only two common themes emerged from the analysis by and agreement among the remaining five raters.

Five subjects commented that rules were cumbersome to generate — “*too many clicks, easy to get lost when the case has multiple rules.*” One case authoring expert suggested that “[...] *you should be able to generate some rules from the data and treat the rest of the rules as time-based event triggers,*” a comment that is addressed in more details in Chapter 7, Section 7.4.

Three subjects stated that there is a chance that models (and the corresponding cases) would get outdated as treatment guidelines and the subsequent clinical outcomes change. One subject provided the following example. *“The data changes when the treatment guidelines change. For example, I work with hepatitis, and the treatments changed drastically in the last 5 years. With new treatments the mortality rates are much lower. So if I had a model based on 5-year-old data, it wouldn’t reflect current treatment practices and outcomes.”* In February of 2018 the guidelines for sepsis treatment have undergone a number of changes [144, 145, 146], meaning that the BN model created for the “Managing Sepsis” case used in this experiment will no longer be reflective of the patients who are treated using newer guidelines.

Two interview responses that DID NOT emerge as common themes still resonated with findings from RQ1 (Chapter 6, Section 6.1.3). One response touched on the shortcomings and advantages of vpSim and ModelPatient and suggested that *“in order to give authors more flexibility and to offer learners a richer learning experience, [I] should consider combining features of a [narrative-branching] system with features of a [BN] model-driven system.”*

Another response addressed a shortcoming of the authoring system that emerged as a common theme of poor formative feedback mechanism in ModelPatient in RQ1 responses (Chapter 6, Section 6.1.3). One of the subjects suggested allowing authors to trigger feedback based on *“a certain number of mistakes — learner has to review summary or complete a tutorial before returning to the case. Even when learner is correct, [authors should be able to] reinforce why [the learner was] correct.”* Another subject mentioned that in order to provide better formative feedback, the authoring system must provide for *“easier ways to allow learners to be wrong.”*

7.0 DISCUSSION AND FUTURE WORK

In this work I investigated the feasibility of designing a virtual patient case authoring and presentation software system that uses Bayesian Network (BN) models learned from EMR data to generate a framework for VP cases. This research experimentally compared a “Managing Sepsis” VP case based on a BN model to one created using a more traditional narrative-branched VP system (RQ1). Furthermore, this work investigated the extent to which ModelPatient facilitated modifying an existing VP case by allowing case authors to manipulate the underlying model in such a way that the VP case would meet alternate learning objectives (RQ2).

7.1 DISCUSSION

7.1.1 Reflections on Research Question 1

Research Question 1: To what extent is it possible to create a virtual patient case using a VP system based on a Bayesian network model that is comparable to a case created using a branched-narrative VP system?

RQ1 study compared subjects’ ratings of two VP cases - a branched-narrative VP case presented in vpSim and a BN-model-based VP case presented in ModelPatient. The two cases were compared on the following criteria:

1. **Q1:** How well did the learning experience reflect the stated learning objectives?
2. **Q2:** How accurately did the case depict the chosen topic?

3. **Q3:** How well did the case represent clinical variations associated with the chosen topic?
4. **Q4:** How well did the case introduce / reinforce relevant clinical skills?
5. **Q5:** How well did the case introduce / reinforce behaviors and attitudes relevant to sepsis treatment?
6. **Q6:** How well did the case provide formative feedback for choices / outcomes?
7. **Q7:** How realistic was the case?
8. **Q8:** Based on the presentation method of the virtual patient case, how engaged were you?
9. **Q9:** Based on the presentation method of the virtual patient case, how effective do you think this case is in teaching the subject matter to novice clinical practitioners?

Analysis of RQ1 responses discussed in Chapter 6, Section 6.1.3 indicates that a case created using a branched-narrative VP system outperformed a BN model-based system on reflecting the stated learning objectives (Q1), introducing / reinforcing clinical skills (Q4), introducing / reinforcing attitudes and behaviors relevant to sepsis treatment (Q5), providing formative feedback for choices / outcomes (Q6), and being more effective in teaching the subject matter to novice clinical practitioners (Q9).

A BN model-based case was rated significantly higher than the branched-narrative version on representing clinical variations associated with sepsis (Q3), on realism (Q7) and on engagement (Q8).

The subjects' indication that the vpSim version of the "Sepsis" case was better at "**reflecting the stated learning objectives (Q1)**" is likely due to the fact that the vpSim version of the case provided a guided learning experience through feedback, ability to review previous choices and with just-in-time learning modules, while the ModelPatient version of the case allowed learners to make choices without any direction or guidance. Even though comparing the two instructional modalities (open-ended vs. guided) wasn't an aim of this dissertation, this finding raises new research questions regarding the effect of VP case modality on learning outcomes. In their

2006 work Kirschner, et al. make an argument for the superiority of guided instruction and provide empirical evidence that unguided or minimally guided instructional approaches are “less effective and less efficient than instructional approaches that place a strong emphasis on guidance of the student learning process [147].” The authors further argue that unguided or minimally guided instructional approaches only work when “learners have sufficiently high prior knowledge to provide ‘internal’ guidance [147],” a sentiment reflected in many of this study’s subjects’ comments.

Subjects’ preference for vpSim’s ability to reflect learning objectives could be further explained by some of the differences in user interfaces and VP case design choices between the two systems and the two versions of the “Managing Sepsis” case. For example, one subject commented regarding the ModelPatient case that *“if ordering labs is not a part of the learning objectives, the system shouldn’t have an option to ‘Add Lab’,”* indicating that some of ModelPatient’s UI elements distracted learners from the stated learning objectives. Another subject reported that *“[ModelPatient] is much more suitable for advanced learners [...], but not for novice clinical practitioners. [The case presented in vpSim] was a nice guided experience with feedback and just in time learning opportunities, which is what you want to see in a case for novice clinical practitioners.”* Several subjects felt that the open-ended nature of the ModelPatient’s case was actually detrimental to learners, stating that *“having an open-ended case like that would not work well for medical students.”* Last, but not least, some subjects expressed that *“[...] having the ability to test learners’ knowledge throughout the simulation is very important to improve learning outcomes.”*

vpSim’s ability to embed quizzes and just-in-time learning modules can provide an explanation as to why the subjects rated vpSim’s version of “Managing Sepsis” as being better at **introducing / reinforcing clinical skills, attitudes, and behaviors relevant to sepsis treatment (Q4, Q5)** than its ModelPatient counterpart. This finding is supported by a number of studies showing that debriefing and just-in-time learning opportunities in clinical simulations lead to improved learning out-

comes [148, 149, 150, 24]. Several subjects expressed their preference for the assessment and just-in-time learning aspects of the vpSim case over the ModelPatient version, reporting that *“asking students questions to make sure that they understand the subject matter and aren’t just guessing is really important in medical education.”*

“Providing formative feedback for choices / outcomes (Q6)” was another criterion in which the vpSim’s VP case was rated significantly higher. The subjects’ preference of vpSim’s feedback mechanism and of how it was implemented in the vpSim’s version of “Managing Sepsis” emerged as one of the common themes (as identified by independent raters). While this theme was already discussed in Chapter 6, it is worth mentioning that many of the subjects reported that ModelPatient needs *“some type of [...] feedback that explains why the probabilities and the values change the way they do.”* Furthermore, while a number of subjects liked that ModelPatient presented feedback in the form of changing vital signs and laboratory test values, they felt that *“getting clinical notes and vitals and labs updates in the EHR view was a really good form of real-time feedback, [but] this type of feedback may not be appropriate for a novice learner.”* Subjects also indicated that in vpSim’s “Managing Sepsis” an incorrect choice would lead the learner down a “wrong” path with additional choices that would allow them to remediate poor decisions. In ModelPatient, when learners made decisions based on distractors (choices that did not exist in the underlying model but were added by the case author), there was no feedback at all (no vital signs or laboratory test values changed). *“When I added hydrocortisone, there was no feedback or new [changes] which [would indicate] that [this choice] would not have done any additional benefit or harm.”*

vpSim’s case’s higher rating in **“being more effective in teaching the subject matter to novice clinical practitioners (Q9)”** can be explained by a combination of multiple factors, including formative feedback, just-in-time learning opportunities, and built-in assessments that are present in vpSim’s version of “Managing Sepsis”, but not in the corresponding ModelPatient’s version. Some subjects felt that “see-

ing changes in labs and vitals is great for advanced users, but novices would have difficulty interpreting the data and selecting appropriate treatment.” Subjects also reported that ModelPatient’s inability to allow learners to change incorrect decisions (e.g., cancel an inappropriate medication order) was detrimental to novices — *“ability to remediate poor choices is especially important for novices — [ability to change orders] would give them an opportunity to change ‘bad’ decisions.”* ModelPatient’s overwhelming number of decision choices and outcomes may have been another contributing factor as to why subjects felt that vpSim was more appropriate to novice learners. As one subject commented, *“when we treat sepsis, we generally only care about 2 - 3 labs, not 15.”* The “too many choices” argument is well-supported by a number of marketing studies that suggest that an excessive number of choices may lead to “a decrease in the motivation to choose, to commit to a choice, or to make any choice at all [151, 152].” Furthermore, in vpSim the temporal order of decisions is enforced through the narrative and through the sequence of presented decision points. The open-ended nature of ModelPatient makes it difficult to enforce the temporal order of decisions and to provide feedback for treatments that were ordered / administered in an incorrect order. As one subject reported, *“I shouldn’t be able to select an antibiotic before hydration in this case — all treatment options are presented at the same time which might be confusing for the learner.”* Another subject suggested that *“enforcing some kind of temporal treatment order either through feedback or through not showing certain options would be very helpful to novices.”*

While there were no explicit responses explaining why the subjects thought that ModelPatient VP case “**represented clinical variations associated with sepsis (Q3)**” better than its vpSim counterpart, this finding could be attributed to the fact that ModelPatient DOES NOT always present learners with the most likely outcome based on the probability distributions in the underlying BN model. ModelPatient case authoring system allows authors to set an option where the learners’ choices result in either “The most statistically likely outcome” or an “Outcome randomly selected from

the model” (Chapter 4, Figure 28). Setting the “Outcome randomly selected from the model” option results in outcomes to be randomly selected from the connected child nodes’ states. In other words, just like not every patient responds to a medication in the same way, in ModelPatient the VP can respond differently to the same medication during different attempts to complete the case.

Another criterion with a significant difference in ratings was the realism of each version of the “Managing Sepsis” VP case (“**How realistic was the case? (Q7)**”). On vpSim’s side, subjects reported that their perception of the case’s realism was negatively affected by the sparsity of decision options. *“I would have liked to see more choices to make the case more challenging, to drive actual critical decision-making rather than [answering] multiple choice questions.”* Another subject echoed this sentiment by stating that *“more decision possibilities would have been nice [and] would make the case a bit more complex and realistic.”*

Lastly, subjects indicated that **based on the presentation method of the virtual patient case (Q8)**, they were more engaged with the VP presented in ModelPatient. Several subjects felt that the vpSim version of the case was not interactive or immersive enough, stating that *“I wish the case was a bit more interactive - more dialogue / Q&A with the patient”* and that *“I did not feel like I engaged with the [virtual] patient..”* Furthermore, subjects’ engagement with the ModelPatient version of the “Managing Sepsis” case could be attributed to an unintentional bias favoring the model-driven VP. This bias is further discussed in Section 7.3 (Limitations). Finally, ModelPatient’s relative novelty might have played a role in the subjects’ preferences — there are numerous studies that describe the effect of novelty on learning and engagement [153, 154, 155]

On ModelPatient’s side, subjects reported that *“the case did a good job reflecting the outcomes of certain treatment choices — that realism would be very useful to teach about conditions rarely seen during medical students rotations.”* ModelPatient’s EMR-like UI had also contributed to a higher perception of realism. Subjects commented

that ModelPatient has a “*much higher degree of realism than vpSim — I liked working in a system that kind of looked like an EMR*” and that “*it was more realistic because I was reviewing a patient’s chart like in a real clinical scenario.*”

These findings suggest that there may be a need for a system that combines actions, responses, and clinical variations in disease states and treatment outcomes made possible by the use of computational models with the narrative, formative feedback, and just-in-time learning typically present in branched-narrative VPs. In other words, a hybrid system that combines the narrative and feedback of vpSim with the realism and clinical variations made possible by ModelPatient.

7.1.2 Reflections on Research Question 2

Research Question 2: To what extent does the proposed system facilitate modifying an existing VP by allowing case authors to change the model’s states in such a way that the VP case meets alternate learning objectives?

One of the original aims of developing ModelPatient was to reduce the case authoring burden by automating the virtual patients’ “behaviors” through the use of BN models. Another aim was provide case authors with an easier way to modify existing BN-model-based cases to address new learning objectives. The study conducted with VP case authoring experts (Chapter 6, Section 6.2.2) showed that developing VP cases based on BN models is still a time-consuming proposition. Almost unanimously the case authoring experts indicated that it is not possible to modify an existing BN-model-based VP case by simply updating the model’s states and the relevant rules. In order for the modified case to meet alternate learning objectives, case authors would also need to update case narrative, VP’s prior medical history and comorbidities.

Creating a new case from scratch would likely reveal problems and obstacles not considered in this study. For the study described in Chapter 6, I provided case authors not only with a working and validated model of sepsis, but also with the initial case content. In a real-world scenario, authors would have to work with modeling experts

to develop computational models to represent relevant disease-states and courses of treatment, which in and of itself is a time-consuming and complex activity that requires specialized expertise.

Moreover, as one of the common themes discovered in the RQ2 study (Section 6.2.3) indicated, a model could represent a possibly outdated state of clinical knowledge (the state of how that clinical knowledge was represented in the data at the time the model was created). As clinical guidelines change, the models would need to be continuously updated and the associated VP cases would have to be revised.

Another important consideration of using BN models to drive virtual patient cases is the models' potential computational complexity. As inference in BN models is NP-hard [156, 157], it may cause the modelers to simplify the models and remove features that exist (and are potentially important) in real-world scenarios.

Last, but not least, it may be challenging for clinicians and health sciences students to understand probabilistic inference. One of the common themes that emerged from RQ1 surveys (Chapter 6, Section 6.1.3) is that the subjects had a difficult time understanding that the the "Managing Sepsis" did not represent a causal relationship and that the learner's decisions/actions reflected the probability distribution that was already in the data. In other words, if the learner prescribed a medication and the system responded with an outcome that given this particular decision, the probability of the virtual patient's death was 55%, it meant that in the data a high percentage of the patients who were diagnosed with sepsis and received this medication did not survive. As one case authoring expert pointed out, *"I lack expertise in creating models and I think this will be an issue for many other [case] authors. [It would also be critical to] understand how the model maps to clinical practice."*

7.2 CASE AUTHORING LESSONS LEARNED: FEASIBILITY CRITERIA / GUIDELINES FOR CREATING VP CASES BASED ON BAYESIAN NETWORK MODELS

The challenges I experienced in creating a BN model representative of the entire DKA treatment scenario (Chapter 5, Section 5.4.2) indicate that at least in its current iteration a VP system based on BN models is more suitable to present VP cases that do not have a temporal component. In other words, BN model-based cases appear particularly suited for VP scenarios that teach about treatment of acute or emergent conditions (conditions that do not require longitudinal treatment) and are more likely to be comparable with branching-narrative VPs.

Table 19 shows a list compiled from the top 25 medical problems commonly encountered by the University of Pittsburgh 3rd and 4th-year medical students during clinical rotations¹, from the list of common medical problems encountered by the Henry Ford Health System residents [1], and from the list of 25 most commonly recorded ICD-10 codes [158].

I worked with three University of Pittsburgh clinical faculty members to review this list and determine which of these conditions can be presented to learners without a temporal treatment component. It is important to note that even though the conditions listed in the left column of Table 19 are chronic and require longitudinal treatment, some of these conditions can have acute presentations. For example, hypertension is a chronic condition that requires long-term treatment; modeling hypertension treatment using a BN model would likely encounter the same difficulties as we experienced attempting to model DKA treatment. However, if the learning objectives included teaching about hypertensive crisis (an acute condition), or about risk assessment for hypertension [159], it would be feasible to design a VP case using the

¹The University of Pittsburgh data was extracted from the Learning Log, an application used by the University of Pittsburgh medical students to log their de-identified patient encounters. This data was extracted by the University of Pittsburgh School of Medicine Laboratory for Educational Technologies (<https://let.pitt.edu>).

methods presented in this work. Colon cancer would be another use case where the entire progression and management of cancer would be challenging to model. However, if one wished to develop a VP case where the learning objectives are to diagnose cancer or estimate patient survival, it would be feasible to develop BN models representative of such case scenarios [160, 161].

Table 19: Medical problems commonly encountered by the University of Pittsburgh 3rd and 4th-year medical students during clinical rotations + common medical problems encountered by the Henry Ford Health System residents [1].

A: CANNOT be presented to learners without a temporal component	B: CAN be presented to learners without a temporal component
Hypertension	Hypertensive crisis
Depression	Anxiety (Panic Attacks)
Arthritis	Asthma
Obesity	Angina
Gastroesophageal reflux	Stroke
Diabetes mellitus	COPD
Hyperlipidemia	Pneumonia
Hernia	Upper respiratory infection
Bipolar disorder	UTI / dysuria / hematuria
ADHD	Otitis media / otitis externa
Sleep apnea	Migrane
Dementia	Diverticulitis
Cancer - colon	Cholangitis
Cancer - lung	Pulmonary embolism
Seizure disorder	Acute renal failure
Pancreatitis	Endocarditis

Pulmonary hypertension	Meningitis
Chronic renal failure	Allergic rhinitis
HIV	Acute bronchitis
Tuberculosis	Acute laryngopharyngitis
Hyperlipidemia	Acute maxillary sinusitis
Back pain	
Osteoarthritis	
Hypothyroidism	
Fibromyalgia / myositis	
Coronary atherosclerosis	

More specifically, case authors who wish to develop VP cases based on the methods presented in this work should consider the following design constraints on data, BN models, and case contents:

- **Acute clinical condition.** As discussed above, the methods presented in this dissertation work better when used for authoring and presenting VP cases for acute clinical conditions. In other words, when selecting a clinical topic for a new virtual patient case based on a BN model, authors should select clinical conditions that do not require time-dependent or longitudinal treatments.
- **Limited treatment options.** Datasets for clinical conditions with limited treatment options will have fewer variables and will likely result in simpler and better performing models.
- **The data is representative of the clinical condition AND of the learning objectives.** While ModelPatient makes it possible to augment a BN model-based VP case with distractors (Chapter 4, Section 4.6.4) and rules (Chapter 4, Section 4.6.6), a number of subjects from the RQ2 study found rule creation to be cumbersome and time-consuming. For a simpler case authoring experience, case

authors need to ensure that the dataset used for training a BN model contains all variables required for the desired case scenario and for the stated learning objectives.

- **The data is reasonably recent.** One of the common themes that emerged from the RQ2 study was that there is a chance that models (and the corresponding cases) could get outdated as treatment guidelines and the subsequent clinical outcomes change. When selecting a patient cohort dataset for a model-driven VP, authors should ensure that patient inclusion criteria take into account the dates of recent changes / updates to treatment guidelines and protocols. In other words, authors should avoid including patients that have been treated according to outdated treatment guidelines and protocols.
- **Treatment homogeneity within patient cohort.** It would be well advised to limit VP authoring to clinical conditions where all patients receive approximately the same treatment(s) in the same sequence. As discussed in Chapter 5, Section 5.4, one of the challenges with developing a BN model representative of DKA was that for some of the patients represented in the data there were as few as 3 time slices of treatments and responses, and one patient had as many as 57 time slices. To complicate things further, the treatments in each time slice were different. Presence of these very different treatments in each time slice makes it challenging to model disease states using static BN models.
- **Limited number of treatment options that directly influence the outcome.** If authors choose to use causal BN models trained on EMR datasets (Chapter 3, Section 3.4), they would need to select a clinical condition and a corresponding model where the number of variables that directly influence the treatment outcome is reasonably small. Figure 54 shows a partial BN model of sepsis trained with the PC algorithm.

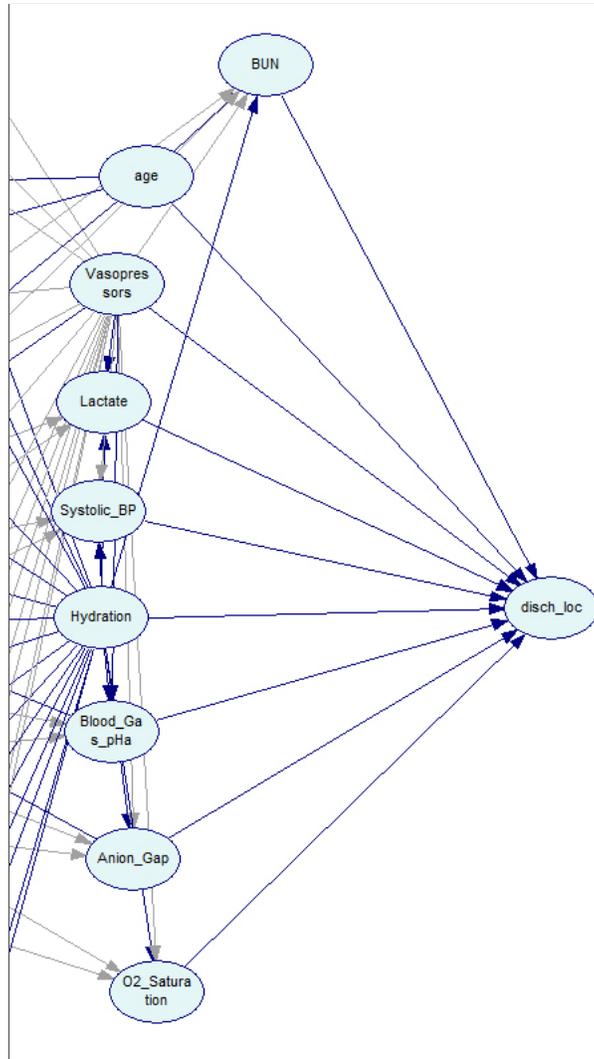


Figure 54: Partial BN model of sepsis trained with PC algorithm (Chapter 3, Section 3.4).

Note that the outcome node has nine parents — BUN, age, vasopressors, lactate, systolic blood pressure, hydration, blood gas pHa, anion gap, and oxygen saturation. The PC algorithm determined that the outcome (patient is discharged from the hospital, patient remains in hospital care, patient dies) is dependent on these variables based on results of a chi-square test. Such a high number of parents results in large conditional probability tables. Moreover, when such models are trained on datasets with a small number of records, the conditional probability distributions can be estimated as uniform (Figure 55). Subsequently, such models tend to perform poorly both in terms of computational performance (it takes a long time to recalculate probability distributions based on learner selections) and in terms of predictive accuracy.

age	☐						
Anion_Gap	☐						
BUN	☐						
Blood_Gas_pHa	☐						
Lactate	☐						
Systolic_BP	☐						
O2_Saturation	☐						
Vasopressors	☐						
Hydration	Epinephrine						
	D10	D5_Half_Sa...	D5W	Half_Saline	Lactated_Ri...	missing	Normal_Saline
▶ Deceased	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Hospital	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Managed_care	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Self_care	0.25	0.25	0.25	0.25	0.25	0.25	0.25

Figure 55: Partial conditional probability table (CPT) of the “discharge location” node from a BN model of sepsis trained with PC algorithm (Chapter 3, Section 3.4).

7.3 LIMITATIONS

In interpreting findings of this work, it is important to acknowledge its limitations.

Three out of seven independent raters noted that while reviewing text responses to RQ1 surveys they noticed that some comments praised ModelPatient as being data-driven and model-driven without providing an explanation as to why this was an advantage to the learning experience. It seems that by telling the subjects in advance that ModelPatient uses computational models to control VP scenarios an unintentional bias favoring the model-driven VP may have been introduced.

Another limitation was related to the topic of the VP case used in the RQ1 experiment. I selected the “Managing Sepsis” VP case for this study because it had been validated and used in teaching and in academic research [118, 119, 120]. When designing the experiment I did not take into account the fact that sepsis presents in only 6% of adult hospitalizations [162] and that not all clinicians regularly encounter sepsis cases during hospital rotations. 5 out of 33 RQ1 subjects reported that they do not often encounter sepsis patients in their clinical practice and are not as familiar

with sepsis treatment guidelines as, for example, an ICU attending would be. While working on either the vpSim or ModelPatient version of the “Managing Sepsis” VP case, these subjects used mobile devices to look up sepsis treatment guidelines. Future studies may limit subject inclusion criteria to physicians, clinical pharmacists, and nurses who routinely encounter sepsis patients as part of their clinical practice.

Yet another limitation of the RQ1 study is that the VP case presented in vpSim and ModelPatient was evaluated by clinicians and educators, not by the actual intended learners (post-graduate trainees with working knowledge of human physiology and clinical medicine - learners who have completed their medical, nursing, or pharmacy degrees but have not completed clinical training). While I can make some inferences about how comparable the educational qualities of the two versions of “Managing Sepsis” are based on the expert subjects’ responses, the study would benefit from an additional experiment with the intended target audience (e.g., 4th year pharmacy students) and an evaluation of the degree of comparability between the two cases with a pre- and a post-test.

Several subjects commented on the importance of being able to retract choices and explore different decision pathways as part of the learning experience. vpSim has the built-in ability to revisit previously-made choices, make different decision, and explore the “what-if” scenarios associated with different decision paths. Implementing similar functionality in ModelPatient by allowing learners to cancel treatment orders would have made the two systems and the two more comparable.

Last, but not least, the RQ1 experiment was conducted in varying environments, including physicians’ offices, hospital cafeterias, coffee shops, a library, and common areas of academic buildings. Most of these locations were noisy and offered a lot of distractions. Furthermore, several times the experiments was interrupted by phone calls or pages. Three of the experiments were interrupted by patient emergencies and had to be split between several sessions. Given the nature of the study participants’ work, I believe that this limitation was unavoidable.

One major limitation of the RQ2 study was its length and complexity. The average time required to complete the RQ2 study was 1 hours and 17 minutes, a time requirement that placed a burden on already busy health sciences educators and clinicians. By the end of the experiment (especially during the interview section), some of the subjects rushed with their responses and made it clear that they have other matters to attend to. One way in which I could have reduced the time requirement for the case authoring experts was to provide them with a pre-recorded system video tutorial, as well as give them access to an online version of the ModelPatient case prior to the interview. This way at least some of the subjects could have reviewed the authoring system and the VP case's contents before the experiment, thus reducing the time investment by approximately 20-30 minutes.

Another limitation was the interview format. For the RQ2 experiment I used a semi-structured interview to elicit responses about strengths and weaknesses of authoring model-driven VP cases. While a more open-ended format of a semi-structured interview allowed me to discover themes and ideas that I did not previously consider, it also generated responses and conversations that were not relevant to the interview questions. In the case of the question "*Describe features of this VP system's case authoring that you struggled with the most?*," most subjects responded by expressing their frustration with authoring text-based and screen-based simulations in general and I found it difficult to direct them to the question at hand. As a result, two out of seven raters who reviewed the text responses reported that they could not find common themes in statements related to this question.

Inconsistencies in the experiment environment and frustrations with technical issues may have affected some of the responses as well. Just like with the RQ1 experiment, the RQ2 experiment was conducted in a wide variety of physical spaces and environments. The experiment suffered from patient-related interruptions as well as from problems associated with teleconferencing. Of the three experiments that were conducted using Microsoft Skype for Business, two experienced several network-

related interruptions, causing subjects to become frustrated and trying to wrap up the experiment as quickly as possible.

Last, but not least, some of the RQ2 subjects were my research collaborators on other projects and studies (Chapter 6, Section 6.1.1). A prior professional relationship with research subjects may have biased some of their responses in favor of ModelPatient.

7.4 FUTURE WORK

7.4.1 Alternate Modeling Approaches

One of the biggest criticisms that ModelPatient received in both studies described in Chapter 6 was the inability to design VP cases that correctly represent and enforce an event sequence for learners. For example, one of the case authoring experts subjects commented on the “Managing Sepsis” case that *“learners should not be able to prescribe a vasopressor before prescribing hydration, or should at least receive corrective feedback if they do.”*

While designing the ModelPatient version of the “Managing DKA” VP case where the disease state and the treatment are presented as 12 sequential time slices in vpSim, we were unable to create a comparable case largely due to the difficulties in modeling long time-sensitive event sequences using BN networks and due to ModelPatient’s inability to enforce the sequences of case presentation and learners’ decisions (Chapter 5, Section 5.4).

In this section, I will discuss several possible approaches that could be explored in future iterations of ModelPatient to address this problem.

7.4.1.1 Representing Temporal VP Cases with Multiple Chained BN Models

One possible approach was inspired by Barclay, et. al. 2013 work titled “Refining a

Bayesian Network using a Chain Event Graph” [163] and Thwaites, et. al. 2010 work titled “Causal Analysis with Chain Event Graphs” [164].

In such approach, each time slice would be treated as a discrete BN model. For example, let us consider a model with the initial parameters of “Age” and “Renal failure”, and observable outcomes (response variables) of “Anion Gap” and “Glucose” (Figure 56).

If a VP case author wanted to add an event that required a learner to order “Saline Bolus,” a corresponding BN model for that event would look like Figure 57, modeling the probabilities of states of “Anion Gap” and “Glucose” given “Saline Bolus.”

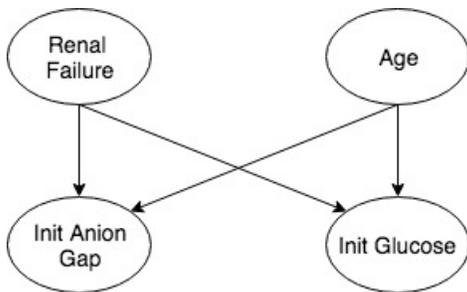


Figure 56: Initial parameters and response variables — first “sub-model” in a chain of BN network models.

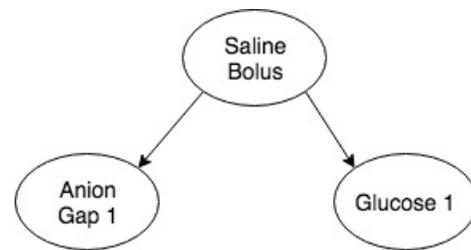


Figure 57: Second “sub-model” in a chain of BN network models representing the “Saline Bolus” event.

In real world, these two events would not be independent, as the initial values of anion gap and glucose lab results in a patient would affect the subsequent values of those same variables after a treatment (or treatments) had been administered. One approach to address this issue would be to chain the two models together, where the initial states of “Anion Gap 1” and “Glucose 1” would be set to the outcome probabilities of the corresponding states of the previous model in the chain (Figure 58).

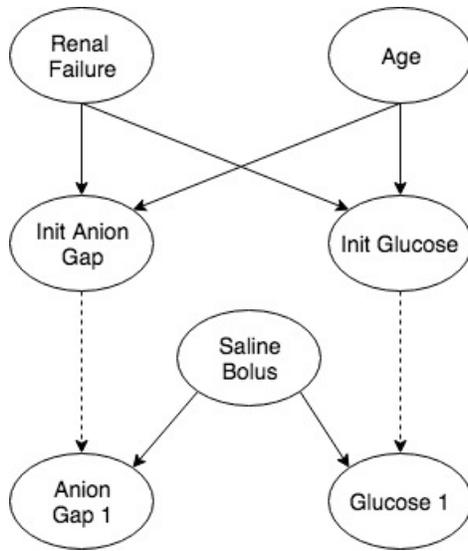


Figure 58: Two “sub-model” chained together. Dashed arrow lines represent setting the initial states of “Anion Gap 1” and “Glucose 1” to the outcome probabilities of the corresponding states of the previous model in the chain.

With this approach, a six-timeslice chain of BN models representing part of the DKA treatment scenario might look like Figure 59.

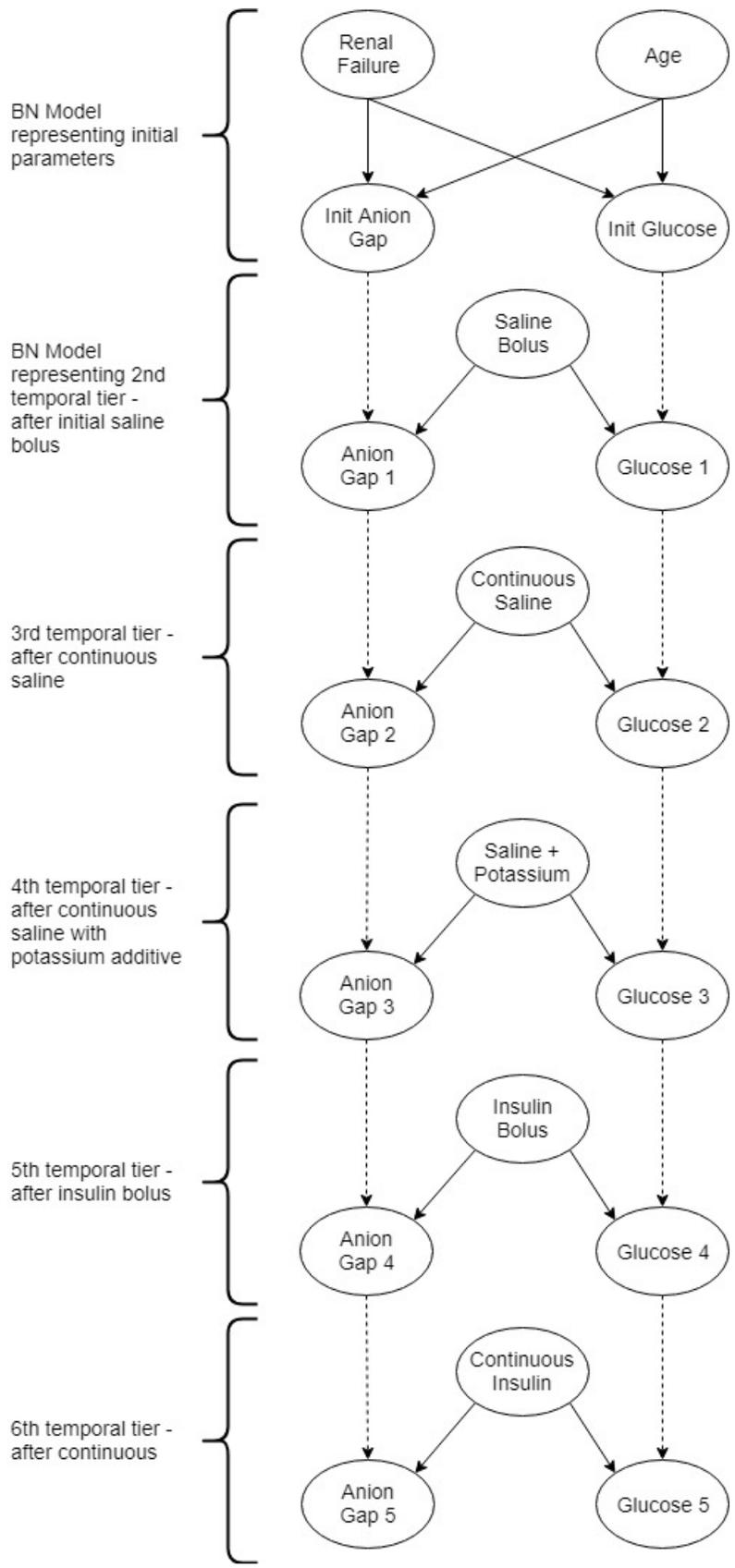


Figure 59: Multiple time slices represented as discrete BN models.

An additional advantage of this approach is that it would allow ModelPatient to support VP cases where multiple clinical conditions need to be represented as part of the case scenario. A example scenario described by one of the RQ2 subjects presented a patient admitted to a hospital with a perforated peptic ulcer. In some cases it is possible for patients with severe gastric reflux to develop peptic aspiration pneumonia [165]. Creation the proposed scenario in ModelPatient would require two models — one for the disease state and treatment of a peptic ulcer patient, and a secondary model representative of an aspiration pneumonia patient. In this case, the case author could link the two models together where the final state probabilities from the “Peptic ulcer” model would become the initial probabilities for the corresponding node states in the “Aspiration pneumonia” model.

Another possible advantage of this approach is that parameters for smaller models (models with fewer features) that comprise a larger scenario could be learned from smaller data sets, which may be easier to obtain than large and feature rich data from EMRs.

A major disadvantage of using a chain of smaller BN models to control a VP scenario is that authoring such a case would create more burden on the author, requiring the author not only to connect multiple BN models into a chain, but also to validate such a case and to ensure that the case narrative is consistent with how the probabilities change from one model to the next.

7.4.1.2 Representing Temporal VP Cases with Hidden Markov Models Another possible approach to modeling disease progression over time and the associated sequence of treatment steps is using Hidden Markov models (HMMs). HMMs are ubiquitously used for modeling time series data and are widely used in speech recognition systems [166, 167], computational biology [168, 169], and disease state models [170, 127].

A patient’s progression through DKA could be modeled as a discrete-time HMM, where each discrete time step (latent state) is associated with vital signs and labo-

ratory test results that were affected by prescribed treatments. In such a model we could account for patient heterogeneity by incorporating variables associated with patient history (age, gender) and comorbidities (renal failure) when calculating state transition probabilities.

The hidden Markov process illustrated in Figure 60 comprises of an initial patient state (S init) representing initial vital signs and laboratory test results at ICU admission, five latent states (S1-S5) representing a patient state after a treatment, and three outcome (absorbing) states (death, hospitalization, discharge). Directional arcs represent transitional probabilities for improvement from state to state (left-to-right arcs), worsening from state to state (right-to-left arcs), or remaining in the same state (loop arc). Details of this approach, including transitional probability matrices, parameter estimation approaches, and model inferences are described in detail in Ghosh, et. al. 2016 work [171] and Petersen, et. al. 2018 work [127].

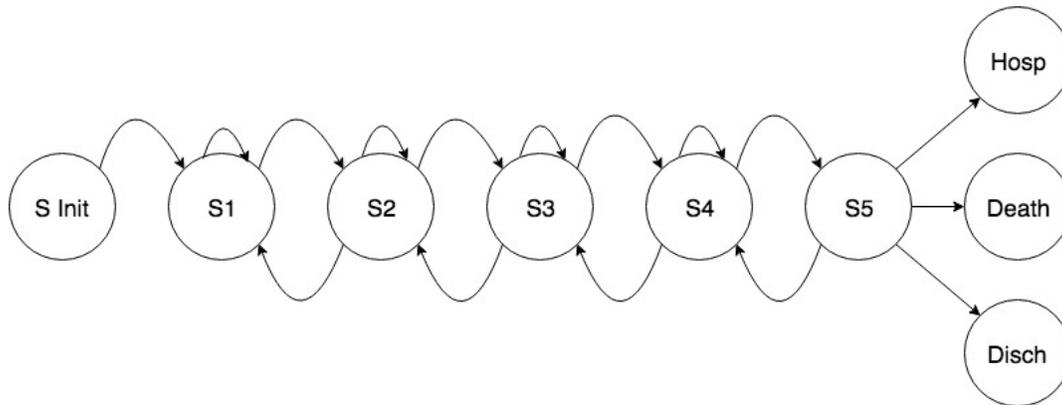


Figure 60: DKA treatment progression modeled with HMM. “S Init” represents initial vital signs and laboratory test results at ICU admission. S1 - S5 are latent states representing a patients’ state given a treatment. Death, Hospitalization, and Discharge are absorbing states.

7.4.1.3 Representing Temporal VP Cases with Interval Temporal Bayesian Networks One of the shortcomings of typical BN models is that they only represent three temporal relationships — precedes, follows, equals [2]. Furthermore, as BN models are essentially probabilistic finite state-machines, their state-space (and subsequently the size of conditional probability tables) grows exponentially in size as the number of parallel events increases [172]. As we saw with the DKA model (Chap-

ter 5, Section 5.4.2), the large state-space quickly became unmanageable for modeling a complex activity such as management of diabetic ketoacidosis.

In their 2013, work Zhang, et. al. proposed to augment BN models with interval algebra representation of 13 atomic time period relations between two events [173]. This approach allowed the authors to represent an interval relationship between two events as a union of these atomic relations relations (Table 20) [2].

Table 20: “Allen’s 13 atomic temporal relations to represent temporal relationships between two events X and Y” [2, 3]

Relation	Symbol	Inverse
Y before X	b	bi
Y meets X	m	mi
Y overlaps X	o	oi
Y starts X	s	si
Y during X	d	di
Y finishes X	f	fi
Y equals X	eq	eq

For example, if we wanted to model (and enforce) the notion that a normal saline hydration should be administered BEFORE vasopressor, Allen’s representation of this relationship would be YbX, where Y is the normal saline hydration and X is the vasopressor. Combined with a traditional graphical BN model, we can model relationships that are not normally represented by the connecting edges (Figure 61). Furthermore, by modeling the “hydration precedes vasopressor” relationship using an interval relation we effectively reduce the size of the conditional probabilities tables and make a complex model more computationally feasible [2].

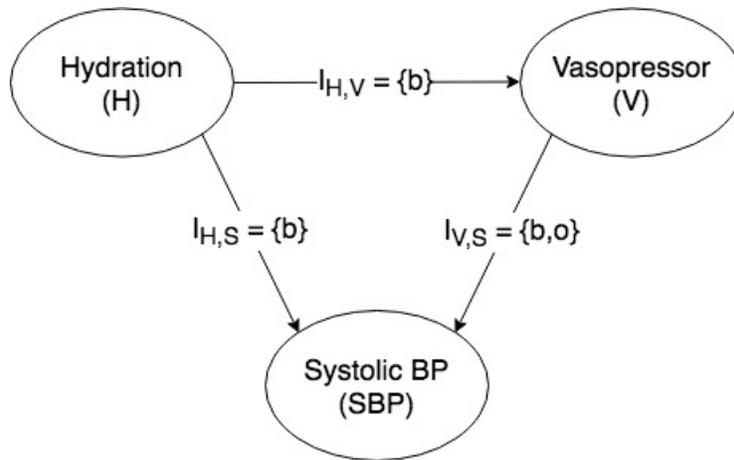


Figure 61: A graphical representation of temporal dependencies among temporal entities Hydration (H), Vasopressor (V), and Systolic BP (SBP). In this example, the temporal relationships are “H occurs before V,” “H occurs before SBP,” and “V occurs before SBP or overlaps SBP.”

7.4.2 Automatically Generating Association Rules

One of the common themes that emerged from the ModelPatient authoring system evaluation is that rules are too cumbersome to generate. One of the subjects suggested that “*rules should be extracted from the data, not manually created by authors.*” There is quite a large body of work dedicated to extracting association rules from medical health records data [174, 175, 176, 177].

In their 2017 work, Lakshmi and Vadivu describe a method of automated extraction of association rules from medical health records using multi-criteria decision analysis [178]. The authors use a combination of Apriori, Close, RElim, Eclat and FP-Growth algorithms and select generated rules with the highest confidence values. Since the example rules generated in this work are similar to the ones used in ModelPatient (in the form of “*if X then Y*”), I would like to try this approach with the DKA patient cohort dataset to see if it is possible to generate rules similar to those manually created by an expert case author.

7.4.3 Automatically Generating Feedback

ModelPatient’s lack of formative feedback emerged as a common theme in both RQ1 and RQ2 studies. More specifically, multiple subjects remarked on difficulties in understanding how and why outcome probabilities changed in response to their decisions. One subject commented: *“After I made the first decision and the VP told me that the outcome of death is 42%. Compared to what? Did I make the patient better or worse?”* Several subjects found the outcome probabilities *“confusing despite the explanation at the beginning of the case”* and *“difficult to understand and interpret.”* In their 2002 review of explanation methods for Bayesian networks, Lacave and Díez classify explanations of Bayesian belief networks into three categories [179]:

1. Explanation of evidence which “consists of determining which values of the unobserved variables justify the available evidence.”
2. Explanation of the model — a verbal or graphical description of the model to aid in understanding its structure and purpose [180]
3. Explanation of reasoning — an explanation of the process by which the model produced (or did not produce) the results/outcomes [180]

Creating a qualitative explanation (i.e. feedback) for every decision and every outcome would place a tremendous burden on case author and invalidate any time-saving advantage that model-based VPs may provide. Automating formative feedback is something I would like to explore further in my future work. Since ModelPatient already provides a graphical explanation of the model (Figure 62), it is the explanation of evidence and the explanation of reasoning that need to be presented as feedback to the learners. There is a small body of work that suggests quantitative approaches to generating explanations for Bayesian network models, Madigan et al. work on graphical display of the weight of evidence [181], qualitative propagation algorithm proposed by Druzdzel and Henrion [182], and enhanced qualitative networks approach [183], but none of those approaches provide a definitive guideline for generating explanations from the model itself or from the underlying data.

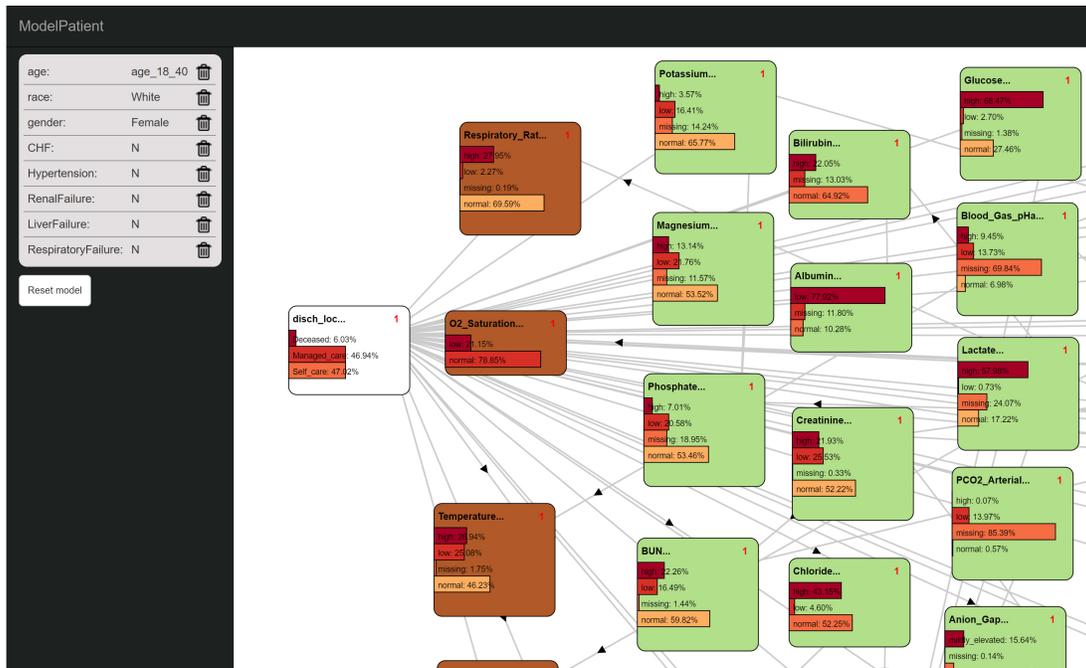


Figure 62: BN model representation (graphical explanation) in ModelPatient case authoring tool.

7.4.4 Future Studies

As mentioned in the description of RQ1 study limitations, the subjects who participated in evaluating the “Managing Sepsis” VP case content were practicing clinicians and educators. To assess if the way the case is presented in vpSim vs. ModelPatient has an effect on learning outcome, would like to conduct a study with more novice practitioners that have not completed their clinical training.

The results of the RQ1 study also suggested that the difference in learning outcomes may be affected by the differences in VP case presentation modalities. While there is evidence presented in the literature that guided instructional design leads to better learning outcomes than open-ended instructional design [147], I would like to conduct a study to evaluate the effect of VP case modality on learning outcomes in novice clinical practitioners.

Last, but not least, I would like to design a study to evaluate the effect of VP case presentation fidelity on learner engagement. In other words, I would like to evaluate relationship between the degree to which the system's UI resembles an EMR and the learner engagement with a VP case.

APPENDIX A

PATIENT COHORT DATA DESCRIPTION

A.1 ORIGINAL PATIENT COHORT DESCRIPTION

The patient cohort data used in this study was originally obtained for a research study titled “Comparison of Normal Saline and Plasma-Lyte for Fluid Resuscitation in Septic Shock,” IRB protocol #PRO14100151. The purpose of this study was “to compare a retrospective cohort of ICU patients receiving normal saline for fluid resuscitation in septic shock with a retrospective cohort of patients receiving Plasma-Lyte for fluid resuscitation in septic shock.” This data set contains de-identified retrospective medical chart data of 71,202 patients admitted to the emergency rooms (ER) or hospitals of the UPMC hospital system between 2001 and 2014. The data was retrieved from UPMC MARS data archive by an honest broker.

A.1.1 Inclusion criteria

- age > 18
- Plasma-Lyte or Normal Saline administration [bolus > 500ml or rate>200cc/hr]
- Administration antibiotics (piperacillin/tazobactam, ampicillin/sulbactam, vancomycin, cefepime, ceftriaxone, ciprofloxacin, levofloxacin, aztreonam, meropenem, imipenem, ertapenem, doripenem, ticarcillin, ceftazidime, azithromycin, gentamicin, tobramycin,

amikacin, fluconazole, caspofungin, daptomycin, linezolid, tigecycline)

- Mean Arterial Pressure (MAP) < 70 mmHg OR systolic BP < 90 mmHg
- Signs of tissue hypoperfusion (output < 0.5 mL/kg for one hour OR serum lactate level > 2 mmol/L)
- Requiring vasopressors (norepinephrine, dopamine, vasopressin, epinephrine, phenylephrine)

A.1.2 Additional variables

- Blood gases
- Common chemistries
- Special chemistries
- Hematology
- Ventilator settings
- Vital signs
- Hemodynamics
- Intake and output
- Hospital length of stay
- ICU length of stay
- Mortality

A.2 SEPSIS PATIENT COHORT

Since all sepsis patients receive hydration treatment, I extracted a sub-cohort of patients who were admitted to UPMC hospitals with with a primary, secondary, or tertiary diagnosis of “sepsis,” ICD-9 codes 995.90, 995.91, and 995.92. The final sepsis cohort dataset contained 3,991 hospital admission records. The final dataset, after

removing partial records and records with missing values, contained 3,974 hospital admission records.

Normal value ranges for discretization of continuous features were obtained from Medscape [184] and Lab Tests Online [185] websites and further refined through interviews with clinical experts to better meet requirements of the associated VP case scenarios.

A.3 DATASET FEATURES

Table 21: Sepsis Dataset — Demographic Information

Feature	Description	Discrete Categories	Counts
Age	Patient’s age at the time of hospital admission	18 - 40 41 - 65 >65	278 1,434 2,262
Gender	Patient’s gender	m - male f - female	2,066 1,908
Ethnicity	Patient’s ethnicity (race)	African American Other White	552 238 3,184

Table 22: Sepsis Dataset — Comorbidities

Comorbidity	ICD-9 Code(s)	Discrete Categories	Counts
Congestive Heart Failure (CHF)	428.0	Y	1,016
		N	2,958
Hypertension	401.0, 401.1, 401.9	Y	1,867
		N	2,107
Liver Failure	573.8	Y	13
		N	3,961
Renal Failure	584, 586	Y	86
		N	3,886
Respiratory Failure	518.81	Y	1,604
		N	2,370

Table 23: Sepsis Dataset — Administered (Bedside) Medications

Medication	Counts
Normal Saline	1,993
Half Saline	43
D5 Half Saline	375
D5W	473
D10	247
Lactated Ringers	437
Bolus - Normal Saline	3,502
Bolus - Plasmalyte A	471
Norepinephrine	1,715
Epinephrine	1,768
Phenylephrine	1,236
Vasopressin	831
Piperacillin-Tazobactam	2,264
Vancomycin	2,984
Meropenem	659
Metronidazole	1,235
Levofloxacin	76

Table 24: Sepsis Dataset — Laboratory Tests

Lab. Test Name	Discrete Category	Counts
Anion Gap	< 13: normal	2,508
	≥ 13 and < 15: mildly elevated	648
	≥ 15 and < 17: moderately elevated	350
	≥ 17: severely elevated	465
Blood Gas pHa	< 7.38: low	1,004
	≥ 7.38 and ≤ 7.42: normal	437
	> 7.42: high	556
BUN	< 7: low	260
	≥ 7 and ≤ 20: normal	1,638
	> 20: high	2,065
Chloride	< 96: low	207
	≥ 96 and ≤ 106: normal	2,071
	> 106: high	1,696
CO2	< 23: low	207
	≥ 23 and ≤ 28: normal	2,071
	> 28: high	1,696
Glucose	< 70: low	132
	≥ 70 and ≤ 100: normal	967
	> 100: high	2,839
Potassium	< 3.5: low	672
	≥ 3.5 and ≤ 5: normal	2,654
	> 5: high	217
Sodium	< 135: low	756
	≥ 135 and ≤ 145: normal	2,845
	> 145: high	370

Table 25: Sepsis Dataset — Vital Signs

Vital Sign	Discrete Category	Counts
Systolic Blood Pressure	< 120: normal	2176
	≥ 120 and ≤ 145: prehypertension	1212
	> 145: hypertension	581
Diastolic Blood Pressure	< 80: 'normal'	3398
	≥ 80 and ≤ 89: prehypertension	360
	> 89: hypertension	208
O ₂ Saturation	< 95: low	857
	≥ 95 and ≤ 100: normal	3117
Pulse	< 60: low	132
	≥ 60 and ≤ 100: normal	2687
	> 100: high	1147
Respiratory Rate	< 12: low	85
	≥ 12 and ≤: normal	2375
	> high	1497

Table 26: Sepsis Dataset — Outcomes

Outcome	Discrete Category	Counts
Discharge Location	Deceased	919
	Managed care	2071
	Self care	984

A.4 DIABETIC KETOACIDOSIS PATIENT COHORT DATA

Table 27: DKA Dataset — Demographic Information

Feature	Description	Discrete Categories	Counts
Age	Patient's age at the time of hospital admission	18 - 40 41 - 65 >65	186 986 1,395
Gender	Patient's gender	m - male f - female	1,232 1,335
Ethnicity	Patient's ethnicity (race)	African American Other White	406 158 2,003

Table 28: DKA Dataset — Comorbidities

Comorbidity	ICD-9 Code(s)	Discrete Categories	Counts
Congestive Heart Failure (CHF)	428.0	Y N	526 2,041
Hypertension	401.0, 401.1, 401.9	Y N	1,282 1,285
Lactic Acidosis	276.2	Y N	2,428 139
Psychiatric Co-morbidities	291.3, 292.1, 292.89, 293.81, 293.84, 295.30, 295.32, 295.34, 295.80, 295.90, 295.92, 296.00, 296.20, 296.23, 296.30, 296.31, 296.33, 296.50, 296.53, 296.54, 296.64, 296.90, 298.9, 300.00, 300.02, 300.3, 300.4, 300.9, 301.4, 304.40, 306.9, 309.24, 309.28, 309.81, 310.9, 311, 368.16, 797, 799.29, 969.3, 969.8, 969.9, E854.2, E939.3, E939.7, E939.8, E950.3, E980.3, V11.0, V11.8, V17.0, V62.89, V70.2	Y N	717 1850
Renal Failure	584, 586	Y N	75 2492

Table 29: DKA Dataset — Administered (Bedside) Medications

Medication	Counts
Normal Saline	1,192
Half Saline	40
D5 Half Saline	250
D5W	455
D10	143
Lactated Ringers	295
Regular Insulin	321
Sliding-Scale Insulin	1,185
Bolus - Normal Saline	2,315
Bolus - Plasmalyte A	241

Table 30: DKA Dataset — Laboratory Tests — Initial Values at the Time of Hospital Admission

Lab. Test Name	Discrete Category	Counts
Anion Gap	< 13 : normal ≥ 13 and < 15 : mildly elevated ≥ 15 and < 17 : moderately elevated ≥ 17 : severely elevated	734 427 356 1,050
Blood Gas pHa	< 7.38 : low ≥ 7.38 and ≤ 7.42 : normal > 7.42 : high	993 220 217
BUN	< 7 : low ≥ 7 and ≤ 20 : normal > 20 : high	79 845 1,643
Chloride	< 96 : low ≥ 96 and ≤ 106 : normal > 106 : high	448 1,413 706
CO2	< 23 : low ≥ 23 and ≤ 28 : normal > 28 : high	1367 854 346
Glucose	< 70 : low ≥ 70 and ≤ 100 : normal > 100 : high	117 524 1926
Potassium	< 3.5 : low ≥ 3.5 and ≤ 5 : normal > 5 : high	343 1,685 539
Sodium	< 135 : low ≥ 135 and ≤ 145 : normal > 145 : high	528 1,773

Table 31: DKA Dataset — Initial Vital Signs at the Time of Hospital Admission

Vital Sign	Discrete Category	Counts
Systolic Blood Pressure	< 120: normal	1,257
	≥ 120 and ≤ 145 : prehypertension	828
	> 145: hypertension	482
Diastolic Blood Pressure	< 80: 'normal'	2,116
	≥ 80 and ≤ 89 : prehypertension	298
	> 89: hypertension	153
O ₂ Saturation	< 95: low	561
	≥ 95 and ≤ 100 : normal	2,004
Pulse	< 60: low	109
	≥ 60 and ≤ 100 : normal	1,362
	> 100: high	1,095
Respiratory Rate	< 12: low	29
	≥ 12 and ≤ 16 : normal	1,753
	> 16: high	785

Table 32: DKA Dataset — Outcomes

Outcome	Discrete Category	Counts
Discharge Location	Deceased	594
	Managed care	1,441
	Self care	532

APPENDIX B

NORMAL VALUE RANGES FOR VITALS AND LABORATORY TESTS

Table 33: Normal Value Ranges for Vitals

Vital	Unit(s)	Ranges
Systolic BP	mmHg	Normal: < 120 Prehypertension: 120 - 140 Hypertension: \geq 140
Diastolic BP	mmHg	Normal: < 80 Prehypertension: 80 - 90 Hypertension: \geq 90
O2 Saturation	mmHg	Low: < 95 Normal: 95-100
Pulse	beats per minute	Low: < 60 Normal: 60 - 100 High: > 100
Respiratory Rate	breaths per minute	Low: < 12 Normal: 12 - 20 High: > 20

Table 34: Normal Value Ranges for Laboratory Tests

Laboratory Test	Unit(s)	Ranges
Anion Gap	mEq/L	Normal: < 13

		Mildly Elevated: 13 - 15 Moderately Elevated: 15 - 17 Severely Elevated: ≥ 17
Blood Gas pHa		Low: < 7.38 Normal: 7.38 - 7.42 High: > 7.42
Blood Urea Nitrogen (BUN)	mg/dL	Low: < 7 Normal: 7-20 High: > 20
Chloride	mEq/L	Low: < 96 Normal: 96 - 106 High: > 106
CO2	mEq/L	Low: < 23 Normal: 23-28 High: > 28
Glucose	mEq/L	Low: < 70 Normal: 70-100 High: > 100
Potassium	mEq/L	Low: < 3.5 Normal: 3.5 - 5.0 High: > 5.0
Sodium	mEq/L	Low: < 135 Normal: 135 - 145 High: > 145
Creatinine	mg/dL	Low: < 0.6 Normal: 0.6 - 1.2 High: > 1.2
Magnesium	mg/dL	Low: < 1.7 Normal: 1.7 - 2.2 High: > 2.2
Lactate	mmol/L	Low: < 0.5 Normal: 0.5 - 1.0 High: > 1
Albumin	g/dL	Low: < 3.5 Normal: 3.5 - 5.5 High: > 5.5
Bilirubin	mg/dL	Low: < 0.1 Normal: 0.1 - 1.2 High: > 5.5
Phosphate	mg/dL	Low: < 2.5 Normal: 2.5 - 4.5 High: > 4.5
Partial pressure of oxygen (PaO2)	mg/dL	Low: < 75 Normal: 75 - 100

		High: > 100
Partial pressure of carbon dioxide (PaCO ₂)	mg/dL	Low: < 38 Normal: 38 - 42 High: > 42

APPENDIX C

MODEL PATIENT CASE PRESENTATION AND AUTHORIZING EVALUATION QUESTIONNAIRES

C.1 QUESTIONNAIRE: MODEL PATIENT CASE PRESENTATION EVALUATION

Questions in the following questionnaire are based on and derived from features and uses of high-fidelity medical simulations that lead to effective learning proposed by Isenberg, et al. [24], Posel's dissertation work on validating criterion-referenced guidelines for virtual patient case authoring [25], and Bateman et. al exploratory study on virtual patient design [26].

Note: The questionnaire distributed to subjects will contain a note stating that both sepsis and DKA VP cases have been designed for post-graduate trainees with working knowledge of human physiology and clinical medicine — learners who have completed their medical, nursing, or pharmacy degrees but have not completed clinical training.

Question 1: Virtual patient case evaluation. Please rate your responses to the following questions on the scale of 1 to 5, with 1 being "Poor" and 5 being "Excellent."

Table 35: Virtual Patient Case Evaluation Questionnaire 1

		<i>Poor</i>				<i>Excellent</i>
How well did the learning experience reflect the stated learning objectives?	N/A	1	2	3	4	5
How accurately did the case depict the chosen topic?	N/A	1	2	3	4	5
How well did the case represent clinical variations associated with the chosen topic?	N/A	1	2	3	4	5
How well did the case introduce / reinforce relevant clinical skills?	N/A	1	2	3	4	5
How well did the case introduce / reinforce behaviors and attitudes relevant to [DKA or sepsis] treatment?	N/A	1	2	3	4	5
How well did the case provide formative feedback for choices / outcomes?	N/A	1	2	3	4	5
How realistic was the case?	N/A	1	2	3	4	5
Based on the presentation method of the virtual patient case, how engaged were you?	N/A	1	2	3	4	5
Based on the presentation method of the virtual patient case, how effective do you think this case is in teaching the subject matter to novice clinical practitioners (learners who have completed graduate medical or pharmacy program but have not completed clinical training)?	N/A	1	2	3	4	5

Question 2: Describe the best/strongest features of this case? Why did you identify a particular feature as best/strongest?

Question 3: What features of this case did you struggle with the most? How would you improve them?

Question 4: Describe the best/strongest features of this VP system? Why did you identify a particular feature as best/strongest?

Question 5: Describe the features of this VP system that could use improvement? How would you improve them?

**C.2 QUESTIONNAIRE: CASE CONTENT AND LEARNING OBJECTIVES
COMPARISON BETWEEN THE “MANAGING SEPSIS” VP CASE
PRESENTED IN VPSIM AND MODELPATIENT**

Question 1: Please rate your responses to the following questions on the scale of 1 to 5, with 1 being “Not comparable at all” and 5 being “Extremely comparable.”

Table 36: Case Content and Learning Objectives Comparison Questionnaire

	<i>Not comparable at all</i>					<i>Extremely comparable</i>
	1	2	3	4	5	
In terms of educational content, how comparable is the case presented in ModelPatient to the case presented in vpSim?	1	2	3	4	5	
in terms of meeting the learning objectives, how comparable is the case presented in ModelPatient to the case presented in vpSim?	1	2	3	4	5	

C.3 QUESTIONNAIRE: MODEL PATIENT CASE AUTHORIZING EVALUATION

Question 1: How long have you been designing virtual patient cases?

Question 2: How many complete virtual patient cases have you designed?

Question 3: What virtual patient (VP) authoring tools do you have most experience with?

- vpSim
- DecisionSim
- OpenLabyrinth
- Twine
- Other

Question 4: Compared to other VP authoring systems that you have used in the past, how easy or difficult was it for you to modify this VP case?

- Much more difficult
- Somewhat more difficult
- About the same level of difficulty
- Somewhat easier
- Much easier

Question 5: Based on your experience with modifying a VP case, please rate your responses to the following questions on the scale of 1 to 5, with 1 being “Difficult” and 5 being “Easy.”

Table 37: Virtual Patient Case Authoring Evaluation

		<i>Difficult</i>			<i>Easy</i>	
How difficult or easy was it to update the model's states?	N/A	1	2	3	4	5

How difficult or easy was it to observe the effects of changes that you made to the model's states?	N/A	1	2	3	4	5
Overall, how difficult or easy was it to understand the use of models in a VP case?	N/A	1	2	3	4	5
How difficult or easy was it to modify the rules?	N/A	1	2	3	4	5
How difficult or easy was it to observe the effect(s) of changes to the rules?	N/A	1	2	3	4	5
Overall, how difficult or easy was it to understand the use of models in a VP case?	N/A	1	2	3	4	5
Overall, how difficult or easy was it to use the VP case authoring system?	N/A	1	2	3	4	5

Question 6: Based on the changes you made in the VP case, please rate your responses to the following questions on the scale of 1 to 5, with 1 being "Poor" and 5 being "Excellent."

Table 38: Virtual Patient Alternate Objective(s) Case Evaluation

		<i>Poor</i>				<i>Excellent</i>
How well did the case reflect the alternate learning objectives?	N/A	1	2	3	4	5
How accurately did the case depict the clinical scenario associated with the new changes?	N/A	1	2	3	4	5
How well did the case represent clinical variations associated with the chosen topic?	N/A	1	2	3	4	5

How well did the case introduce / reinforce relevant clinical skills?	N/A	1	2	3	4	5
How realistic was the case?	N/A	1	2	3	4	5

Question 7: Based on the VP case that you have modified in this VP system, can you list additional learning objectives that could be taught with the case?

Question 8: What other parts of the VP case would you modify in order for the case to reasonably reflect the learning objectives stated in your previous answer?

Question 9: Describe the best/strongest features of this VP system's case authoring. Why did you identify a particular feature as best/strongest?

Question 10: Describe features of this VP system's case authoring that you struggled with the most? How would you improve them?

APPENDIX D

SAMPLE VP CASE COMPLETION REPORTS

D.1 SAMPLE VPSIM COMPLETION REPORT

Table 39: Sample vpSim Completion Report

Step/Action	Date	Time	Metrics
Entered node Welcome to vpSim 'Sepsis''	4/15/2018	3:45:44 AM	
Clicked "Continue" button on Welcome to vpSim 'Sepsis''	4/15/2018	3:45:53 AM	
Entered node "PATIENT CASE: Hannah Foore"	4/15/2018	3:45:53 AM	
Clicked "Continue" button on "PATIENT CASE: Hannah Foore"	4/15/2018	3:47:45 AM	
Entered node "Definitions"	4/15/2018	3:47:45 AM	
Selected choice "Hannah meets criteria for sepsis"	4/15/2018	3:48:04 AM	Score: 99
Revisited node "PATIENT CASE: Hannah Foore"	4/15/2018	3:48:14 AM	
Revisited node "Definitions"	4/15/2018	3:48:24 AM	
Selected choice "Hannah only meets criteria for single organ dysfunction"	4/15/2018	3:48:27 AM	
Entered node "Definitions"	4/15/2018	3:48:54 AM	
Clicked "Continue" button on "Definitions"	4/15/2018	3:49:00 AM	
Entered node "Early antibiotics"	4/15/2018	3:49:00 AM	
Selected choice "Within the 1st hour from severe sepsis or septic shock diagnosis"	4/15/2018	3:49:12 AM	

Clicked "Continue" button on "Early antibiotics"	4/15/2018	3:49:15 AM	
Entered node "Shock"	4/15/2018	3:49:15 AM	
Selected branch "Administer 0.9% NaCl as an IV bolus of 500ml q 30 min as needed to maintain stable CVP "	4/15/2018	3:50:12 AM	
Entered node "Crystalloids vs Colloids (1)"	4/15/2018	3:50:12 AM	
Selected choice "The absence of any clear benefit following the administration of colloid solutions compared to crystalloid solutions, together with the expense associated with colloid solutions, supports a high-grade recommendation for the use of crystalloid solutions in the initial resuscitation of patients with severe sepsis and septic shock."	4/15/2018	3:50:35 AM	
Clicked "Continue" button on "Crystalloids vs Colloids (1)"	4/15/2018	3:50:39 AM	
Entered node "Fluid bolus (1)"	4/15/2018	3:50:39 AM	
Selected branch "Start IV vasopressor"	4/15/2018	3:51:16 AM	
Entered node "Vasopressor options (1)"	4/15/2018	3:51:16 AM	
Selected choice "Start Phenylephrine 30 mcg/min"	4/15/2018	3:51:40 AM	Score: 97
Selected choice "Start Norepinephrine 0.1 mcg/kg/min"	4/15/2018	3:51:43 AM	
Clicked "Continue" button on "Vasopressor options (1)"	4/15/2018	3:51:46 AM	
Entered node "Vasopressin"	4/15/2018	3:51:46 AM	
Selected choice "Low-dose vasopressin can be added to norepinephrine to raise MAP or decrease norepinephrine dosage. It is not recommended as the single initial vasopressor for treatment of severe sepsis or septic shock"	4/15/2018	3:52:26 AM	
Clicked "Continue" button on "Vasopressin"	4/15/2018	3:52:30 AM	
Entered node "Tachycardia mechanisms (1)"	4/15/2018	3:52:30 AM	
Selected choice "Beta-1 agonism"	4/15/2018	3:52:42 AM	
Clicked "Continue" button on "Tachycardia mechanisms (1)"	4/15/2018	3:52:46 AM	
Entered node "Vasopressor tachycardia (1)"	4/15/2018	3:52:46 AM	
Selected branch "Discontinue norepinephrine and start phenylephrine 30 mcg/min IV infusion"	4/15/2018	3:53:36 AM	
Entered node "CVP and MAP rise (1,1)"	4/15/2018	3:53:36 AM	

Clicked "Continue" button on "CVP and MAP rise (1,1)"	4/15/2018	3:53:50 AM	
Entered node "Adrenal Insufficiency (1,1)"	4/15/2018	3:53:50 AM	
Selected choice "No, 'stim' tests are not recommended for septic shock. Hydrocortisone therapy is not an option either since Hannah has responded to her vasopressor and fluid therapy."	4/15/2018	3:56:49 AM	
Clicked "Continue" button on "Adrenal Insufficiency (1,1)"	4/15/2018	3:56:56 AM	
Entered node "GAME OVER (1,1)"	4/15/2018	3:56:56 AM	

Table 40: Sample ModelPatient Completion Report

Date / Time	Choice
2018-04-13T15:42:51.940Z	Demographic
2018-04-13T15:43:02.693Z	Allergies
2018-04-13T15:43:08.009Z	Demographic
2018-04-13T15:43:10.866Z	Diagnoses
2018-04-13T15:43:37.743Z	Notes
2018-04-13T15:43:41.419Z	Vitals
2018-04-13T15:43:57.055Z	Laboratory Tests
2018-04-13T15:44:07.105Z	Drugs
2018-04-13T15:44:21.303Z	Procedures
2018-04-13T15:44:28.602Z	Radiology
2018-04-13T15:44:40.772Z	Social History
2018-04-13T15:44:43.369Z	Vitals
2018-04-13T15:44:48.925Z	Outcome
2018-04-13T15:44:55.181Z	Diagnoses
2018-04-13T15:45:45.045Z	Drugs
2018-04-13T15:46:27.759Z	Laboratory Tests
2018-04-13T15:47:09.623Z	Vitals
2018-04-13T15:47:14.584Z	Outcome
2018-04-13T15:47:36.586Z	Drugs
2018-04-13T15:47:40.354Z	Notes
2018-04-13T15:48:25.554Z	Notes
2018-04-13T15:48:30.380Z	Outcome
2018-04-15T12:12:35.015Z	Demographic
2018-04-15T12:12:42.430Z	Allergies

2018-04-15T12:13:31.030Z	Diagnoses
2018-04-15T12:13:59.826Z	Laboratory Tests
2018-04-15T12:14:47.022Z	Notes
2018-04-15T12:14:55.043Z	Vitals
2018-04-15T12:15:37.164Z	Diagnoses
2018-04-15T12:16:02.565Z	Notes
2018-04-15T12:16:13.905Z	Diagnoses
2018-04-15T12:16:23.783Z	Drugs
2018-04-15T12:18:41.844Z	Laboratory Tests
2018-04-15T12:19:11.928Z	Vitals
2018-04-15T12:19:21.170Z	Outcome
2018-31-13 11:45:10	Patient meets criteria for sepsis
2018-31-13 11:45:34	Meningitis
2018-31-13 11:45:41	Patient meets criteria for single organ dysfunction
2018-31-13 11:46:1	Vancomycin
2018-31-13 11:46:7	Norepinephrine
2018-31-13 11:46:23	Normal Saline
2018-31-15 8:14:22	Chem 7:
2018-31-15 8:15:42	Meningitis
2018-31-15 8:16:19	Patient meets criteria for single organ dysfunction
2018-31-15 8:16:31	Normal Saline
2018-31-15 8:16:55	Vasopressin
2018-31-15 8:18:29	Vancomycin
2018-31-15 8:18:37	Norepinephrine

APPENDIX E

SUBJECT RECRUITMENT MATERIALS

E.1 RECRUITMENT EMAIL FOR RESEARCH QUESTION 1 EXPERIMENT

I hope this email finds you well. My name is Dmitriy Babichenko and I am working on developing a novel virtual patient simulation system in which the simulation is controlled by a computational model instead of a predefined case scenario. This study is designed to assess the feasibility of the system and to understand how it compares with other systems (such as vpSim) in terms of case presentation and case authoring.

1. To what extent is it possible to create a virtual patient case using a VP system based on a Bayesian Network model that is comparable to a case created using a branched-narrative VP system?
2. To what extent does the proposed system facilitate modifying an existing VP by allowing case authors to change the model's states in such a way that the VP case meets alternate learning objectives?

As part of this study, I would like you to complete 2 cases, one in vpSim and another one in ModelPatient. After completing each case, I will ask you to complete a short online survey. I am also asking for your permission to capture your interactions with the two virtual patient systems via a screen capture recording. Please note that your participation in this study is completely voluntary and that you may stop or withdraw

at any time.

Please let me know if you have any questions or would like any additional information.

E.2 RECRUITMENT EMAIL FOR RESEARCH QUESTION 2 EXPERIMENT

I hope this email finds you well. My name is Dmitriy Babichenko and I am working on developing a novel virtual patient simulation system in which the simulation is controlled by a computational model instead of a predefined case scenario. This study is designed to assess the feasibility of the system and to understand how it compares with other systems (such as vpSim) in terms of case presentation and case authoring.

This study aims to answer the following research questions:

1. To what extent is it possible to create a virtual patient case using a VP system based on a Bayesian Network model that is comparable to a case created using a branched-narrative VP system?
2. To what extent does the proposed system facilitate modifying an existing VP by allowing case authors to change the model's states in such a way that the VP case meets alternate learning objectives?

As part of this study, I would like you to modify one existing virtual patient case using the ModelPatient VP authoring system. The purpose of this experiment is to assess how easy or difficult it would be for an expert VP case author to modify an existing case based on a computational model to address different learning objectives. I am asking for your permission to record the interview (audio recording) in order to accurately transcribe and analyze the responses of all participants in this study. I would also like to capture your interactions with the two virtual patient systems via a screen capture recording. Please know that your name and other identifying information will not be associated with this transcript. If you say anything that could

identify you, we will redact this information from the transcript. Please note that your participation in this study is completely voluntary and that you may stop or withdraw at any time.

Please let me know if you have any questions or would like any additional information.

APPENDIX F

PYTHON SCRIPT USED TO SELECT RESEARCH QUESTION 1 SUBJECTS INTO TWO GROUPS

```
import random

# File subjects.txt is a plain text file that
# contains a list of subjects' names.
# Each name is on a new line
f = open('subjects.txt', 'rb')

# Read subjects' names into a list
subject_list = f.read().split('\n')

# Randomly shuffle the list
random.shuffle(subject_list)

# Select subjects into two groups
group_A = subject_list[:17]
group_B = subject_list[17:]
```

APPENDIX G

RESEARCH QUESTION 1 STUDY LOG

Table 41: Research Question 1 Study Log

ID	First System	Date	Start Time	End Time	Location	Time in vp-Sim	Time in MP	Notes
L01	vpSim	2/18/2018	1:38 PM	2:56 PM	Hospital cafeteria	10:58	12:33	vpSim images did not load during the interview
L02	MP	2/19/2018	7:45 AM	9:10 AM	Hospital cafeteria	07:53	19:02	vpSim images did not load during the interview
L03	vpSim	2/21/2018	7:45 AM	8:23 AM	Academic building common area	10:20		vpSim image issue has been fixed. Subject was very thorough, took notes while going through the case (on paper). Kept asking about the fact that radiology images were mentioned in the case, but you could not review them. Looked up Glasgow coma score on a phone app. Kept returning to previous screens to review information. Did not get right away that the patient has meningitis - was upset with self. Got a call right as he started the MP case and had to leave - will have to reschedule for the second part of the interview
L03	MP	2/21/2018	2:25 PM	2:50 PM	Faculty office		17:34	

L04	MP	2/24/2018	2:54 PM	3:38 PM	Coffee shop	5:38	15:16	Virtual machine (VM Ware) crashed as soon as the subject completed the case. The screencast is corrupted towards the end, but since the subject completed the case and I was able to save the completion reports, I don't think this will be a problem. On another note, the subject met with me after a bike ride (still wearing cycling cloths) and seems tired and distracted.
L05	vpSim	2/26/18	10:15 AM	10:49 AM	Coffee shop	5:54	8:04	Subject seemed rushed, especially with the MP case
L06	MP	3/1/2018	8:35 AM	9:14 AM	Academic building common area	4:54	16:07	
L07	vpSim	3/1/2018	10:10 AM	10:58 AM	University library quiet study area	7:45	15:24	I forgot to clear data from the last session and had to stop in the middle of the interview to save the previous session and clear out old data.
L08	MP	3/5/2018	7:45 AM	8:23 AM	Academic building common area	13:47		Subject seemed to have issues using the laptop's trackpad. Interview was interrupted after subject completed the MP case. The subject asked to meet again at the College of Pharmacy at 2PM
L09	vpSim	3/5/2018	11:30 AM	12:17 PM	Hospital, private office	7:31	8:51	Issues with hospital guest wireless - the connection is too slow to load vpSim. Had to use my phone as a WIFI hotspot. The subject looked up some answers on their phone - might skew completion time a bit
L08	vpSim	3/5/2018	2:00 PM	2:21 PM	Academic building common area	5:40		
L10	vpSim	3/5/2018	6:10 PM	7:20 PM	Hospital cafeteria	20:04	14:03	
L11	MP	3/6/2018	9:30 AM	10:20 AM	Coffee shop	8:09	19:02	
L12	vpSim	3/6/2018	11:05	12:09 PM	Hospital patient waiting area	5:47	15:37	
L13	MP	3/16/2018	9:48 AM	10:39 PM	Hospital lobby	5:29	10:28	Forgot to close Outlook - several emails popped up during the pass through MP - was a bit distracting. The subject does not have a faculty appointment, but they are a preceptor.
L14	vpSim	3/20/2018	8:52 AM	9:30 AM	University library quiet study area	4:43	6:56	With the MP case the subject explored around rather than trying to solve the case

L15	vpSim	3/21/2018	9:03 AM	9:45 AM	Academic building common area	8:33	8:42	
L16	vpSim	3/21/2018	11:00 AM	12:15 PM	Academic building, faculty office	5:37	11:54	
L17	MP	3/25/18	3:25 PM	4:11 PM	Coffee shop	6:30	9:25	
L18	vpSim	3/26/18	1:20 PM	2:06 PM	Academic building common area	6:50	7:12	
L19	MP	3/29/2018	10:45 AM	11:38 AM	Academic building, faculty office	5:46	7:20	
L20	vpSim	3/29/2018	2:00 PM	2:33 PM	Hospital patient waiting area	5:50	6:36	
L21	MP	4/12/2018	1:50 PM	2:51 PM	Hospital, private office	6:00	10:02	Forgot to reset MongoDB before the subject entered MP - took about 30 seconds to remove old data
L22	vpSim	4/13/2018	8:30 AM	9:23 AM	Academic building cafeteria	8:06	11:07	I think I will have to discard this interview - the subject answered 4's for all questions and left open-ended questions blank
L23	MP	4/13/2018	11:45 AM	12:33 PM	Hospital cafeteria	5:06	10:00	Had to start and stop the interview twice - the subject was called in to discuss a patient. Stopped after completing MP case. Came back, completed survey, took another call, came back, finished vpSim case and survey
L24	vpSim	4/15/2018	7:30 AM	8:35	Coffee shop	11:32	10:14	Asked a lot of questions, looked up a lot of treatment guidelines on their phone. Found a spelling error and did not want to continue until I fixed it.
L25	MP	4/18/2018	8:00 AM	9:10 AM	Academic building cafeteria	6:37	15:06	
L26	MP	4/23/2018	10:00 AM	11:15 AM	Hospital, private office	5:47	12:07	
L27	vpSim	4/24/2018	9:00 AM	9:54 AM	Hospital cafeteria	11:13	7:07	Did not complete text responses in vpSim survey
L28	vpSim	4/27/2018	1:00 PM	1:45 PM	Hospital, research offices	10:39	9:13	
L29	MP	5/3/2018	8:50 AM	10:07 AM	Academic building, faculty office	9:35	11:03	Had to stop after subject completed MP case. Met again at 12:30 at the Presby 11th floor cafeteria
L30	vpSim	5/7/2018	8:30 AM	9:40 AM	Coffee shop	9:47	16:13	Started at [hospital building], had issues connecting to the WIFI, moved to coffee shop

L31	MP	5/11/2018	10:45 AM	11:32 AM		7:11	10:06	
L32	vpSim	5/16/2018	1:30 PM	2:30 PM	Medical office building, conference room	34:15	4:37	Spent a lot of time in vpSim, mostly asking questions about simulation. The actual time spent on case is much shorter
L33	MP	5/17/2018	8:50 AM	10:03 AM	Coffee shop	7:19	8:54	

- Mean study duration: 57 minutes
- Median study duration: 53 minutes
- Study duration standard deviation: 14 minutes
- Mean time spent in vpSim: 8 minutes 34 seconds
- Median time spent in vpSim: 7minutes 11 seconds
- Time spent in vpSim standard deviation: 6 minutes 8 seconds
- Mean time spent in ModelPatient: 11 min 30 seconds
- Median time spent in ModelPatient: 10 min 28 seconds
- Time spent in ModelPatient standard deviation: 3 min 49 seconds

APPENDIX H

RESEARCH QUESTION 2 STUDY LOG

Note that the “In-person?” column designates whether the study was conducted in person (‘Y’) or using Skype (‘N’).

Table 42: Research Question 2 Study Log

ID	Date	Start Time	End Time	Location	In person?	Audio
A01	02/09/2018	11:00 AM	11:52 AM	Medical building, private office	Y	Y
A02	02/21/2018	8:45 AM	9:50 AM	Academic building, faculty office	Y	Y
A03	02/26/2018	12:30 PM	1:30 PM	Academic building, faculty office	N	Y
A04	02/27/2018	9:00 AM	10:45 AM	Academic building, conference room	Y	Y
A05	03/05/2018	8:00 AM	9:37 AM	Academic building, public area	Y	Y
A06	03/05/2018	8:45 AM	9:30 AM	Coffee shop	Y	Y
A07	03/05/2018	10:00 AM	11:00 AM	Hospital, research offices	Y	Y
A08	03/16/2018	6:00 AM	7:47 AM	Academic building, faculty office	N	N
A09	03/23/2018	6:30 PM	8:05 PM	Academic building, faculty office	Y	N
A10	04/03/2018	12:00 PM	1:40 PM	Coffee shop	Y	N
A11	04/05/2018	1:30 PM	3:10 PM	Academic building cafeteria	Y	Y
A12	05/11/2018	4:45 PM	6:17 PM	Hospital cafeteria	Y	Y

- Mean study duration: 1 hour 17 minutes
- Median study duration : 1 hour 18 minutes
- Standard deviation: 23 minutes

Note that the interview and the data from subject A05 was not used in this study. The subject did not read the selection criteria listed in the recruitment email and did

meet the required inclusion criteria. The subject authored several VP cases but did not co-author any publications. Instead, the subject was asked to play the role of a "learner". The resulting data is designated with a subject ID of "L08" in the Research Question 1 study (Appendix G).

The data from subject A08 was discarded due to numerous technical issues with conducting the study using Skype.

BIBLIOGRAPHY

- [1] Inpatient Rotations. Henry Ford Health Systems, Medical Education. <https://www.henryford.com/hcp/med-ed/residencies-fellowships/hfh/internal-med/inpatient-rotations>.
- [2] Yongmian Zhang, Yifan Zhang, Eran Swears, Natalia Larios, Ziheng Wang, and Qiang Ji. Modeling temporal interactions with interval temporal bayesian networks for complex activity recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(10):2468–2483, 2013.
- [3] John Allen, Louis Buffardi, and Robert Hays. The relationship of simulator fidelity to task and performance variables. Technical report, DTIC Document, 1991.
- [4] Peter Krogh. Non-destructive imaging: An evolution of rendering technology. Technical report, Adobe Systems Incorporated, 12 2007.
- [5] H. Barrows and R. Tamblyn. *Problem-based Learning: An Approach to Medical Education*. New York: Springer, 1980.
- [6] E. J. Wood. The Problems of Problem-Based Learning. *Biochemical Education*, 22:78–82, 1994.
- [7] D. E. Allen, R. S. Donham, and S. A. Bernhardt. Problem-based learning. *New Directions for Teaching and Learning*, 2011:21–29, 2011.
- [8] J. Murray and A. Summerlee. The impact of problem-based learning in an interdisciplinary first-year program on student learning behaviour. *Canadian Journal of Higher Education*, 37:87–107, 2007.
- [9] Kathleen R. Rosen. The history of medical simulation. *Journal of Critical Care*, 23:157–166, June 2008.
- [10] N. Zary, G. Johnson, J. Boberg, and U. G. Fors. Development, implementation and pilot evaluation of a web-based virtual patient case simulation environment - Web-SP. *BMC Medical Education*, 6, 2006.

- [11] M. A. Smith, R. A. Mohammad, and N. Benedict. Use of virtual patients in an advanced therapeutics pharmacy course to promote active, patient-centered learning. *American Journal of Pharmaceutical Education*, 78, 2014.
- [12] N. Saleh. The value of virtual patients in medical education. *Annals of Behavioral Science and Medical Education*, 16:29–31, 2010.
- [13] Greenberg A.D. Wainhouse Research White Paper: Blended learning technology: navigating the challenges of large scale adoption. Technical report, Wainhouse Research, 2012.
- [14] David Cook, Patricia Erwin, and Marc Triola. Computerized Virtual Patients in Health Professions Education: A Systematic Review and Meta-Analysis. *Academic Medicine*, 85:1589–1602, October 2010.
- [15] David A Cook and Marc M Triola. Virtual patients: a critical literature review and proposed next steps. *Medical Education*, 43(4):303–311, 2009.
- [16] Rachel Ellaway, Terry Poulton, Uno Fors, James B McGee, and Susan Albright. Building a virtual patient commons. *Medical Teacher*, 30(2):170–174, 2008.
- [17] Gary Tabas, Mary Kosytkowski, and Harsha Rao. *Diabetic Ketoacidosis [Online patient case]*. American College of Physicians, Philadelphia, PA, 2012.
- [18] Amy L Alexander, Tad Brunyé, Jason Sidman, and Shawn A Weil. From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in PC-based simulations and games. *DARWARS Training Impact Group*, 5:1–14, 2005.
- [19] Jason B Moats, Thomas J Chermack, and Larry M Dooley. Using scenarios to develop crisis managers: Applications of scenario planning and scenario-based training. *Advances in Developing Human Resources*, 10(3):397–424, 2008.
- [20] Maurice Hendrix, Ali Al-Sherbaz, and Bloom Victoria. Game based cyber security training: are serious games suitable for cyber security training? *International Journal of Serious Games*, 3(1):53–61, 2016.
- [21] Hyunjin Shin and Mia K Markey. A machine learning perspective on the development of clinical decision support systems utilizing mass spectra of blood samples. *Journal of Biomedical Informatics*, 39(2):227–248, 2006.
- [22] Senthil K Nachimuthu and Peter J Haug. Early detection of sepsis in the emergency department using dynamic Bayesian networks. In *AMIA Annual Symposium Proceedings*, volume 2012, page 653. American Medical Informatics Association, 2012.

- [23] Daniel-Ioan Curiac, Gabriel Vasile, Ovidiu Baniias, Constantin Volosencu, and Adriana Albu. Bayesian network model for diagnosis of psychiatric diseases. In *Information Technology Interfaces, 2009. ITI'09. Proceedings of the ITI 2009 31st International Conference on*, pages 61–66. IEEE, 2009.
- [24] S Barry Issenberg, William C Mcgaghie, Emil R Petrusa, David Lee Gordon, and Ross J Scalese. Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher*, 27(1):10–28, 2005.
- [25] Nancy Posel. *Making a Case: Validating Criterion-Referenced Guidelines for Virtual Patient Case Authoring*. PhD thesis, McGill University, 2011.
- [26] James Bateman, Maggie Allen, Dipti Samani, Jane Kidd, and David Davies. Virtual patient design: Exploring what works and why. A grounded theory study. *Medical Education*, 47(6):595–606, 2013.
- [27] Janet L Kolodner, Paul J Camp, David Crismond, Barbara Fasse, Jackie Gray, Jennifer Holbrook, Sadhana Puntambekar, and Mike Ryan. Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design (tm) into practice. *The journal of the learning sciences*, 12(4):495–547, 2003.
- [28] Dorothy H Evensen. To group or not to group: Students’ perceptions of collaborative learning activities in law school. *S. Ill. ULJ*, 28:343, 2003.
- [29] Tarja Susi, Mikael Johannesson, and Per Backlund. Serious games: An overview. 2007.
- [30] J. Wan. Playing doctor: Learning about slips of the knife better on “patients” than patients. *The Washington Post*, November 2010.
- [31] David Crookall. Serious games, debriefing, and simulation/gaming as a discipline. *Simulation & Gaming*, 41(6):898–920, 2010.
- [32] D Crookall and D Saunders. Towards an integration of communication and simulation. *Communication and Simulation: From Two Fields to One Theme*, pages 3–29, 1989.
- [33] Rosemary Garris, Robert Ahlers, and James E Driskell. Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4):441–467, 2002.
- [34] Barry S. Issenberg, W. C. Mcgaghie, E. R. Petrusa, D. Lee Gordon, and R. J. Scalese. Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher*, 27:10–28, 2005.

- [35] C. A. Roldan, B. K. Shivley, and M. H. Crawford. Value of the cardiovascular physical examination for detecting valvular heart disease in asymptomatic subjects. *American Journal of Cardiology*, 77:1327–1331, 1996.
- [36] R. S. Haluck, R. L. Marshall, T. M. Krummel, and M. G. Melkonian. Are surgery training programs ready for virtual reality? A survey of program directors in general surgery. *Journal of the American College of Surgeons*, 193:660–665, 2001.
- [37] S. Mangione and L. Z. Nieman. Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency. *The Journal of the American Medical Association (JAMA)*, 278:717–722, 1997.
- [38] N. Benedict. Virtual patients and problem-based learning in advanced therapeutics. *American Journal of Pharmaceutical Education*, 74, 2010.
- [39] J. A. Gordon, W. M. Wilkerson, D. W. Shaffer, and E. G. Armstrong. “Practicing” medicine without risk: students’ and educators’ responses to high-fidelity patient simulation. *Academic Medicine: Journal of the Association of American Medical Colleges*, 76:469–472, 2001.
- [40] W. G. Harless, M. A. Zier, J. E. Smith, R. Dube, R. C. Duncan, and W. R. Ayers. TIME project interactive patient simulations: experiential learning in the medical school classroom. *Journal of Medical Education Technologies*, 2:3–8, 1992.
- [41] Susan R Swing. The acgme outcome project: retrospective and prospective. *Medical Teacher*, 29(7):648–654, 2007.
- [42] R. A. Bergin and U. G. H. Fors. Interactive simulated patient — an advanced tool for student-activated learning in medicine and healthcare. *Computers and Education*, 40:361–376, 2003.
- [43] Douglas R Danforth, Mike Procter, Richard Chen, Mary Johnson, and Robert Heller. Development of virtual patient simulations for medical education. *Journal For Virtual Worlds Research*, 2(2), 2009.
- [44] Rachel Ellaway, Chris Candler, Peter Greene, et al. An architectural model for medbiquitous virtual patients. *MedBiquitous*, 2006.
- [45] Neal Benedict, Kristine Schonder, and James McGee. Promotion of self-directed learning using virtual patient cases. *American Journal of Pharmaceutical Education*, 77(7):151, 2013.

- [46] Neal Benedict. Virtual patients and problem-based learning in advanced therapeutics. *American Journal of Pharmaceutical Education*, 74(8):143, October 2010.
- [47] BB Barone, H Yeh, CF Snyder, and et al. Long-term all-cause mortality in cancer patients with preexisting diabetes mellitus: A systematic review and meta-analysis. *JAMA*, 300(23):2754–2764, 2008.
- [48] Edward Giovannucci, David M. Harlan, Michael C. Archer, Richard M. Bergental, Susan M. Gapstur, Laurel A. Habel, Michael Pollak, Judith G. Regensteiner, and Douglas Yee. Diabetes and cancer: A consensus report. *CA: A Cancer Journal for Clinicians*, 60(4):207–221, 2010.
- [49] Samir Abd El-Razek, Hazem El-Bakry, Wael Abd El-Wahed, and Nikos Mastorakis. Collaborative Virtual Environment Model for Medical ELearning. In *9th WSEAS Int. Conference on Applied Computer and Applied Computational Science*, December 2009.
- [50] Randolph A Miller, Melissa A McNeil, Sue M Challinor, Fred E Masarie Jr, and Jack D Myers. The internist-1/quick medical reference project—status report. *Western Journal of Medicine*, 145(6):816, 1986.
- [51] T Allan Pryor, Reed M Gardner, Paul D Clayton, and Homer R Warner. The help system. *Journal of medical systems*, 7(2):87–102, 1983.
- [52] MS Blois, MS Tuttle, and DD Sherertz. Reconsider: A program for generating differential diagnoses*. In *Proceedings. Symposium on Computer Applications in Medical Care*, pages 263–268. American Medical Informatics Association, 1981.
- [53] G Octo Barnett, James J Cimino, Jon A Hupp, and Edward P Hoffer. Dxplain: an evolving diagnostic decision-support system. *Jama*, 258(1):67–74, 1987.
- [54] Garfield G Duncan, Carl F Clancy, John R Wolgamot, and Barkley Beidleman. Neomycin: results of clinical use in ten cases. *Journal of the American Medical Association*, 145(2):75–80, 1951.
- [55] Ted Shortliffe and Randy Davis. Some considerations for the implementation of knowledge-based expert systems. *ACM SIGART Bulletin*, (55):9–12, 1975.
- [56] William J Clancey. Tutoring rules for guiding a case method dialogue. *International Journal of Man-Machine Studies*, 11(1):25–49, 1979.
- [57] William J Clancey. From guidon to neomycin and heracles in twenty short lessons. *AI Magazine*, 7(3):40, 1986.

- [58] Homer L Chin. Case-based tutoring from a medical knowledge base. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 396. American Medical Informatics Association, 1988.
- [59] Bonnie L Webber, Ron Rymon, and John R Clarke. Flexible support for trauma management through goal-directed reasoning and planning. *Technical Reports (CIS)*, page 146, 1991.
- [60] Sandra Carberry, John Clarke, and Abigail Gertner. Automatic construction of medical cases for training and testing using the knowledge base of an existing decision support system. In *AAAI Spring Symposium on Artificial Intelligence in Medicine*. AAAI Press, Stanford, Calif, pages 16–20, 1996.
- [61] Ronald Ghossein and Virginia A. Livolsi. Papillary thyroid carcinoma tall cell variant. *Thyroid: Official Journal of the American Thyroid Association*, 18(11):1179–1181, November 2008.
- [62] Allen B. Downey. *Think Bayes*, pages 27–28. O’Reilly Media, 2013.
- [63] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, September 1988.
- [64] Sharon Bertsch McGrayne. *The Theory That Would Not Die: How Bayes’ Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. Yale University Press, May 2011.
- [65] Adam Zagorecki, Piotr Orzechowski, and Katarzyna Holownia. A system for automated general medical diagnosis using Bayesian networks. In *MedInfo - World Congress on Health and Biomedical Informatics*, pages 461–465, 2013.
- [66] Y Ye, FR Tsui, M Wagner, JU Espino, and Q. Li. Influenza detection from emergency department reports using natural language processing and Bayesian network classifiers. *Journal of the American Medical Informatics Association*, 21:815–823, September 2014.
- [67] M. Ashcroft. Bayesian Networks in Business Analytics. In *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*, pages 955–961, Sept 2012.
- [68] Steven E. Rigdon, Jason J. Sauppe, and Sheldon H. Jacobson. Forecasting the 2012 and 2014 elections using bayesian prediction and optimization. *SAGE Open*, 5(2), 2015.

- [69] P. Langley, W. Iba, and K. Thompson. An analysis of Bayesian classifiers. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 223–228. MIT Press, 1992.
- [70] J. Cheng and R. Greiner. Comparing Bayesian network classifiers. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 101–108. Morgan Kaufmann Publishers Inc, July 1999.
- [71] P. Spirtes, C.N. Glymour, and R. Scheines. *Causation, Prediction, and Search*. Cambridge, MA: MIT Press, 2000.
- [72] Peter Spirtes, Clark N. Glymour, and Richard Scheines. *Causation, Prediction, and Search.*, chapter 9.2 Selecting Variables. Cambridge, MA: MIT, 2000.
- [73] Dynamic Bayesian networks - an introduction. <https://www.bayesserver.com/docs/introduction/dynamic-bayesian-networks>.
- [74] Kevin P Murphy. Dynamic Bayesian networks. *Probabilistic Graphical Models, M. Jordan*, 7, 2002.
- [75] Christopher J Langmead. Generalized queries and Bayesian statistical model checking in dynamic Bayesian networks: application to personalized medicine. *Life Sciences Society*, 2009.
- [76] Micol Sandri, Paola Berchiolla, Ileana Baldi, Dario Gregori, and Roberto Alberto De Blasi. Dynamic Bayesian networks to predict sequences of organ failures in patients admitted to ICU. *Journal of biomedical informatics*, 48:106–113, 2014.
- [77] Ara V Nefian, Luhong Liang, Xiaobo Pi, Xiaoxing Liu, and Kevin Murphy. Dynamic bayesian networks for audio-visual speech recognition. *EURASIP Journal on Advances in Signal Processing*, 2002(11):783042, 2002.
- [78] Bruno-Edouard Perrin, Liva Ralaivola, Aurelien Mazurie, Samuele Bottani, Jacques Mallet, and Florence d’Alche Buc. Gene networks inference using dynamic Bayesian networks. *Bioinformatics*, 19(suppl_2):ii138–ii148, 2003.
- [79] Le An, Mehran Kafai, and Bir Bhanu. Dynamic Bayesian network for unconstrained face recognition in surveillance camera networks. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 3(2):155–164, 2013.
- [80] Sibel Yaman, Dilek Hakkani-Tür, and Gokhan Tur. Social role discovery from spoken language using dynamic Bayesian networks. In *Eleventh Annual Conference of the International Speech Communication Association*, 2010.
- [81] Dmitriy Babichenko, Marek Druzdzel, Lorin Grieve, Ravi Patel, Jonathan Velez, Taylor Neal, James McCray, Rae-Djamaal Wallace, and Sean Jenkins. Designing

the model patient: Data-driven virtual patients in medical education. In *Proceedings of IEEE SeGAH 2016 4th International Conference on Serious Games and Applications for Health*. IEEE, May 2016.

- [82] Valerie Smothers, Ben Azan, and Rachel Ellaway. Medbiquitous virtual patient specifications and description document. Technical report, 0.48', Technical report, MedBiquitous Consortium, 2008.
- [83] Ms Nijole Bernaitis, Lyndsee Baumann-Birkbeck, Sean Alcorn, Michael Powell, Devinder Arora, and Shailendra Anoopkumar-Dukie. Simulated patient cases using decisionsim™ improves student performance and satisfaction in pharmacotherapeutics education. *Currents in Pharmacy Teaching and Learning*, 2018.
- [84] Daniel Nystrom, Linda Williams, and Douglas Paull. Investigating medical diagnosis: Qualitative results from a virtual patient simulation pilot study. In *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, volume 4, pages 92–96. SAGE Publications Sage India: New Delhi, India, 2015.
- [85] Carol C Dudding and Danika L Pfeiffer. Clinical decision-making in speech-language pathology graduate students: Quantitative findings. *Teaching and Learning in Communication Sciences & Disorders*, 2(1):2, 2018.
- [86] Rachel H Ellaway. Openlabyrinth: An abstract pathway-based serious game engine for professional education. In *Digital Information Management (ICDIM), 2010 Fifth International Conference on*, pages 490–495. IEEE, 2010.
- [87] Tudor Călinici and Valentin Muntean. Open labyrinth—a web application for medical education using virtual patients. *Applied Medical Informatics*, 27(4):15–20, 2011.
- [88] Nabil Zary, Inga Hege, Jörn Heid, Luke Woodham, Jeroen Donkers, and Andrzej A Kononowicz. Enabling interoperability, accessibility and reusability of virtual patients across europe-design and implementation. In *MIE*, pages 826–830. Citeseer, 2009.
- [89] Zoe Paskins, Jo Kirkcaldy, Maggie Allen, Colin Macdougall, Ian Fraser, and Ed Peile. Design, validation and dissemination of an undergraduate assessment tool using simman® in simulated medical emergencies. *Medical teacher*, 32(1):e12–e17, 2010.
- [90] Amy L Seybert, Karen K Laughlin, Neal J Benedict, Christine M Barton, and Rhonda S Rea. Pharmacy student response to patient-simulation mannequins to teach performance-based pharmacotherapeutics. *American Journal of Pharmaceutical Education*, 70(3):48, 2006.

- [91] James D Klein. Design, implementation and evaluation of a nursing simulation: a design and development research study. *Guest Editorial*, 157, 2012.
- [92] Nikki J Maran and Ronnie J Glavin. Low-to high-fidelity simulation—a continuum of medical education? *Medical Education*, 37:22–28, 2003.
- [93] Sarah Yardley, Caragh Brosnan, and Jane Richardson. The consequences of authentic early experience for medical students: creation of mētis. *Medical Education*, 47(1):109–119, 2013.
- [94] Roger Kneebone and Sarah Baillie. Contextualized simulation and procedural skills: a view from medical education. *Journal of Veterinary Medical Education*, 35(4):595–598, 2008.
- [95] Peggy A Ertmer and James D Russell. Using case studies to enhance instructional design education. *Educational Technology*, pages 23–31, 1995.
- [96] Sören Huwendiek, Bas A De Leng, Nabil Zary, Martin R Fischer, Jorge G Ruiz, and Rachel Ellaway. Towards a typology of virtual patients. *Medical teacher*, 31(8):743–748, 2009.
- [97] Nancy Posel, David Fleischer, and Bruce M Shore. 12 Tips: Guidelines for authoring virtual patient cases. *Medical Teacher*, 31(8):701–708, 2009.
- [98] James B McGee, Jane Neill, Leon Goldman, and Edward Casey. Using multimedia virtual patients to enhance the clinical curriculum for medical students. *Studies in health technology and informatics*, 52:732–735, 1998.
- [99] Sören Huwendiek, Friedrich Reichert, Hans-Martin Bosse, Bas A De Leng, Cees PM Van Der Vleuten, Martin Haag, Georg F Hoffmann, and Burkhard Tönshoff. Design principles for virtual patients: a focus group study among students. *Medical education*, 43(6):580–588, 2009.
- [100] M Triola, H Feldman, AL Kalet, S Zabar, EK Kachur, C Gillespie, M Anderson, C Griesser, and M Lipkin. A randomized trial of teaching clinical skills using virtual and live standardized patients. *Journal of general internal medicine*, 21(5):424–429, 2006.
- [101] Sara Kim, William R Phillips, Linda Pinsky, Doug Brock, Kathryn Phillips, and Jane Keary. A conceptual framework for developing teaching cases: a review and synthesis of the literature across disciplines. *Medical education*, 40(9):867–876, 2006.
- [102] Tiffany Grunwald and Charisse Corsbie-Massay. Guidelines for cognitively efficient multimedia learning tools: educational strategies, cognitive load, and interface design. *Academic medicine*, 81(3):213–223, 2006.

- [103] Conor Linehan, Ben Kirman, Shaun Lawson, and Gail Chan. Practical, appropriate, empirically-validated guidelines for designing educational games. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1979–1988. ACM, 2011.
- [104] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. *Game analytics*. Springer, 2016.
- [105] Changchun Liu, Pramila Agrawal, Nilanjan Sarkar, and Shuo Chen. Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction*, 25(6):506–529, 2009.
- [106] Marc M Triola, Ned Champion, James B McGee, Susan Albright, and Peter Greene. An XML standard for virtual patients: exchanging case-based simulations in medical education. *Computer*, 4:5, 2007.
- [107] Marcus Gallagher and Amanda Ryan. Learning to play pac-man: An evolutionary, rule-based approach. In *Evolutionary Computation, 2003. CEC'03. The 2003 Congress on*, volume 4, pages 2462–2469. IEEE, 2003.
- [108] Julian Togelius, Noor Shaker, Sergey Karakovskiy, and Georgios N Yannakakis. The mario ai championship 2009-2012. *AI Magazine*, 34(3):89–92, 2013.
- [109] Miller R.A., Pople H. E. Jr, and Myers J.D. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. *New England Journal of Medicine*, 307:468–476, August 1982.
- [110] Guilan Kong, Dong-Ling Xu, Richard Body, Jian-Bo Yang, Kevin Mackway-Jones, and Simon Carley. A belief rule-based decision support system for clinical risk assessment of cardiac chest pain. *European Journal of Operational Research*, 219(3):564–573, 2012.
- [111] Adrian Doloca, Oana ȚĂNCULESCU, Iulia Ciongradi, Laura Trandafir, Simona Stoleriu, and Gabriela Ifteni. Comparative study of virtual patient applications. *Proc. of the Romanian Academy Series A-Mathematics Physics Technical Sciences*, 16(3):466–473, 2015.
- [112] Sinthana Umakanthan, Alyce Anderson, Dmitriy Babichenko, Claudia Ramos-Rivers, Benjamin Click, Jana Hashash, Michael Dunn, Marc Schwartz, Jason Swoger, Arthur Barrie, Miguel Regueiro, and David Binion. Development of an Inflammatory Bowel Disease Research Registry Derived from Observational Electronic Health Record Data for Comprehensive Clinical Phenotyping. *American Journal of Gastroenterology*, 2016.

- [113] Benjamin Click, Claudia Ramos Rivers, Ioannis E Koutroubakis, Dmitriy Babichenko, Alyce M Anderson, Jana G Hashash, Michael A Dunn, Marc Schwartz, Jason Swoger, Leonard Baidoo, Arthur Barrie, Miguel Regueiro, and David G Binion. Demographic and clinical predictors of high healthcare use in patients with inflammatory bowel disease. *Inflammatory Bowel Diseases*, March 2016.
- [114] Marek J. Druzdzal. SMILE: Structural Modeling, Inference, and Learning Engine and GeNIe: A development environment for graphical decision-theoretic models (Intelligent Systems Demonstration). In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pages 902–903. AAAI Press/The MIT Press, Menlo Park, CA, 1999.
- [115] Alexandru Boicea, Florin Radulescu, and Laura Ioana Agapin. MongoDB vs. Oracle-database comparison. In *EIDWT*, pages 330–335, 2012.
- [116] Olivier Curé, Robin Hecht, Chan Le Duc, and Myriam Lamolle. Data integration over NoSQL stores using access path based mappings. In *International Conference on Database and Expert Systems Applications*, pages 481–495. Springer, 2011.
- [117] Chitra Babu and G Gunasingh. Desh: Database evaluation system with Hibernate ORM framework. In *Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on*, pages 2549–2556. IEEE, 2016.
- [118] Neal Benedict. Virtual patients and problem-based learning in advanced therapeutics. *American Journal of Pharmaceutical Education*, 74(8):143, 2010.
- [119] Michael A Smith, Laura A Siemianowski, and Neal Benedict. Virtual patient case sharing across two schools of pharmacy. *American Journal of Pharmaceutical Education*, 80(9):153, 2016.
- [120] Thomas B Talbot, Kenji Sagae, Bruce John, and Albert A Rizzo. Sorting out the virtual patient: How to exploit artificial intelligence, game technology and sound educational practices to create engaging role-playing simulations. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 4(3):1–19, 2012.
- [121] American College of Physicians, virtual patients annals. diabetic ketoacidosis. <https://www.acponline.org/cme-moc/online-learning-center/diabetic-ketoacidosis>.
- [122] Gary H Tabas and Gary Tabas. A novel approach to teaching the management of diabetic ketoacidosis using virtual patient technology. In *Journal of General*

Internal Medicine, volume 26, pages S582–S583. Springer 233 Spring Street, New York, NY 10013 USA, 2011.

- [123] International guidelines for management of severe sepsis and septic shock: 2016. 2016.
- [124] Dyanne P Westerberg. Diabetic ketoacidosis: evaluation and treatment. *American Family Physician*, 87(5), 2013.
- [125] Henry Chueh and G Octo Barnett. “Just-in-time” clinical information. *Academic Medicine: Journal of the Association of American Medical Colleges*, 72(6):512–517, 1997.
- [126] David A Cook. The research we still are not doing: An agenda for the study of computer-based learning. *Academic Medicine*, 80(6):541–548, 2005.
- [127] Brenden K Petersen, Michael B Mayhew, Calvin OE Ogbuefi, John D Greene, Vincent X Liu, and Priyadip Ray. Modeling sepsis progression using hidden Markov models. *arXiv preprint arXiv:1801.02736*, 2018.
- [128] Roger C Bone, Robert A Balk, Frank B Cerra, R Phillip Dellinger, Alan M Fein, William A Knaus, Roland MH Schein, and William J Sibbald. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest*, 101(6):1644–1655, 1992.
- [129] Mayo Clinic. Sepsis — causes and symptoms. <https://www.mayoclinic.org/diseases-conditions/sepsis/symptoms-causes/syc-20351214>.
- [130] Mervyn Singer, Clifford S Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R Bernard, Jean-Daniel Chiche, Craig M Coopersmith, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*, 315(8):801–810, 2016.
- [131] What is qsofa? <http://www.qsofa.org/what.php>.
- [132] Philip M Podsakoff, Scott B MacKenzie, Jeong-Yeon Lee, and Nathan P Podsakoff. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5):879, 2003.
- [133] Jum C Nunnally and Ira H Bernstein. *Psychometric theory*. McGraw-Hill New York, 1978.

- [134] Jamie L Estock, Holly K Curinga, Airan Li, Lorin B Grieve, and Christopher R Brackney. Comparison of chest compression interruption times across two automated devices: A randomized, crossover simulation study. *The American Journal of Emergency Medicine*, 34(1):57–62, 2016.
- [135] Alberto Bacchelli and Christian Bird. Expectations, outcomes, and challenges of modern code review. In *Proceedings of the 2013 international conference on software engineering*, pages 712–721. IEEE Press, 2013.
- [136] Jianxin Jiao and Chun-Hsien Chen. Customer requirement management in product development: a review of research issues. *Concurrent Engineering*, 14(3):173–185, 2006.
- [137] Scott M Shafer, H Jeff Smith, and Jane C Linder. The power of business models. *Business horizons*, 48(3):199–207, 2005.
- [138] Jean Anderson, F Fleak, Kathi Garrity, and Fred Drake. Integrating usability techniques into software development. *IEEE Software*, 18(1):46–53, 2001.
- [139] Len Bass and Bonnie E John. Linking usability to software architecture patterns through general scenarios. *Journal of Systems and Software*, 66(3):187–197, 2003.
- [140] John L Campbell, Charles Quincy, Jordan Osserman, and Ove K Pedersen. Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3):294–320, 2013.
- [141] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.
- [142] Nancy Posel, Bruce M Shore, and David Fleiszer. Virtual patient cases: A qualitative study of the requirements and perceptions of authors. *International Journal of Medical Education*, 3:175–182, 2012.
- [143] Robert K Yin. *Case study research: Design and methods*. Sage Publications, 2013.
- [144] Lena M Napolitano. Sepsis 2018: Definitions and guideline changes. *Surgical Infections*, 19(2):117–125, 2018.
- [145] 2018 update to surviving sepsis guidelines: Cue backlash. *PulmCMM*.
- [146] Mitchell M Levy, Laura E Evans, and Andrew Rhodes. The surviving sepsis campaign bundle: 2018 update. *Intensive Care Medicine*, 44(6):925–928, 2018.

- [147] Paul A Kirschner, John Sweller, and Richard E Clark. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2):75–86, 2006.
- [148] Barry Issenberg, Michael S Gordon, David Lee Gordon, Robert E Safford, and Ian R Hart. Simulation and new learning technologies. *Medical teacher*, 23(1):16–23, 2001.
- [149] Ruth M Fanning and David M Gaba. The role of debriefing in simulation-based learning. *Simulation in healthcare*, 2(2):115–125, 2007.
- [150] Gunjan Kamdar, David O Kessler, Lindsey Tilt, Geetanjali Srivastava, Kajal Khanna, Todd P Chang, Dorene Balmer, and Marc Auerbach. Qualitative evaluation of just-in-time simulation-based learning: The learners’ perspective. *Simulation in Healthcare*, 8(1):43–48, 2013.
- [151] Sheena Sethi-Iyengar, Gur Huberman, and Wei Jiang. How much choice is too much? contributions to 401 (k) retirement plans. *Pension Design and Structure: New Lessons From Behavioral Finance*, 83:84–87, 2004.
- [152] Sheena S Iyengar and Mark R Lepper. When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6):995, 2000.
- [153] A Houillon, RC Lorenz, W Boehmer, MA Rapp, A Heinz, J Gallinat, and K Obermayer. The effect of novelty on reinforcement learning. In *Progress in Brain Research*, volume 202, pages 415–439. Elsevier, 2013.
- [154] Marilyn P Arnone, Ruth V Small, Sarah A Chauncey, and H Patricia McKenna. Curiosity, interest and engagement in technology-pervasive learning environments: A new research agenda. *Educational Technology Research and Development*, 59(2):181–198, 2011.
- [155] Graham Heaslip, Paul Donovan, and John G Cullen. Student response systems and learner engagement in large classes. *Active Learning in Higher Education*, 15(1):11–24, 2014.
- [156] Gregory F Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2-3):393–405, 1990.
- [157] Paul Dagum and Michael Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60(1):141–153, 1993.

- [158] Robert Rowley. The 25 most common medical diagnoses. <https://www.practicefusion.com/blog/25-most-common-diagnoses/>, February 2011.
- [159] Junghye Lee, Wonji Lee, Il-Su Park, Hun-Sung Kim, Hyeseon Lee, and Chi-Hyuck Jun. Risk assessment for hypertension and hypertension complications incidences using a Bayesian network. *IIE Transactions on Healthcare Systems Engineering*, 6(4):246–259, 2016.
- [160] Alexander Stojadinovic, Anton Bilchik, David Smith, John S Eberhardt, Elizabeth Ben Ward, Aviram Nissan, Eric K Johnson, Mladjan Protic, George E Peoples, Itzhak Avital, et al. Clinical decision support and individualized prediction of survival in colon cancer: Bayesian belief network model. *Annals of surgical oncology*, 20(1):161–174, 2013.
- [161] Moritz Gerstung, Michael Baudis, Holger Moch, and Niko Beerenwinkel. Quantifying cancer progression with conjunctive bayesian networks. *Bioinformatics*, 25(21):2809–2815, 2009.
- [162] Chanu Rhee, Raymund Dantes, Lauren Epstein, David J Murphy, Christopher W Seymour, Theodore J Iwashyna, Sameer S Kadri, Derek C Angus, Robert L Danner, Anthony E Fiore, et al. Incidence and trends of sepsis in us hospitals using clinical vs claims data, 2009-2014. *JAMA*, 318(13):1241–1249, 2017.
- [163] Lorna M Barclay, Jane L Hutton, and Jim Q Smith. Refining a Bayesian network using a chain event graph. *International Journal of Approximate Reasoning*, 54(9):1300–1309, 2013.
- [164] Peter Thwaites, Jim Q Smith, and Eva Riccomagno. Causal analysis with chain event graphs. *Artificial Intelligence*, 174(12-13):889–909, 2010.
- [165] Paul E Marik. Aspiration pneumonitis and aspiration pneumonia. *New England Journal of Medicine*, 344(9):665–671, 2001.
- [166] Xuedong D Huang, Yasuo Ariki, and Mervyn A Jack. Hidden Markov models for speech recognition. 1990.
- [167] Lalit Bahl, Peter Brown, Peter De Souza, and Robert Mercer. Maximum mutual information estimation of hidden Markov model parameters for speech recognition. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86.*, volume 11, pages 49–52. IEEE, 1986.
- [168] Anders Krogh, Michael Brown, I Saira Mian, Kimmen Sjölander, and David Haussler. Hidden Markov models in computational biology: Applications to protein modeling. *Journal of Molecular Biology*, 235(5):1501–1531, 1994.

- [169] Alexander Schliep, Alexander Schönhuth, and Christine Steinhoff. Using hidden Markov models to analyze gene expression time course data. *Bioinformatics*, 19(suppl_1):i255–i263, 2003.
- [170] Christopher H Jackson, Linda D Sharples, Simon G Thompson, Stephen W Duffy, and Elisabeth Couto. Multistate markov models for disease progression with classification error. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(2):193–209, 2003.
- [171] Shameek Ghosh, Jinyan Li, Longbing Cao, and Kotagiri Ramamohanarao. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. *Journal of Biomedical Informatics*, 66:19–31, 2017.
- [172] Claudio S Pinhanez. *Representation and recognition of action in interactive spaces*. PhD thesis, Massachusetts Institute of Technology, 1999.
- [173] James F Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
- [174] Ilhem Feddaoui, Faïçal Felhi, and Jalel Akaichi. Extract: New extraction algorithm of association rules from frequent itemsets. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*, pages 752–756. IEEE, 2016.
- [175] Hany Mahgoub. Mining association rules from unstructured documents. In *Proc. 3rd Int. Conf. on Knowledge Mining, ICKM, Prague, Czech Republic*, pages 167–172, 2006.
- [176] Cherif Chiraz Latiri and Sadok Ben Yahia. Textmining: Generating association rules from textual data. In *INFORSID*, pages 27–39, 2001.
- [177] Hany Mahgoub, Dietmar Rösner, Nabil Ismail, and Fawzy Torkey. A text mining technique using association rules extraction. *International Journal of Computational Intelligence*, 4(1):21–28, 2008.
- [178] KS Lakshmi and G Vadivu. Extracting association rules from medical health records using multi-criteria decision analysis. *Procedia Computer Science*, 115:290–295, 2017.
- [179] Carmen Lacave and Francisco J Díez. A review of explanation methods for bayesian networks. *The Knowledge Engineering Review*, 17(2):107–127, 2002.
- [180] Marek J Druzdzel and Max Henrion. Efficient reasoning in qualitative probabilistic networks. In *Proceedings of the 11th National Conference on Artificial Intelligence, Washington*, pages 548–553, 1993.

- [181] David Madigan, Krzysztof Mosurski, and Russell G Almond. Graphical explanation in belief networks. *Journal of Computational and Graphical Statistics*, 6(2):160–181, 1997.
- [182] Marek J Druzdzel. *Probabilistic reasoning in decision support systems: from computation to common sense*. PhD thesis, Carnegie Institute of Technology, 1993.
- [183] Silja Renooij, Linda C Van Der Gaag, Simon Parsons, and Shaw Green. Pivotal pruning of trade-offs in qpns. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, pages 515–522. Morgan Kaufmann Publishers Inc., 2000.
- [184] Medscape. <https://emedicine.medscape.com>.
- [185] Lab Tests Online. <https://labtestsonline.org/>.