

**Evaluation of a Prototype for a Diagnostic Decision Support Tool for
Diagnosing Respiratory Distress in Infants**

A thesis submitted to
The Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree
Masters of Arts

by

Janette Folkens

Department of Psychology
Carleton University
Ottawa, Ontario, Canada
September, 2009

©2009 Janette Folkens



Library and Archives
Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-64450-8
Our file *Notre référence*
ISBN: 978-0-494-64450-8

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

When medical clinicians are faced with a patient presenting with ambiguous symptoms, the likelihood of misdiagnosis is high. This is a recurring problem in the medical field. In the present study participants were presented with 24 case scenarios each describing an infant suffering from one of two different diseases. Case scenarios included base rates for each disease and individual information about each case. One half of the participants were exposed to a *Diagnostic Support System (DSS)* to aid in decision making. The task was to estimate the probability that each infant suffered from one of two diseases. The purpose was to investigate the extent to which a DSS may help reduce the number of errors in diagnosis. Results indicate that exposure on the DSS had a small effect on the number of correct diagnoses, and both base rates and individuating information was being used to estimate probability estimates.

Acknowledgments

I would like to thank my wonderful husband Norbert for his tireless efforts supporting me through my long journey to complete my Masters. Without his patience, understanding, and continual encouragement this would not be possible. I hope that this thesis is encouragement to my children to achieve higher learning and to not give up even when the road ahead appears long and twisted.

Many thanks to my advisor Dr. Gitte Lindgaard for absolutely everything she has done to make this thesis possible. Her persistence, early morning meetings, late night emails, and guidance taught me professionalism and how to write clearly for an audience. Additionally I would like to thank Dr. Lindgaard for her many revisions, allowing me the opportunity to learn through this iterative process.

Thanks to Cathy Dudek for opening up her home to me when the statistics were overwhelming. Her direction was greatly appreciated and helped make the statistics comprehensible. Thank you to Dr. Satid Thammasitboon for his constant willingness to help with the vignettes, and participant recruitment; and to Dr. Erika Bariciak and Dr. Robin Walker for their many contributions to the case scenarios and symptoms.

Without the hard work from both Catherine Pyper and Dr. Monique Frize NeoPeDSS would not be possible. Catherine continued to work with NeoPeDSS during this thesis, and provided assistance whenever and however possible, thank you very much for all your help and dedication.

Last but not least I would like to thank all my participants and my committee members. Without the tireless efforts of my participants, who were no doubt already worn out, this thesis would not be possible. The insight that I gained from my committee

members helped me to learn more about designing a study and looking at the data critically. In particular, Dr. Jo-Anne LeFevre spent large amounts of selfless time helping and teaching me about statistical analysis and write up, I am very thankful for all the time she spent.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
Table of Contents.....	v
List of Tables.....	vii
List of Figures.....	viii
List of Acronyms	ix
Introduction.....	1
Decision Making Errors.....	3
Diagnosticity.....	11
Bayes Theorem.....	13
Diagnostic Decision Support Systems.....	16
Hypothesis.....	24
Method.....	24
Participants.....	25
Material	25
Design	28
Procedure.....	31
Analysis	32
Results	33
Discussion	45
Theoretical Implications of the Findings.....	47
Future Research and Limitations of the Present Research	51

Conclusion	51
References	52
Appendix A- Announcement for Recruitment.....	59
Appendix B- Informed Consent Form.....	61
Appendix C- Interview Protocol Participant Information.....	63
Appendix D- Case Scenario Information.....	64
Appendix E- Phase 1 Survey	67
Appendix F- NeoPeDDS Instructions.....	68
Appendix G- Phase 2 NeoPeDDS Case Scenarios.....	70
Appendix H- Phase 3 Survey.....	71
Appendix I- Debriefing	72
Appendix J- Test of Normality	73
Appendix K- Test for Absence of Outliers.....	74

List of Tables

	Page
Table 1. Design- Outline of the Experiment	29
Table 2. Probability Estimates for HH, LL, HL, and LH Vignettes for RDS and TTN	37
Table 3. Phase 1 Mean and Standard Deviations for 2-way Interaction	39
Table 4. Phase 3 Mean and Standard Deviations for 2-way Interaction	40
Table 5. Phase 1 and 3 Diagnosis Methods	44

List of Figures

	Page
Figure 1. NeoPeDDS Input Screen Prior to Data Entry	26
Figure 2. NeoPeDDS Panel Showing Further Description Options Available Through a Drop-Down Menu After the Apgar 5 Score has been Selected	27
Figure 3. Output Probabilities Upon Entry of an Apgar 5 with a Score of 3	27
Figure 4. Mean Probability Estimates for High and Low Base Rate Groups	35
Figure 5. Mean Probability Estimates for Base Rate Groups with and without Exposure For Phase 1 and 3	36
Figure 6. Probability Estimates in Phase 1 by Serial Position of the Highly Diagnostic Symptoms in each Vignette	38
Figure 7. Probability Estimates in Phase 3 by Serial Position of the Highly Diagnostic Symptoms in each Vignette	39
Figure 8. The Average Number of Correct Diagnoses in Phase 2 Before and After use of NeoPeDDS	41
Figure 9. Estimated Frequency of Occurrence for List of Symptoms Presented as either a Symptom of RDS or TTN	43

List of Acronyms

General Terms

APGAR	Appearance, Pulse, Grimace, Activity, Respiration
CHEO	Children's Hospital of Eastern Ontario
DDS	Diagnostic Decision Support System
DDSS	Diagnostic Decision Support System
NeoPeDSS	Neonatal Pediatric Decision Support System
PDA	Personal Digital Assistant
RDS	Respiratory Distress Syndrome
SMDM	The Society for Medical Decision Making
TTN	Transient Tachypnoea
WHO	World Health Organization

Specific Terms

HH	Symptom High for RDS and TTN (RDS-H & TTN-H)
HBR	High Base Rate Group
LL	Symptoms Low for RDS and TTN (RDS-L & TTN-L)
LBR	Low Base Rate Group
RDS-H	Symptom High for Respiratory Distress Syndrome
RDS-L	Symptom Low for Respiratory Distress Syndrome
TTN-H	Symptom High for Transient Tachypnoea
TTN-L	Symptom Low for Transient Tachypnoea

Evaluation of a Prototype for a Diagnostic Decisions Support Tool for Diagnosing Respiratory Distress in Infants

Introduction

Diagnostic uncertainty, and hence the possibility of misdiagnosing patients presenting with ambiguous symptoms, is a recurring problem faced by medical clinicians (Croskerry, 2006; Johnson & Feldman, 1995; Ramnarayan, Winrow et al., 2006). The incidence of misdiagnosis is much higher than one might expect -- over the past 65 years, clinical autopsies have revealed that approximately 40% of deaths are due to misdiagnosis (Burton, Troxclair, & Newman, 1998; Nichols, Aronica, & Babe, 1998). In other studies misdiagnoses were estimated to account for somewhere between 26% (Schiff et al., 2006) and 78% (Amy, Lyman, & Borowitz, 2006; Larissa, Lyman, & Borowitz, 2006) of all medical errors. Unfortunately, clinical autopsies are no longer carried out routinely; therefore, medical practitioners receive little or no feedback on the correctness of their diagnoses, especially if they work in the Emergency Room (ER). This is because patients are transferred to regular wards once they have been assessed and diagnosed, and once this has happened, no further reports on the patients' condition are fed back to the ER clinicians.

A high number of misdiagnoses could be prevented (Croskerry & Sinclair, 2001) by improving clinicians' decision making processes, in particular, by reducing the incidence of various well-studied cognitive biases. Despite the high occurrence of diagnostic error, ways to improve decision making among medical practitioners has received little attention to date (Schiff et al., 2006). Accordingly, the goal of this thesis is to determine the degree to which a computerized Decision Support System (DSS) may

help to reduce medical diagnostic decision errors. To do that, the decision making performance of pediatric residents who were faced with fictitious patients presenting with ambiguous clinical pictures that are equally indicative of more than one hypothesis, was assessed.

Decision making under maximum uncertainty was explored because research has found that diagnosticians tend to formulate a diagnosis based predominantly on the first, generally also the most salient, item of diagnostic information they encounter. Because of this, they tend to ignore the possibility of alternative diagnoses that may be equally likely. A decision support system was used to enhance decision making because these systems are designed to overcome decision making biases, in particular by presenting the available information in ways that avoid the biases that people seem to find difficult to avoid. Because a common cause of error in clinical decision making is a failure to consider alternative diagnoses after an initial diagnosis had been made (Graber, Franklin, & Gordon, 2005), DSSs may potentially help by increasing diagnosticians' awareness of their cognitive shortcomings. Hence, the main hypothesis tested in this project was whether exposure to a DSS might help diagnosticians overcome their cognitive biases and subsequently lead to more accurate decisions.

Overview of the thesis

The next section discusses a range of decision making errors showing how these can affect medical diagnosis. This is followed by a discussion of diagnosticity as related to both the vignettes presented in this research and to Bayes' Theorem. Next, Bayes' Theorem is discussed with its applicability to DSSs, specifically in the medical field. Then DSSs are introduced with emphasis on NeoPeDSS as this is the tool being used in

the present study. The hypotheses are discussed in the next section, which is followed by an outline of the methods. Then, the results are presented, which leads into a discussion of the results. Finally, the limitations of the study and suggestions for future research are discussed.

Decision Making Errors

In this section human error in decision making will be discussed, specifically showing how decision heuristics can lead to decision errors, as well as why medical residents are prone to decision errors. Also discussed are the specific types of error that occur in medical diagnosis as well as in other decision making contexts. Decisions in the medical field cannot be made in a 'fast and frugal' manner because they may have deadly consequences (Graber, 2007). Unfortunately, preventable errors and misdiagnosis can result in death or permanent disability (Croskerry & Sinclair, 2001). Often, experts in a variety of fields display overconfidence in their decision making abilities, which can affect the correctness of their decisions. Overconfidence is the tendency to be more certain in the correctness of a decision than is warranted (Wallsten, 1981). Because of this tendency to be overconfident and the lack of feedback on the correctness of their decisions, physicians continue to apply flawed decision processes (Graber, 2007). Unfortunately, few medical schools provide training in formal decision theory or insight into the kinds of decision heuristics and biases that can have devastating results in the context of medical diagnosis.

Heuristics, which are mental shortcuts (Croskerry, 2000), are often used to make decisions quickly and can be very helpful in most everyday life situations. However,

these shortcuts should not be relied heavily upon in certain areas of decision making such as medical diagnosis (Croskerry, 2006a) precisely because they could lead to wrong diagnoses. Pediatrics is characterized by a very high level of diagnostic uncertainty because the patients are unable to communicate where it hurts and how they feel. In addition, many infants display ambiguous clinical pictures in the sense that the symptoms they present point to different diseases. Consequently, it is one area in which diagnostic error is a major problem (Ramnarayan, Winrow et al., 2006). This thesis presents cases that maximize ambiguity in that the symptoms are equally indicative of two different diseases, considered within the framework of Bayes' Theorem, which will be discussed later. This is done in part to replicate earlier work (Lindgaard, 1985; Lindgaard & Triggs, 1990; 2009; Lindgaard & Ralph, 2009; Ralph, 2003), in part to test a prototype DSS (Lindgaard, Pyper et al., 2009; Pyper, 2007; Pyper, Frize et al., 2008), and in part to learn more about these decision making biases, especially with a view to learning if these biases can be reduced by relatively simple exposure using a DSS.

The lack of formal training in decision making processes and biases is one reason for the continued problems in the field of medical decision making. When training in clinical decision making is provided, it does not typically reflect the pressures of making critical decisions in the real world that can seriously affect the well-being of patients (Croskerry & Campbell, 2004). The Society for Medical Decision Making (SMDM) promotes a course designed for practitioner clinicians. However, the SMDM states that the course will aid, but is not designed to train, future medical diagnosticians to make decisions (Society for Medical Decision Making, 2007). To date, there is no standard curriculum for medical decision making, although short courses are offered by the

SMDM, and a few textbooks do exist. Resident clinicians therefore still tend to learn about medical decision making through mentoring, conversation, and practice (Elstein, Dawson-Sauders, & Belzer, 1985).

Decisions in medicine are thus typically made without a thorough understanding of the process by which clinicians are likely to arrive at conclusions (Gigerenzer & Hoffrage, 1995). One assumption here is that it would be beneficial to prompt clinical decision makers and students to become aware of the fallacies that occasionally guide their cognitive strategies (Sandhu & Carpenter, 2006), to help them better to understand the faulty procedures they tend to apply in their own decision making. The research in this thesis is designed to raise awareness that competing hypotheses should be considered when making decisions, under the assumption that such awareness could reduce potentially fatal diagnostic errors (Lindgaard, 2005). Medical education is still directed towards memorizing large numbers of facts, and not on teaching sound decision strategies, even though there is little knowledge as to how these facts are recalled or used in the field (Round, 1999; Wood, 1999).

Ramnarayan, Winrow et al. (2006) trained junior doctors on a Web-based diagnostic reminder system one month before starting their resident training with patients. While working, the junior doctors would consult the Web-based diagnostic aid at which time they would decide to apply or reject the advice offered by the system. Results revealed that when junior doctors were offered web-based diagnostic aid, the number of 'unsafe' diagnostic errors decreased from 45.2% to 32.7% over a five-month period, a difference that is significant ($p < .001$). Furthermore, Ramnarayan, Winrow et al. (2006) stated that hospitals, in which more computers with Internet access were

available and readily accessible, had the highest number of users. Shanteau, Grier, Johnson, and Berner (1991) taught student nurses decision making skills in a course that focused on inappropriate information use, biased risk assessment, and suboptimal alternative evaluation. They found that the students acquired skills that was on par with that of an expert nurse upon completion of the course (Baron et al., 1991). These studies indicate that if appropriate teaching and support is made available in the form of a DSS, junior practitioners are likely to use it and benefit from it.

Decision making research indicates that people are able to store only a limited amount of information in working memory at any given time (Kalat, 2004; Wallsten, 1981; Wood, Wood, Wood, & Desmarais, 1999). Therefore, to compensate for the limited amount of storage of information, people will adopt various techniques and biases, which will be discussed shortly. However, not all of these methods are accurate or produce reliable results at all times (Wallsten, 1981).

There are several heuristics or decision making processes that could or might affect clinicians. These include serial position effects (which is important for the work in this thesis), that is, primacy and recency effects, confirmation biases, anchoring biases, the availability heuristic, and the algorithm method. Each of these is discussed in turn.

Problems such as overconfidence and representativeness have also been found to affect decision making. A selection of these heuristics and biases are discussed here, as they are among the more popular heuristics and biases discussed in the literature.

Primacy Effect

The primacy effect is defined as items that are presented first have a greater impact and are more memorable than items presented later (Anderson, 1981). Therefore,

the order in which clinicians receive information (symptoms) has a large impact on their initial diagnosis. In the presence of a primacy effect, the information presented first is more easily recalled, and is given more attention, than information presented in the middle or at the end of a string of items. A primacy effect may occur in the medical field when clinicians rely on new information that supports their current hypothesis while failing to consider alternative information that could lead to an alternative hypothesis (Ofir, 1988).

Wallsten (1981) found that fourth-year medical students and physicians distorted new information provided to them by using it to support their initial hypothesis. Participants were each presented with a booklet presenting hypothetical patients with different clinical histories. The information in the history is intended to rule out diagnoses leaving only one diagnosis, out of a list of 50. The information that was given at a later time was more reliable than information provided earlier, however the later information was used only to the degree in which it supported a hypothesis already formed and based on earlier, not as reliable, information. Information usage was assessed by a rating system whereby the participants rated the information as applicable or not applicable to the case. Additionally, diagnostically neutral information was perceived as positive if it supported the clinician's initial working hypothesis. Elstein (1976) found that physicians were more likely to perceive new information as relevant if it supported the initial working hypothesis; otherwise new information was not seen as relevant.

Lindgaard (1985) investigated the information integration strategies and base rate use by nurses when diagnosing case-specific information. Participants were estimating the probability of particular diagnoses for nondiagnostic case information and for cases

which supported one of two hypotheses. Nondiagnosticity exists when the information given provides equal support for more than one hypothesis. Lindgaard (1985) found a strong and very robust primacy effect in decisions in several disease-pairs tested. This thesis used a similar method using pediatric case information designed to enable the generation of nondiagnostic case information. Contrary to Lindgaard's study in which the case information was not based on actual patient data, the case information in the present study was based on probabilities calculated from actual cases on which the symptoms, the test information, and the final diagnoses were based. However, it was predicted that a primacy effect would also be found in this thesis, at least prior to exposure on the decision support system prototype tested here.

Confirmation Bias

Confirmation Bias occurs when a decision maker places an unjustifiably large amount of confidence in one hypothesis. Confirmation biases can be distinguished from primacy effects by the fact that the primacy effect is concerned with evidence presented at the beginning of a sequence of items, whereas a confirmation bias can be based on information presented in any order or in any serial position. Therefore, if new information does not concur with the initial hypothesis, the new information may not be seen as relevant. In a medical context, confirmation biases can occur when clinicians fail to consider diagnoses that are not in line with their initial diagnosis but which the case-specific evidence supports. As such, unwarranted weight is placed on one hypothesis even though conflicting evidence supports an alternative hypothesis. In Bayes' Theorem, which will be discussed in more detail later, two hypotheses are always presented, both of which should be considered in the evaluation of the case-specific information. In the

presence of a confirmation bias, the decision maker searches for cues or symptoms that support the initial diagnosis and ignores those symptoms that fit the alternative diagnosis, regardless of their serial position (Klayman & Ha, 1987). Confirmation bias can also occur when a clinician fails to alter or update their hypothesis in light of incoming non-supportive data. Although not tested in this thesis, when the confirmation bias is combined with an anchoring bias, the results can be extremely harmful (Sandhu & Carpenter, 2006) such that, the physician would pursue one hypothesis while ignoring symptoms that do not fit with it, thereby increasing the likelihood of an incorrect diagnosis.

Anchoring Bias

An anchoring bias occurs when a decision is based on the most salient item of information a decision maker receives. As is true for the confirmation bias, this item need not be presented first, which distinguishes it from the primacy effect. Rather it is an item that the decision maker has determined is the most important in the array on which to base a decision. This decision may have been made because the symptom was just being discussed, or because it was seen the patient history or for any other number of reasons. In the case of an anchoring effect, the decision maker would make a decision prematurely without consideration of other potentially even more important items (Sandhu & Carpenter, 2006). Consequently, anchoring can occur when the decision maker fails to adjust their hypothesis in the light of conflicting information. In the case of medical decision making, an anchoring effect could thus lead to an incorrect diagnosis.

Availability Heuristic

This is the propensity of a physician to miss less commonly known illnesses or diseases because these particular diseases are not in their repertoire. The opposite can also be true in that less commonly known, but perhaps recently occurring (such a recent outbreaks), diseases are typically discussed at great lengths in the literature; therefore these diseases have the propensity to be over-diagnosed. In the presence of the availability heuristic, diagnoses are based on diseases that are encountered more frequently, diseases that have been encountered recently, or diseases that have been discussed at great length due to their rarity (Croskerry, 2006; Wood, 1999). In a similar fashion, events that are easiest to recall appear the most probable to us (Combs & Slovic, 1979). It is claimed that the availability heuristic is likely to be observed more frequently in resident physicians than in their experienced colleagues (Sandhu & Carpenter, 2006) as they are not as well versed in differential diagnosis and are more likely to miss less frequently occurring diagnoses.

Algorithm Method

Another type of clinical decision making is described by Sandhu and Carpenter (2006) as the Algorithmic Method. This method uses flowcharts or an algorithm such as Bayesian probability theory, to determine the probability, or likelihood, that a diagnosis is correct. When decision makers are using Bayesian Theory, it would be beneficial for them to understand the principles underlying it, as failure to consider all symptoms presented into the algorithm is likely to affect the diagnosis. Sandhu and Carpenter (2006) describe the Algorithm methods as inflexible and as removing independent thinking. However, depending on the context it can be very valuable. It was applied

successfully in a recent study in which expert cardiac nurses provided tele-triage to cardiology and cardiac surgery patients on the phone. The algorithms enabled the nurses reliably to distinguish between very similar cardiac conditions requiring different actions to be taken. These varied between calling 911 immediately and resting for a few days at home (Somoza, Momtahan, & Lindgaard, 2006). Thus, the relative inflexibility is precisely what makes this method so advantageous in that biases are not brought into the equation and human error can be reduced. Therefore, even though it may not be perfect, this method is more advantageous than other approaches discussed. It would appear that clinicians, at least on occasion, have difficulty separating what is diagnostically relevant or important and what is not. Therefore, the issue of diagnosticity is discussed next.

Diagnosticity

Diagnosticity is the relative frequency with which a given symptom occurs under a range of diseases and also the degree to which a symptom exclusively supports one disease only, even when other diseases are being considered. Theoretically, a diagnostic symptom should influence a clinician's decision. When there are several items of information to consider, diagnosticity should be assessed by assigning a value, or weight, to each item of information. The combined value should then be used as evidence supporting one or more hypotheses. The weighting, once applied to a hypothesis, should shift in a systematic way in the light of new incoming evidence (Beyth-Marom & Fischhoff, 1983). From the perspective of Bayes' Theorem, the degree of diagnosticity of all case-specific information assessed in terms of both competing hypotheses determines if and how much the hypothesis currently entertained is supported (Fischhoff & Beyth-Marom, 1983). Therefore, one can distinguish a highly diagnostic item of information

from a low diagnostic item of information because information that is very high in diagnosticity will appear frequently under one condition but not at all, or at least very infrequently, under another. Both, high- and low-diagnostic information was incorporated into this study.

Nondiagnosticity occurs when the patient's symptoms are equally indicative of several hypothesis or diseases, in which case theoretically, the clinician's decision should not change (Lindgaard, 2005). Nondiagnostic information that supports several hypotheses can only be detected if the candidate hypotheses are being entertained; otherwise one of the above mentioned cognitive biases may come into play. In the present study, one half of the vignettes contained only highly diagnostic or low diagnostic information about the diseases related to the individual infants. As the symptoms in these vignettes equally supported the two competing diseases in the Bayesian equation, the vignettes were nondiagnostic, if Bayesian reasoning was being applied. Therefore, if a vignette contains a symptom highly diagnostic for disease A and a symptom highly diagnostic for disease B, then the two cancel each other out and the Bayesian reasoner should conclude that the symptoms are not useful in making a diagnosis. Thus, one prediction is that vignettes with a pair of symptoms there are both either highly diagnostic or nondiagnostic, an individual using Bayesian logic should discount that information and rely on base rate information.

The other half of the vignettes contained one symptom that was highly diagnostic for one of the diseases and another symptom that was low-diagnostic for the other disease. These vignettes were thus slightly diagnostic in the sense that the information supported both diseases but possibly should be interpreted as more supportive of the

disease with the highly diagnostic symptom. As shown by Fischhoff and Beyth-Marom (1983), participants in Bayesian decision tasks tend to consider the evidence provided by the individuating information (that is, the specific symptoms) only in terms of one of the two competing hypotheses instead of considering it in terms of both hypotheses. This, they argue, is because people generally have a poor understanding of the concept of diagnosticity. Therefore, nondiagnosticity of individuation information (in this case, the lack of agreement between symptoms) tends not to be detected unless it is palpably useless, in this case a vignette would have two symptoms that clearly indicate two different diseases.

Bayes' Theorem

Bayes' Theorem has been described as the optimal way to make decisions under uncertainty, including in the area of medicine (Berry, 2006; Lucas, 2001). It is a way of calculating the probability of a disease using an algebraic equation based on the prevalence of the disease and the symptoms presented by the patient (Johnson & Feldman, 1995). It provides the probability of one hypothesis being true, given a second hypothesis under consideration. Therefore, in light of new evidence or test results, the probability of an existing disease should be altered in a predictable way (Round, 1999). So, if a clinician wanted to know the probability of a hypothesis being correct, according to Bayes' Theorem, they should consider the probability of the current hypothesis over the probability of both competing hypotheses, including the alternative hypothesis, $(p(H)/P(H_1))$, where 'H' represents the current hypothesis and 'H₁' represents the alternative (Ofir, 1988). Taken together, the two hypotheses, 'H' and 'H₁', add to a probability of 1.0.

Bayes' Theorem comprises three components: the posterior probability, the prior probabilities, also called the base rates, and the individuating information. The purpose of using Bayes' Theorem is to calculate the posterior probability. This is the probability of Hypothesis 'H' being correct given the data 'D', represented as $P(H|D)$. The posterior probability is the combination of base rates and the individuating information. In the present context, it is information about a particular patient that produces probabilities distributed over a set of hypotheses (Berry, 2006; Lindgaard, 2005). Therefore, the posterior probability is the likelihood that a hypothesis is true given a set of observations (Cosmides & Tooby, 1996).

A base rate is the frequency with which an attribute or event is objectively present in a population (Koehler, 1996). The prior probabilities, or base rates, are thus the known distributions of the two competing hypotheses, 'H' and 'H₁', that are being considered and compared, represented as $P(H)$ and $P(H_1)$ respectively. In an experimental setting, the prior probability distribution contains the values assigned by the experimenter (Berry, 2006).

The individuating information is specific information pertaining to symptoms observed in an individual patient. The diagnosticity of all the information given about the patient determines the weight of the evidence in support of the hypothesis in terms of which diagnosticians provide a subjective probability. Individuating information is presented as $P(D|H)$ and $P(D|H_1)$. That is, the available data about an individual in terms of support for both the competing hypotheses. This evidence is thus supposed to be weighed in terms of its support for both hypotheses. Individuating information can be either diagnostic or nondiagnostic. Diagnostic evidence is more indicative of one

hypothesis than the other and nondiagnostic evidence supports both hypotheses equally or none of the two hypotheses.

Individuating information can have a large impact on decision making in that base rates are often neglected in favor of individuating information (Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1973; Koehler, 1996). In Kahneman and Tversky's (1973) original study, subjects were divided into two groups, a high- and a low base-rate group. The high base-rate group was informed that, in the sample from which the vignettes they would assess in the experiment, 70/100 were lawyers and 30/100 were engineers. These figures were reversed for the low base-rate group in which consequently 30/100 were lawyers and 70/100 were engineers. Both groups were given the same vignettes for which the individuating information was intended to be nondiagnostic, equally indicative of a lawyer and an engineer. Results indicated that the individuating information determined the outcome; the base rates were not used. This so-called base-rate-neglect phenomenon has been demonstrated in numerous studies since then (Kahneman et al., 1973, 1982; Koehler, 1996). For example, in a study by Ginosar and Trope (1980) participants were asked to assess the likelihood the information given actually described a target person. Three of the four groups were given nondiagnostic information and the fourth group was given diagnostically relevant information. Ginosar and Trope found that participants relied on the individuating information for their probability estimates and ignored the base rate information. Base rate information was only used when the individuating information presented was of no use. Base rate neglect was therefore also expected in the present study before, but not after, exposure on the decision support system prototype tested in this thesis.

It has been shown repeatedly that neither novice physicians nor medical students are competent in applying Bayes' Theorem in clinical situations. Eddy (1982) and Eddy and Clanton (1982) have shown that seasoned physicians have difficulty in applying the concept of Bayes' Theorem as well. More specifically, physicians have difficulty distinguishing between prior probabilities and posterior probabilities. However, an overwhelming majority of SMDM members agree that Bayes' Theorem is an important concept needing to be taught (Elstein et al., 1985). However, the possibility that a Diagnostic Decision Support System (DDSS) that incorporates Bayes' Theorem into its output could help future medical practitioners to propose accurate diagnoses faster by providing a narrow range of probable diseases medicine without also offering formal training in decision making was tested here.

Diagnostic Decision Support Systems

A DDSS is a computerized clinical aid that provides an output of possible diagnoses when it is provided with a set of symptoms. The purpose of a DDSS is to increase the accuracy of diagnosticians' decision making ability and decrease error rates. DDSSs have been developed in many areas of medicine; they offer a means of providing both quick and diagnostically accurate information to clinicians, including pediatricians, when time is of the essence and diagnostically ambiguous cases present themselves (Ramnarayan, Winrow et al., 2006).

These support systems are proliferating rapidly because they have the ability to store, sort through huge amounts of data and then quickly and accurately provide feedback to the clinician (Johnson & Feldman, 1995). Additionally, computerized DDSS's are on the rise as the need to keep up with the latest trends in medicine can

sometimes seem impossible for busy clinicians (Ramnarayan & Britto, 2002). Therefore, as the need to manage these increasingly large amounts of data increases, so does the need for an increase in the number of computerized DDSS's. This can provide some comfort to the patients who will benefit from the input, as DDSS's can provide advice faster that is also more accurate than a clinician's in the sense that, unlike human beings, the computer does not fall prey to decision biases (Lindgaard, 2005).

DDSSs provide medical decision makers with a range of possible diagnoses that might not otherwise have been considered and therefore, have the ability to reduce decision errors. It is hoped that by using a DDSS, and thereby becoming aware of that alternative diagnoses do exist, that when not using the DDSS the clinician will have learnt that more than one possible diagnosis is available. Because of this, DDSSs are said to be a technological solution to decision making error (Graber, Gordon, & Franklin, 2002) and as such, the term expert system has been used to describe DDSSs (Johnson & Feldman, 1995). Four DDSSs are discussed next, all of which relate specifically to DDSSs in pediatrics although one of these is more widely known and offers diagnostic advice on diseases in adults and older children as well.

Isabel

Isabel is a DDSS invented by a pediatrician whose infant daughter, Isabel, suffered from a disease that was very difficult to diagnose due to an ambiguous clinical picture. It is a web-based diagnostic tool that aims to reduce error in decision making by aiding clinicians in their decision making process. Isabel provides a checklist of possible diagnoses in response to the symptoms and other clinical findings entered by a

pediatrician. Larissa et al. (2006) found that the use of Isabel led pediatricians to consider diagnoses they would otherwise not have thought of.

Several studies have found that users of Isabel report this tool to be very helpful at supporting medical decision making (Larissa et al., 2006; Maffei, Nazarian, Ramnarayan, Thomas, & Rubenstein, 2004). However, there are also drawbacks of Isabel. Firstly, Isabel has not yet had an impact on the education of medical students (Maffei et al., 2004). This is because the output can be overwhelming in the sense that Isabel provides up to 10 diagnostic categories for the pediatrician to search through, each of which may point to up to 35 different diseases. These diagnostic categories are gastrointestinal disorders, nervous system disorders, shock states, urologic disorders, infection diseases, neoplastic diseases, endocrine system, metabolic diseases, respiratory system disorders, cardiac disorders, liver disorders, nephrology, hematology, allergic disorders, and toxicology. Additionally, for each diagnosis it provides a range of medical articles on each individual disease. Therefore, although it is claimed to be very popular, Isabel does not exactly provide quick diagnostic aid in a high-pressure clinical setting, but rather supplemental reading material for those who have the time. Secondly, Isabel does not provide the probability associated with different diagnoses related to a set of symptoms. Therefore, there is no way of knowing if one diagnosis is correct, or more probable, than another. This is because symptoms, signs, laboratory findings, and so on, are not quantified in medicine. Since Isabel's database is derived from a cross-section of medical texts, it has no more accurate information than these. Medical text books simply provide undifferentiated lists of symptoms associated with each disease. The frequency with which each symptom may be expected to occur under a given disease, that is, its

diagnosticity, is not stated. The fact that many symptoms occur in a range of diseases is not stated explicitly. Lastly, Isabel does not provide a list of further symptoms to be explored or tests to be performed to increase the probability of a specific diagnosis. In an effort to investigate how a DDSS like Isabel could be improved, the current study provided a short list of the five most probable diagnoses (described in more detail later on) thereby reducing the search time to find the correct diagnosis.

ePocrates

ePocrates is a handheld personal digital assistant (PDA) decision support tool used to provide guidelines for pediatric dosing of medication (Longhurst & Hahn, 2005). As mentioned in Longhurst and Hahn (2005), the ePocrates manufacturers conducted a survey of ePocrates users to determine its use and effectiveness. Results indicate that 75% of respondents reported use of their PDA more than six times a day, most of which was to access information on drugs. Furthermore, 80% of the pediatricians reported that the use of their PDA aided safer medical practice. However, these results may be skewed due to a Hawthorne effect and/or a social desirability bias, because the ePocrates owners conducted the survey themselves.

A second survey was conducted of 870 physicians and medical students who had used hand held devices and who use ePocrates (Rothschild et al., 2000). Of the respondents, 85% had used ePocrates for at least two months, in some instances more. The survey addressed questions relating to: product evaluation, impact on physician practice and efficiency, impact on physician behavior for information-seeking and decision-making, and impact on patient care. Questions were asked regarding the participants' work habits (i.e., search time for drug related information) and patient

practice (i.e., parent satisfaction) both before and after using ePocrates. Results indicate that clinicians saved time searching for drug-related information: 90% of participants required less than one minute to find information with ePocrates whereas 55% required more than a minute with previous information finding techniques. It also improved the quality of care of their patients by improving drug related decisions: 83% reported that patients were better informed, and 81% reported improved drug related decisions. Of notes is that Rothschild et al., in conducting this study, were representing ePocrates Inc, therefore to the extent that these results are positive, they may also be skewed due to a bias such as the Hawthorne effect.

It is apparent then that a tool that can provide relevant information quickly would be used by physicians if it were available. Furthermore, easily accessible information was seen as helpful and provided perceived safer working conditions for physicians as they were given more accurate information to with which to diagnose. Therefore, if the DDS being used in the current study is seen as useful, then it appears that it may aid in safer working conditions.

MEDITEL Pediatric Diagnostic System

MEDITEL Pediatric Diagnostic System (MEDITEL-PEDS) is a DDSS for use in pediatric care which provides a list of alternate diagnoses to clinicians once a list of symptoms is entered into the database. The system differentiates one disease from another on the basis of abnormal findings, therefore highly diagnostic symptoms. The data were entered into the system from literature reviews of various states focusing on abnormal results for each disease (Barness, Tunnessen, Worley, Simmons, & Ringe, 1974). A diagnosis is arrived at by observing abnormal symptoms, physical findings, and

laboratory results from a list provided. An output, which takes approximately five minutes to generate because information was sent through a teletypewriter over phone lines, was then provided that can have anywhere from four to 70 different diagnoses (Barness et al., 1974).

It is not clear exactly why MEDITEL-PEDS is not in use today (Johnson & Feldman, 1995). The literature on this DDSS appears to stop suddenly. One would suspect it is because of the lengthy input process and the time-consuming wait of the output process. As a web-based tool, Isabel provides output much faster, but both these DDSSs suffer from the potentially very long lists of possible diagnoses, which may lead to greater, but not speedier, diagnostic accuracy. MEDITEL-PEDS is similar to NeoPeDSS, the prototype tested in this study, in that they both use a Bayesian approach to estimate the probability of a disease based on the patient's symptoms (Johnson & Feldman, 1995; Ramnarayan & Britto, 2002). However, where MEDITEL-PEDS provides an output of up to 70 different diagnoses, NeoPeDSS provides an output only of the five most probable diseases. Some advantages of a short list of probable diseases include the dramatically reduced reading time and that the user must only consider the most probable diseases. By contrast, when used in a training setting, one advantage of a longer list of diagnoses is that it shows all possibilities of diseases, thereby offering the opportunity to read up on them at a later time.

The waiting time for a diagnosis output of MEDITEL-PEDS was approximately five minutes, while NeoPeDSS provides the most probable diagnoses in seconds. The MEDITEL-PEDS database is derived from a literature review conducted by the database

contributors whose disciplinary background is not known, while NeoPeDSS data is based on actual patient data, thereby providing actual, rather than hypothetical, probabilities.

Wexler, Swender, Tunnessen, and Oski (1975) showed that when residents were presented with results from MEDITEL-PEDS the number of diagnostic errors decreased. Although the participants did not use the MEDITEL-PEDS directly, the output of the system was given to the participants and, based on this; participants would decide whether to alter their diagnosis. Participants were not made aware of the lengthy input process, which, it is believed, ultimately led to clinicians' infrequent use of the system (Ramnarayan, Winrow et al., 2006). Technological developments since the MEDITEL-PEDS was first released have allowed future generations of DDSSs to deliver information much faster and universally to desktops or hand held equipment. Lastly, MEDITEL-PEDS also suffers from the problem that its database relies on medical texts, which still do not provide quantified symptom diagnosticity information. The major advantage of NeoPeDSS is that the information provided in the output are probabilities based on the symptoms the diagnostician has entered.

NeoPeDSS

NeoPeDSS differs from the previously mentioned DDSS by including the diagnosticity of signs and symptoms in its database, which is used to calculate an output. NeoPeDSS therefore provides both the five topmost likely diagnoses as well as the probabilities of these, based on input the pediatrician provides. Therefore, although other DDSSs provide an output of possible conditions, NeoPeDSS offers the relevant probability of each of the five diagnoses. Pediatricians are thus made aware of the possibility that more than a single disease could account for the findings entered for a

given sick infant. One of the main functions of NeoPeDSS is an attempt subtly to raise awareness of the concept of diagnosticity and to show how probabilities are related to diagnosticity.

Data. NeoPeDSS database is based on 1200 sick infants' records from the Children's Hospital of Eastern Ontario (CHEO) Neonatal Intensive Care Unit (NICU). Of these 1200 cases, 97 match the symptoms and test results depicting respiratory distress as defined by the World Health Organization (WHO) International Classification of Diseases, version 10 (ICD10). The target diseases were selected by Dr. Walker, the highly experienced pediatrician who is also a member of the NeoPeDSS prototype development team. The aim was for the target diseases to represent a suitably ambiguous medical condition and one that occurs reasonably frequently in infants. A medical record librarian identified these cases based on a list of signs, symptoms, and laboratory findings, provided by Dr. Walker, which were indicative of respiratory distress. For cases in which information was missing in the original database, such as the final diagnosis, the relevant patient records were obtained to ensure that the medical records for the NeoPeDSS' database would be complete. As such, the diagnosticity of each sign, symptom, and laboratory finding was obtainable allowing for the calculation of the value of the individuating information, $P(D|H)$ and $P(D|H_1)$, and hence of the posterior probability, $P(H|D)$.

By encouraging clinicians to consider a wider range of possible diagnoses when presented with a set of ambiguous symptoms, it was anticipated that this should improve the accuracy and speed of their diagnostic decisions.

The Present Study

The purpose of the present study was to demonstrate that a DDSS may provide valuable assistance to diagnosticians for cases in which the clinical picture is ambiguous. More specifically it is anticipated that, after using the DDSS, clinicians will realize that there may be more than one diagnosis to consider and thus, would increase their decision making abilities. Accordingly, that the following Hypotheses were tested:

- 1) Hypothesis 1: In accordance with numerous previous findings, base rates will be ignored in the subjective probability estimates made before exposure but not after exposure on NeoPeDDS.
- 2) Hypothesis 2: Vignettes containing high-diagnostic information for both diseases will lead to higher probability estimates than vignettes containing low-diagnostic information. This will show, in accordance with previous findings, that participants are using the individuating information and ignoring the base rates.
- 3) Hypothesis 3: Subjective probability estimates will vary by serial position of the symptoms presented in the individuating information. If supported, this would indicate that participants are largely affected by the individuating information, therefore are not paying attention to the base rates.

Method

In preparation for this thesis, a DSS called NeoPeDDS (Neonatal Pediatric DSS) was designed and developed by a biomedical engineering student, Catherine Pyper (Pyper, 2007) at Carleton University, under the supervision of Dr. Monique Frize. The

database contained information from infants who had been admitted to the Neonatal Intensive Care ward at CHEO, and the aim was to provide diagnostic support to residents. Respiratory distress was the condition chosen by Dr. Robin Walker, an expert pediatrician, because it occurs frequently in infants, the symptoms are ambiguous, and can therefore be indicative of many different diseases. NeoPeDSS relies on Bayes' Theorem for its probability calculations. Because it is based on real patient data, it yields actual and correct probabilities.

Participants

A sample of 40 resident pediatric participants was recruited from hospitals with Neonatal Intensive Care Units from Canada and the United States. Participants were recruited through posted announcements and via email (see Appendix A for recruitment notice). They were paid \$100 each upon completion of the experiment which took approximately one hour. The experiment was presented on the Internet, and participants were able to complete the study in their own time.

Materials

Interface. NeoPeDSS (Pyper, 2007; Pyper et al. 2008) is very straightforward and easy to operate. It consists of an input screen (see Figure 1 and 2) and an output screen (see Figure 3). On the input screen, the clinician selects the first relevant symptom the patient is exhibiting from a drop-down menu under *symptom* (see Figure 1) at which point another more detailed subset list of values associated with that symptom is provided, for example the APGAR score, under the heading *description* (see Figure 2). The APGAR score is based on observations of an infant's Appearance, Pulse, Grimace, Activity, and Respiration. It is taken very shortly after birth, after 1 and 5 minutes

respectively and may be repeated perhaps at 10 and 15 minutes if it is or remains very low. A value between 0 and 2 is assigned to each variable; the final score ranges from 0 to 10. A score of 7-10 is considered ideal. The clinician selects the relevant value from the drop down list. At this point she can either press the 'diagnosis' button or enter additional symptoms as they see fit before diagnosing. The 'Diagnose' button can be pressed as many times as the clinician wishes to check on the disease probabilities. When the button is pressed the information is analyzed using Bayes' Theorem. An output screen then appears providing a list of the five most likely diagnoses and the probability of each (see Figure 3). The probability of each diagnosis is presented in a column labeled *Likelihood* in the lower part of the screen. As more symptoms are entered into the database, the list of possible diagnoses that is generated will decrease, with the optimal output resulting in only one diagnosis and a probability of 1.0, expressed as a percentage (i.e. 100%).

The screenshot shows the NeoPeDSS input screen. The window title is "ReDSS" and the menu bar includes "File", "Edit", and "Help". The main section is titled "Patient Case" and contains a table with two columns: "Symptom" and "Description". The "Symptom" column has a dropdown menu with the following options: AbdLiver, Activity, AirEntry, Appar 1, Appar 5, Birth Weight, Blood Pressure, and Breath. To the right of the table are buttons for "Delete" and "DeleteAllButton". At the bottom right of the "Patient Case" section is a "Diagnose" button. Below this is the "Diagnostics" section, which includes a "History" dropdown menu and a table with three columns: "ICD10 Dx", "Description", and "Likelihood".

Figure 1. NeoPeDSS input screen prior to data entry.

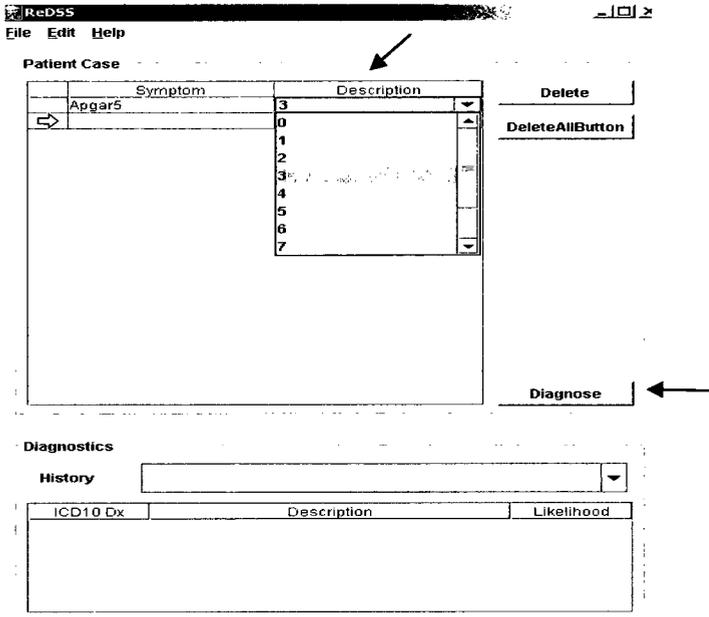


Figure 2. NeoPeDSS panel showing further description options available through a drop-down menu after the Apgar5 score has been selected

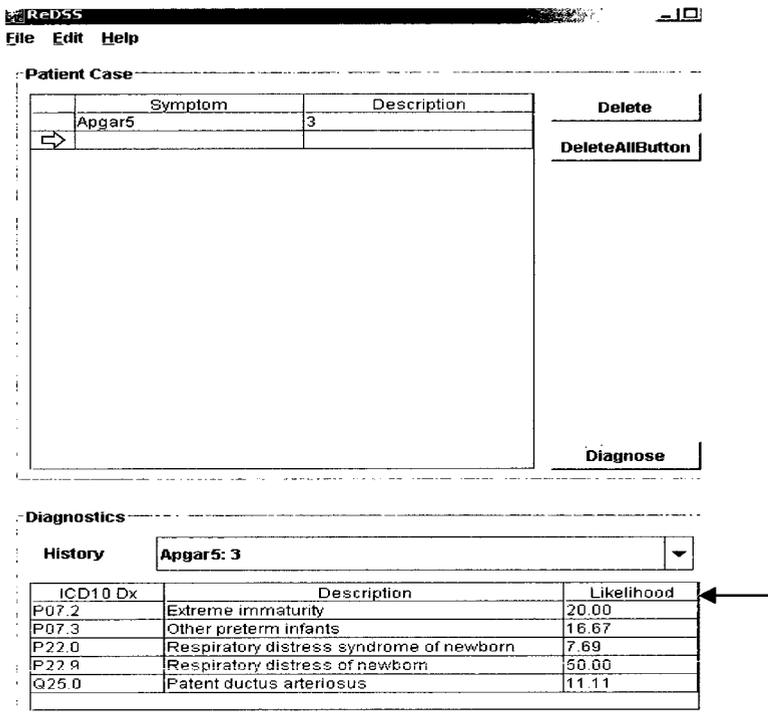


Figure 3. Output probabilities upon entry of an Apgar5 with a score of 3

Each participant received a login to the experiment and an informed consent form by email (see Appendix B). A brief questionnaire collecting a limited amount of demographic information about the participants was prepared (see Appendix C). A cover story and 24 fictitious vignettes were also prepared (Appendix D). The cover story provided background information justifying the limited amount of information provided about the individual patients in the vignettes. Another questionnaire collecting information on their method of diagnoses is presented in Appendix E. Introduction and learning materials were prepared for the 20 participants in the experimental group and they were given introductory exposure on NeoPeDDS and then asked to complete a set of five cases using NeoPeDSS (see Appendix G). The cases consisted of three symptoms each, with each case having one symptom either high or low for RDS, one symptom high or low for TTN and one symptom being non-diagnostic (neutral) for each of the two diseases. Provisions were made for participants to propose an initial diagnosis (a diagnosis before using NeoPeDSS) as well as a final diagnosis (a diagnosis made with the aid of NeoPeDSS) for each case as well as indicating their level of confidence in each diagnosis.

Design

In Phase 1, participants were asked to assess the probability of respiratory distress in a set of vignettes each describing symptoms and signs of a fictitious infant presenting with an ambiguous clinical picture. In Phase 2, exposure and hands-on practice on NeoPeDSS was provided for half the participants, and in Phase 3, the same ambiguous vignettes, albeit with different names, were presented again in a different random order to determine the effects of the NeoPeDSS exposure. The exposure allowed participants to

learn about decision making and be acquainted with NeoPeDDS .The experimental design is outlined in Table 1 below.

Table 1

Outline of the experiment

Group	N	Phase 1	Phase 2	Phase 3
Residents	20	Demographic questionnaire, 24 vignettes, and a diagnosis questionnaire	NeoPeDSS exposure and 5 cases to be diagnosed	24 vignettes and post-experimental questionnaire
Residents	20	Demographic questionnaire, 24 vignettes, and a diagnosis questionnaire	Read an article on DSS's	24 vignettes and post-experimental questionnaire

The demographic questionnaire was designed to gather information such as age, if in the past they used a DDSS, what year they were at in their studies etc. All 40 participants were residents and only two had taken a decision making course. The diagnosis questionnaire was designed to collect information regarding what information they used that lead to their final decision or diagnosis. Lastly, the post-experimental questionnaire, administered in Phase 3, was designed first, to collect information regarding what information they used that led to their final diagnosis and second, to

gather information regarding the estimated rate of frequency of occurrence of the symptoms.

One half of the participants were assigned at random to the High Base Rate group (HBR) in which the majority of patients (72/100) were said to have been diagnosed with Respiratory Distress Syndrome (RDS) and the remainder (28/100) with Transient Tachypnoea (TTN). The other half were assigned to the Low Base Rate group (LBR) in which 28% of cases were said to have RDS and 72% TTN. All participants estimated the probability that each patient described in a vignette has RDS. The same vignettes were used in Phases 1 and 3 but the infants described were given different names. Vignettes were presented in a different random order for each participant and in each Phase.

Cover Story

After an extremely violent hurricane, several falling trees hit and caved in the roof of a large hospital for sick children and infants in the vicinity of where you work as a pediatrician. Patients rescued from there are arriving at your hospital by the dozen. As a pediatrician, your job is to attend to the newborn infants immediately upon their arrival, to propose a preliminary diagnosis, and to commence initial treatment where needed. Patient records have not yet arrived. Consequently, you know very little about each infant. In fact, all you have been told is that, of the 100 infants expected to arrive in the next few hours, 72 (or 28) have been diagnosed with respiratory distress syndrome and 28 (or 72) with transient tachypnoea.

Case example

Infant Gonzales, born about one hour ago, has just been admitted with increased interstitial markings. She appears slightly restless, and her breathing shows clear intercostals retractions.

How likely do you think it is that infant Gonzales is one of the 72 (or 28) cases of respiratory distress syndrome? _____%

Procedure

Upon agreeing to participate in the experiment, participants first received an informed consent form with an introduction to the experiment. Following this, they were asked to complete a short demographic questionnaire.

In phase 1 each participant read the first cover story, followed by an illustrative example showing how to complete the 24 vignettes, which were then presented, one at a time, with no opportunity to look back over a completed vignette. Participants were then asked to complete the questionnaire regarding their diagnosis methods. In Phase 2, one half of the participants were instructed on how to use NeoPeDSS and shown an illustrative example as well given an opportunity to interact with it (see Appendix F). When ready, they received five cases, one at a time (see Appendix G) that they diagnosed both with and without using NeoPeDSS. The other half of the participants were presented with an article to read. In Phase 3 participants received the same cover story and again were presented with the 24 vignettes, one at a time until all had been completed. Participants were asked to describe how they diagnosed the cases depicted in the vignettes, the level of confidence they had in their diagnoses, and the estimated frequency of occurrence of each symptom used in the vignettes under both diseases (see Appendix

H). Lastly, following the completion of the study, all participants were debriefed (see Appendix I) and excused.

Data analysis

Phase 1 and 3. A $2 \times 2 \times 2 \times 2 \times 2$ mixed ANOVA with repeated measures was conducted for base rate group (HBR, LBR), phase (1, 3), exposure (with exposure, without exposure), diagnosticity RDS (high, low), and diagnosticity TTN (high, low). This ANOVA was done to test hypothesis 1 and 2. If participants neglect the base rates in Phase 1 before they were exposed to NeoPeDSS, but not in Phase 3 after exposure (Hypothesis 1), then there should be an interaction between base rate and phase, such that estimates in Phase 3 should be closer to the base rate in that condition than estimates in Phase 1. If vignettes containing highly diagnostic information for RDS and TTN lead to higher probability estimates than vignettes that have low diagnosticity for both the diseases (Hypothesis 2), then there should be an interaction of the diagnosticity of the RDS and TTN conditions. Lastly, to test hypothesis 3, that a primacy effect would be evident in phase 1 but not in phase 3, a 2×2 mixed-design ANOVA for type of symptom (RDS-H, TTN-H), and serial position (1, 3) was conducted. This was done for phase 1 and 3 separately to avoid having each scenario being counted more than once in the same analysis

Phase 2. A qualitative data analysis identified the number and types of symptoms participants entered into NeoPeDSS before diagnosing, the sequence in which they entered the symptoms, whether they entered all the symptoms, and their final diagnosis with its probability.

Results

Following some brief statements below, the results are then presented in three parts. First results are discussed in terms of the three hypotheses. Second, diagnoses offered by participants who had exposure in Phase 2 are dealt with. Third, an analysis of the estimated frequency of occurrence of all the symptoms used in both the diseases and provided in Phase 3 is presented.

Initially it was intended to analyze data from participants who had taken decision making courses separately and compare them with data from participants who had no formal background in decision making. However, as only two participants had taken decision making courses, that analysis was not carried out.

Participants

Participants were 12 males and 28 females, 70% of whom were between the ages of 25-30, 10% were ages 30 or older, and 5% were ages 20-25. Eighty-five percent of participants rated themselves as intermediate in terms of computer use while 12.5% rates themselves as expert, and 2.5% as novice. Regarding decision making and knowledge of DDSSs, 92.5% of participants reported they would use a DDSS if one were available; however 97.5% have never used one. Additionally, only 27.5% of participants were familiar with any decision making models which is not surprising given the lack of formal training on decision making.

All data analyses were completed ensuring that all the assumptions underlying the Analysis of Variance (ANOVA) had been met. The test of Normality and Absence of Outliers were satisfied and were not violated. This can be seen in Appendix J and K

respectively. There were no missing data; Test of Homogeneity of Variance and Sphericity for each analysis are presented with the respective analyses.

Testing the Hypotheses

Hypothesis 1 predicted that base rates would be ignored in the subjective probability estimates (dependent variable) made in Phase 1 but not in Phase 3. Mauchley's Test of Sphericity using the Huynh-Feldt formula indicated that Sphericity was not violated $p = 1.000$, therefore the variances for all the set's of scores were sufficiently similar to trust the results (Tabachnick and Fidell, 2007). Levene's test of Equality of Variances indicates that the assumption of homogeneity of variance has been met.

The main effect for base rate group was significant, $F(1, 36) = 15.81, p < .001$. As can be seen in Figure 4, the probability estimates of the High Base Rate ($M = 68.4, SE = 2.08$) group were higher than those of the Low Base Rate group ($M = 56.6, SE = 2.08$), indicating participants did not ignore the base rate information. The LBR group estimates were somewhat higher than would be expected if participants had been relying on the base rates exclusively. The main effect for exposure was non-significant, $F(1, 36) = 1.80, p > .05$. Thus, although estimates given by the two base rate groups differed significantly, exposure to the DDSS had no effect on the probability estimates in Phase 3 as there was also no significant effect of Phase.

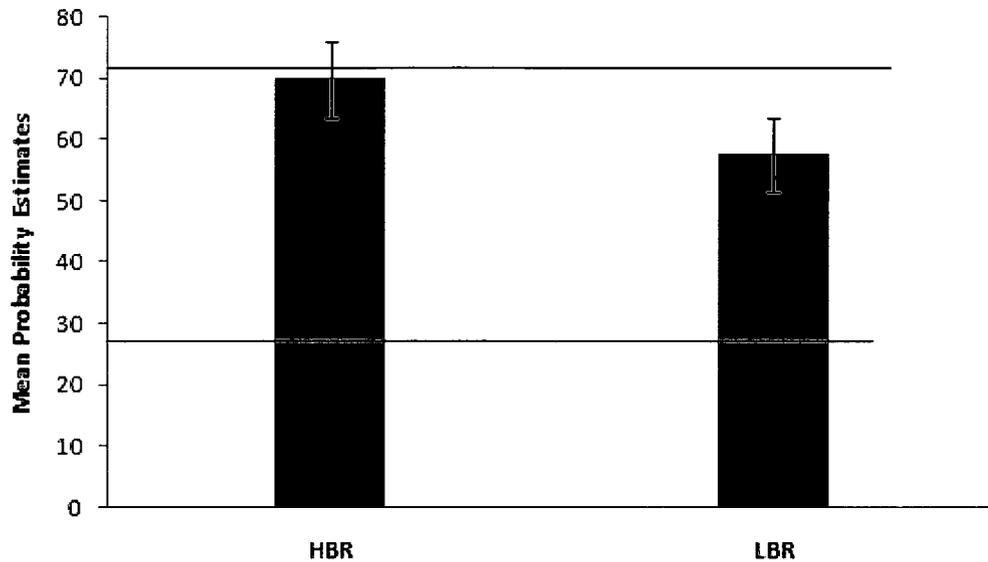


Figure 4. Mean probability estimates for High and Low Base Rate group

As can be seen in Figure 5 below, even though there was no significant interaction, the High Base Rate group with exposure remained closer to the base rate values in Phase 3 than the High Base Rate group without exposure, which is in the predicted direction. However, estimates for the Low Base Rate group with no exposure were somewhat closer to their base rates values in Phase 3 than the Low Base Rate group with exposure, which is not in the predicted direction. Therefore, hypothesis 1 was refuted in that base rates were being used and exposure was shown to have no effect.

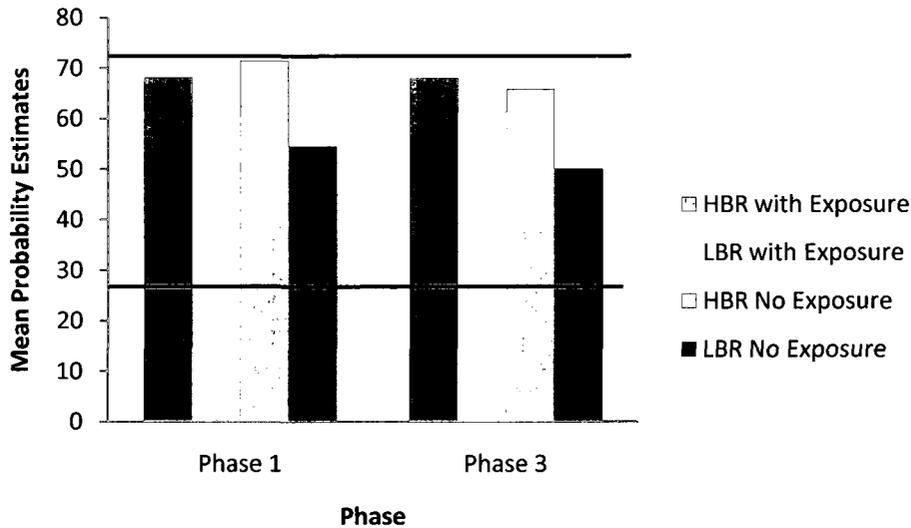


Figure 5. Mean probability estimates for base rate groups, with and without exposure for Phases 1 and 3.

Hypothesis 2 stated that vignettes containing high-diagnostic information for RDS and TTN (HH) would lead to higher probability estimates than vignettes containing low-diagnostic information (LL), data are shown in Table 2 below. Results indicate that there was a significant main effect for RDS, $F(1, 36) = 15.89, p < .001$, while the main effect for TTN approached significance, $F(1, 36) = 3.73, p = .061$. Thus, diagnosticity was a predictor of estimates when it was high for both diseases. When a highly diagnostic symptom for RDS was included in the vignette, it had a greater influence on estimates than a highly diagnostic symptom for TTN. Because participants were always asked to estimate the probability of TTN, it is likely they had an availability bias towards RDS.

As shown in Table 2, participants gave the lowest estimates of RDS when both symptoms were low in diagnosticity and gave the highest estimates when both symptoms were highly diagnostic. Overall these patterns show that participants were using the

individuating information about each case (i.e., the specific information from the symptoms) but in somewhat complex ways.

Table 2

Probability estimates for HH, LL, HL, and LH vignettes for RDS and TTN

		RDS	
		High	Low
TTN	High	66.5	61.6
	Low	64.2	57.4

Hypothesis 3 predicted a serial position effect for type of symptom (RDS-H, TTN-H), and serial position (1, 3) of the symptoms. This was performed separately for Phases 1 and 3. Because exposure was found to make no difference as stated earlier, the data were collapsed across exposure in the next two analyses.

Phase 1

Mauchly's test of Sphericity using the Huynh-Feldt indicated that Sphericity has not been violated as symptom, serial position, and symptom by serial position were all $p = 1.00$.

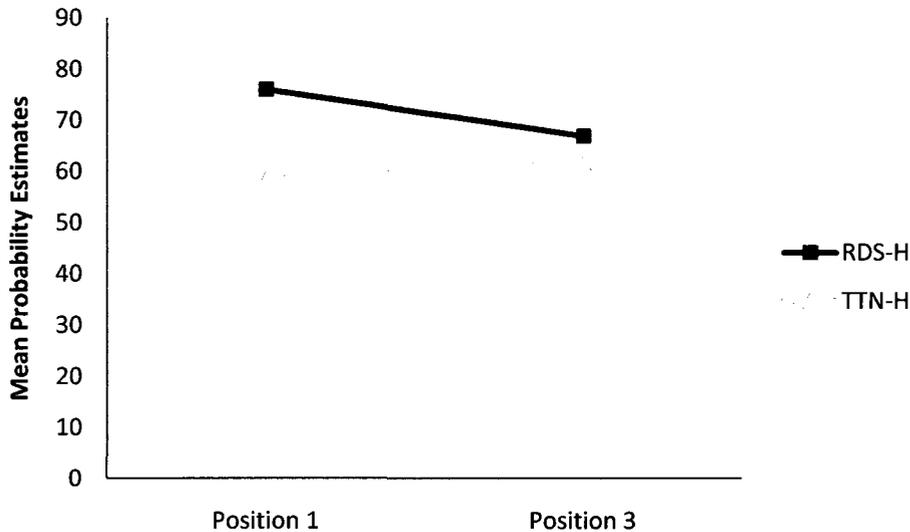


Figure 6. Probability estimates in Phase 1 by Serial position of the highly-diagnostic symptoms in each vignette

As can be seen in Figure 6, estimates of the probability of RDS were higher when the RDS-H condition was the first symptom and lower when the TTN-H symptom was first, supporting the hypothesis that clinicians put more weight on the first symptom they encounter than on later information. This interpretation was supported by the results of the analysis. Estimates of RDS were higher in position 1 than position 3 (67.2 vs. 64.2), $F(1, 39) = 5.45, p < .05$, and higher for the RDS-H condition than for the TTN-H condition (71.4 vs. 59.9), $F(1, 39) = 40.71, p < .001$. Lastly, the 2-way interaction of symptom and serial position was significant $F(1, 39) = 11.62, p < .01$, see data in Table 3 below and figure 6 above.

The interaction suggests that the RDS-H symptom had the largest effect on the decision, as the estimates in the TTN-H conditions were quite similar. Again, these results show that participants do use individuating information and that they show a primacy effect.

Table 3*Phase 1 mean and standard deviations for 2-way interaction*

	Position 1		Position 3	
	M	SD	M	SD
RDS-H	76.0	15.1	66.9	17.7
TTN-H	58.3	16.0	61.5	13.4

Phase 3

As in Phase 1, a 2 x 2 ANOVA was carried out for type of symptom (RDS-H, TTN-H) and serial position (one, three).

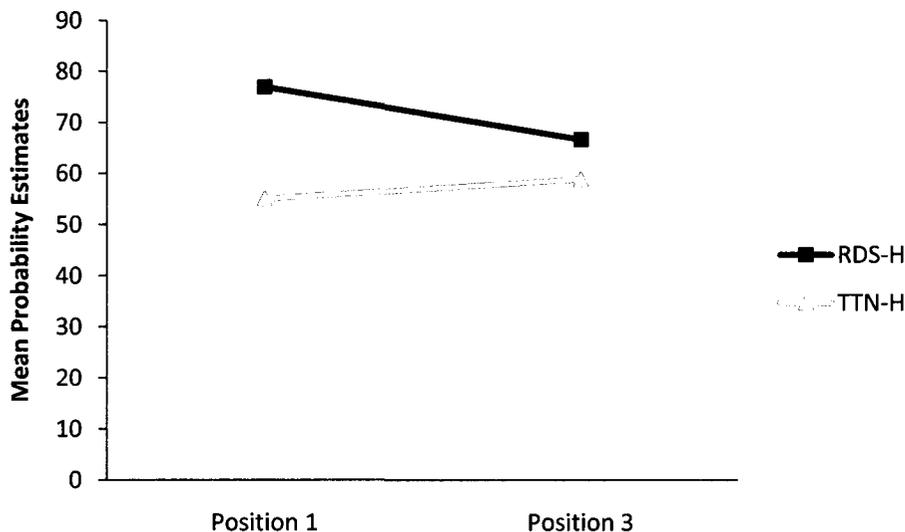


Figure 7. Probability estimates in Phase 3 by Serial position of the highly-diagnostic symptoms in each vignette

Results indicate that the main effect for serial position (position 1 $M = 66.0$, $SE = 2.1$, position 3 $M = 62.7$, $SE = 2.5$) was significant $F(1, 39) = 10.86$, $p < .01$. As in Phase 1, this suggests the presence of a primacy effect which contradicts Hypothesis 3 for Phase

3. Main effect for symptoms (RDS-H $M = 71.8$, $SE = 2.4$, TTN-H $M = 56.9$, $SE = 2.4$) was also significant $F(1, 39) = 70.06$, $p < .001$. The 2-way interaction between symptom and serial position was significant $F(1, 39) = 28.89$, $p < .001$, see table 4 below for means and standard deviations, and figure 7 above for interaction.

Table 4

Phase 3 mean and standard deviations for 2-way interaction

	Position 1		Position 3	
	M	SD	M	SD
RDS-H	76.9	13.9	66.6	19.2
TTN-H	55.0	16.9	58.9	15.2

Taken together, the above results indicate that hypothesis 3 is refuted as participants did show a primacy effect in both Phases 1 and 3. Therefore, participants were affected by the serial position of the symptoms which also indicates they were again using the individuating information.

Diagnoses proposed in Phase 2

This analysis was performed to see if participants who had been given exposure to NeoPeDDS had benefited from it by making better decisions while using the tool. It will be recalled that, in Phase 2, one half of the participants were exposed to NeoPeDDS. Following this, they were asked to read five vignettes, one at a time, which had not been presented before. For each they proposed an initial diagnosis first. Then they used NeoPeDDS to help determine the most likely diagnosis upon which they proposed a final diagnosis. A paired samples t-test was completed on the first and final diagnoses. Due to

an oversight when preparing the study, those who did not have exposure to NeoPeDDS did not complete this part of Phase 2. Figure 8 shows the number of correct diagnoses in Phase 2 before and after use of NeoPeDDS. The t-test for related samples was significant $t(19), p < .008$, suggesting that NeoPeDDS aided in increasing the number of correct diagnoses. Therefore, the NeoPeDDS exposure did help participants to make better decisions. It did not have an effect on their overall decision making accuracy, however, given that there were no changes in decision making in Phases 1 and 3.

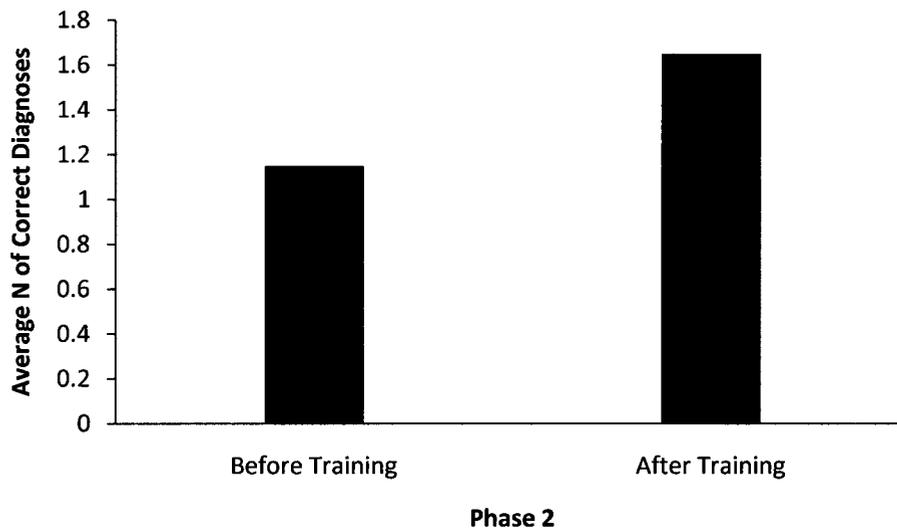


Figure 8. The average number of correct diagnoses in Phase 2 before and after use of NeoPeDDS

Estimated Frequencies of Occurrence

Lastly, to determine the degree to which participants agreed with the clinicians who had selected the High and Low diagnostic symptoms for the two diseases, the estimated frequency of occurrence of each symptom was analyzed under each of the two diseases. Participants were given the same list of all symptoms supportive of both diseases, presented in random order. In one instance they asked to estimate the frequency

of occurrence of these symptoms in a sample of 100 infants all diagnosed with Respiratory Distress Syndrome, and in the other instance they were asked to estimate the frequency of occurrence of the symptoms in sample of 100 infants all diagnosed with Transient Tachypnoea. Participants were asked to judge in how many of the 100 infants all diagnosed with disease X they would see each symptom in the list. These estimates were analyzed in a 2 (symptom: RDS, TTN) by 2 (diagnosticity: High, Low) repeated measures ANOVAs, separately for each disease.

Mauchly's test of Sphericity using the Huynh-Feldt indicates that Sphericity was not violated. Symptom, diagnosticity, and symptom by diagnosticity are $p = 1.00$, $p = 1.00$, and $p = 1.00$ respectively. As seen in Figure 9 it appears that when presented as a symptom of RDS, the high-diagnostic RDS symptom was seen as occurring more frequently than the low-diagnostic RDS symptom. Accordingly, the high-diagnostic TTN symptom was seen to occur less frequently when presented as a symptom of RDS whereas the low-diagnostic TTN symptom was seen to occur somewhat more frequently. The reverse was true when the same symptoms were judged under TTN.

For Respiratory Distress Syndrome (RDS), the main effect for symptom was significant $F(1, 33) = 67.44$, $p < .000$ with RDS being higher than TTN. Thus, symptoms indicative of RDS, when presented under RDS, were thought to occur more frequently than TTN symptoms. Similarly, the main effect for diagnosticity was significant, $F(1, 33) = 35.41$, $p < .000$ with the symptoms high in diagnosticity perceived to occur more frequently than the symptoms low in diagnosticity. Therefore, participants were sensitive to the diagnosticity of the symptoms.

The interaction of symptom and diagnosticity was significant $F(1, 33) = 61.32, p < .000$ as can be seen clearly in Figure 9. When the list of symptoms was presented under RDS (left data plot), residents judged that the high-diagnostic RDS symptom would occur more frequently than the low-diagnostic RDS symptom. The high-diagnostic TTN symptom was seen to occur less frequently than the high-diagnostic RDS symptom, whereas the low-diagnostic TTN symptom was deemed to occur as frequently as the low-diagnostic RDS symptom. The low diagnostic TTN symptom was also seen to occur more frequently than the high-diagnostic TTN symptom.

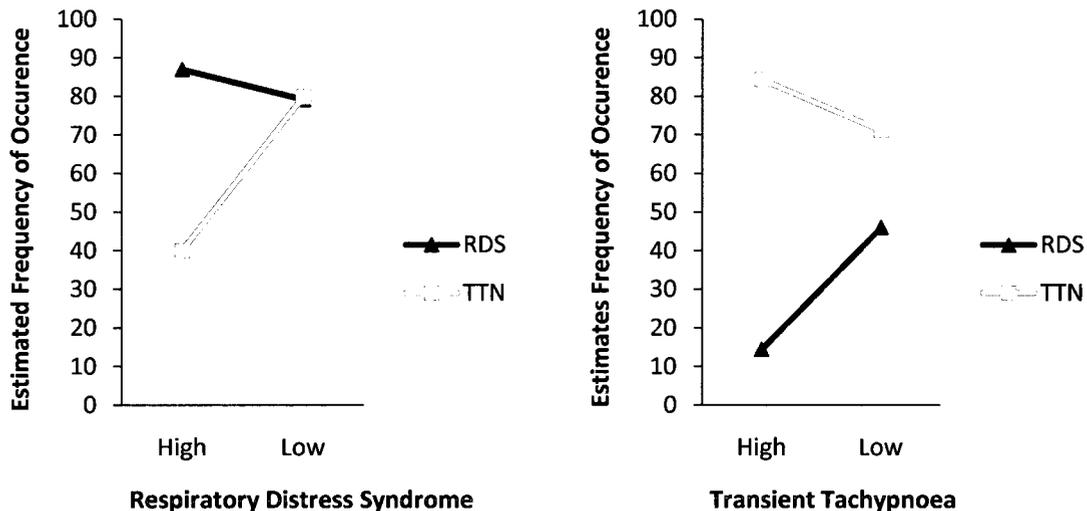


Figure 9. Estimated frequency of occurrence for list of symptoms presented as either a symptom of RDS (left panel) or of TTN (right panel).

The same analysis conducted for TTN indicated that the main effect for symptom was significant $F(1, 33) = 228.61, p < .000$, as was the main effect for diagnosticity $F(1, 33) = 6.58, p < .000$. Most importantly, however, the interaction of symptom and diagnosticity was highly significant, $F(1, 33) = 62.74, p < .000$. As can be seen in Figure 9, when the list of symptoms was presented under TTN we see the opposite effect in that the

symptom high for TTN was estimated to occur more frequently than the low-diagnostic TTN symptom. The high-diagnostic RDS symptom was estimated to occur less frequently than the high-diagnostic TTN symptom, as was true for the low-diagnostic RDS symptom. Taken together, these results suggest that participants agreed with the diagnosticity as well as with the symptoms selected for these diseases.

To determine the method participants claimed to use when making their decisions, they were all asked at the end of Phases 1 and 3 respectively to describe their method and to judge how confident they were in their diagnosis. As can be seen in Table 5, in Phase 1, 60% of participants reported using only the patient information as compared to 40% in Phase 3. Accordingly, although non-significant, in Phase 3 the number of participants claiming to use both the patient information and the base rate statistics increased from 40% to 62.5%, regardless of whether or not they had received exposure and practice with NeoPeDDS. Thus it appears that participants may have realized that they should rely on more than just the patient information; this issue is taken up in the discussion.

Table 5

Phase 1 and 3 Diagnosis Methods

Method Used to Diagnose	Phase 1	Phase 3
Patient Information	60%	37.5%
Population Statistics	0%	0%
Both	40%	62.5%
Guess	0%	0%

Lastly, participants were asked on a scale from 0-100%, in both Phases 1 and 3, how confident they were in their predictions of the probability estimates. A paired samples t-test indicates that between Phase 1 ($M = 55.0$, $SD = 23.1$) and 3 ($M = 61.2$, $SD = 17.6$) participants confidence levels increased significantly $t(39) = -2.60$, $p < .013$. Therefore, although there were no effects of Phase, it appears that participants increased their use of base rates and individuating information while increasing their confidence levels increased.

Discussion

The goal of this research was to determine the extent to which a DSS could aid in decreasing the amount of diagnostic error by increasing awareness of alternative hypotheses. The focus was on Respiratory Distress syndrome in infants as the symptoms are often very ambiguous and therefore difficult to diagnose.

Presenting the main findings first, Hypothesis 1 predicted that participants would ignore base rates in Phase 1, before exposure, and not in Phase 3, after exposure. Results indicate that participants did pay attention to the base rates in Phase 1 but that base rate usage did not increase in Phase 3 as predicted. It is interesting that the estimates of the HBR group were quite close to the base rate ($M = 68.4\%$) -- that is 72%, whereas the LBR groups estimates were very different from the base rates ($M = 56.6\%$) -- that is 28%. Regardless, the modest base rate usage in Phase 3 still indicates that Hypothesis 1 was refuted.

It is interesting to note the dramatic increase in the number of participants who claimed to use both the individuating information and the base rate between Phase 1 and Phase 3. This dramatic increase in claims that they were using both the base rates and

individuating information is supported by the first analysis first Hypothesis 1 showing that a significant number of participants were using base rates while the analysis done for Hypothesis 2 indicates that participants were using the individuating information.

Therefore it would appear that participants are doing exactly what they say they are doing and that is using both the base rates and individuating information, more specifically in Phase 3 as this is where the claimed increase of base rate and individuating information use occurred. However, if awareness of which components to use had been a consequence of increased exposure to NeoPeDDS, than the exposure-by-phase interaction should have been significant, which did not happen. It is therefore possible that the claimed difference in information usage between the two Phases is due to an experimental demand characteristic. That is, because participants were asked the question twice, some of them may simply have assumed that a different answer would be appropriate when encountering the question the second time. Either way, it is safe to conclude that the probability estimates per se yielded no evidence of the success of the exposure on the NeoPeDDS prototype.

Hypothesis 2 predicted that vignettes that were provided high-diagnostic evidence in support of both diseases RDS and TNN (HH) would result in higher probability estimates than similar vignettes providing low-diagnostic evidence for the two diseases (LL). The significant main effect of the HH-based and the LL-based estimates supported Hypothesis 2 and indicate that individuating information was being used to make probability estimates.

The prediction of Hypothesis 3 was that there would be a serial position effect in Phase 1 but not in Phase 3 for participants who had received exposure on the NeoPeDDS

prototype. They were expected predominantly to rely on the base rates for their probability estimates in Phase 3, thereby eliminating the serial position effect predicted to occur in Phase 1. By contrast, probability estimates produced by those participants who had not received exposure were expected to result in serial position effects in both Phases 1 and 3. Results indicate that even though there was a serial position effect in Phase 1, this was also present in Phase 3 and as such, Hypothesis 3 was refuted.

Theoretical Implications of the Findings

Base rate usage

The fact that base rates were found to be used in both non-diagnostic and near-nondiagnostic cases is unusual. The literature indicates very clearly that, in such cases, people tend to ignore base rates and instead rely exclusively on the individuating information (the symptoms presented by the patient) (Hamm, 1994; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1973; Ginosar & Trope, 1980). Kahneman, Slovic, and Tversky (1982) showed that participants used base rates in nondiagnostic cases only when no information was provided about an individual. Since such information was provided here, one possible explanation for the base rate usage in this experiment is that participants may have perceived the cases to be nondiagnostic but that they may not have understood how to deal with these as they had received no training in Bayesian decision theory. The observation that roughly 75% (29/40) of participants indicated that they were not familiar with decision models supports this possibility.

Beyth-Marom and Fischhoff (1983) claim people will use the base rates provided the individuating information is palpably useless, as in the studies reported by Kahneman, Tversky, and Slovic (1982). This is because, Beyth-Marom and Fischhoff argue,

individuals generally have a poor understanding of the concept of diagnosticity. The above results show clearly that participants knew, and agreed on, the frequencies of occurrence of the various symptoms presented in the vignettes. Frequency of occurrence is one, but not the only, indicator of diagnosticity, as discussed earlier (pp. 11-12). Despite this knowledge, participants in the present experiment apparently did not know how to combine ambiguous information into an optimal probability estimate. The issue of diagnosticity is taken up in more detail later.

Beyth-Marom and Fischhoff further claim that people tend to consider information only in terms of the degree to which it supports Hypothesis H, $[P(D|H)]$, thereby neglecting also to consider its support for the alternative Hypothesis H_1 , $[P(D|H_1)]$. The possibility exists that participants here behaved in that manner. The nondiagnostic vignettes (HH and LL) supported both diseases to an approximately equal extent. Therefore, the presence of the high-diagnostic disconfirmatory symptom (TTN-H) in HH-based vignettes should attenuate the relevant probability estimates to the same extent as the presence of the high-diagnostic confirmatory symptom (RDS-H) should increase it. This would cancel out the effect of both symptoms. The same would be true for the nondiagnostic LL-based vignettes. Therefore, even if participants in both groups (HBR and LBR) had performed optimally, the mean probability of all nondiagnostic HH- and LL-based estimates would then have been .50 or 50% for both the HH- and the LL-based vignettes. In that case, these should not have differed significantly which they did, therefore this indicates that participants were using the individuating information.

Serial position effects

The serial position effect provides some support for Beyth-Marom and Fischhoff's (1983) assertion that the individuating information may have been considered mainly in terms of its support for Hypothesis H. Estimates in which the H-diagnostic confirmatory symptom (RDS-H) was presented in the first serial position were significantly higher than estimates in which the H-diagnostic disconfirmatory symptom (TTN-H) was presented in the first serial position. The observation that the effect appeared to decrease across the two serial positions supports the presence of a primacy effect. It also supports the argument of symptom-weighting favoring Hypothesis H. This primacy effect is consistent with the above results. It is also consistent with results from other Bayesian experiments using clinical diagnosticians (Hamm, 1994; Lindgaard & Triggs, 1990; Lindgaard, Pyper, et al., 2009; Lindgaard & Ralph, 2009; Ralph, 2003). None of these studies provided training in decision making or access to a decision support system. However, the above results show that mere exposure to the probability that more than a single disease may account for a particular constellation of symptoms is insufficient reliably also to affect probability estimates. Even though NeoPeDDS usage did improve the accuracy of diagnoses proposed in Phase 2, additional training in formal decision making is apparently required for DDSSs to have the desired effect.

Future Research and Limitations of the Present Research

The possibility that NeoPEDSS could provide a suggestion of remaining symptoms needed to confirm a diagnosis could drastically increase the probability that a correct diagnosis will be made faster. For example, a prompt may be displayed saying

“check for the presence of symptoms X, Y, and Z to increase the probability of disease W”. While that will not increase the likelihood of users learning about Bayes’ Theorem and how to use it, it could ultimately increase diagnostic accuracy. It would be interesting to learn how availability of such information could facilitate the task of diagnosing ambiguous diseases in future studies.

One long-term goal of NeoPeDSS is to recommend specific tests or encourage the diagnostician to obtain specific additional information to increase the probability of diagnosing the most probable disease. Eventually, it is also anticipated that NeoPeDSS will encompass diseases that are not confined to infants, such that physicians in various areas of specialty may be taught better decision making strategies.

Some limitations to the present study are worth noting. First, all participants should have completed the diagnostic exercise in Phase 2. That would have allowed for a better comparison of participants who and who had not received NeoPeDDS exposure. That will be done in a future experiment.

The selection of symptoms is another possible limitation. Although the study did aim to provide ambiguous vignettes, the selection of symptoms may have been perceived to be more indicative of the respective diseases than anticipated. This is evident in the estimates of the frequency of occurrence for the respective diseases that the residents completed at the end of the survey. Residents estimated that the symptoms high and low for respiratory distress syndrome would occur quite frequently. Therefore it would appear that these symptoms were not as ambiguous as previously thought and in fact were quite indicative of the respective diseases as the same was true for Transient Tachypnoea. This may have diluted the effect these could have had on the probability estimates.

Conclusion

The goal of this thesis was to determine the extent to which brief exposure with a computerized decision support system may help to reduce medical diagnostic decision errors. As the results of Phase 2 revealed, diagnostic accuracy did improve; apparently, the provision of probabilities was enough for participants to change their initial diagnosis to the correct one. Although base rates were used there was no increased base rate usage between Phases 1 and 3 which is most likely due to a lack of appreciation of how to deal with base rates even though it is possible that the relative worthlessness of the case-specific evidence was recognized. Apparently, the assumption that such recognition alone would inspire participants to rely more on the base rates for their probability estimates, was not entirely reasonable. In order for a DSS such as NeoPeDDS to provide optimal assistance to aspiring pediatricians, formal training on conditional probabilities as in Bayes' Theorem would thus seem to be necessary.

References

- Amy, L.R., Lyman, J., & Borowitz, S. (2006). Impact of a web-diagnosis reminder system on errors of diagnosis. *American Medical Informatics Association* (pp. 2-6).
- Anderson, N. (1981). Meaning constancy in person perception. *Foundations of Information Integration Theory*. New York: Academic Press.
- Barness, L.A., Tunnessen, W.W., Worley, W.E., Simmons, T.L., & Ringe, T. (1974). Computer-assisted diagnosis in pediatrics. *American Journal of Diseases of Children*, 127(6), 852-858.
- Berry, D. (2006). Bayesian Statistics. *Medical Decision Making*, 429-430.
- Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology*, 45(6), 1185-1195.
- Burton, E., Troxclair, D., & Newman, W. (1998). Autopsy diagnoses of malignant neoplasms: How often are clinical diagnoses incorrect. *The Journal of the American Medical Association*, 280, 1245-1248.
- Chesley, G.R. (1978). Subjective probability elicitation techniques: A performance comparison. *Journal of Accounting Research*, 16(2), 225-241.
- Combs, B., & Slovic, P. (1979). Causes of death: Biased newspaper coverage and biased judgments. *Journalism Quarterly*, 56, 837-843.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58(1), 1-73.
- Croskerry, P. (2000). The cognitive imperative: Thinking about how we think. *Academic Emergency Medicine*, 7(11), 1223-1231.

- Croskerry, P. (2006). Critical thinking and decision making: Avoiding the perils of thin-slicing. *Annals of Emergency Medicine*, 48(6).
- Croskerry, P., & Campbell, S. (2004). Canadian adverse events study. *Canadian Medical Association Journal*, 8, 171.
- Croskerry, P., & Sinclair, D. (2001). Emergency medicine: A practice prone to error? *Canadian Journal of Emergency Medicine*, 3, 271-276.
- Eddy, D.M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In P. Slovic, A. Tversky & D. Kahneman (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.
- Eddy, D.M., & Clanton, C.H. (1982). The art of diagnosis: Solving the clinicopathological exercise. *New England Journal of Medicine*, 306(21), 1263-1268.
- Elstein, A. (1976). Clinical Judgment: Psychological research and medical practice. *Science*, 194, 696-700.
- Elstein, A., Dawson-Saunders, B., & Belzer, L. (1985). Instruction in medical decision making. *Medical Decision Making*, 3(2), 229-233.
- Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, 90(3), 239-260.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4), 684-704.
- Ginosar, Z., & Trope, Y. (1980). The effects of base rates and individuating information on judgments about another person. *Journal of Experimental Social Psychology*, 16, 228-242.

- Graber, M. (2007). Diagnostic errors in medicine: What do doctors and umpires have in common? *Morbidity & Mortality Rounds on the Web*. Retrieved from <http://www.webmm.ahrq.gov/about.aspx>
- Graber, M., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine, 165*(13).
- Graber, M., Gordon, R., & Franklin, N. (2002). Reducing diagnostic errors in medicine: What's the goal? *Academic Medicine, 77*(10), 981-992.
- Hamm, R. (1994). Underweighting of base-rate information reflects important difficulties people have with probabilistic inference. *Psychology, 5*(3).
- Horowitz, I. (1970). *Decision Making and the Theory of the Firm*. New York: Rinehart & Winston.
- Johnson, K., & Feldman, M. (1995). Medical informatics and pediatrics. *Archives of Pediatrics and Adolescent Medicine, 149*(12).
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237-251.
- Kalat, J. (2004). The biology of learning and memory. *Biological Psychology* (8th ed.). Toronto, Ontario, Canada: Wadsworth/Thomson Learning.
- Koehler, J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19*, 1-53.
- Larissa, A., Lyman, J., & Borowitz, S. (2006). *Impact of a web-based diagnosis reminder system on errors of diagnosis*. Poster session presented at American Medical Informatics Association.

- Levin, I.P. (1975). Information integration in numerical judgments and decision processes. *Journal of Experimental Psychology*, 104(1), 39-53.
- Lindgaard, G. (1985). *Weighting of Individuating Elements and Base Rates in a Nursing Decision-Making Task Involving Non-Diagnostic Case Information*. Unpublished manuscript, Monash University.
- Lindgaard, G., Pyper, C., Frize, M., Walker, R., Boutilier, C., Hui, B., Narasimhan, S., Folkens, J. (2008). Intelligent decision support in Medicine: Back to Bayes? *Journal of Universal Computer Science*, 14(16), 2720-2736.
- Lindgaard, G. & Triggs, T.J. (1990). Can artificial intelligence outperform real people?: The potential of computerised decision aids in medical diagnosis, in W. Karwowski, A. Genaidy & S.S. Asfour (Eds.), *Computer-Aided Design: Applications in Ergonomics and Safety*, Taylor & Francis, .
- Lindgaard, G. (2005). Human judgments, decision theory, and technology: Applications of Bayes' theorem. *History and Philosophy of Psychology Bulletin*, 17(2), 29-39.
- Lindgaard, G., Pyper, C., Frize, M. & Walker, R. (2009). Does Bayes have it?: Decision support systems in diagnostic medicine, *International Journal of Industrial Ergonomics* 39(3), 524-532.
- Lindgaard, G. & Ralph, S. (2009). Diagnosing ambiguous cases: A study of physicians' information integration strategies, submitted to the *Journal of Medical Decision Making*, January, 2009.
- Longhurst, C., & Hahn, J. (2005). Clinical decision-support systems in pediatrics. *The Steering Committee on Clinical Information Technology News*, 3(1).

- Lucas, P. (2001). Bayesian networks in medicine: A model-based approach to medical decision making. *Proceedings of the EUNITE Workshop on Intelligent Systems in Patient Care* (pp. 73-91).
- Maffei, F., Nazarian, E., Ramnarayan, P., Thomas, N., & Rubenstein, J. (2004). Use of a web-based tool to enhance medical student learning in the pediatric intensive care unit and inpatient wards. *Pediatric Critical Care Medicine*, 6(1).
- Nichols, L., Aronica, P., & Babe, C. (1998). Are autopsies obsolete? *American Journal of Clinical Pathology*, 110(2), 210-218.
- Ofir, C. (1988). Pseudodiagnosticity in judgment under uncertainty. *Organizational Behavior and Human Decision Processes*, 42, 343-363.
- Poulton, E.C. (1975). Range effects in experiments on people. *The American Journal of Psychology*, 88(1), 3-32.
- Pyper, C. (2007). *Final Report for the Pediatric Resident Decision Support System*. Unpublished honours thesis, Carleton University.
- Pyper, C., Frize, M., Lindgaard, G. & Walker, R. (2008). Bayesian-based diagnostic decision-support for pediatrics (PedBDS), *Proceedings 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Vancouver, August 20-24.
- Ralph, S. (2003). *Heuristics and biases in medical diagnosis using Bayes' theorem: Are physicians error prone?* Unpublished Honours Thesis, Carleton University, Ottawa, Canada.
- Ramnarayan, P., & Britto, J. (2002). Pediatric clinical decision support systems. *Archives of Disease in Childhood*, 87(5).
- Ramnarayan, P., Winrow, A., Coren, M., Nanduri, V, Buchdahl, R., Jacobs, B., et al. (2006). Diagnostic omission errors in acute paediatric practice: Impact of a

- reminder system on decision-making. *BMC Medical Informatics and Decision Making*, 6(37).
- Rothschild, J., Lee, T., Bae, T., Yamamoto, R., Horsky, J., & Bates, D. (2000). Survey of physicians' experience using a hand held drug reference guide. *American Medical Informatics Association*.
- Round, A. (1999). Teaching clinical reasoning- A preliminary controlled study. *Medical Education*, 33.
- Sandhu, H., & Carpenter, C. (2006). Clinical Decision making: Opening the black box of cognitive reasoning. *Annals of Emergency Medicine*, 48(6).
- Schiff, G., Kim, S., Abrams, R., Cosby, K., Lambert, B., Elstein, A., et al. (2006). Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project. *Advances in Patient Safety*, 2, 255-278.
- Shanteau, J., Grier, M., Johnson, J., & Berner, E. (1991). Teaching decision making skills to student nurses. In J. Baron & R.V. Brown (Eds.), *Teaching Decision Making to Adolescents* (pp. 185-206). Hillsdale, NJ.: Lawrence Erlbaum Associates.
- Society for Medical Decision Making. (2007). *Education Modules: Basic Diagnostic Test Evaluation Course*. Retrieved 2007, from <http://www.smdm.org/>
- Somoza, K.C., Momtahan, K., & Lindgaard, G. (2006). *Effects of a computerized cardiac teletriage Decision Support System on Nurse Performance: Results of a Controlled Human Factors Experiment Using a Mid-Fidelity Prototype*. Proceedings Medinfo 2007 August 24-27, Brisbane, Australia.
- Tabachnick, B., Fidell, L. (2007). *Experimental Designs Using ANOVA*. U.S: Thomson Brooks.
- Wallsten, T. (1981). Physician and medical student bias in evaluating diagnostic information. *Medical Decision Making*, 1(2), 145-164.

- Wexler Jr., Swender, PT., Tunnessen, WW., Jr., & Oski, FA. (1975). Impact of a system of computer-assisted diagnoses. Initial evaluation of the hospitalized patient. *American Journal of Diseases of Children*, 129(2), 203-205.
- Wood, B. (1999). Decision making in radiology. *Radiology*, 211, 601-603.
- Wood, S.E., Wood, E.R., Wood, E., & Desmarais, S. (1999). Memory. *The World of Psychology* (2nd ed.). Scarborough, Ontario, Canada: Prentice-Hall Canada Inc.

Appendix A

Announcement for Recruitment

A Diagnostic Decision Support Tool for Diagnosing Respiratory Distress in Infants

Diagnostic uncertainty increases the probability of misdiagnosing patients. This is a serious problem faced by medical clinicians. We are currently conducting an experiment for a master's thesis on the effectiveness of a diagnostic decision support system (DDSS) to help pediatricians diagnose respiratory distress syndrome in infants. The objective is to learn more about how physicians make diagnoses under intense time pressure and when faced with a minimum amount of ambiguous information about their patient.

Potential Risk or Discomfort

There are no potential risks or discomforts in this study. You will be allowed to take rest breaks after each phase.

Confidentiality

All data will be coded such that your name will not be associated with the data. The data will be seen only by the researchers involved in this project. Full anonymity is thus ensured. You will not be identifiable in any publications or presentations resulting from this study.

Benefits to Participation

There is no direct medical benefit to you from participation in this study. The knowledge gained from this study may help to improve the medical care of newborn infants in an NICU setting. We will be providing you with financial compensation in the amount of a \$100.00 Amazon.ca gift certificate upon completion of your online involvement in this study.

Task Requirements

Briefly, you will be asked to complete a few questionnaires, read some case scenario and determine the probability that the children in the scenarios have respiratory distress syndrome (RDS) or transient tachypnoea (TTN). The study should take approximately 1 hour 15 minutes for which you will be compensated with a gift voucher to the value of \$100 to Amazon.Com

Right to Withdraw

You have the right to withdraw from the study at any time. Financial compensation will only be provided to those participants completing all three phases of the study.

Questions About the Study

The following personnel are involved in this project and may be contacted at any time. Janette Folkens, j_folkens@rogers.com (Principal Investigator), 613-520-2600 ext. 6628, and Dr. Gitte Lindgaard (Faculty Sponsor), 613-520-2600 ext. 2255, gitte_lindgaard@carleton.ca.

If any ethical concerns about this study should arise please contact Dr. Avi Parush (Chair, Dept. of Psychology Ethics Committee), 613-520-2600, ext. 6026 avi_parush@carleton.ca, or the Chair of the CHEO Research Ethics Board at 613-737-2393 ext.3272. Should you have any other concerns about this study, please contact Dr. A. Bowker (Chair, Dept of Psychology,) 520-2600 ext. 2648, psychchair@carleton.ca.

As only residents and fellows have the necessary background to complete the experimental tasks, I would really appreciate it if you would take part in this experiment.

Interested? Please e-mail me at j_folkens@rogers.com

Thank you very much in advance

Janette Folkens, MA candidate

Appendix B

Consent Form

Title of Study: "A Diagnostic Decision Support Tool for Diagnosing Respiratory Distress in Infants"

Introduction

Diagnostic uncertainty, and hence the probability of misdiagnosing patients presenting with ambiguous symptoms, is a recurring problem faced by medical clinicians. It has been suggested that a large proportion of medical errors are the result of patient misdiagnosis. These errors may be prevented by changing the cognitive processes involved in making a diagnosis. The purpose of this study is to determine whether the use of a diagnostic decision support system will improve the ability of physicians trained in pediatrics to diagnose newborn infants presenting with respiratory distress.

Duration and Locale

The study will take approximately one hour to complete and will take place online. You may complete it over one or more sessions, at your discretion.

*Task Requirements **

You will first be asked to complete a brief questionnaire. You will then be asked to read and estimate the probability of a particular condition in a number of case scenarios describing individual infants, proceeded by a second short questionnaire. Following this will be a brief learning session on a diagnostic decision support system in which you will again be asked to diagnose cases using the system. You will then be asked to complete step one again.

Potential Risk or Discomfort

There are no potential risks or discomforts in this study. You will be allowed to take rest breaks after each phase.

Benefits to Participation

There is no direct medical benefit to you from participation in this study. The knowledge gained from this study may help to improve the medical care of newborn infants in an NICU setting. We will be providing you with financial compensation in the amount of a \$100.00 Amazon.com gift card upon completion of your online involvement in this study.

Confidentiality

All data will be coded such that your name will not be associated with the data. The data will be seen only by the researchers involved in this project. Full anonymity is thus ensured. You will not be identifiable in any publications or presentations resulting from this study.

Right to Withdraw

You have the right to withdraw from the study at any time. Financial compensation will only be provided to those participants completing all three phases of the study.

Questions About the Study

The following personnel are involved in this project and may be contacted at any time. Janette Folkens, j_folkens@rogers.com (Principal Investigator), 520-2600 ext. 6628, and Dr. Gitte Lindgaard (Faculty Sponsor), 520-2600 ext. 2255, gitte_lindgaard@carleton.ca. If any ethical concerns about this study should arise please contact Dr. Avi Parush (Chair, Dept. of Psychology Ethics Committee), 520-2600, ext. 6026 avi_parush@carleton.ca. Should you have any other concerns about this study, please contact Dr. A. Bowker (Chair, Dept of Psychology,) 520-2600 ext. 2648, psychchair@carleton.ca.

Print

I have read the above description of the study and understand the conditions of my participation. I agree to participate in this research project.

Next

* For one half of the participants this will be the case:

You will first be asked to complete a brief questionnaire. You will then be asked to read and estimate the probability of a particular condition in a number of case scenarios describing individual infants, proceeded by a second short questionnaire. Following this will be a short article which you will be asked to read. You will then be asked to complete step one again.

Appendix C

Interview Protocol- Participants Information

Please select a response to each question below to help us understand who you are.

Have you ever taken any decision-making courses?

- Yes
- No

Are you familiar with any decision-making models?

- Yes
- No

Have you used a Diagnostic Decision Support System in the past?

- Yes
- No

If yes for how many years?

- ___ years

If one were made available for you, would you use it?

- Yes
- No

You are a...

- Fellow
- Senior Resident

How do you rate yourself as a computer user?

- Novice
- Intermediate
- Expert

What is your current age in years?

- 20-24,
- 25-29,
- 30+

Your gender?

- M
- F

Appendix D

Case Scenario Information

Cover story:

After an extremely violent hurricane, several falling trees hit and caved in the roof of a large hospital for sick children and infants in the vicinity of where you work as a pediatrician. Patients rescued from there are arriving at your hospital by the dozen. As a pediatrician, your job is to attend to the newborn infants immediately upon their arrival, to propose a preliminary diagnosis, and to commence initial treatment where needed. Patient records have not yet arrived. Consequently, you know very little about each infant. In fact, all you have been told is that, of the 100 infants expected to arrive in the next few hours, 72 have been diagnosed with respiratory distress syndrome and 28 with transient tachypnoea.

Presentation of cases (High Base Rate group):

Infant Gonzales, born about one hour ago, has just been admitted with increased interstitial markings. She appears slightly restless, and her breathing shows clear intercostals retractions.

How likely do you think it is that infant Gonzales is one of the 72 cases of respiratory distress syndrome? _____%

Presentation of cases (Low Base Rate group):

Infant Gonzales, born about one hour ago, has just been admitted with increased interstitial markings. She appears slightly restless, and her breathing shows clear intercostals retractions.

How likely do you think it is that infant Gonzales is one of the 28 cases of respiratory distress syndrome? _____%

- 1) Baby Smith, was admitted to the NICU with breathing problems. A chest x-ray reveals `ground glass`. The infant is slightly pale and has increased interstitial markings.
- 2) Baby Hudson was admitted to the NICU for observation. She has developed `ground glass` changes and increased interstitial markings on her chest x-ray. She is slightly restless.

- 3) Baby Burrows is slightly restless. He has reticulo-granular markings on his chest x-ray and wet lungs.
- 4) Baby Hollinger is slightly pale. Her x-ray shows signs of increased interstitial markings as well as reticulo-granular markings.
- 5) Baby Browne was admitted to the NICU after a chest x-ray showed increased interstitial markings with a reticulo-granular pattern. He is slightly pale.
- 6) Baby Gonzales has just been admitted after a chest x-ray showed increased interstitial markings. She appears slightly restless and has a reticulo-granular pattern.
- 7) Baby Flowers was admitted with severe subcostal retractions. His tone is normal, and he has a respiratory rate of 110/min.
- 8) Baby Markus has severe subcostal retractions. Her respiratory rate is 90/min, and she appears slightly pale.
- 9) Baby Hansen appears slightly restless. He has severe subcostal retractions and has a respiratory rate of 85/min.
- 10) Baby Hesterton is a little pale. His respiratory rate is 89/min, and he is showing severe subcostal retractions.
- 11) Baby Blewett's respiratory rate of 95/min is of some concern, as is her severe subcostal retractions. She appears slightly pale.
- 12) Baby Martini has a respiratory rate of 100/min, is slightly pale and has severe subcostal retractions.
- 13) Baby Cartier's chest x-ray shows reticulo-granular markings. He is slightly restless and has a respiratory rate of 98/min.
- 14) A chest x-ray taken a few minutes ago of baby Chessari had a distinct 'ground glass' appearance. She had respiratory rate of 109/min and she now seems a little restless.
- 15) Baby O'Neill has had a normal pulse since this morning. His chest x-ray now shows 'ground glass' changes and his respiratory rate is 82/min.
- 16) Baby Khan appears rather pale. She has increased interstitial markings on her x-ray and severe subcostal retractions.

- 17) A chest x-ray of baby Petruilis showed clear signs of increased interstitial markings. He has severe subcostal retractions and appears slightly restless.
- 18) Baby Haas` chest x-ray shows wet lungs and she is slightly pale. She is also showing severe subcostal retractions.
- 19) Baby Sorensen is showing severe subcostal retraction. She has a normal pulse, and her chest x-ray shows increased interstitial markings.
- 20) Baby Mcdonald`s severe subcostal retractions are causing some concern. Her chest x-ray is showing wet lungs. Right now she appears a little restless.
- 21) Baby Halling is very restless. He has severe subcostal retractions and his x-ray taken earlier this morning shows clear signs of increased interstitial markings.
- 22) Infant Maso is a little pale and has a respiratory rate of 82/min. X-rays taken show signs of reticulo-granular markings.
- 23) With a respiratory rate of 110/min, and a chest x-ray having the appearance of `ground glass`, baby Petrovsky was admitted to the NICU. He is slightly pale.
- 24) After a difficult delivery, Baby Kruus has a respiratory rate of 98/min despite her normal tone. Of more concern is the distinct reticulo-granular appearance of her chest x-ray.

Appendix E

Phase I Survey

Please indicate what best describes the method you used when diagnosing that the children in the vignette had respiratory distress syndrome:

- Patient Information
- Population Statistics
- Both patient information and statistics
- Guess

Please describe how confident you were in diagnosing that the children in the vignettes had respiratory distress syndrome:

_____ %

Appendix F

NeoPeDDS Instructions

Introduction to ReDSS

What it is

The Resident Decision Support System (ReDSS) is a diagnostic decision support system (DDSS) for pediatricians designed to assist diagnostic decision processes when faced with infants suffering from some form of respiratory distress syndrome (RDS). RDS occurs frequently in infants and can often be difficult to diagnose due to the ambiguous nature of the signs and symptoms. ReDSS' database is based on actual infant data; the output probabilities it provides are therefore real.

What it does

ReDSS was designed to provide diagnostic support to pediatric residents faced with diagnosing many conditions including respiratory distress. When using ReDDS, the symptoms, signs, and the related values observed in a particular infant are entered into the database. ReDSS will then provide the five most probable conditions based on the information thus entered. Most importantly, a probability is provided for each condition enabling the diagnostician to hone in on the most probable condition and perhaps provide additional information to further increase its probability.

How it works

ReDSS is straight-forward and easy to operate. It consists of input screens as shown in Figure 1 below, and an output screen as shown in Figure 2 below. On the input screen you select one symptom at a time that you have observed in the infant with respiratory distress who is to be diagnosed. The drop-down menu, shown on the left (see number 1), contains a list of all the signs and symptoms associated with respiratory distress as taken from the original patient database from which ReDDS' database originates. Once a symptom is selected, the drop-down menu on the right (see number 2) opens. It displays the range of possible values of the sign or symptom you have just entered. After selecting the relevant value, you can either select the next sign or symptom from the drop-down box on the left as you just did in step one, or you can press the button called 'diagnose' (see number 3). You may continue to select signs and symptoms until all the information you currently have about the to-be-diagnosed infant has been entered, or you may press the 'diagnose' button repeatedly, at any time in the process. Upon pressing the 'diagnosis' button, a list of the five most probable diagnoses appears in the output (see number 4). A probability is given for each of the five diagnoses, showing clearly if one diagnosis is more likely than another, or if several conditions are equally probable. At this point, you can accept one of the diagnoses provided, refine the input by entering more information, or start over.

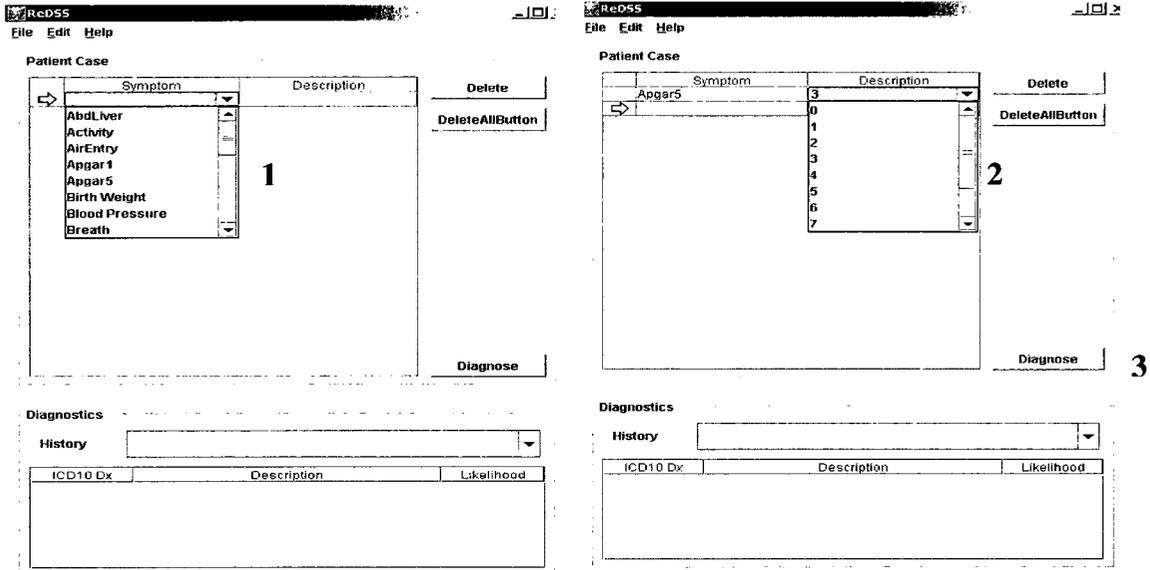


Figure 1. Input screens where the left panel shows a ReDSS screen prior to data entry with a list of all the signs and symptoms of respiratory distress. The right panel shows the possible range of values available through a second drop-down menu.

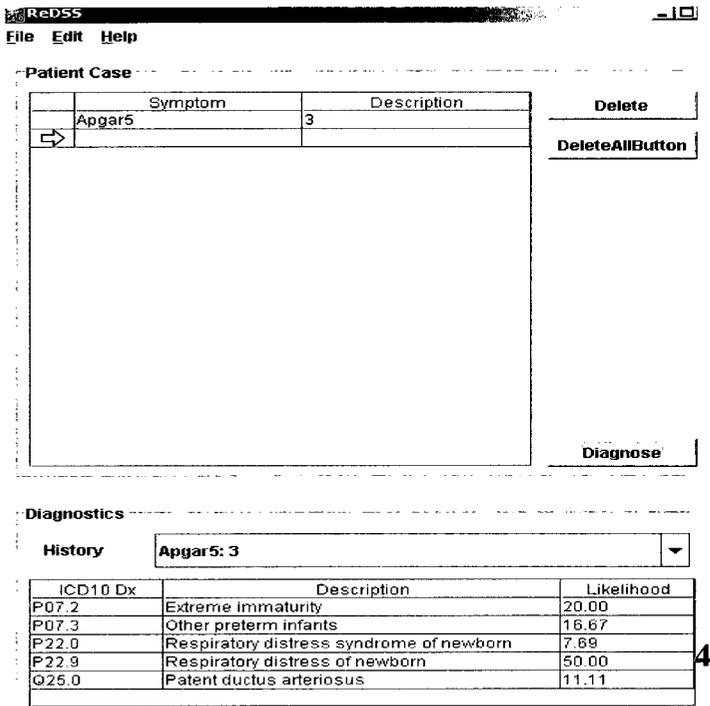


Figure 2. Output probabilities upon entry of an Apgar5 with a score of 2

Appendix G

NeoPeDDS Case Scenarios

1. Baby Ely is 4-day old infant who was born at 30-week gestational age. He has respiratory distress with crackling breath sound. His respiratory rate is 70.
2. Baby Johnson has increased work of breathing with intercostal retraction. She appears irritable. Her systolic blood pressure is 50.
3. Baby Thomson is a 36-week infant who present with acute respiratory distress. His respiratory rate is 70. The auscultation reveals diminished breath sound.
4. Baby Cortes is a full-term infant. His O₂ saturation is 88%. He has equal air entry on auscultation. He has normal muscular tone. The pulses are somewhat weak.
5. Baby Williams is a premature infant. He appears ill today. Her temperature is 35 C. Her systolic blood pressure is 50. Her O₂ saturation is at 95% on O₂ of 1 l/min

Appendix H

Phase 3 Survey

Please indicate what best describes the method you used when diagnosing that the children in the vignette had respiratory distress syndrome:

- Patient Information
- Population Statistics
- Both patient information and statistics
- Guess

Please describe how confident you were in diagnosing that the children in the vignettes had respiratory distress syndrome:

_____ %

In how many of the 100 infants all diagnosed with respiratory distress syndrome would you expect to see each of the following symptoms? Please note, each percentage is independent of the others, they do not add up to 100%

- 1) Chest X-ray with “ground glass” appearance (reticulo-granular)
- 2) Severe subcostal retractions
- 3) Normal tone
- 4) Normal pulses
- 5) Slightly pale skin
- 6) Restless
- 7) Increased interstitial markings (wet lungs) on chest x-ray
- 8) Respiratory rate of 80-110/min

In how many of the 100 infants diagnosed with Transient Tachypnoea would you expect to see each of the following symptoms? Please note, each percentage is independent of the others, they do not add up to 100%

- 1) Respiratory rate of 80-110/min
- 2) Normal Tone
- 3) Slightly pale skin
- 4) Normal Pulses
- 5) Increased interstitial markings (wet lungs)
- 6) Severe subcostal retractions
- 7) Restless
- 8) Chest x-ray with “ground glass” appearance (reticulo granular)

Appendix I

Debriefing

Thank you for participating in this study. Your time and effort are greatly appreciated!

The outcome of this research will be used to demonstrate that a diagnostic decision support system may provide valuable assistance to physicians when faced with ambiguous cases in which the patient's signs and symptoms point to several possible diseases. Specifically, it is hoped that the information about probabilities associated with a range of possible, even equiprobable, diagnoses, may help physicians to consider a limited, but focused, range of likely conditions, and that this may lead to faster and more accurate diagnoses.

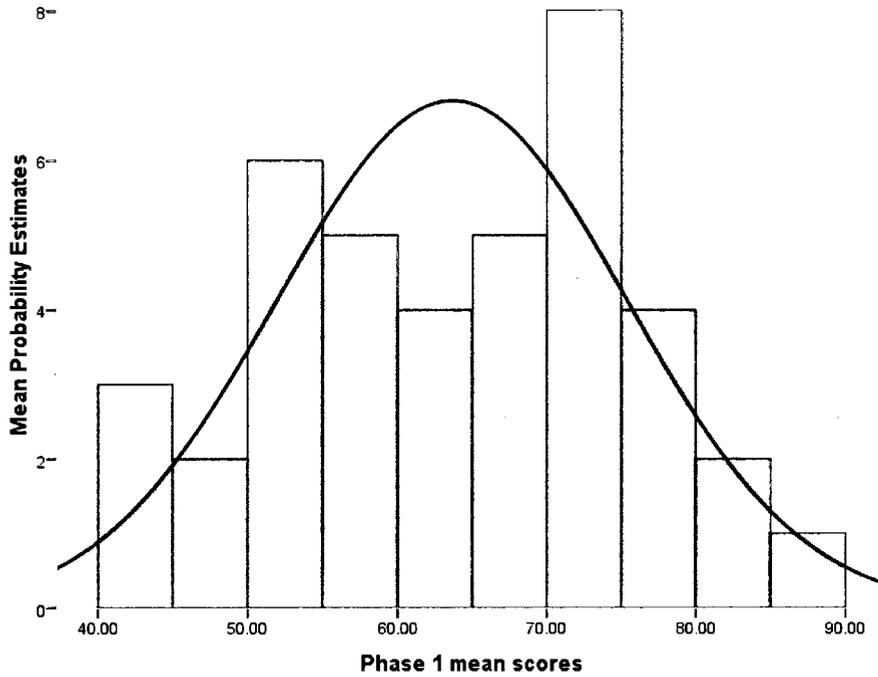
Please feel free to contact us at any time if you wish to discuss any aspects of the research. Janette Folkens, j_folkens@rogers.com (Principal Investigator), 520-2600 ext. 6628, and Dr. Gitte Lindgaard (Faculty Sponsor), 520-2600 ext. 2255, gitte_lindgaard@carleton.ca. If any ethical concerns about this study should arise please contact Dr. Avi Parush (Chair, Dept. of Psychology Ethics Committee), 520-2600, ext.6026, avi_parush@carleton.ca. Should you have any other concerns about this study, please contact Dr. A. Bowker (Chair, Dept of Psychology,) 520-2600 ext. 2648, psychchair@carleton.ca.

A \$100.00 Amazon.ca gift certificate will be sent shortly to the email address you used to participate in this study. If you do not receive it in your Inbox, please check your spam.

Thank you again very much for taking time to help us out with this experiment.

Appendix J

Test of Normality



Additionally, the Test of Normality using the Kolmogorov-Smirnov was non-significant $p = .200$ indicating that Normality has not been violated.

Appendix K

Test for Absence of Outliers

