2011

CLINICAL PREDICTION MODELS FOR DIAGNOSIS OF APPENDICITIS IN CHILDREN WITH ABDOMINAL PAIN

Dale W Steele
University of Rhode Island, dale_steele@brown.edu

CLINICAL PREDICTION MODELS FOR DIAGNOSIS OF APPENDICITIS IN CHILDREN WITH ABDOMINAL PAIN

BY

DALE W. STEELE

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

DEPARTMENT OF COMPUTER SCIENCE AND STATISTICS

UNIVERSITY OF RHODE ISLAND

2011

MASTER OF SCIENCE THESIS

OF

DALE W. STEELE

APPROVED:

Thesis Committee:

Major Professor Liliana Gonzalez

John P. Burkett

Alfred Ordoñez

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2011

ABSTRACT

Appendicitis is common in children, but remains difficult to diagnose accurately. The clinician must integrate information from the history, physical examination and screening laboratory tests to decide whether to reassure, order diagnostic imaging, or proceed to the operating room. This process is best framed as a decision problem with two thresholds; a lower threshold, below which further testing may be unnecessary, and an upper threshold, above which further testing need not delay appendectomy. The goal of this analysis was to model the probability of appendicitis.

This project analyzes observations by 23 physicians on 143 children with abdominal pain evaluated in a Pediatric Emergency Department. Clinicians recorded the presence or absence of various signs and symptoms, and provided their gestalt estimate of the probability of appendicitis (*priorprob*) prior to obtaining screening laboratory tests such as white blood cell count (*wbc*). A final diagnosis of appendicitis was confirmed pathologically in 45 (31.5%) patients.

Exploratory plots utilize nonparametric exploratory kernel density and locally weighted scatterplot smoothing. Missing data is imputed using both single and multiple imputation. Receiver Operator Characteristic curves illustrate the superior discrimination of a logistic clinical factors model vs. the Pediatric Appendicitis Score which dichotomizes *wbc*. The Akaike Information criteria provide support for a model that substitutes gestalt clinical probability (*priorprob*) for individual clinical factors. The bootstrap is used to produce bias-corrected calibration plots for each model and to estimate confidence intervals for coefficients. To account for the correlation within physicians, Generalized Linear Mixed models with clinician specific random effect(s) were fit using maximum likelihood and Bayesian methods.

The apparent importance of *gender* in exploratory plots is confirmed using

parametric models. Contrary to prior studies, the presence of fever reduces the probability of appendicitis. Conditional predictions from the preferred (random intercept) Bayesian model suggest that one can most confidently omit imaging in girls with low clinical suspicion (*priorprob*) and low white blood cell counts (*wbc*). Conversely, the best case for proceeding directly to the operating room can be made for boys with both high *priorprob* and high *wbc*. When levels of *priorprob* and *wbc* are discordant, imaging, or further observation, will be necessary.

ACKNOWLEDGMENTS

To all those who helped me learn a little Statistics — Shira Dunsiger, who patiently re-taught high school algebra — Jo Hogan for "If you wanted to study French Literature, you need to speak French." — Orlando Merino, who never seemed to skip steps — Liliana Gonzalez for her warmth, patience and attention to detail — John Burkett, for his elegant lectures, insightful critiques and sense of humor — Profs Ordonez and Akhlaghi for being members of my defense committee — the wizards who created the Open Source tools used to produce this thesis— children with abdominal pain who inspire a desire to do better.

DEDICATION

To my mother, Margaret O'Shaughnessy Steele — a teacher with a lifelong passion for learning.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND STUDY DESIGN

1.1 Statement of the Problem

Appendicitis is common in childhood, but remains difficult to diagnose accurately. After obtaining a history and physical examination, the treating clinician(s) may decide to send the child home without further testing, perform diagnostic imaging using ultrasound or computed axial tomography (CT scan), admit for observation, or proceed to the operating room. The costs of unnecessary surgery are balanced by an increased risk of perforation if the diagnosis is missed on initial evaluation. Perforated appendicitis results in greater morbidity. Radiologic imaging increases cost, and may introduce further delays. Concerns have been raised about possible long term effects of radiation from CT scan.

Most prior studies have treated the diagnosis of appendicitis as a classification/discrimination problem. The result has been the creation of a number of "clinical scores" and associated proposed decision rules. Most recently, the focus of these decision rules has been to define a patient at "low risk" for appendicitis [1, 2].

1.2 Justification and Significance of the Problem

Harrell and others make a compelling argument that suggests many of these attempts are flawed [3, 4].

Harrell argues that many decision rules ignore subject heterogeneity and categorize predictions as either diseased or normal, letting fear of probabilities and costs/utilities lead the analyst, not the treating physician, to be the provider of the utility function. Middle probabilities allow for gray zones and deferred decisions pending further testing. One such decision analysis has been published [5].

Harrell further argues that many physician investigators exhibit "dichotomania", attempting to find cutpoints in continuous predictor variables using improper scoring rules such as sensitivity and specificity. Mathematically, such cutpoints waste information, and cannot exist unless the relationship with outcome is discontinuous.

With this perspective in mind, the goals of this study were to develop and compare clinical prediction models [4] as follows:

1. Using multiple logistic regression, develop a prediction model from history, physicial examination and laboratory tests. I will explore methods to avoid resorting to stepwise variable selection techniques, use model validation techniques based on the bootstrap,and use graphical methods to aid in understanding model predictions [4, 6].

2. Develop a prediction model using the clinician's estimate of the prior probability of appendicitis, calibrated for gender and adjusted for the subsequently obtained white blood cell (wbc) count.

3. Compare the potential predictive utility of a model based on a clinical score vs. use of the clinician's subjective assessment of prior probability.

4. Extend the second model to account for dependency of clinical prior probability within clinicians using Generalized Linear Mixed Models (GLMM).

5. Replicate the logistic models and the GLMM from a Bayesian perspective and explore incorporation of an informative prior using data from a study performed at the Children's Hospital of Philadelphia (CHOP).

1.3 Methodology

1.3.1 Study Design

Existing data from two completed, IRB approved studies of appendicitis in children presenting to the Emergency Department (ED) are available.

In the primary study, performed at Hasbro Children's Hospital (HCH) in Providence, RI we enrolled a prospective cohort of children presenting to a pediatric emergency department with abdominal pain in whom the treating physician considered a diagnosis of appendicitis. Faculty and fellows recorded potentially predictive information obtained during a structured history and physical examination. Clinicians were also asked to mark a vertical hash on a 10cm line to express their clinical estimate of the probability of appendicitis. These clinical variables were recorded prior to availability of laboratory tests (eg. wbc count) or results of abdominal imaging (ultrasound or computed tomography).

In those children who had surgery, the final pathology report was used as the criterion for a diagnosis of appendicitis. In non-operative cases, telephone follow-up was done after seven days to ensure that symptoms had resolved.

1.3.2 Descriptive Statistics

At Hasbro Children's Hospital, 143 children were evaluated by 23 phyisicians. Individual physicians saw as few as 1 patient, to as many as 17 patients each. Overall, 45 (31.5%) children had a final diagnosis of appendicitis, leaving 98 (68.5%) without appendicitis.

A colleague kindly provided de-identified data from a similar study done at the Children's Hospital of Philadelphia (CHOP). This study enrolled 217 patients greater than five years old, of whom 86 (39.6%) had appendicitis. This supplemental dataset will be used to develop an informative prior in the Bayesian analysis.

1.4 Reproducible Research

The goal of reproducible research is to tie specific instructions to data analysis and experimental data so that scholarship can be recreated, better understood and verified [7].

The source documents for this thesis were written using the GNU Emacs text editor using the add-on package Emacs Speaks Statistics(ESS)[8]. R: A Language and Environment for Statistical Computing, was used for all statistical analyses [9]. The `Sweave` function [10] executes R code chunks embedded in the source file, producing a LaTeX document which incorporates statistical analyses and graphics with prose for typesetting using the URI thesis format [11].

1.5 Descriptive Statistics

The **reporttools** package creates LaTeX tables with descriptive statistics for the variables collected [12].

Factors presumed to be associated with appendicitis were recorded. A history of fever at home or in the ED was defined as *fever* if body temperature was ever $\geq 38\,^{\circ}\mathrm{C}$. The variable *migrate* was coded yes if pain had migrated to the right lower quadrant (RLQ). Each child was asked: "What is your favorite food?" "If we had some here, would you want to eat it now?" If not, *anorexia* was coded as yes. Finally, *emesis* was defined as any history of vomiting.

Variable	**Levels**	**n**	%
Appendicitis (*appy*)	no	98	68.5
	yes	45	31.5
	all	143	100.0
Gender (*gender*)	Female	78	54.5
	Male	65	45.5
	all	143	100.0
Fever (*fever*)	no	106	74.1
	yes	37	25.9
	all	143	100.0
Migration of Pain (*migrate*)	no	94	65.7
	yes	47	32.9
	missing	2	1.4
	all	143	100.0
Anorexia (*anorexia*)	no	65	45.5
	yes	77	53.9
	missing	1	0.7
	all	143	100.0
Vomiting (*emesis*)	no	70	49.0
	yes	72	50.4
	missing	1	0.7
	all	143	100.0

Table 1. Diagnosis, Gender and History Variables

The next table summarizes the physical examination variables. If the patient was tender to palpation in the RLQ of the abdomen, the *rlqpain* was present. The variables *hoppain*, *coughpain*, *shakepain* and *percpain* were coded as present if the patient reported pain with attempts to hop or cough, or in response to a gentle pelvic shake or manual percussion of the abdomen. Rebound tenderness (*rebound*) was considered present if the patient complained of pain after sudden release of abdominal compression. Urinary ketones were measured using a point-of-care dipstick test which provide semiquantitative levels. Thus, *ketones* were at least ordinal, and were treated as a continuous variable in regression models.

The 4 category scale for urinary ketones *ketones* was reduced to three levels (none, small, medium to large) because of low frequencies of observations with values of 2 (moderate) and 3 (large).

Variable	**Levels**	**n**	%
RLQ tenderness (*rlqpain*)	no	17	11.9
	yes	123	86.0
	missing	3	2.1
	all	143	100.0
Pain with Hopping (*hoppain*)	no	54	37.8
	yes	87	60.8
	missing	2	1.4
	all	143	100.0
Pain with Cough (*coughpain*)	no	81	56.6
	yes	58	40.6
	missing	4	2.8
	all	143	100.0
Pain with Shaking (*shakepain*)	no	89	62.2
	yes	54	37.8
	all	143	100.0
Pain with Percussion (*percpain*)	no	60	42.0
	yes	82	57.3
	missing	1	0.7
	all	143	100.0
Rebound Tenderness (*rebound*)	no	91	63.6
	yes	52	36.4
	all	143	100.0
Urinary Ketones (*ketones*)	0	89	62.2
	1	32	22.4
	2	7	4.9
	3	7	4.9
	missing	8	5.6
	all	143	100.0

Table 2. Physical Examination and Bedside Urinary Ketones

After taking a history and performing a physicical examination, but prior to knowledge of the white blood cell count, clinicians expressed their gestalt clinical estimate of the percent probability of appendicitis (*priorprob*) by making a vertical hash mark on a 10 cm line. Total white blood cell count (*wbc*) is often measured as an indicator of inflammation. The percentage of polymorphonuclear cells multiplied by the total count is the absolute neutrophil count (*anc*), and may represent a more specific indicator of inflammation. Information is available regarding the duration of pain in hours (*durpain*), but it was not considered in the current models. If a larger dataset were available, it might be useful to model interactions between

durpain and other clinical variables, since patients may develop increasingly severe signs of inflammation over time.

Variable	**n**	**Min**	$\mathbf{q_1}$	$\widetilde{\mathbf{x}}$	$\bar{\mathbf{x}}$	$\mathbf{q_3}$	**Max**	**s**	**IQR**	**#NA**
Clinical Probability (*priorprob*)	142	1.0	21.5	52.0	50.5	75.5	94.0	28.5	54.0	1
White Blood Cell Count (*wbc*)	139	2.2	7.2	9.4	11.2	14.4	30.0	5.4	7.2	4
Absolute Neutrophil Count (*anc*)	139	1.2	4.3	6.5	8.6	11.5	26.4	5.6	7.1	4
Duration of Pain (hours)	143	2.0	12.0	24.0	38.3	48.0	168.0	36.9	36.0	0

Table 3. Continuous variables.

1.6 Exploratory Data Analysis

This section is philosophically somewhat at odds to the recommendations not to perform variable selection based on the relationship of predictors to the outcome. However, nonparametric exploratory graphics such as those offer a nonparametric approach to visualizing relationships between predictors and outcomes.

The distribution of the continous variables *wbc* and *priorprob* are graphically displayed as empirical kernel density plots conditional on the final diagnosis and gender using the R function `densityplot` in package **lattice** [13].

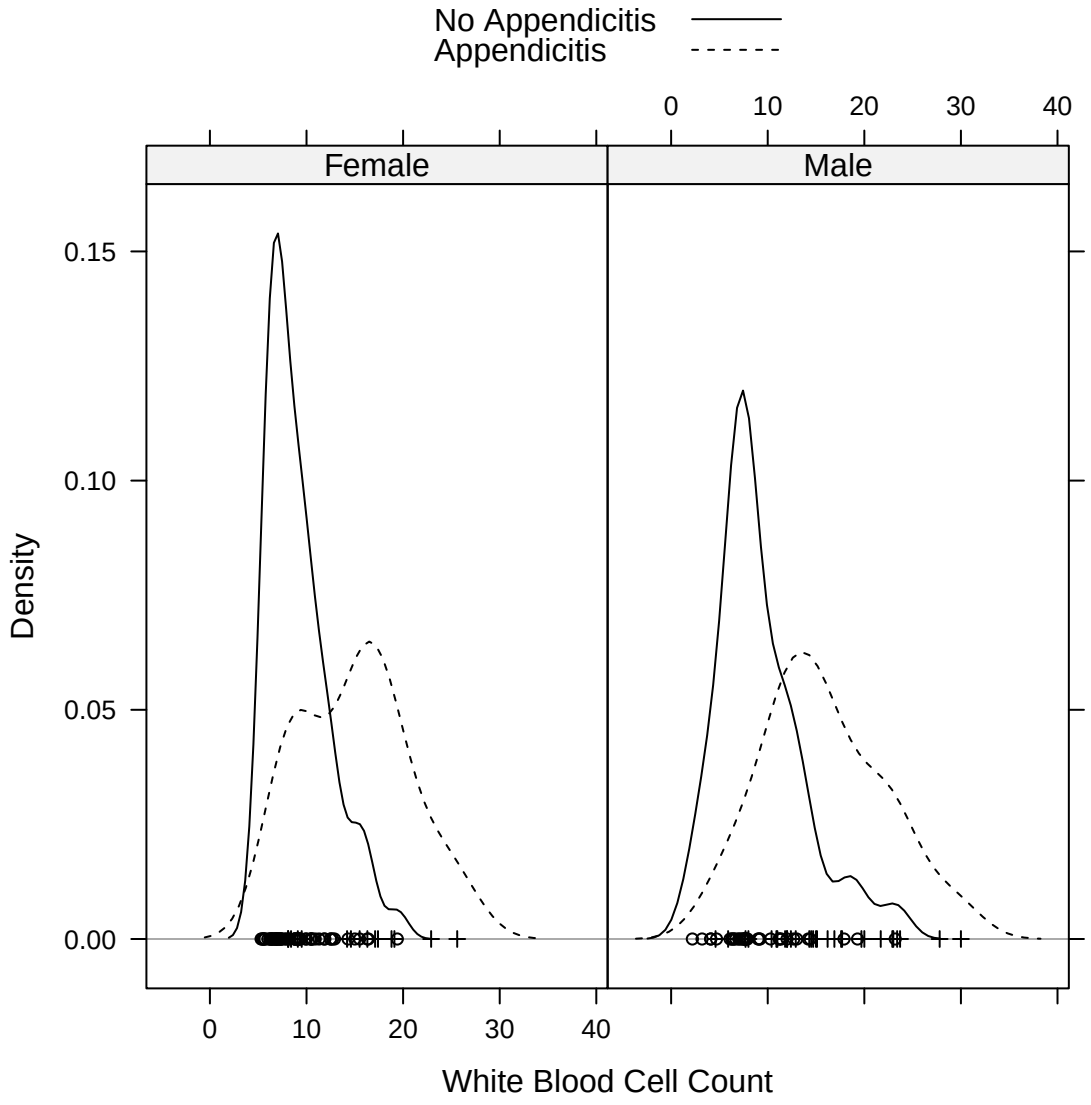

Figure 1. Empirical Conditional Density Plots: White Blood Cell Count by Diagnosis and *gender*

Note that the variance of *wbc* count is larger for patients with appendicitis, and there is a suggestion of bimodality in the distribution of the *wbc* in girls with appendicitis.

The distributions (shown below) of the clinical probability of appendicitis (*priorprob*) for patients who ultimately were found not to have appendicitis appears

to be (at least) bimodal, whereas the distribution of *priorprob* is unimodal for those with appendicits.

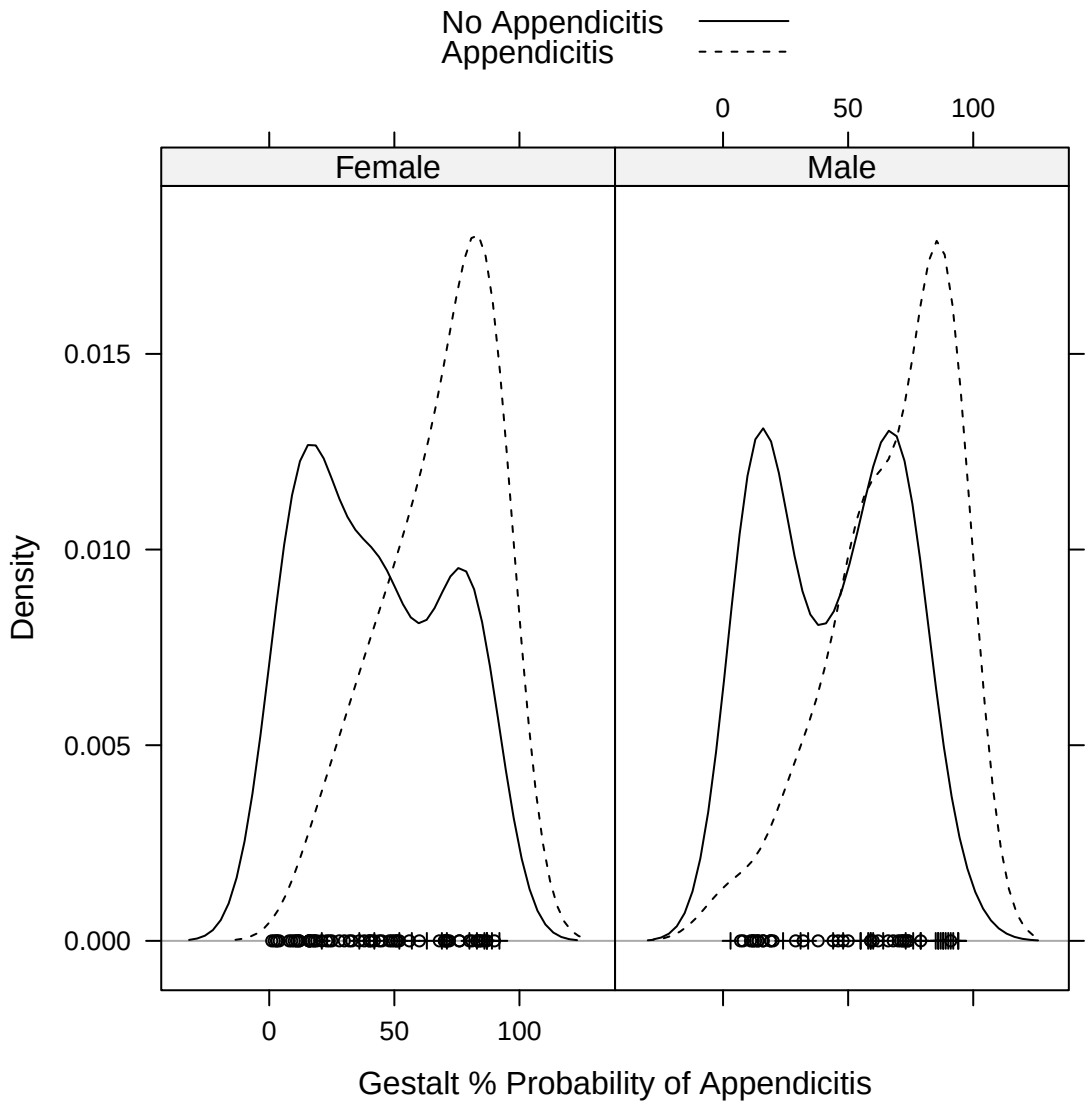

Figure 2. Empirical Conditional Density Plots: White Blood Cell Count by Diagnosis and *gender*

The plot below uses Loess (locally weighted scatterplot smoothing) to estimate the smoothed proportion of children with appendicitis, conditional on gender and white blood cell count. The slope of the relationship between *wbc* and smoothed

predicted probability appears to be similar for boys and girls, suggesting a *gender* by *wbc* interaction term will not be needed.

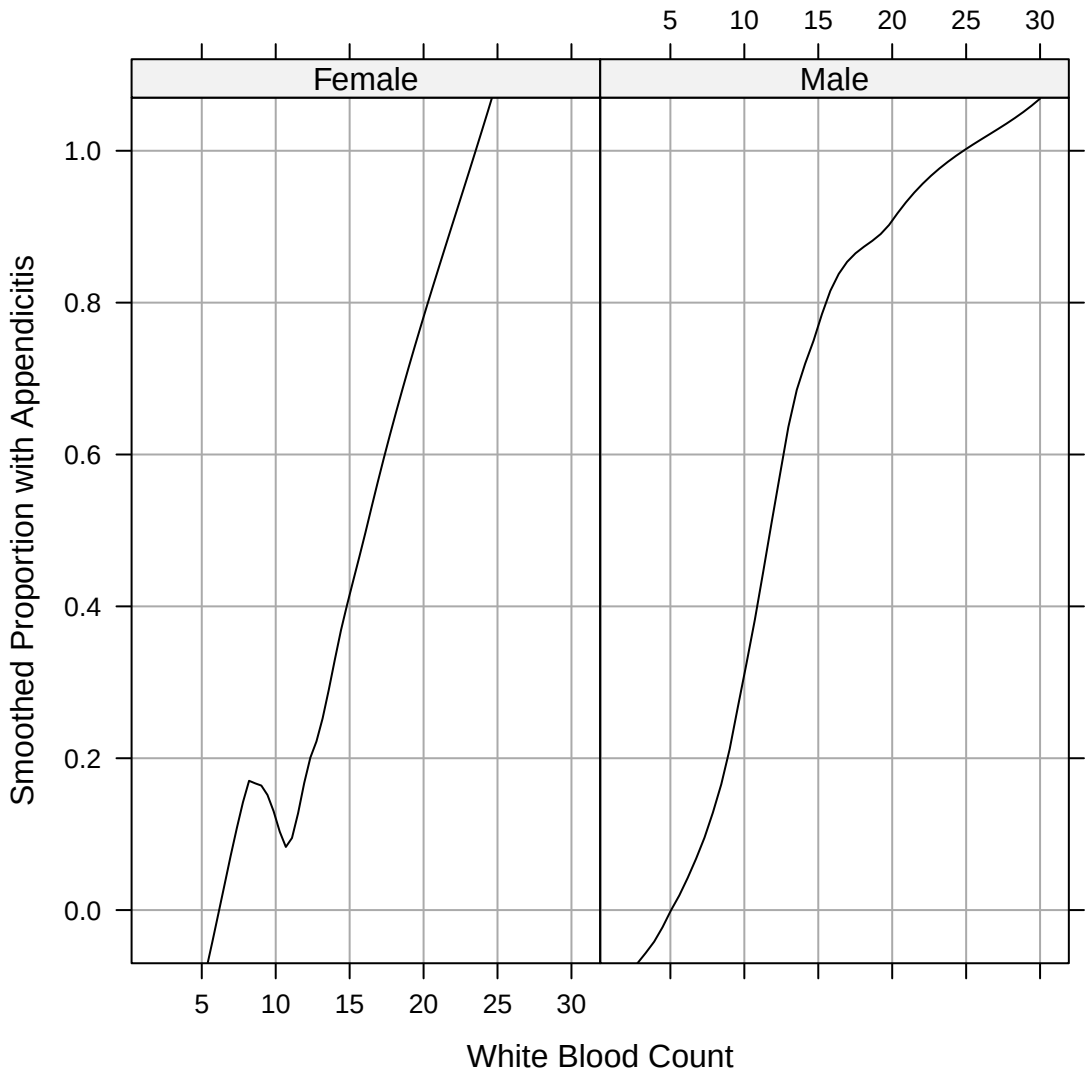

Figure 3. Smoothed Proportion of Children with Appendicitis, conditional on *gender* and *wbc*

The next plot depicts the smoothed proportion of children with appendicitis conditional on gender and estimated prior probability. Note that clinicians appear to greatly overestimate the probability of appendicitis in girls. Only one of the

previously published clinical prediction rules has included gender [14]. Thus, I will consider including *gender* as a covariate in my candidate prediction models.

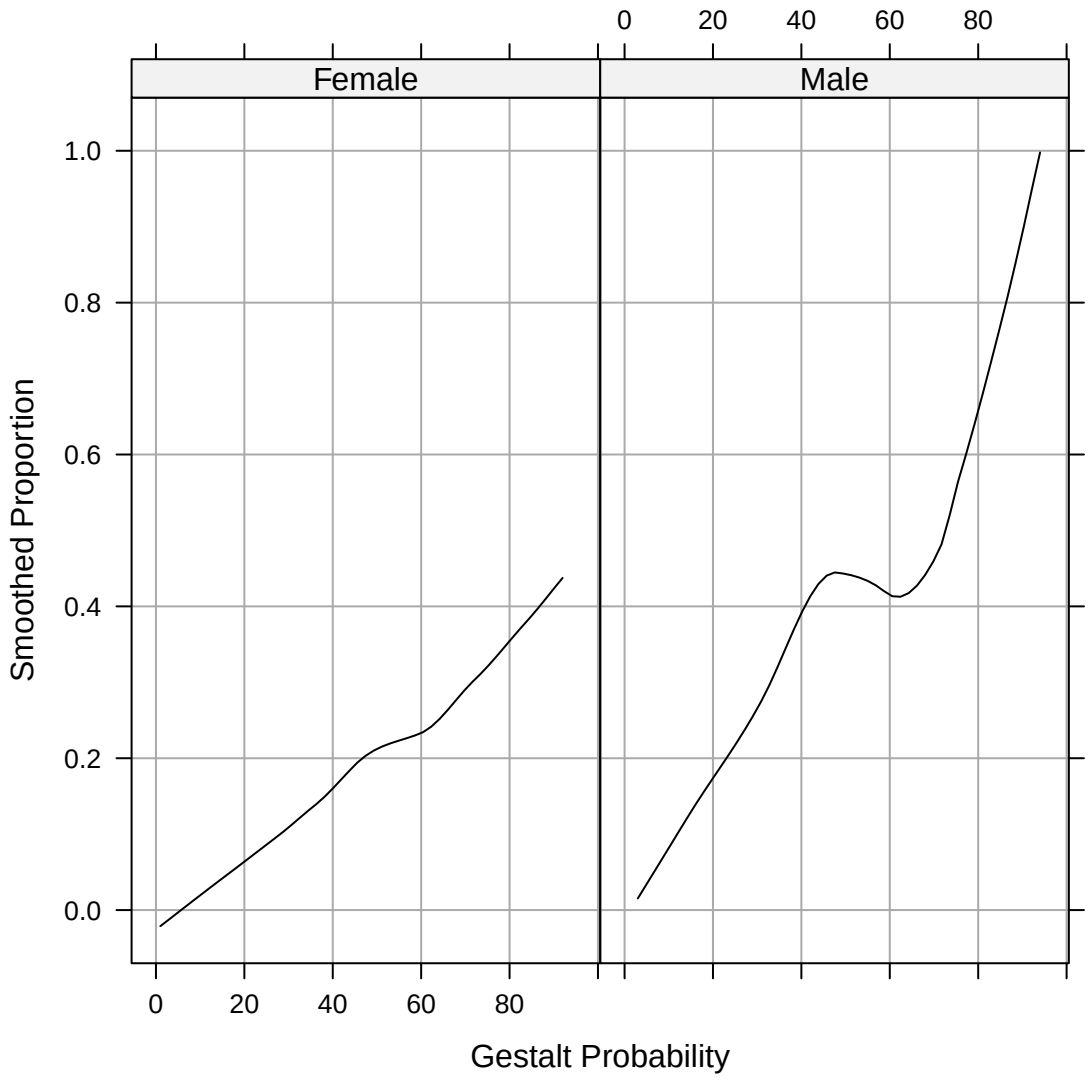

Figure 4. Smoothed Proportion of Children with Appendicitis, Conditional on *priorprob* and *gender*

1.7 Missing Data Management

1.7.1 Missing Values

These data have relatively few missing values for each variable. The software default is to discard any row (patient) with any missing data (listwise deletion). If all available variables were included in the analysis, case-wise deletion would reduce the sample size from 143 to 122, reducing the precision of predictions. Such listwise deletion assumes that the data are missing completely at random (MCAR), ie. complete observations are a random subsample of the full dataset. To the degree that the MCAR assumption is violated, and the missing data mechanism is missing at random (MAR), the resulting regression parameters will be biased. Thus, in many common scenarios, imputation of missing values may be superior to complete case analysis [15].

1.7.2 Simple Imputation

As a first approach, I have set the value of each missing to the sample (unconditional) median for that variable. Missing values of categorical variables are assigned the most common category for that variable.

1.7.3 Multiple Imputation

In Chapter 2, the R package **mice** is used to perform multiple imputation [16]. Missing values are imputed by Gibbs sampling. By default, each variable containing missing values is predicted from all other variables in the dataset. These prediction equations are used to impute plausible values for the missing data. The process iterates until convergence over the missing values is achieved. By default, predictive mean matching is used to replace missing data on continuous variables, while logistic regression is used for target variables that are dichotomous [17].

In Multiple imputation (MI), this process is used to complete five completed datasets from the existing dataset that contains missing values. Standard methods

are applied to each of the simulated datasets, and the estimates from individual models are combined to provide estimated results and confidence intervals that take into account the uncertainty introduced by the missing values.

List of References

[1] A. B. Kharbanda, G. A. Taylor, S. J. Fishman, and R. G. Bachur, "A clinical decision rule to identify children at low risk for appendicitis." *Pediatrics*, vol. 116, no. 3, pp. 709–716, Sep 2005.

[2] C. Schneider, A. Kharbanda, and R. Bachur, "Evaluating appendicitis scoring systems using a prospective pediatric cohort," *Annals of Emergency Medicine*, vol. 49, no. 6, pp. 778–84, Jun 2007.

[3] F. E. Harrell, Jr., "UseR!2010. invited lecture: Information allergy," 2010, Accessed 26 Nov 2011. [Online]. Available: http://www.r-project.org/conferences/useR-2010/Invited/harrell_talk.pdf

[4] E. W. Steyerberg, *Clinical prediction models : a practical approach to development, validation, and updating.* New York, NY: Springer, 2009. [Online]. Available: http://www.clinicalpredictionmodels.org/

[5] B. A. Hagendorf, J. R. Clarke, and R. S. Burd, "The optimal initial management of children with suspected appendicitis: a decision analysis," *Journal of Pediatric Surgery*, vol. 39, no. 6, pp. 880–885, Jun 2004.

[6] F. E. Harrell, Jr., *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis.* New York: Springer, 2001. [Online]. Available: http://biostat.mc.vanderbilt.edu/wiki/Main/RmS

[7] M. Kuhn, "CRAN Task View: Reproducible Research," Accessed 15 Nov 2011. [Online]. Available: http://cran.r-project.org/web/views/ReproducibleResearch.html

[8] A. Rossini, R. Heiberger, R. Sparapani, M. Maechler, and K. Hornik, "Emacs Speaks Statistics: A multiplatform, multipackage development environment for statistical analysis," *Journal of Computational and Graphical Statistics*, vol. 13, no. 1, pp. 247–261, 2004.

[9] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, Accessed 26 Nov 2011. [Online]. Available: http://www.R-project.org/

[10] F. Leisch, "Sweave: Dynamic generation of statistical reports using literate data analysis," in *Compstat 2002 — Proceedings in Computational*

Statistics, W. Härdle and B. Rönz, Eds. Physica Verlag, Heidelberg, 2002, pp. 575–580, iSBN 3-7908-1517-9. [Online]. Available: http://www.stat.uni-muenchen.de/~leisch/Sweave

[11] T. M. Toolan. University of Rhode Island. "A guide to producing your thesis with latex." Accessed 14 Nov 2011. June 2008. [Online]. Available: http://www.ele.uri.edu/info/thesis/guide

[12] K. Rufibach, "reporttools: R functions to generate latex tables of descriptive statistics," *Journal of Statistical Software, Code Snippets*, vol. 31, no. 1, pp. 1–7, 8 2009. [Online]. Available: http://www.jstatsoft.org/v31/c01

[13] D. Sarkar, *Lattice : multivariate data visualization with R.* New York: Springer, 2008. [Online]. Available: http://lmdvr.r-forge.r-project.org/figures/figures.html

[14] H. Lintula, E. Pesonen, H. Kokki, K. Vanamo, and M. Eskelinen, "A diagnostic score for children with suspected appendicitis." *Langenbecks Arch Surg*, vol. 390, no. 2, pp. 164–170, Apr 2005.

[15] G. J. van der Heijden, A. Rogier, T. Donders, T. Stijnen, and K. G. Moons, "Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example." *J Clin Epidemiol*, vol. 59, no. 10, pp. 1102–1109, Oct 2006.

[16] S. van Buuren and K. Groothuis-Oudshoorn, "Mice : Multivariate imputation by chained equations," *Journal Of Statistical Software*, vol. To appear., 2011, Accessed 14 Nov 2011. [Online]. Available: http://www.stefvanbuuren.nl/publications.html

[17] R. I. Kabakoff, *R in Action: Data Analysis and Graphics with R.* Manning, 2011. [Online]. Available: http://www.manning.com/kabacoff/

CHAPTER 2

CLINICAL PREDICTION MODEL DEVELOPMENT

This study suffers from a common challenge — many candidate predictor variables, and relatively small sample size.

Published studies have attempted to address this challenge by selecting predictors using formal or informal stepwise methods after univariate screen for statistically significant predictors. Harrell and others condemn using the predictor-outcome relation in the data under study to select candidate predictors, since models developed in this fashion have been shown to perform poorly with future patients. Rather, their suggestion is to select predictors based on subject knowledge, formal meta-analysis from prior published research and statistical data reduction methods [1, 2].

2.1 Previous Studies

The Pediatric Appendicitis Score (PAS) score was developed using a prospective sample of 1,170 children with abdominal pain, of whom 63% had appendicitis [3]. The author of this paper employed stepwise logistic regression to create a final model from which he developed a weighted additive score, composed of 8 variables. Rather than using a model to directly predict the probability of appendicitis, he reported the sensitivity and specificity of various cutpoints of the (max 10 point) PAS score. The author of the original paper does not explain why or how he chose to dichotomize the continuous variables (white blood cell count and increased absolute neutrophil count).

The variables composing the PAS and, in parentheses, their respective weights are: cough/percussion/hop tenderness (2), anorexia (1), fever (temperature $\geq$ 38 °C (1), nausea/vomiting (1), tenderness in right lower quadrant (2), leukocyto-

sis, defined as white blood cell count $\geq$ 10,000 (1), neutrophilia, defined as absolute neutrophil count $\geq$ 7.5 (1) and history of migration of pain (1). The PAS score has been independently evaluated by at least four separate groups [4, 5, 6, 7].

A group at Boston Children's Hosptital has defined a different decision rule to identify patients at "low-risk" for appendicitis [8]. They analyzed 24 potential predictors recorded in a derivation set of 425 patients with abdominal pain, of whom 157 (37%) had appendicitis. They selected variables with less than 10% missing data and a p-value of ≤ 0.001 for a bivariate association with appendicitis. These 12 variables were then entered into a backward stepwise logistic regression analysis. Six variables were retained, and used to define a weighted score. The variables, with weights in parenthesis, were; nausea (2), history of focal RLQ pain (1), migration of pain (1), difficulty walking (1), rebound tenderness or pain with percussion (2) and absolute neutrophil count > 6.75 (6).

Lintula [9] reported a small study which enrolled 131 children with suspected appendicitis. He recorded information about 35 clinical variables, retained 19 with a univariate association with the outcome with a p-value ≤ 0.05. These 19 variables were entered into a backward stepwise logistic model. The final model contained 8 variables; gender, intensity of pain, migration of pain, vomiting, fever, pain in RLQ, bowel sounds and rebound tenderness (pain upon removal of pressure).

In an early adult study, Alvarado studied 305 adults with abdominal pain and proposed a clinical score with the variables; migration, anorexia-acetone, nausea-vomiting, tenderness in RLQ, rebound, elevation of temperature, leukocytosis ($>$ 10,000),and shift to the left ($>$ 75% neutrophils) [10].

2.2 Models Using Factors from the Pediatric Appendicitis Score

Steyerberg argues that overfitting is a major problem in regression modeling. If potential predictors are included in the model based on univariate associations

with the outcome, the effect of such predictors is overestimated, a phenomenon known as *testimation bias* [2, p. 88]. Given how they were developed, it is very likely that the prediction rules suffer from overfitting.

Nonetheless, I will first consider the factors included in the Pediatric Appendicitis Score; white blood cell count, absolute neutrophil count, pain with cough/percussion/hop, anorexia, vomiting/nausea, fever, tenderness in RLQ and history of migration of pain to the RLQ.

Harrell [1, 11] and Senn [12] make multiple criticisms of the common practice of dichotomizing continuous variables. White blood count (*wbc*) and absolute neutrophil count (*anc*) are kept as continuous variables. Rather than dichotomizing *wbc* at some arbitrary cut-point, which makes the strong assumption that there is a piecewise uniform relationship, I have kept it as a continuous variable. Given the small dataset, I will assume *wbc* enters the model linearly. In a larger dataset, Harrell recommends use of restricted cubic spline functions to avoid linearity assumption [1].

2.2.1 Redundancy Analysis

As a first step, a redundancy analysis is done using Harrell's `R redun` function in package **Hmisc** to determine if any variables can be predicted from a combination of the remaining variables using flexible parametric additive models [13]. The absolute neutrophil count can be predicted from the other variables with an R^2 of 0.96. This is not surprising since absolute neutrophil count is defined as total white blood cell count (*wbc*) times the proportion of neutrophils. These two variables are strongly collinear, as shown in Figure 5 below.

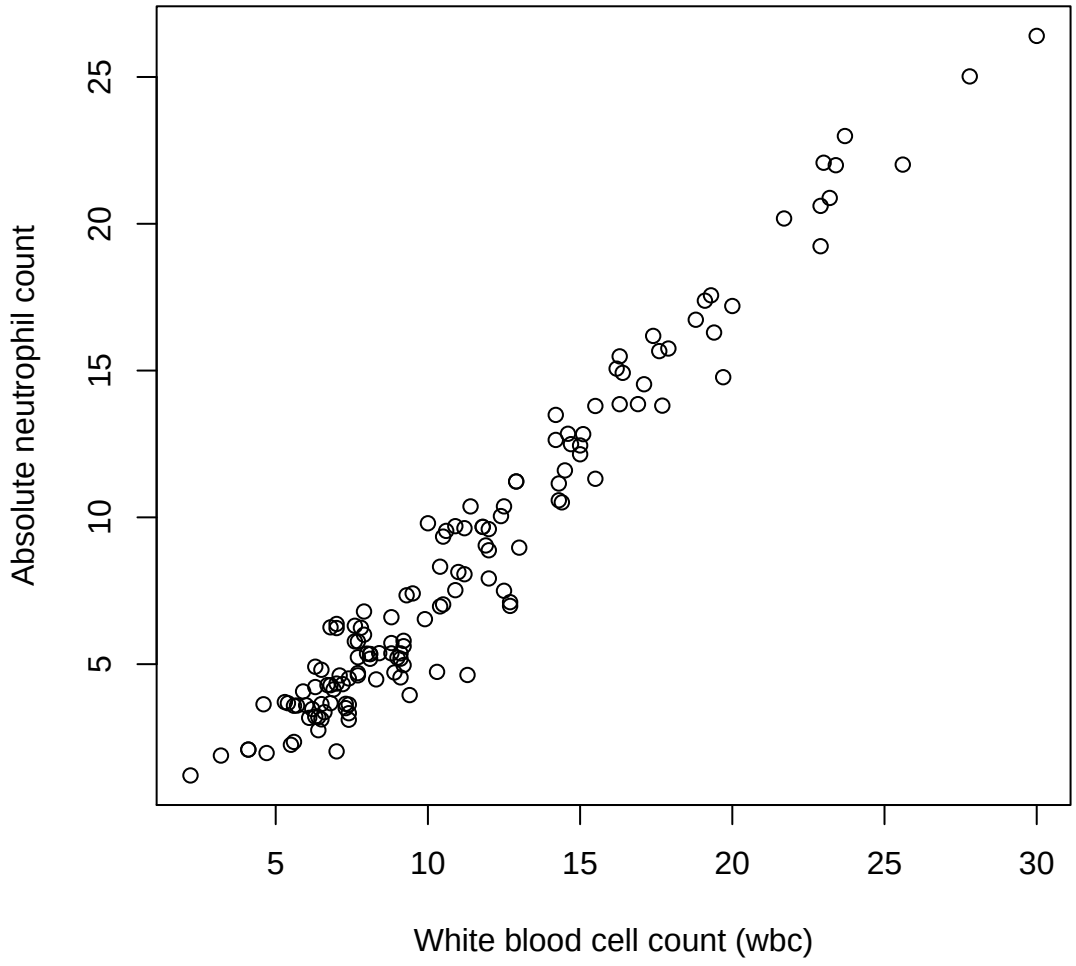

Figure 5. Plot of White blood count vs. Absolute neutrophil count

White blood cell count is familiar to clinicians and readily available. Although others have found *anc* to be a slightly better univariate predictor of appendicitis [14], redundancy analysis suggests that *anc* can be dropped from the model with little loss of predictive information.

2.2.2 Variable Clustering

As a prelude to data reduction, a hierarchical cluster analysis on the variables is done, using squared Spearman correlations as a similarity measure and plotted in Figure 6.

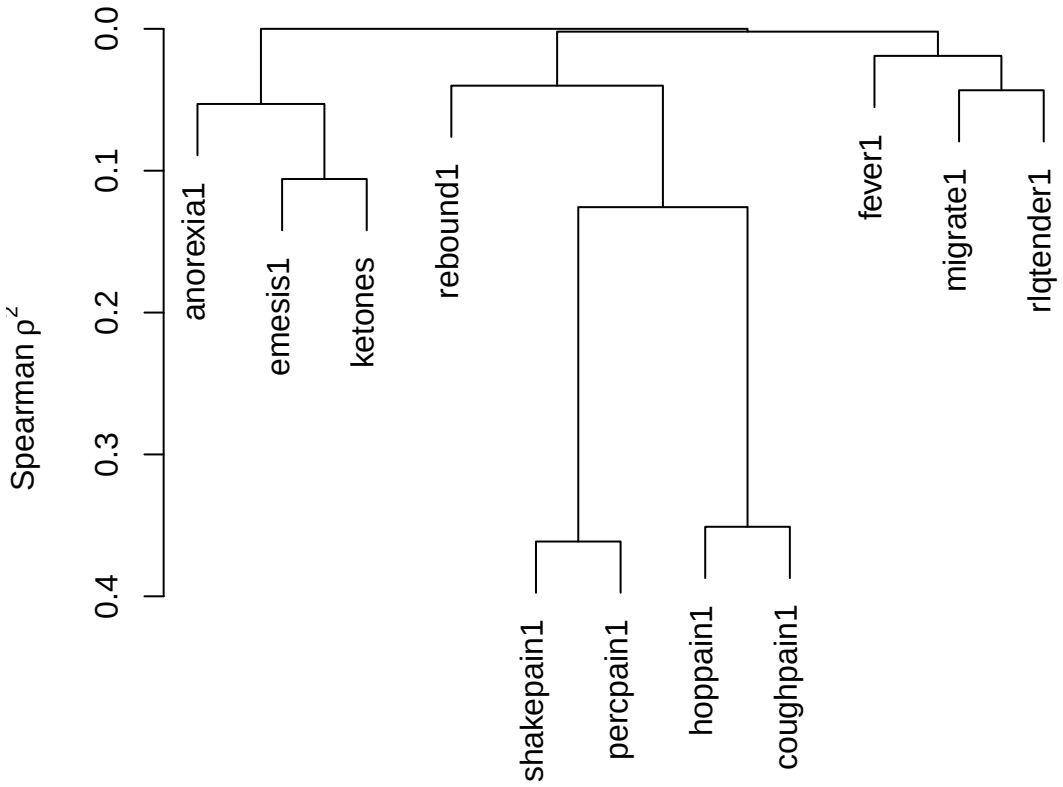

Figure 6. Hierarchical Cluster Analysis on the Variables

The first cluster; anorexia, vomiting and ketosis, makes clinical sense since ketones are a consequence of fasting resulting from nausea. Kharbanda fit a re-

cursive partitioning model of appendicitis, and found that vomiting and anorexia were surrogate variables for nausea [8]. It should be noted that young children have difficulty describing the sensation of nausea. Determination of anorexia is also difficult. In this study, we asked each child "What is your favorite food?" "If we had some here, would you want to eat it now?" Thus, a parsimonious choice might be to substitute urinary *ketones* for the variables *emesis* and *anorexia*.

A second cluster, pain with shaking, percussion, hopping or cough, likely reflects how techniques represent alternative ways to elicit signs of peritoneal irritation. I've created a new dichotomous variable, *periton*, which evaluates to 'yes' if the patient had pain with cough, percussion, hopping, or a gentle shake of the pelvis.

The inter-rater reliability of all of these measures may be limited. Cohen's κ measures the chance corrected agreement between two raters who each classify patients into mutually exclusive categories [15]. Rebound tenderness, *rebound* was found to have moderate reliability (Cohen's $\kappa = 0.54$), compared to less than moderate agreement for tenderness to percussion and palpation [16]. However, as noted by Samuel, a pediatric surgeon who developed the Pediatric Appendicitis Score, "Rebound tenderness is a particularly painful clinical feature to elicit and results in undue pain, loss of confidence and trust, and ultimately leads to loss of cooperation. Hence this sign should not be elicited in children" [3]. For me, this is a compelling argument for not including this sign in a prediction model.

Rather than perform "testimation" and step-wise selection, I chose to create a preliminary model using the variables found in the Pediatric Appendicitis Score; wbc, cough/percussion/hop, anorexia, vomiting/nausea, fever, tenderness in RLQ and history of migration of pain to the RLQ.

Since redundancy analsysis suggests that *anc* can be predicted from *wbc* and

other variables, I removed *anc* from the model.

Finally, given the apparent influence of gender seen in the exploratory plots, *gender* is included in the model, with female gender as the reference category. Since the outcome variable is binary, a generalized linear model with *logit* link (logistic model) is appropriate.

2.3 Generalized Linear Models

A generalized linear model is a statistical model in which the *linear predictor* for the ith response, $\eta_i = \boldsymbol{x}_i\boldsymbol{\beta}$ where $\boldsymbol{x}_i$ is the ith row of the $n \times p$ model matrix $\boldsymbol{X}$ derived from the form of the model and the values of any covariates, is related to the *expected value of the response*, μ_i, through an invertible *link function*, g. That is

$$\boldsymbol{x}_i\boldsymbol{\beta} = \eta_i = g(\mu_i) \quad i = 1, \ldots, n \tag{1}$$

and

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\boldsymbol{x}_i\boldsymbol{\beta}) \quad i = 1, \ldots, n \tag{2}$$

When the distribution of y_i given μ_i is from the exponential family there is a *natural* link function for the family. For a binomial response the natural link is the *logit* link defined as

$$\eta_i = g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) \quad i = 1, \ldots, n \tag{3}$$

with inverse link

$$\mu_i = g^{-1}(\eta_i) = \frac{1}{1+\exp(-\eta_i)} \quad i = 1, \ldots, n \tag{4}$$

Because μ_i is the probability of the ith observation being a "success", η_i is the log of the odds ratio. With the *logit* link, this is the multiple logistic regression

model. Models will be fitted using the `lrm` function (Logistic Regression Model) provided by the **rms**: Regression Modeling Strategies Package [17].

2.3.1 Multiple Logistic Regression Models

To avoid the loss of precision and possible bias of a complete case analysis, and to ensure that nested models are fitted on the same data, I have used the full dataset with simple (median) imputation for the preliminary models.

A likelihood ratio test is used to determine if the model that includes gender adds predictive information. The LR χ^2 is 5.562 with a p-value of 0.018. Therefore, we retain *gender*.

Akaike's information criterion (AIC) provides a method which can be used to compare two competing, non-nested models of different complexity (lower is better). A model which uses the *ketones* as a proxy for gastrointestinal distress is supported as it has an AIC of 121.63, compared to a higher AIC of 124.03 when *anorexia* and *emesis* are substituted. Thus, the model with *ketones* is preferred. Thus, our preliminary model using clinical factors is:

$$\text{Prob}\{\text{dx} = \textit{appendicitis}\} = \frac{1}{1+\exp(-\boldsymbol{X}\boldsymbol{\beta})}, \quad \text{where}$$

$$\begin{aligned}\boldsymbol{X}\hat{\boldsymbol{\beta}} = & \\ & -7.17 + 0.96\ \textit{ketones} + 0.79\ \textit{periton} + -2.02\ \textit{fever} \\ & +0.27\ \textit{migrate} + 1.94\ \textit{rlqtender} \\ & +0.28\ \textit{wbc} + 1.17\textit{gender}\end{aligned}$$

The Type II analysis of deviance table below displays the change in deviance and significance of removing each variable from a model containing all variables. Not all of the variables are significant at the p=0.05 level, but should probably be retained in a prediction model.

	LR Chisq	Df	Pr(>Chisq)
ketones	5.80	1	0.0160
periton	1.68	1	0.1950
fever	8.05	1	0.0045
migrate	0.28	1	0.5981
rlqtender	3.37	1	0.0663
wbc	31.15	1	0.0000
gender	5.56	1	0.0184

Table 4. Preliminary Clinical Factors (Type II Analysis)

We can conclude that that *ketones*, *fever*, *wbc* and *gender* are significant predictors of appendicitis after adjusting for the presence or absences of a history of migration of pain, signs of peritonitis and right lower quadrant tenderness. Surprisingly, the coefficient for fever is negative (-2.02). Thus, children with fever in this sample are less likely to have appendicitis. This finding contradicts the implicit assumption in the PAS score that fever is a positive predictor.

The direction and importance of the relationship between fever and appendicitis may depend on the duration of illness. Fever early in the clinical course may point to other bacterial or viral illnesses. However, fever in patients with longer duration of symptoms may be associated with intra-abdominal inflammation. Future studies should investigate a possible interaction between symptom duration and fever.

A Receiver Operating Characteristic curve (ROC curve) is a plot of the true positive rate against the false positive rate over the range of a predictor. The area under the ROC curve, or c-statistic, is a measure of predictive discrimination. A useless predictor has an ROC area of 0.5. The package **ROCR** is used for the ROC plots [18].

The ROC method provides a way to compare the discrimination ability of the PAS score to that for the predictions from clincical factors model.

The solid curve in the ROC plot below utilizes the predictions from our clinical factors logistic prediction model, with c-statistic (area under the curve) of 0.895. The dotted curve represents the operating characteristics of the original PAS score, calculated for each patient in the current dataset, with a c-statistic of 0.755. Applying DeLong's test, provided by the package **pROC** for correlated ROC curves, we reject the null hypothesis of equal areas under the curve, p=0.0006 [19].

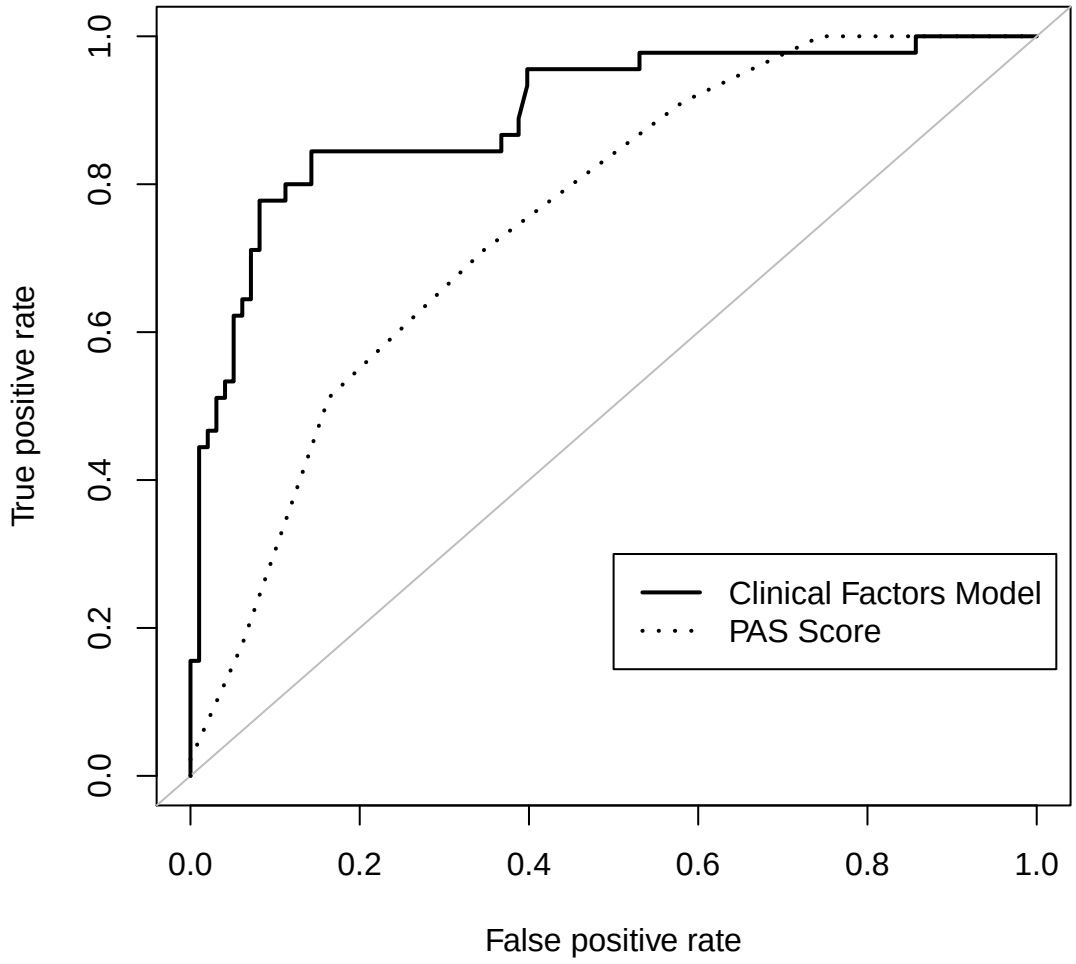

Figure 7. ROC Curves for Clinical Factors Model and Original PAS Score

2.4 Model Using Gestalt Estimate of Probability

It is apparent from the analysis of deviance for our preliminary model (Table 4) that most of the predictive power comes from *wbc*. Yen et al. found relatively poor inter-rater reliability for dichotomized physical exam variables in pediatric patients with abdominal pain [16].

An alternative approach would be to substitute each clinicians' gestalt estimate of the probability of appendicitis for individual history and physical examination factors. Recall that clinicians provided this estimate using a visual analog scale (VAS), by making a mark on a 10cm line (recorded in the variable *priorprob* on a scale of zero to 100).

Because the linear predictor η is intended to approximate the logit of the probability of appendicitis, the component of $\boldsymbol{X}$ for the clinician's gestalt estimate is expressed in the analagous form as logit(*priorprob*). In the ideal case in which clinicians are perfect diagnosticians, the coeffient of logit(*priorprob*) would be one [20].

The exploratory plot strongly suggest that clinicians overestimate the probability of appendicitis *priorprob* in girls. Thus, gender is again included in the preliminary model.

This relatively simple model with predictors has a c index (area under the ROC curve of 0.904, compared to a c index of 0.895 for the previous model with 7 predictors. I prefer the more parsimonious clinical gestalt model, which has a lower AIC of 112.317 compared to the clinical predictor model with an AIC of 121.63.

Subsequent analyses will consider this model. First, we determine the pooled fit of this model for five multiply imputed datasets.

	est	se	t	df	Pr(>|t|)	lo 95	hi 95
β_0	-5.07	0.87	-5.83	135.01	0.00	-6.79	-3.35
$\beta_{logit(priorprob)}$	0.93	0.23	4.12	131.65	0.00	0.49	1.38
β_{wbc}	0.29	0.06	4.80	132.62	0.00	0.17	0.41
β_{gender}	1.16	0.50	2.31	136.41	0.02	0.17	2.16

Table 5. Pooled Model Fit - Multiple Imputation

The next table compares the coefficients and standard errors for the model fit from single non-conditional median imputation to the multiple imputation model. As expected, given the small amount of missing data, there are only minor differences between the estimates.

	SI:coef	SI:se	MI:coef	MI:se
β_0	-4.93	0.83	-5.07	0.87
$\beta_{logit(priorprob)}$	0.86	0.22	0.93	0.23
β_{wbc}	0.28	0.06	0.29	0.06
β_{gender}	1.15	0.50	1.16	0.50

Table 6. Comparison of Coefficients: Single (Median) vs. Multiple Conditional Imputation

2.5 Bootstrap Estimation and Validation

The nonparametric bootstrap is used to estimate the regression coefficients for our model by refitting the model repeatedly on samples, with replacement, from these data. This approach does not rely on asymptotic sampling distributions to estimate the standard errors of the regression coefficients, and provides more accurate estimates for small datasets such as this one.

The bootstrap distributions of the regression coefficients found using the **boot** package are shown below [21]. Note that several of these distributions appear to be significantly skewed.

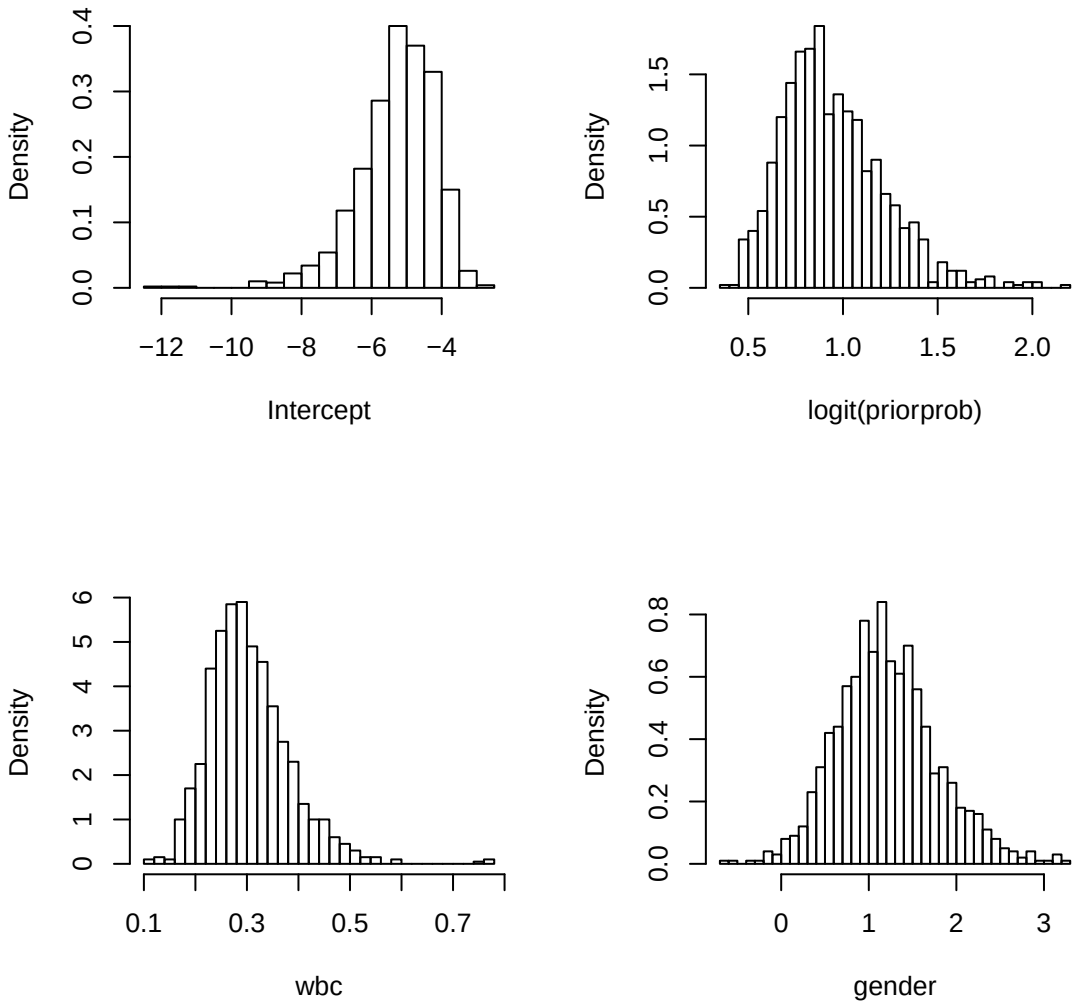

Figure 8. Bootstrap Distributions of Regression Coefficients

The table below compares the confidence intervals obtained by profiling the likelihood, with those found using the bootstrap. The bootstrap confidence intervals are somewhat wider.

	2.5 %	97.5 %
Intercept	-6.75	-3.45
logit(priorprob)	0.47	1.33
wbc	0.18	0.41
gender	0.19	2.16

Table 7. Profile Likelihood Confidence Intervals

	2.5%	97.5%
Intercept	-6.90	-3.11
logit(priorprob)	0.46	1.38
wbc	0.13	0.42
gender	0.09	2.38

Table 8. Bootstrap Confidence Intervals

We can also use the bootstrap to assess internal validity calculating indices of discrimination and plotting a calibration curve for a set of 1000 bootstrap samples. Figure 9 and 10 below, illustrate the calibration curves for the preliminary clinical factors model and the clinical gestalt model.

2.6 Model Checking

2.6.1 Using the Bootstrap to Assess Internal Validity

The **rms** package function `calibrate` is used to produce calibration plots for the two models [17]. The function uses the bootstrap to get overfitting-corrected estimates of predicted vs. observed values using nonparametric smoothers. Note that the first model exhibits greater problems with calibration, particularly for small predicted probabilities. Secondly, there appear to be greater discrepencies between the 'apparent' and bootstrap 'bias-corrected' lines in the clinical predictors model, suggesting some overfitting.

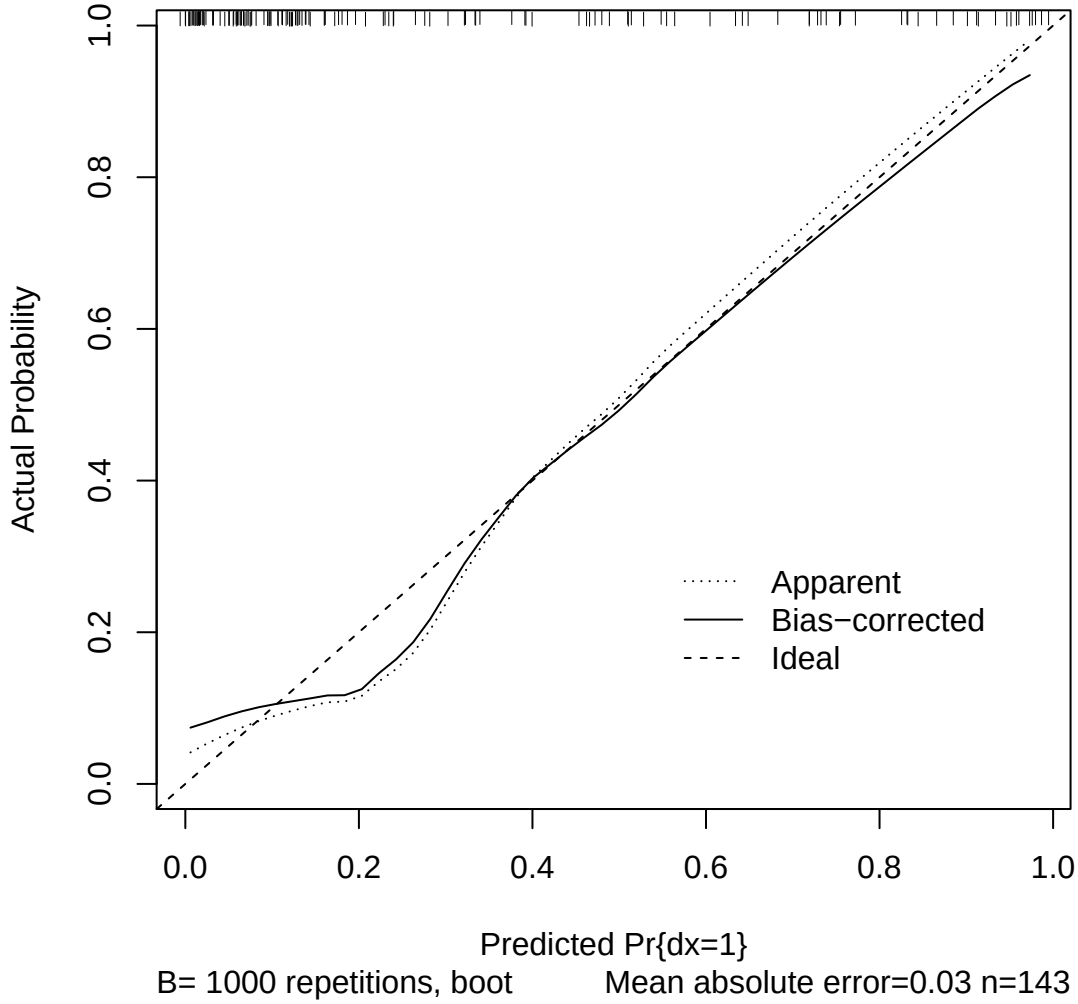

Figure 9. Clinical Factors Model: Calibration Plot

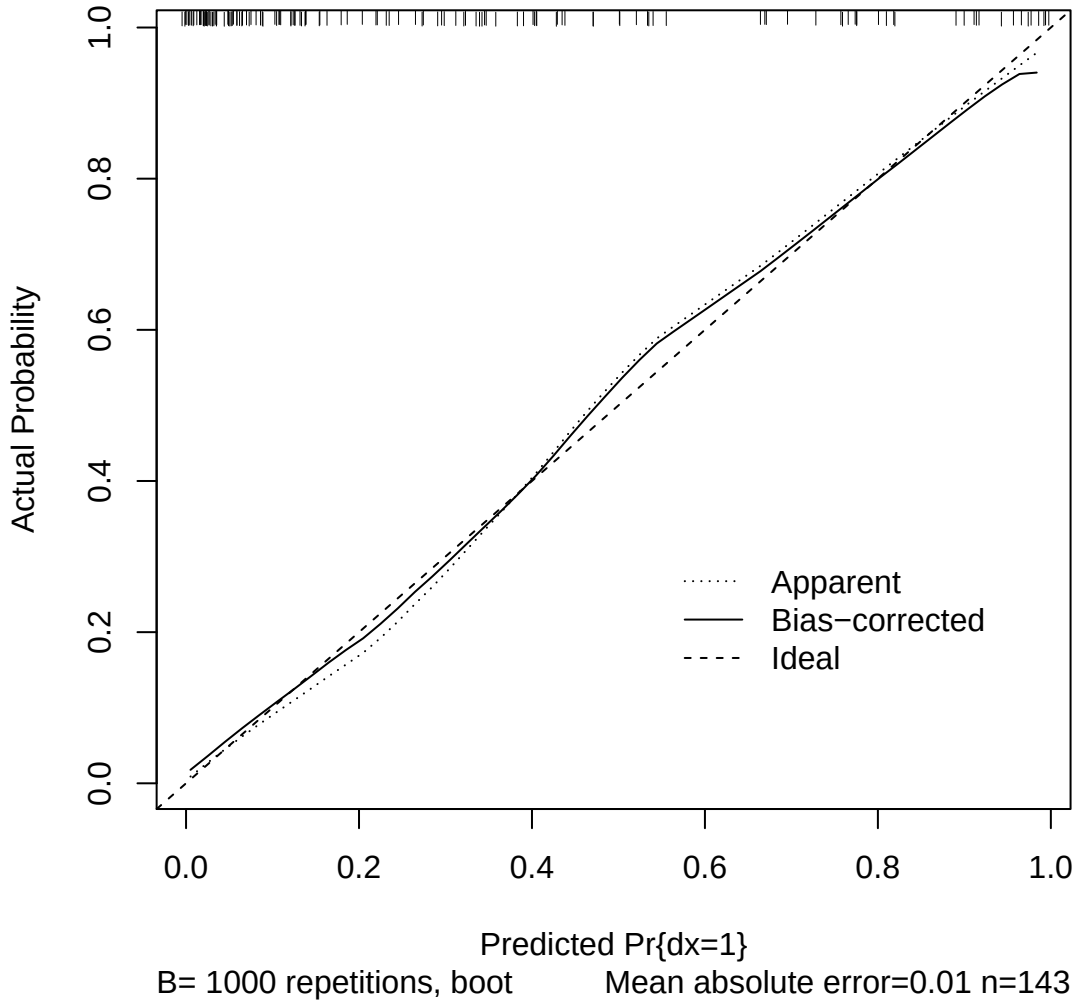

Figure 10. Clinical Gestalt Model: Calibration Plot

Another suggested way of displaying the ability of the model to discriminate is to plot side-by-side box plots of the predicted probabilities for the two possible outcomes as shown in Figure 11. This plot highlights four patients without appendicitis that were predicted to have a high probability of appendicitis by the gestalt probability model.

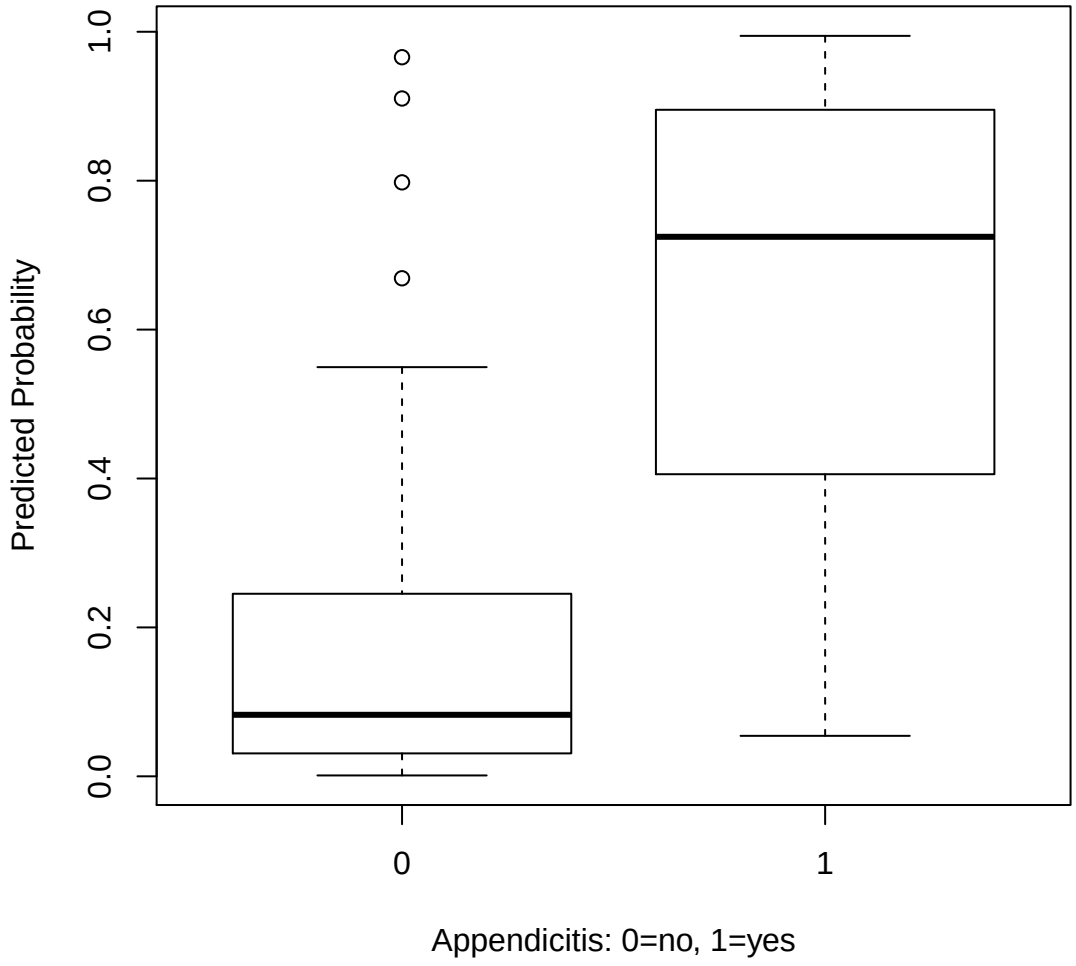

Figure 11. Predicted Probabilities by Final Diagnosis

2.6.2 Residual Analysis

There are a number of possible residuals that can be defined for the logistic model. The standardized deviance residuals from the gestalt probability model are plotted for sequential patients in Figure 12 below.

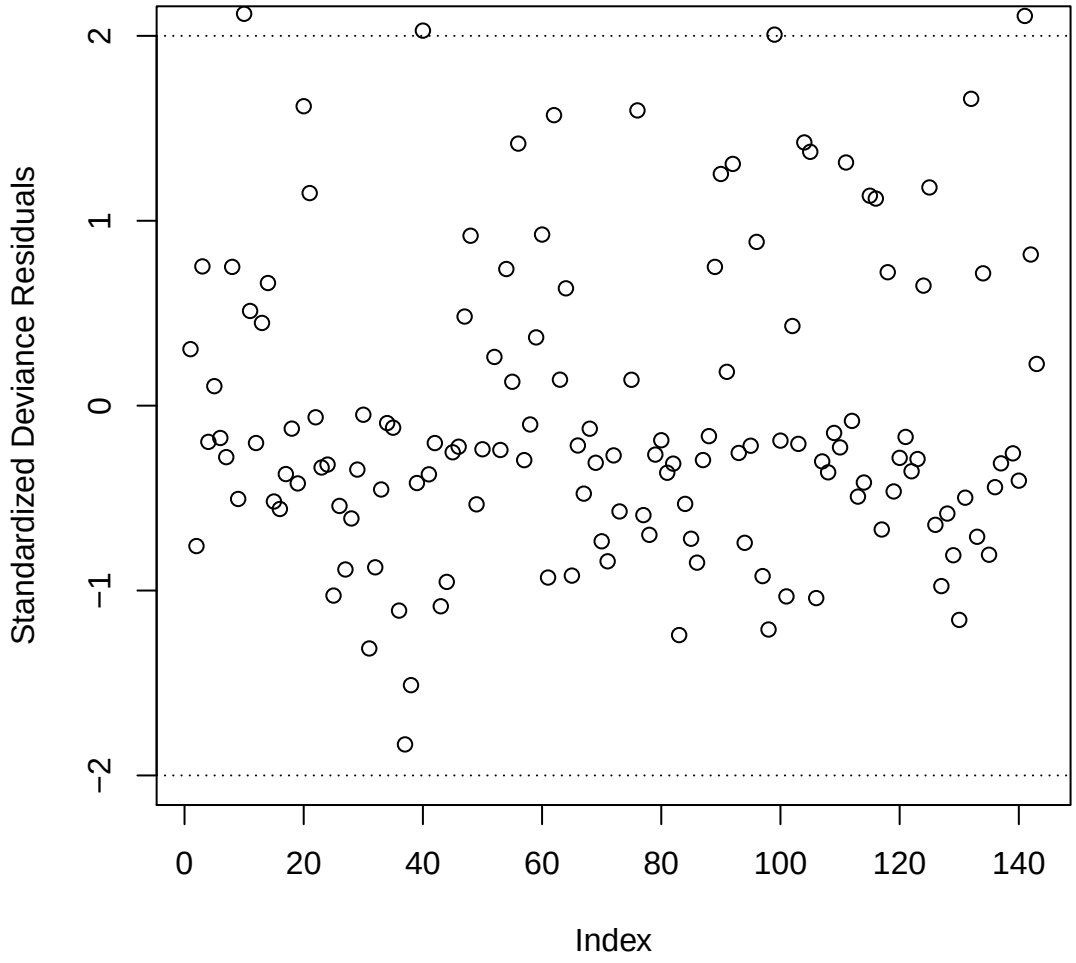

Figure 12. Standardized Deviance Residuals

The p-value for the le Cessie-van Houwelingen test is 0.57, therefore we cannot reject the null hypothesis of global goodness of fit [22].

2.7 Clinician Effects

Thus far, we have assumed that patients are independent. However, all patients were evaluated by a group of 23 clinicians who saw between 1 and 17 pa-

tients each. It is likely that each clinician used the estimated probability scale in a systematically different way. Thus, we would expect some degree of dependency between the prior probability estimates in patients who were evaluated by the same physician.

One simple approach to evaluate this would be to add 23-1 dummy variables, treated the evaluating clinician *doc* as a fixed effect. However, the coefficients for individual clinician effects (not shown) have very large standard errors, reflecting the small number of patients evaluated by each doctor. Gelman and Hill refer to this as the varying intercept model and the model omitting clinicians as the "complete pooling"or "constant intercept" model [23]. Given we are not primarily interested in comparing the diagnoses of specific doctors, a better approach may be to treat *doc* as a random effect.

List of References

[1] F. E. Harrell, Jr., *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis*. New York: Springer, 2001. [Online]. Available: http://biostat.mc.vanderbilt.edu/wiki/Main/RmS

[2] E. W. Steyerberg, *Clinical prediction models : a practical approach to development, validation, and updating*. New York, NY: Springer, 2009. [Online]. Available: http://www.clinicalpredictionmodels.org/

[3] M. Samuel, "Pediatric appendicitis score." *J Pediatr Surg*, vol. 37, no. 6, pp. 877–881, Jun 2002.

[4] R. D. Goldman, S. Carter, D. Stephens, R. Antoon, W. Mounstephen, and J. C. Langer, "Prospective validation of the pediatric appendicitis score." *J Pediatr*, vol. 153, no. 2, pp. 278–282, Aug 2008.

[5] M. Bhatt, L. Joseph, F. M. Ducharme, G. Dougherty, and D. McGillivray, "Prospective validation of the pediatric appendicitis score in a Canadian pediatric emergency department." *Acad Emerg Med*, vol. 16, no. 7, pp. 591–596, Jul 2009.

[6] A. Escribá, A. M. Gamell, Y. Fernández, J. M. Quintillá, and C. L. Cubells, "Prospective validation of two systems of classification for the diagnosis of

acute appendicitis." *Pediatr Emerg Care*, vol. 27, no. 3, pp. 165–169, Mar 2011.

[7] K. Mandeville, T. Pottker, B. Bulloch, and J. Liu, "Using appendicitis scores in the pediatric ED," *Am J Emerg Med*, vol. 29, pp. 972–977, Nov 2011.

[8] A. B. Kharbanda, G. A. Taylor, S. J. Fishman, and R. G. Bachur, "A clinical decision rule to identify children at low risk for appendicitis." *Pediatrics*, vol. 116, no. 3, pp. 709–716, Sep 2005.

[9] H. Lintula, E. Pesonen, H. Kokki, K. Vanamo, and M. Eskelinen, "A diagnostic score for children with suspected appendicitis." *Langenbecks Arch Surg*, vol. 390, no. 2, pp. 164–170, Apr 2005.

[10] A. Alvarado, "A practical score for the early diagnosis of acute appendicitis." *Ann Emerg Med*, vol. 15, no. 5, pp. 557–564, May 1986.

[11] F. E. Harrell, Jr., "UseR!2010. invited lecture: Information allergy," 2010, Accessed 26 Nov 2011. [Online]. Available: http://www.r-project.org/conferences/useR-2010/Invited/harrell_talk.pdf

[12] S. Senn and S. Julious, "Measurement in clinical trials: A neglected issue for statisticians?" *Statistics in Medicine*, vol. 28, no. 26, pp. 3189–3209, 2009.

[13] F. E. Harrell, Jr., *Hmisc: Harrell Miscellaneous*, 2011, Accessed 26 Nov 2011. [Online]. Available: http://CRAN.R-project.org/package=Hmisc

[14] D. G. Bundy, J. S. Byerley, E. A. Liles, E. M. Perrin, J. Katznelson, and H. E. Rice, "Does this child have appendicitis?" *JAMA*, vol. 298, pp. 438–451, Jul 2007.

[15] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[16] K. Yen, A. Karpas, H. J. Pinkerton, and M. H. Gorelick, "Interexaminer reliability in physical examination of pediatric patients with abdominal pain." *Arch Pediatr Adolesc Med*, vol. 159, no. 4, pp. 373–376, Apr 2005.

[17] F. E. Harrell, Jr., *rms: Regression Modeling Strategies*, 2011, Accessed 26 Nov 2011. [Online]. Available: http://biostat.mc.vanderbilt.edu/wiki/Main/Rrms

[18] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer, *ROCR: Visualizing the performance of scoring classifiers.*, 2009, Accessed 4 Dec 2011. [Online]. Available: http://CRAN.R-project.org/package=ROCR

[19] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, p. 77, 2011.

[20] J. Burkett, Personal communication, 2011.

[21] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press, 1997, iSBN 0-521-57391-2. [Online]. Available: http://statwww.epfl.ch/davison/BMA/

[22] S. le Cessie and J. C. van Houwelingen, "A goodness-of-fit test for binary regression models, based on smoothing methods," *Biometrics*, vol. 47, pp. 1267–1282, 1991.

[23] A. Gelman and J. Hill, *Data analysis using regression and multilevel/hierarchical models*. Cambridge ; New York: Cambridge University Press, 2007. [Online]. Available: http://www.stat.columbia.edu/˜gelman/arm/

CHAPTER 3

GENERALIZED LINEAR MIXED MODELS (GLMM)

In a generalized linear mixed model (GLMM) the n-dimensional vector of linear predictors, $\boldsymbol{\eta}$, incorporates both fixed effects, $\boldsymbol{\beta}$, and random effects, $\boldsymbol{b}$, as

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} \tag{5}$$

where $\boldsymbol{X}$ is an $n \times p$ model matrix and $\boldsymbol{Z}$ is an $n \times q$ model matrix.

The distribution of the random effects is modeled as as a multivariate normal (Gaussian) distribution with mean $\mathbf{0}$ and $q \times q$ variance-covariance matrix $\boldsymbol{\Sigma}$. That is,

$$\boldsymbol{b} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{6}$$

Generalized linear mixed models add random effect(s) to the model, allowing for clinician specific variation in intercepts, and possibly slopes. Models with a random intercept only will be compared to models with correlated random intercept and slopes using Likelihood Ratio (LR) tests.

The methods used by **lme4** integrate over random effects to compute the likelihood using either a Laplace approximation, or in some situations adaptive Gauss-Hermite quadrature (GHQ), which is more accurate, but more computationally intensive [1, 2].

3.1 Random Slope and Intercept Model

The first model assumes random slope of *priorprob* within *doc* and a correlated random intercept. The Laplace approximation is used for these fits.

	Estimate	Std. Error	z value	Pr(>|z|)
β_0	-5.79	1.04	-5.54	0.00
$\beta_{logit(priorprob)}$	0.99	0.26	3.75	0.00
β_{wbc}	0.35	0.07	4.89	0.00
β_{gender}	1.43	0.59	2.43	0.02

Table 9. Random Intercept/Random Slope Model

3.2 Random Intercept Model

The next model, again fit using the Laplace approximation, assumes a random group intercept only.

	Estimate	Std. Error	z value	Pr(>|z|)
β_0	-5.68	1.02	-5.58	0.00
$\beta_{logit(priorprob)}$	0.98	0.24	4.04	0.00
β_{wbc}	0.33	0.07	4.77	0.00
β_{gender}	1.40	0.57	2.48	0.01

Table 10. Random Intercept Model

The estimated variance σ^2 of the random effect for *doc* is 0.961. Thus, σ is 0.98.

	Df	logLik	Chisq	Chi Df	Pr(>Chisq)
Random Intercept	5	-50.98			
Random Intercept/Slope	7	-50.38	1.20	2	0.5475

Table 11. Likelihood Ratio Test

From the likelihood ratio test above, we can conclude that the random intercept-random slope model is not significantly better than the random intercept model. Thus, I'll retain the simpler random intercept model. The next table summarizes model fit when Adaptive Gaussian Quadrature is utilized. Note that the estimates are identical.

	Estimate	Std. Error	z value	Pr(>|z|)
β_0	-5.68	1.02	-5.58	0.00
$\beta_{logit(priorprob)}$	0.98	0.24	4.04	0.00
β_{wbc}	0.33	0.07	4.77	0.00
β_{gender}	1.40	0.57	2.48	0.01

Table 12. Random Intercepts Model (using Adaptive Gaussian Quadrature)

The standard deviation of the random intercept term is 0.982. The predicted probability for a doctor not in the study will be based on the fixed-effects only, because it applies to a doctor not in this study, the expected value for the random effect is zero in the absence of any information on that doctor.

There is a simple random effect for each *doc*. To get a prediction for a specific physician one adds the random effect to the value of η before transforming it.

In the next chapter, the preceding generalized linear mixed models are simulated using Bayesian methods.

List of References

[1] D. Bates, M. Maechler, and B. Bolker, *lme4: Linear mixed-effects models using S4 classes*, 2011, Accessed 26 Nov 2011. [Online]. Available: http://R-Forge.R-project.org/projects/lme4/

[2] B. M. Bolker, M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. Stevens, and J. S. White, "Generalized linear mixed models: a practical guide for ecology and evolution," *Trends in Ecology & Evolution*, vol. 24, no. 3, pp. 127–135, Mar 2009.

CHAPTER 4

BAYESIAN APPROACHES

Bayesian inference is based on summary statistics of the posterior distribution, which is proportional to the product of the prior distribution and the likelihood [1]. Markov Chain Monte Carlo (MCMC) methods generate random samples from the posterior distributions of parameter values.

Just Another Gibbs Sampler (JAGS) is a program for analysis of Bayesian hierarchical models MCMC simulation, written in C++ by Martyn Plummer as an extensible, multi-platform clone of BUGS [2].

The JAGS glm module used for all models implements samplers for efficient updating of generalized linear and generalized linear mixed models. According to Plummer, block updating of the parameters in the linear predictor frees the user from the need to center predictor variables, without affecting the mixing of the Markov chain.

The model is defined in a text file using a dialect of the BUGS language. Two types of relations are defined. A stochastic relation ($\sim$) defines a stochastic node which represents a random variable in the model. A deterministic relation (<-) defines a deterministic node, the value of which is determined exactly by the values of its parents.

4.1 Bayesian Logistic Regression

One of the advantages of the Bayesian approach is that the posterior from one model can provide prior information for subsequent models. The data from a similar study at the Children's Hospital of Philadelphia (CHOP) allow us to model the final diagnosis of appendicitis as a function of *wbc* and *gender*.

4.1.1 Logistic Regression Model: Children's Hospital of Philadelphia Data

In this section, I fit a logistic model to the CHOP dataset, with covariates *wbc* and *gender*. Diffuse, non-informative independent normal priors with mean zero and large variance (small precision) were assumed for the coefficient vector $\boldsymbol{\beta}$.

The BUGS code for model is shown below:

```
model {
# Likelihood
 for (i in 1:n) {
    appy[i] ~ dbin(p[i], 1)
    logit(p[i]) <- b.0 + b.wbc*wbc[i]+ b.gender*gender[i]
    }


# independent normal priors
 b.0 ~ dnorm(0,0.0001)
 b.gender ~ dnorm(0,0.0001)
 b.wbc ~ dnorm(0,0.0001)
}
```

A summary of posterior distributions follows:

	mean	sd	2.5%	50%	97.5%
β_0	-3.810	0.557	-4.950	-3.792	-2.776
β_{gender}	1.147	0.327	0.516	1.143	1.795
β_{wbc}	0.236	0.041	0.160	0.234	0.319

Table 13. CHOP data: Posterior Summary

4.1.2 Logistic Regression Models - HCH data
Diffuse Priors

We now fit a model which includes *gender*, *wbc*, *priorprob* as covariates, using diffuse priors.

One of the advantages of Bayesian modeling is the ability to directly sample from functions of the parameters such as odds-ratios (OR).

The BUGS code for the model is shown below:

```
model {
# Likelihood
 for (i in 1:n) {
    appy[i] ~ dbin(p[i], 1)
    logit(p[i]) <- b.0 +  b.logitpp*logitpp[i] + b.wbc*wbc[i] + b.
        gender*gender[i]
     }


# Diffuse priors
 b.0 ~ dnorm(0,0.0001)
 b.logitpp ~ dnorm(0, 0.0001)
 b.wbc ~ dnorm(0,0.0001)
 b.gender ~ dnorm(0,0.0001)


# posterior predictions
 or.gender <- exp( b.gender )
 or.wbc <- exp( b.wbc )
 or.logitpp <- exp( b.logitpp)
}
```

	mean	sd	2.5%	50%	97.5%
β_0	-5.207	0.859	-7.024	-5.164	-3.653
$\beta_{logit(priorprob)}$	0.913	0.227	0.492	0.903	1.387
β_{wbc}	0.300	0.060	0.189	0.297	0.424
β_{gender}	1.199	0.509	0.225	1.193	2.211
$OR_{logit(priorprob)}$	2.559	0.607	1.636	2.467	4.002
OR_{wbc}	1.352	0.082	1.209	1.345	1.527
OR_{gender}	3.781	2.106	1.252	3.296	9.123

Table 14. HCH data: Posterior Summary

4.1.3 Logistic Model for HCH with priors from CHOP dataset

We can now compare the estimates and predictions from the model which utilizes the prior information from the CHOP data. The credible intervals for coefficients are notably smaller when prior information is utilized.

The mean of the posterior coefficient for *wbc*, conditional on gender from the CHOP data model, is used to provide prior information. To account for uncertainty due to data from different populations, and the fact that the coefficient for *wbc* is now conditional on logit(*priorprob*), I have doubled the standard deviation (precision/4). As expected, the precision of the posterior standard deviation for 'b.wbc' improves, with little change in the posterior mean coefficient.

	mean	sd	2.5%	50%	97.5%
β_0	-4.916	0.730	-6.410	-4.893	-3.566
$\beta_{logit(priorprob)}$	0.890	0.220	0.478	0.883	1.341
β_{wbc}	0.276	0.048	0.188	0.275	0.371
β_{gender}	1.187	0.504	0.248	1.167	2.216
$OR_{logit(priorprob)}$	2.502	0.802	1.613	2.419	3.821
OR_{wbc}	1.319	0.064	1.207	1.316	1.449
OR_{gender}	3.737	2.115	1.281	3.214	9.169

Table 15. Posterior Summary - priors from CHOP

4.2 Bayesian Generalized Linear Mixed Models

The models that follow can be defined as multilevel logistic models. These can be thought of as a generalization of generalized linear models, where intercepts, and possibly slopes, are allowed to vary by group [3].

4.2.1 Varying-intercept

The BUGS code for the varying intercept model is shown below:

```
model {
# Likelihood
 for (i in 1:n) {
    appy[i] ~ dbin(p[i], 1)
    logit(p[i]) <-  a[doc[i]] + b.logitpp*logitpp[i] + b.wbc*wbc[i] +
         b.gender*gender[i]
    }
for (j in 1:ndoc){
 a[j] ~ dnorm(a.hat[j], tau.a)
 a.hat[j] <- mu.a
}

mu.a ~ dnorm(0, 0.0001)
tau.a <- pow(sigma.a, -2)
sigma.a ~ dunif(0, 100)
b.logitpp ~ dnorm(0, 0.0001)
b.gender ~ dnorm(0, 0.0001)

b.wbc ~ dnorm(0.235753240832236 , 148.6807331366)
}
```

As shown in the figure below, the three chains appear to show good mixing, suggesting that the Markov chains are fully exploring the posterior densities of the

coefficients. The posterior densities of the coefficients are plotted below.

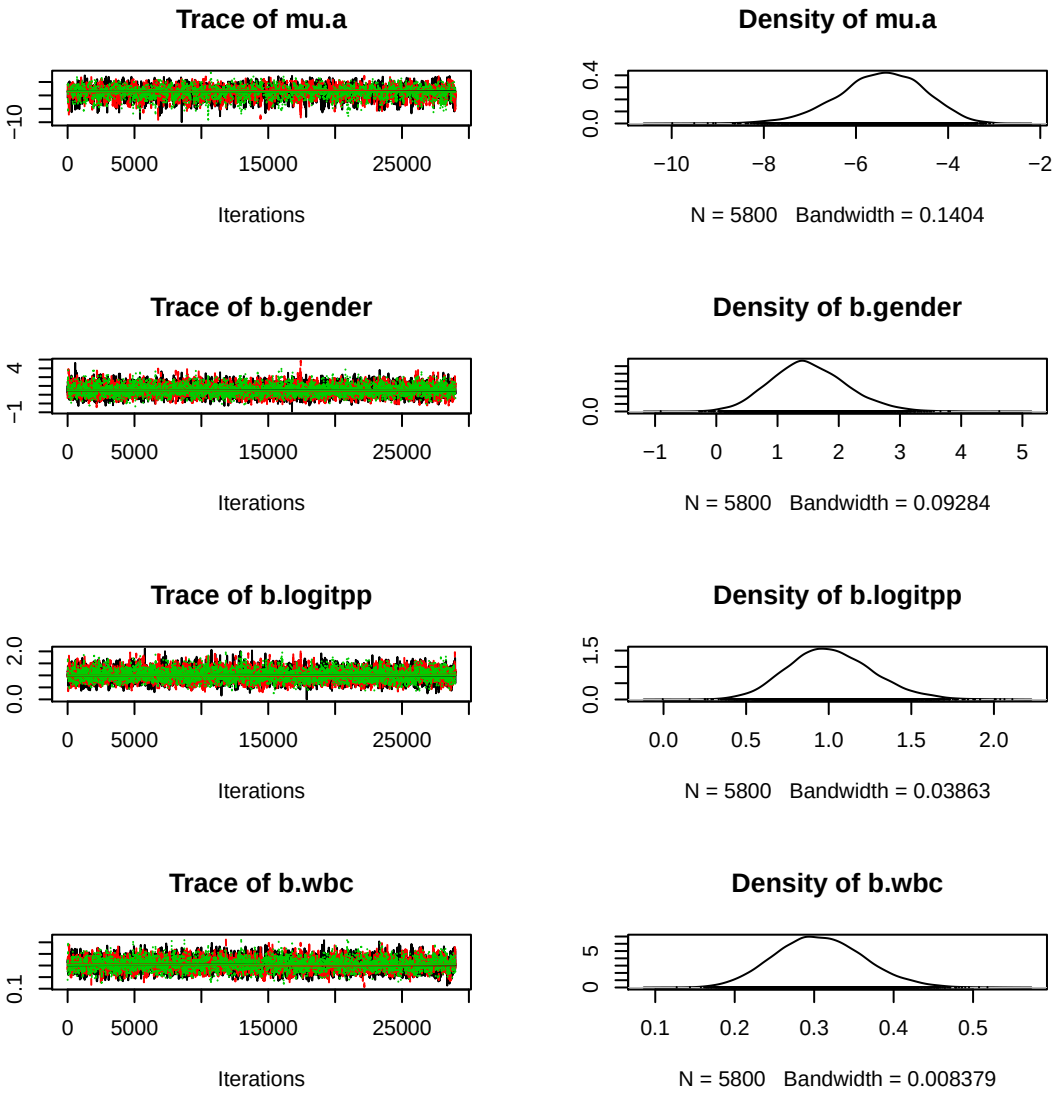

Figure 13. Trace Plots and Posterior Density Plots

The distribution of the posterior standardard deviation for the clinician specific intercepts (σ_{doc}) is shown below.

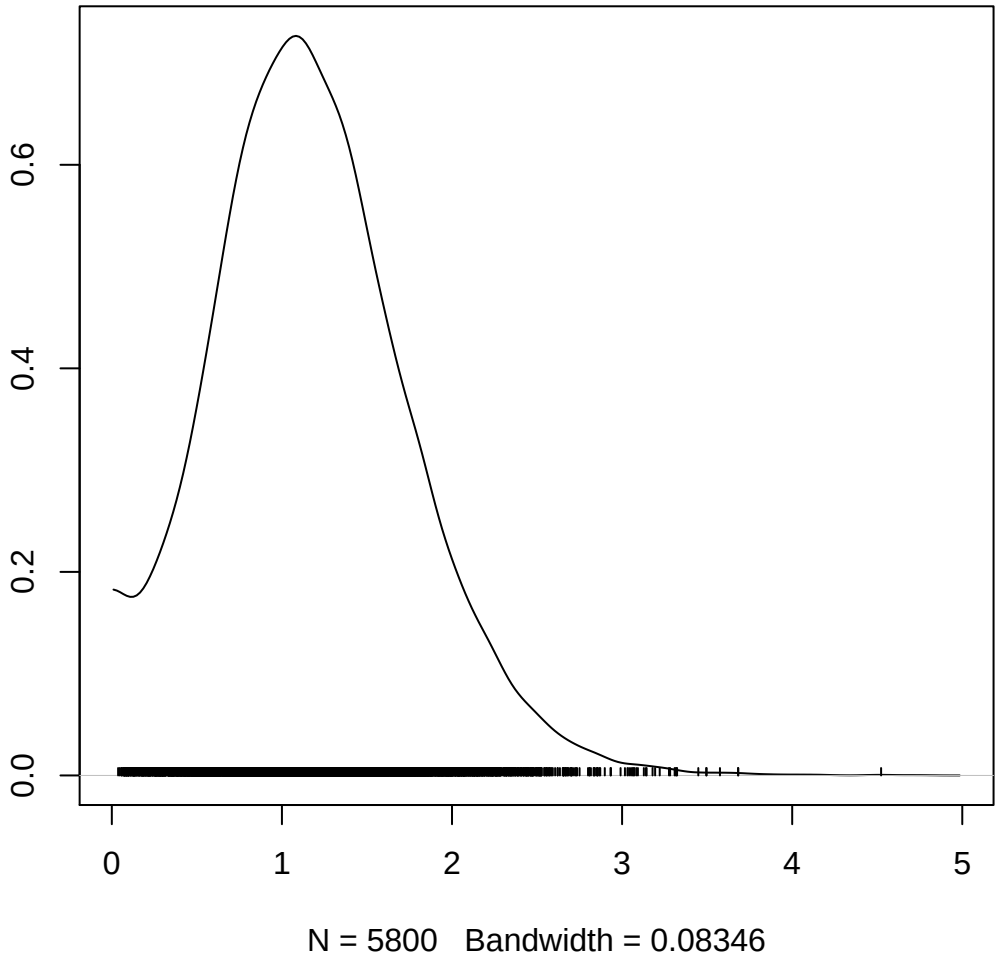

Figure 14. Posterior Density of the Standard Deviation of Clinician Specific Intercept Coefficients

The clinician specific random coefficient model is:

$$\eta_i = \alpha_{j[i]} + \beta_{logit(priorprob)} \cdot logit(priorprob)_i + \beta_{wbc} \cdot wbc_i + \beta_{gender} \cdot gender_i$$

$$\alpha_j \sim N(\mu_\alpha, \sigma^2_{doc}), \text{ for } j = 1, \ldots, n^{doc}$$

The coefficient μ_α is the expected intercept for a randomly chosen clinician. The coefficient for an individual clinician (α_j) follows a normal distribution with mean μ_α and a standard deviation of σ_{doc}.

	mean	50%	sd	2.5%	97.5%
μ_α	-5.47	-5.42	0.94	-7.46	-3.81
σ_{doc}	1.18	1.13	0.57	0.14	2.42
$\beta_{logit(priorprob)}$	1.01	1.00	0.26	0.55	1.56
β_{wbc}	0.31	0.31	0.06	0.20	0.42
β_{gender}	1.51	1.48	0.62	0.38	2.83

Table 16. Random Coefficient Model

The random intercept α_j for each clinician is plotted below. The dotted horizontal line is μ_α.

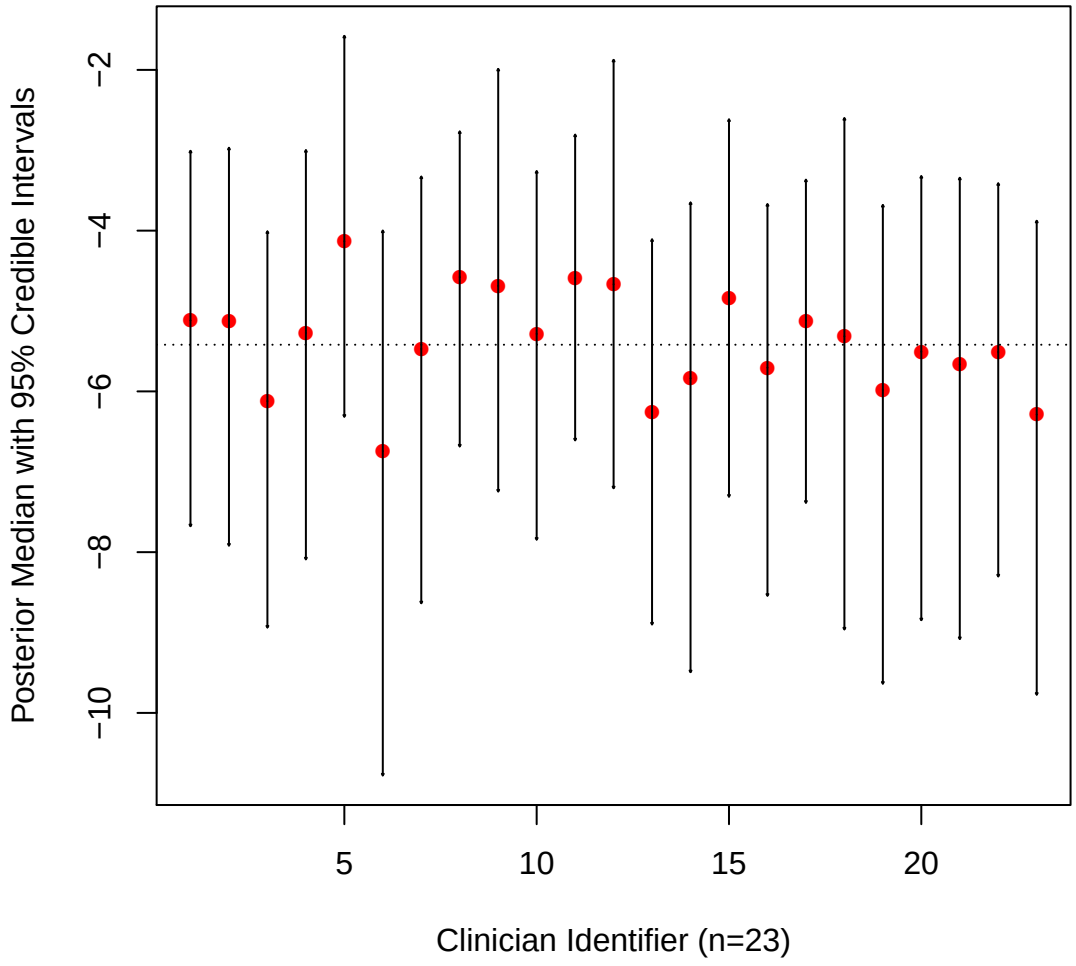

Figure 15. Caterpillar Plot of Median and 95 % Credible Invervals for Clinician Specific Intercept

4.2.2 Varying-intercept, varying-slope, no correlation between intercepts and slopes

In this model, we assume a random intercept and slope for each physician, the following BUGS code. The model statement for this model is as follows:

```
model {
```

```
# Likelihood
 for (i in 1:n) {
    appy[i] ~ dbin(p[i], 1)
    logit(p[i]) <-  a[doc[i]] + b[doc[i]]*logitpp[i] + b.wbc*wbc[i]
                    + b.gender*gender[i]
    }
for (j in 1:ndoc){
 a[j] ~ dnorm(a.hat[j], tau.a)
 b[j] ~ dnorm(b.hat[j], tau.b)
 a.hat[j] <- mu.a
 b.hat[j] <- mu.b
}

mu.a ~ dnorm(0, 0.0001)
mu.b ~ dnorm(0, 0.0001)
tau.a <- pow(sigma.a, -2)
tau.b <- pow(sigma.b, -2)
sigma.a ~ dunif(0, 100)
sigma.b ~ dunif(0,100)
b.logitpp ~ dnorm(0, 0.0001)
b.gender ~ dnorm(0, 0.0001)

b.wbc ~ dnorm(0.235753240832236 , 148.6807331366)
}
```

The random intercepts were assumed to have a normal distribution. The prior distribution of the mean of this normal distribution was assumed to be a normal distribution with mean 0 and variance 10,000. The prior distribution for the precision (the inverse of the variance) of this normal distribution was assumed to be a uniform distribution over the range of 0 to 100. The posterior mean of each regression parameter was determined from the monitored samples from the

posterior distribution.

The clinician specific random coefficient and intercept model is:

$$\eta_i = \alpha_{j[i]} + \beta_{j[i]} \cdot logit(priorprob)_i + \beta_{wbc} \cdot wbc_i + \beta_{gender} \cdot gender_i$$

$$\alpha_j \sim N(\mu_\alpha, \sigma^2_{\alpha[doc]}, \text{ for } j = 1, \ldots, n^{doc}$$

$$\beta_j \sim N(\mu_\beta, \sigma^2_{\beta[doc]}), \text{ for } j = 1, \ldots, n^{doc}$$

The coefficient μ_α is the expected intercept and μ_β is the expected slope coefficient for *logit(priorprob)* for a randomly chosen doctor.

The intercept and slope coefficients for individual clinicians (α_j) and (β_j) follow a normal distributions with means μ_α and $\mu\beta$ and standard deviations $\sigma_{\alpha[doc]}$ and $\sigma_{\beta[doc]}$.

The Bayesian posterior summary estimates for the random coefficients model are shown below. The standard deviation of the clinician specific slopes is relatively small, but still greater than zero.

	mean	50%	sd	2.5%	97.5%
μ_α	-5.707	-5.656	0.943	-7.700	-4.019
μ_β	1.168	1.129	0.349	0.582	1.957
β_{wbc}	0.315	0.314	0.055	0.212	0.426
β_{gender}	1.585	1.556	0.649	0.387	2.933
$\sigma_{\alpha[doc]}$	1.133	1.084	0.595	0.146	2.448
$\sigma_{\beta[doc]}$	0.617	0.560	0.410	0.037	1.572

Table 17. Random Slope and Intercept

4.2.3 Varying-intercept, varying-slope, Correlation ρ between intercepts and slopes

The BUGS code for model is shown below:

```
model {
# Likelihood
 for (i in 1:n) {
    appy[i] ~ dbin(p[i], 1)
    logit(p[i]) <-  a[doc[i]] + b[doc[i]]*priorprob[i] + b.wbc*wbc[i]
                  + b.gender*gender[i]
    }
for (j in 1:ndoc){
    a[j] <- xi.a*B.raw[j,1]
    b[j] <- xi.b*B.raw[j,2]
    B.raw[j,1:2] ~ dmnorm (B.raw.hat[j,], Tau.B.raw[,])
    B.raw.hat[j,1] <- mu.a.raw
    B.raw.hat[j,2] <- mu.b.raw
  }
  mu.a <- xi.a*mu.a.raw
  mu.b <- xi.b*mu.b.raw
  mu.a.raw ~ dnorm (0, .0001)
  mu.b.raw ~ dnorm (0, .0001)

  xi.a ~ dunif (0, 100)
  xi.b ~ dunif (0, 100)

  Tau.B.raw[1:2,1:2] ~ dwish(W[,], df)
  df <- 3
  Sigma.B.raw[1:2,1:2] <- inverse(Tau.B.raw[,])
  sigma.a <- xi.a*sqrt(Sigma.B.raw[1,1])
  sigma.b <- xi.b*sqrt(Sigma.B.raw[2,2])
  rho <- Sigma.B.raw[1,2]/sqrt(Sigma.B.raw[1,1]*Sigma.B.raw[2,2])
  b.wbc ~ dnorm(0, 100)
  b.gender ~ dnorm(0, 100)
}
```

	mean	50%	sd	2.5%	97.5%
μ_α	-3.95	-3.91	0.74	-5.52	-2.64
μ_β	0.87	0.86	0.24	0.45	1.40
β_{wbc}	0.25	0.24	0.05	0.15	0.36
β_{gender}	0.05	0.05	0.10	-0.14	0.24
$\sigma_{\alpha[doc]}$	0.45	0.34	0.36	0.07	1.36
$\sigma_{\beta[doc]}$	0.25	0.15	0.25	0.02	0.92
ρ	0.12	0.15	0.52	-0.85	0.93

Table 18. Correlated Random Slope and Intercept

JAGS can fit models with both a random intercept and slope that varies by clinician. The posterior median of ρ suggests a low correlation between the random effects. The posterior estimates of the standard deviation of the random slope term suggest a random intercept only model may be sufficient to account for clinician effects.

4.3 Predictions from Bayesian Random Intercepts Model

The primary use of this model will be to predict the probability of appendicitis when evaluated by a 'new' clinician with similar diagnostic ability to those who evaluated patients in this study. The Bayesian random intercept model is my preferred choice to provide predictions.

It is informative to plot predictions for three groups of patients; Figure 16, high gestalt clinical probability (*priorprob*=90%), Figure 17, lower gestalt clinical probability (*priorprob*= 10%) and Figure 18, those those with *priorprob* = 50%, conditional on *gender* and *wbc*. For each sample from the posterior distribution of coefficients, a linear predictor η can be calculated for a specified design matrix. An adaptation of the function `bprobit.probs` in the package **LearnBayes** is used to calculate predicted probabilities [4]. In each plot the median and 95% credible intervals of the posterior predictive distribution are plotted in subsequent graphs over a range of *wbc*.

In the first plot, the predicted probability of appendicitis with 95% credible intervals is plotted for boys (in blue) and girls (in pink) for a patient with a gestalt clinical predicted probability (*priorprob*=90%). Note that precise, high probability predictions are possible only for boys.

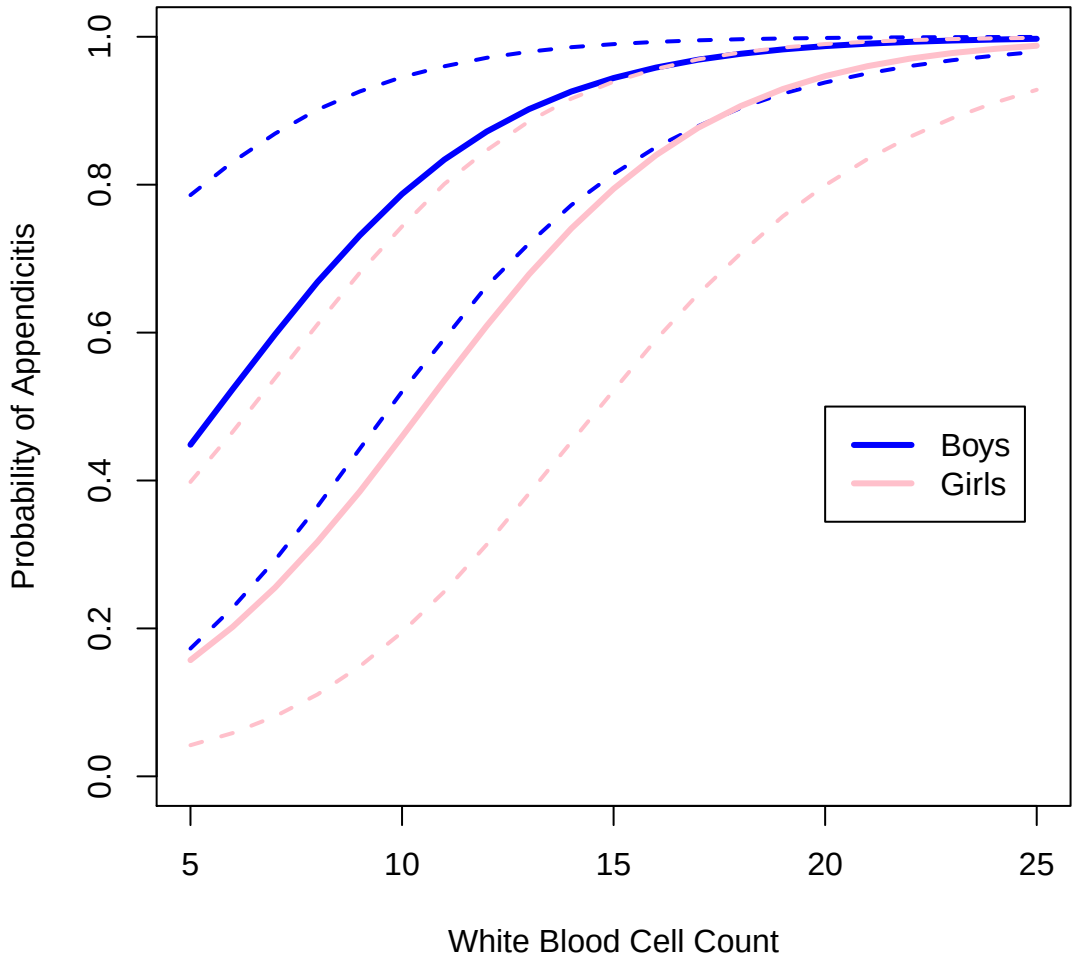

Figure 16. Probability of Appendicitis given 90% Clininical Prediction

In the next plot, the predicted probability of appendicitis with 95% credible intervals, is plotted for boys (in blue) and girls (in pink) for a patient with a gestalt

clinical predicted probability (*priorprob*=10%). Note that precise, low probability predictions are only possible for girls.

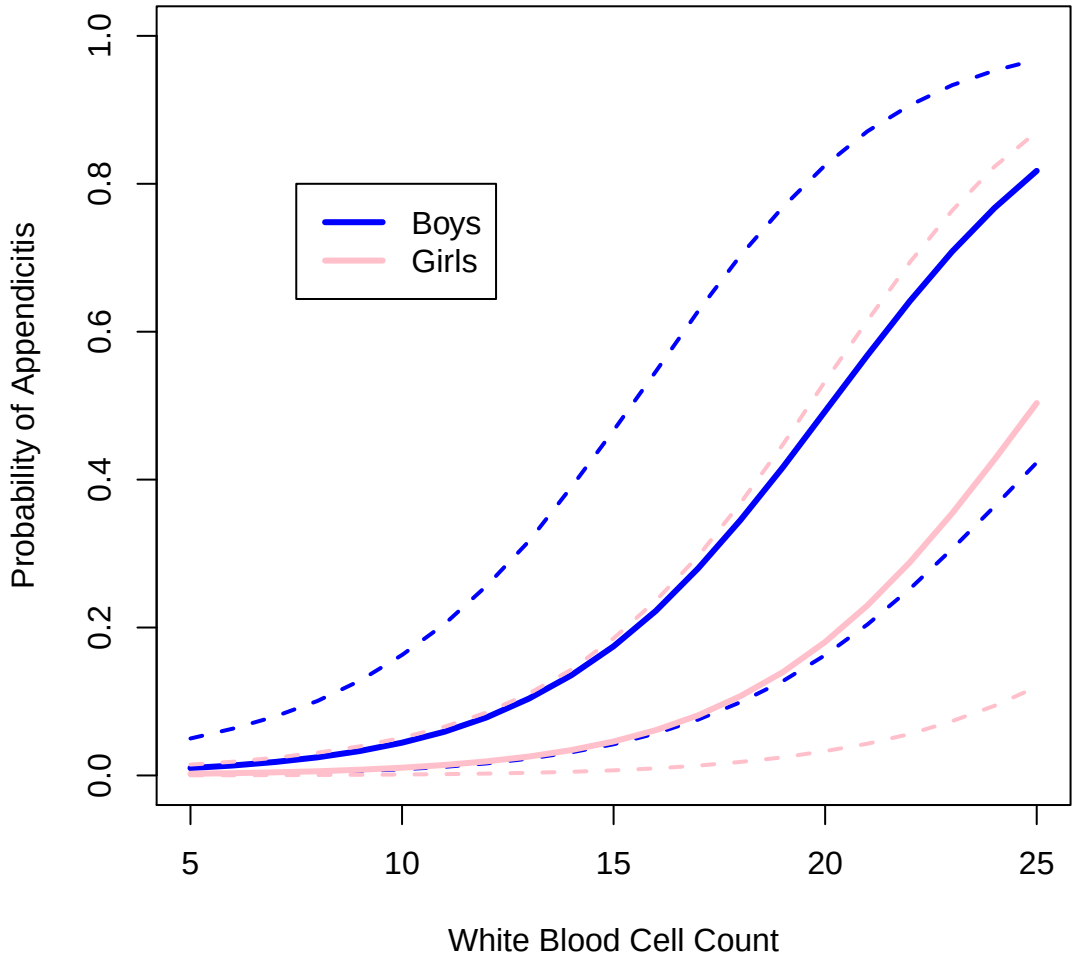

Figure 17. Probability of Appendicitis given 10% Clininical Prediction

In the next plot, the predicted probability of appendicitis, with 95% credible interval, is plotted for boys (in blue) and girls (in pink) for a patient with a gestalt clinical predicted probability (*priorprob*=50%). No precise high or low predicted probabilites are found for boys or girls over the range of *wbc*.

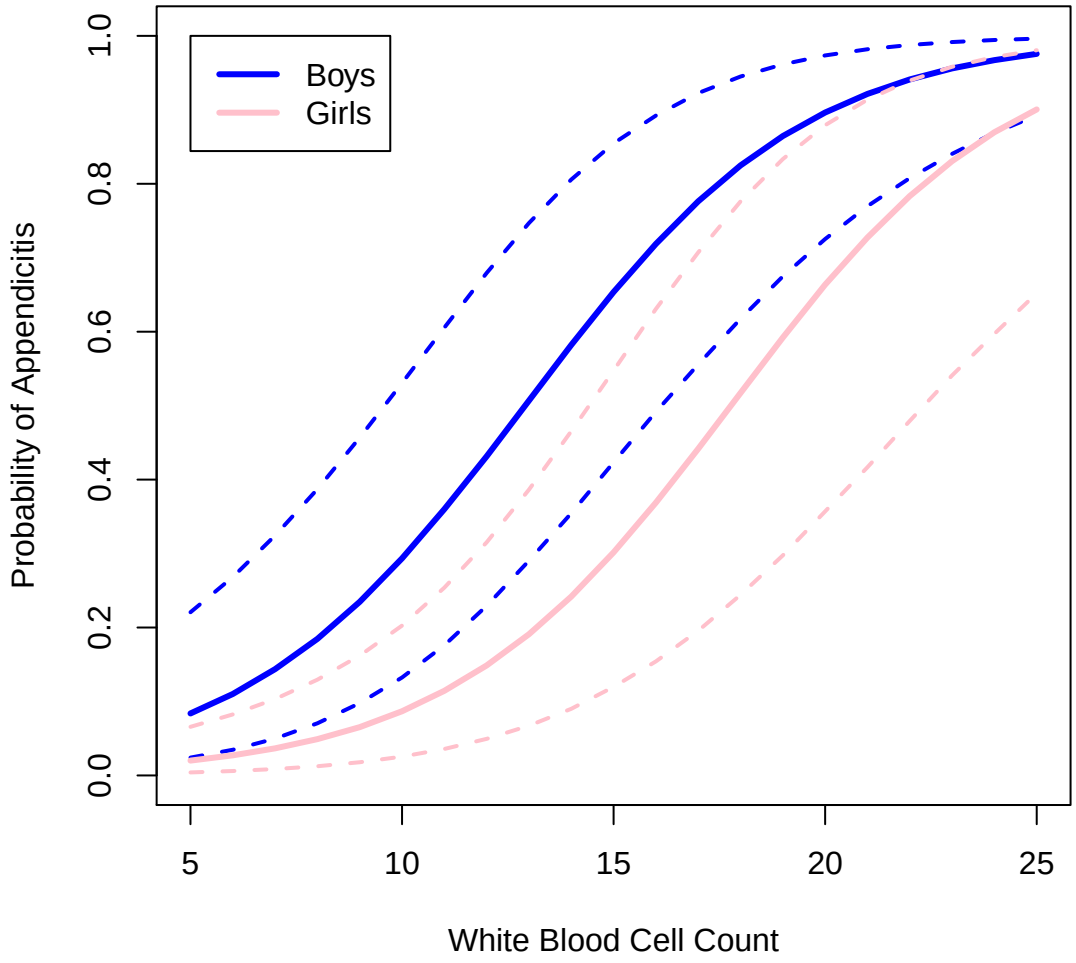

Figure 18. Probability of Appendicitis given 50% Clininical Prediction

List of References

[1] I. Ntzoufras, *Bayesian modeling using WinBUGS*. Hoboken, N.J.: Wiley, 2009. [Online]. Available: http://stat-athens.aueb.gr/~jbn/winbugs_book/

[2] M. Plummer, "Jags: A program for analysis of Bayesian graphical models using Gibbs sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, 2003. [Online]. Available: http://www.ci.tuwien.ac.at/Conferences/DSC-2003/

[3] A. Gelman and J. Hill, *Data analysis using regression and multilevel/hierarchical models*. Cambridge ; New York: Cambridge University Press, 2007. [Online]. Available: http://www.stat.columbia.edu/~gelman/arm/

[4] J. Albert, *Bayesian computation with R*. New York: Springer, 2009.

CHAPTER 5

CONCLUSIONS AND FUTURE RESEARCH

In their paper, 'Why Clinicians are Natural Bayesians', Gill and colleagues conclude that clinical decision making is fundamentally bayesian, and that all clinical history questions and physical examination maneuvers constitute diagnostic tests [1]. Bayesian methods explicitly use probability models to measure uncertainty [2]. It is therefore surprising that most of the recent efforts to assist the clinician with the diagnosis of appendicitis rely only indirectly on probability models, but instead propose clinical decision rules.

Feinstein was among the first to point out the inadequacy of binary models for the clinical reality of three-zone diagnostic decisions [3]. The evaluation and management of children at risk for appendicitis is best framed as a decision problem with at least two thresholds; a lower threshold, below which further testing may be unnecessary, and an upper threshold where it may be most appropriate to remove the appendix [4]. Thus, the decision makers need a probability model.

How best to create a probability model? One approach is to build on prior research, utilizing multiple dichotomous factors from the history, physical examination and screening laboratory tests. The practices of 'testimation' and stepwise selection have been routinely applied to small datasets. The result has been overfitted models which are unlikely to predict well in future patients. The deeply entrenched practice of dichotomizing continuous variables, and the recent enthusiasm for recursive partitioning, encourage 'dichotomania' and throw away much of the predictive information contained in continuous variables. In Chapter 2, the area under the ROC curve, an index of discrimination, is much greater for a logistic model with *wbc* as a continuous variable, than for the PAS score. In the process of

developing a clinical factors model, it became apparent that it was important to include *gender* in the model. Surprisingly, I found the presence of *fever* predicted a significantly LOWER risk of appendicitis for the children in this sample.

An alternative, and arguably preferable approach, is to take advantage of each clinicians ability to provide a 'gestalt' clinical probability estimate between zero and 100%. I attempted to specify a clinical factors model without regard to outcomes (except in preliminary exploratory plots). Nonetheless, internal validation plots with optimism estimated from bootstrap samples suggest that the smaller clinical probability model, adjusted for gender, may perform better in a new sample. Each patient arrives with a nuanced and ideosyncratic story, surely providing more information to an experienced clinician than a sum of binary variables.

Experience suggests that individual physicians may vary in their diagnostic ability or in how they use the probability scale.

Patients in clinical studies such as this are evaluated by a finite group of clinicians, and it is likely that probability estimates by a particular clinician will be more similar than those by a different clinician, introducing a within-physician correlation. Given the small number of patients evaluated by a each doctor, fixed-effects estimates are very poorly determined. The generalized linear mixed models in Chapters 3 and 4 allow a variance component to be estimated. The models fit by maximum likelihood and Bayesian MCMC approaches suggest that a clinician specific random intercepts model is adequate to account for clinician variation. The Bayesian models are particularly well suited for estimating the expected variability of the model predictions.

The conditional prediction plots in Chapter 4 suggest several clinical heuristics. One can feel most confident omitting imaging (CT or ultrasound) in girls with low clinical suspicion AND low white counts (Figure 17). Conversely, the

best case for proceeding directly to the operating room can be made for boys with both high clinical suspicion and high white blood cell counts (Figure 16). When there is equipoise after the history and physical exam, imaging will be necessary, irrespective of *wbc* (Figure 18). When clinical probability and white blood count are discordant (one high, the other low), further evaluation should always be considered.

It is likely that there will be more false positive imaging studies in patients with low prior probability of appendicitis, and more false negative studies in patients with a high prior probability of appendicitis. Novel inflammatory markers are currently under active development. It is naive to expect that a single threshold value of a new marker will be equally useful in all patients. Rather, the appropriate threshold for a continuous marker will depend on the covariate pattern in a given patient.

It is not easy to define clear probability thresholds. They should reflect costs (risks and benefits) and may differ between patients. Although I have treated appendicitis as binary, in reality the pathology ranges from pain due to obstruction of a hollow tube, to advanced peritonitis in ruptured appendicitis. Decision making is complicated by the fact that the emergency physician often is most focused on the lower threshold, and greatly regrets (and may be sued for) missing the early diagnosis of appendicitis. The surgeon must make the decision to remove the appendix, and has an interest in minimizing the number of patients without appendicitis who are taken to the operating room unnecessarily.

In future, it is clear that much larger sample sizes are necessary. This will require multicenter investigations. Hierarchical models become particularly attractive in this setting, as one can add covariates at multiple levels. For example, clinician level predictors might include level of training, years of experience, type of

training (surgical/pediatrics/emergency medicine). Hospital level predictors might include patient volume and resource availability. Regional and national level predictors may also be relevant.

The Bayesian paradigm seems particularly relevant as it allows a specification of priors which reflect information from previous studies. Journal policies which encourage reproducible research and availability of data will facilitate incorporation of prior information.

List of References

[1] C. J. Gill, L. Sabin, and C. H. Schmid, "Why clinicians are natural Bayesians," *BMJ*, vol. 330, no. 7499, pp. 1080–1083, 2005.

[2] A. Gelman, *Bayesian Data Analysis*, 2nd ed. Boca Raton, Fla.: Chapman & Hall/CRC, 2004.

[3] A. R. Feinstein, "The inadequacy of binary models for the clinical reality of three-zone diagnostic decisions." *J Clin Epidemiol*, vol. 43, no. 1, pp. 109–113, 1990.

[4] B. A. Hagendorf, J. R. Clarke, and R. S. Burd, "The optimal initial management of children with suspected appendicitis: a decision analysis," *Journal of Pediatric Surgery*, vol. 39, no. 6, pp. 880–885, Jun 2004.

APPENDIX

Appendix A

A.1 Chapter 1 R Code

```
### R code from vignette source 'chapter1.Rnw'

###################################################
### code chunk number 1: data.input
###################################################
require(rms)
require(reporttools)
require(xtable)
load("../data/appy.Rdata")
n.doc <- length(table(appy$doc))
most <- max(table(appy$doc))
least <- min(table(appy$doc))
n <- nrow(appy)
n.appy <- sum(appy$dx == 1); n.not <- sum(appy$dx == 0)
pappy <- round(n.appy/n * 100, 1)
nappy <- 100 - pappy
mean.wbc <- round(mean(appy$wbc),1)
range.wbc <- range(appy$wbc)
mean.priorprob <- round(mean(appy$priorprob, na.rm=TRUE),1)
range.priorprob <- range(appy$priorprob, na.rm=TRUE)


###################################################
### code chunk number 2: chop
###################################################
load("../data/appyLB.RData")
# drop rows if any NA
```

```
appyLB <- appyLB[-which(apply(appyLB,1,function(x) any(is.na(x)))),]
tabdx <- table(appyLB$appy)
appyLB.n <- sum(tabdx)
pct.appy <- round((tabdx[2]/appyLB.n* 100),1)


##################################################
### code chunk number 3: HXvars
##################################################

vars0 <- with(appy, data.frame(
          "Appendicitis (\\emph{appy})" = factor(appy$dx, levels=0:1,
              label=c("no", "yes")),
          "Clinical Probability (\\emph{priorprob})" = appy$priorprob,
          "White Blood Cell Count (\\emph{wbc})" = appy$wbc,
          "Absolute Neutrophil Count (\\emph{anc})" = appy$anc,
          "Duration of Pain (hours)" = appy$durpain,
          "Gender (\\emph{gender})" =    factor(appy$gender),
          "Fever (\\emph{fever})" = factor(appy$fever,levels = 0:1,
              label =c("no", "yes")),
          "Migration of Pain (\\emph{migrate})" = factor(appy$migrate,
               levels = 0:1, label =c("no", "yes")),
          "Anorexia (\\emph{anorexia})" = factor(appy$anorexia, levels
               = 0:1, label =c("no", "yes")),
          "Vomiting (\\emph{emesis})" = factor(appy$emesis, levels =
              0:1, label =c("no", "yes")),
          "RLQ tenderness (\\emph{rlqpain})" = factor(appy$rlqtender,
              levels = 0:1, label =c("no", "yes")),
          "Pain with Hopping (\\emph{hoppain})" = factor(appy$hoppain,
               levels = 0:1, label =c("no", "yes")),
          "Pain with Cough (\\emph{coughpain})" = factor(appy$
              coughpain, levels = 0:1, label =c("no", "yes")),
```

```
        "Pain with Shaking (\\emph{shakepain})" = factor(appy$
            shakepain, levels = 0:1, label =c("no", "yes")),
        "Pain with Percussion (\\emph{percpain})" = factor(appy$
            percpain, levels = 0:1, label =c("no", "yes")),
        "Rebound Tenderness (\\emph{rebound})" = factor(appy$rebound
            , levels = 0:1, label =c("no", "yes")),
        "Urinary Ketones (\\emph{ketones})" = factor(appy$ketones),
        check.names = FALSE))
attach(vars0, warn.conflicts = FALSE)
vars1 <- vars0[, c(1, 6, 7, 8, 9, 10)]
cap1 <- "Diagnosis, Gender and History Variables"
for (i in 4:6) {
vars1[,i] <- NAtoCategory(vars1[,i], label="missing")
}


#####################################################
### code chunk number 4: chapter1.Rnw:173-176
#####################################################
source("mytableNominal.R")
mytableNominal(vars = vars1, cap=cap1, cumsum=FALSE, vertical = TRUE,
            lab = "tab: HXvars", longtable = FALSE)


#####################################################
### code chunk number 5: PEvars
#####################################################
vars2 <- vars0[, c(11, 12, 13, 14, 15, 16, 17)]
cap2 <- "Physical Examination and Bedside Urinary Ketones"

for (i in 1:7){
vars2[,i] <- NAtoCategory(vars2[,i], label="missing")
```

```
}
mytableNominal(vars = vars2, cap=cap2, cumsum=FALSE, vertical = FALSE
    , lab = "tab: PEvars", longtable = FALSE)


##################################################
### code chunk number 6: contVars
##################################################
vars3 <- vars0[, c(2:5)]
cap3 <- "Continuous variables."
source("mytableContinuous.R")
mytableContinuous(vars = vars3, cap = cap3, lab = "tab: contVars",
    longtable = FALSE, font.size="scriptsize")


##################################################
### code chunk number 7: chapter1.Rnw:219-223
##################################################
require(car)
appy$ketones <- recode(appy$ketones, "0=0; 1=1; c(2,3)=2")
detach(package:car)
label(appy$ketones) <- "Urinary ketones"


##################################################
### code chunk number 8: wbc
##################################################
require(lattice)
appy$appy <- factor(appy$dx)
levels(appy$appy) <- c("No Appendicitis", "Appendicitis")
lattice.options(default.theme = standard.theme(color = FALSE))
print(densityplot( ~ wbc | gender, plot.points=TRUE, groups= appy,
```

```
        ref=TRUE, xlab="White Blood Cell Count", auto.key=TRUE,
            data=appy))


#####################################################
### code chunk number 9: priorprob
#####################################################
 print(densityplot( ~ priorprob | gender, plot.points=TRUE, groups=
     appy,
              xlab="Gestalt % Probability of Appendicitis", auto.key=
                  TRUE,
              ref=TRUE, data=appy))


#####################################################
### code chunk number 10: predictwbc
#####################################################
print(xyplot(ifelse(dx == 1, 1, 0) ~ wbc|gender, appy,
             type = c("g", "smooth"),
             auto.key = list(space = "top", points = FALSE,
             lines = TRUE, columns = 4),
             ylab = "Smoothed Proportion with Appendicitis", xlab = "
                 White Blood Count"))


#####################################################
### code chunk number 11: predictpriorprob
#####################################################
print(xyplot(ifelse(dx == 1, 1, 0) ~ priorprob|gender, appy,
             type = c("g", "smooth"),
             auto.key = list(space = "top", points = FALSE,
             lines = TRUE, columns = 4),
```

```
        ylab = "Smoothed Proportion", xlab = "Gestalt
            Probability"))


#######################################################
### code chunk number 12: na.examine
#######################################################
n.miss <- nrow(appy[!complete.cases(appy),]) #number of rows with at
    least one missing value
n.complete <- n - n.miss                    #complete cases across
    all variables


#######################################################
### code chunk number 13: single.imputation
#######################################################
# list of variable names to impute with NA set to  median: if no
    missing, unchanged
# WORKAROUND: e1071 package also contains impute(), detach it, or use
     Hmisc::impute as workaround
# impute.list <- c("dx", "wbc", "priorprob", "doc", "shakepain",  "
    coughpain", "percpain", "hoppain", "anorexia",
#                 "fever", "emesis", "ketones", "rlqtender", "migrate
    ", "gender" )

## NOTE: lapply creates a list
# appy.si.list <- lapply(appy[, impute.list], Hmisc::impute)
                  #list of imputed* variables
appy.si <- as.data.frame(lapply(appy, Hmisc::impute))
                        #all variables imputed
save(appy.si, file= '../data/appy.si.Rdata')
```

A.2 Chapter 2 R Code

```
### R code from vignette source 'chapter2.Rnw'

###################################################
### code chunk number 1: readData
###################################################
require(rms)
require(car)
require(xtable)
load("../data/appy.Rdata")


###################################################
### code chunk number 2: redundancy
###################################################
 # redundancy analysis
redund <- redun(~ gender + priorprob + wbc + anc + fever + migrate +
    anorexia +
                emesis + ordered(ketones) + rlqtender + hoppain +
                   coughpain +
                shakepain + percpain + rebound, data=appy)
anc.r <- round(redund$rsquared, 2)


###################################################
### code chunk number 3: ancVwbc
###################################################
plot(appy$wbc, appy$anc, xlab="White blood cell count (wbc)", ylab="
    Absolute neutrophil count")


###################################################
```

```
### code chunk number 4: vc
#####################################################
v <- varclus(~ fever + migrate + anorexia + emesis + ketones +
    rlqtender
            + hoppain + coughpain + shakepain + percpain + rebound,
                data=appy)
plot(v)


#####################################################
### code chunk number 5: m2.1.si
#####################################################
load("../data/appy.si.Rdata")
# table(appy.si$ketones) #verify that ketones are collapsed
appy.si$periton <- ifelse(appy.si$coughpain ==1 | appy.si$shakepain
    ==1 |appy.si$percpain ==1 |
                              appy.si$hoppain==1, 1, 0)
m2.1 <- glm(dx ~ ketones + periton + fever + migrate + rlqtender +
    wbc + gender, family='binomial',  data=appy.si)
m2.2 <- update(m2.1, . ~ . - gender)
m2.3 <- update(m2.1, . ~ . - ketones + emesis + anorexia)
lrt.1 <- anova(m2.2, m2.1, test='Chisq') #LR test(reduced, full)
lrt.1.dev <- lrt.1$"Deviance"[2]
lrt.1.pv <- lrt.1$"Pr(>Chi)"[2]
m2.1.aic <- round(AIC(m2.1),2)
m2.3.aic <- round(AIC(m2.3),2)
m2.4 <- update(m2.1, . ~ . - fever, data=appy.si)


#####################################################
### code chunk number 6: chapter2.Rnw:229-241
#####################################################
```

```
options(digits=3)
m2.1tab <- xtable(summary(m2.1))
rownames(m2.1tab) <- c("$\\beta_{0}$", "$\\beta_{ketones}$", "$\\beta
    _{periton}$", "$\\beta_{fever}$",
                        "$\\beta_{migrate}$", "$\\beta_{rlqtender}$",
                            "$\\beta_{wbc}$", "$\\beta_{gender}$")
# print(m2.1tab, caption="Pooled Model Fit - Multiple Imputation",
    table.placement="H",
#       caption.placement="bottom", sanitize.rownames.function =
    function(x) {x})

lrm2.1 <- lrm(dx ~ ketones + periton + fever + migrate + rlqtender +
    wbc + gender, data=appy.si)
# latex(lrm2.1, caption="Preliminary Clinical Factors Model", caption
    .loc="bottom", file='')
B.0 <- round(m2.1$coef[1],2); B.ketones <- round(m2.1$coef[2],2); B.
    periton <- round(m2.1$coef[3],2);
B.fever <- round(m2.1$coef[4],2); B.migrate <- round(m2.1$coef[5],2);
     B.rlqtender <- round(m2.1$coef[6],2);
B.wbc <- round(m2.1$coef[7],2); B.gender <- round(m2.1$coef[8],2)



#####################################################
### code chunk number 7: chapter2.Rnw:257-274
#####################################################
# drop1(m2.1, test="LRT")    #-->gives type II tests as well, and is
    equivalent to Anova(m2.1)
options(digits=3, show.signif.stars = FALSE)
m2.1devtab <- xtable(Anova(m2.1))
# m2.1devtab <- xtable(drop1(m2.1, test="LRT"))
rownames(m2.1devtab) <- c("\\emph{ketones}", "\\emph{periton}", "\\
    emph{fever}", "\\emph{migrate}",
```

```
                                "\\emph{rlqtender}", "\\emph{wbc}", "\\emph
                                    {gender}")

xtable::label(m2.1devtab) <- "m2.1dev"
xtable::caption(m2.1devtab) <- "Preliminary Clinical Factors (Type II
    Analysis)"
print(m2.1devtab, table.placement="H",
      caption.placement="bottom", label="m2.1dev", sanitize.rownames
          .function = function(x) {x})

# latex(anova(lrm2.1), caption='Preliminary Clinical Factors Deviance
    Table', caption.loc="bottom", here="TRUE",
#       booktabs=TRUE, type="Sinput",   file='')

cf <- coef(lrm2.1)["fever"]
c.m2.1 <- round(lrm2.1$stats["C"],3)


#####################################################
### code chunk number 8: chapter2.Rnw:288-294
#####################################################
library(ROCR)
lrm2.1.yhat <- predict(lrm2.1, type="fitted.ind")
m2.1.scores <- prediction(lrm2.1.yhat, appy.si$dx)
# the area under the ROC curve == c statistic
m2.1.auc <- performance(m2.1.scores, measure="auc")@y.values[[1]]
auc.m2.1 <- round(m2.1.auc,3)  #agrees with c-statistic


#####################################################
### code chunk number 9: PASscore
#####################################################
```

```
cph <- ifelse(appy.si$coughpain ==1 |appy.si$percpain ==1 | appy.si$
    hoppain==1, 1, 0)
anorexia <- appy.si$anorexia
fever <- appy.si$fever
emesis <- appy.si$emesis
rlqpain <- appy.si$rlqpain
leuk <- ifelse(appy.si$wbc >= 10, 1, 0)
neut <- ifelse(appy.si$anc >= 7.5, 1, 0)
migrate <- appy.si$migrate
PASmat <- cbind(cph, anorexia, fever, emesis, rlqpain, leuk, neut,
    migrate)
PAS <- apply(PASmat, 1, sum)
pas.scores <- prediction(PAS, appy.si$dx)
pas.auc <- round(performance(pas.scores, measure="auc")@y.values
    [[1]],3)


#####################################################
### code chunk number 10: compareROC
#####################################################
require(pROC); require(xtable)
roc1 <- roc(appy.si$dx, PAS)
roc2 <- roc(appy.si$dx, lrm2.1.yhat)
rt1 <- roc.test(roc1, roc2)
# rt2 <- roc.test(roc1, roc2, method="venkatraman", paired=TRUE)
# The latter used Delong's test. To use bootstrap test:
# rt3 <- roc.test(roc1, roc2, method="bootstrap", boot.n=10000)


#####################################################
### code chunk number 11: roc1
#####################################################
```

```
plot(performance(m2.1.scores, "tpr", "fpr"), col = "black", lty=1,
    lwd=2)
abline(0,1, col = "grey")
plot(performance(pas.scores, "tpr", "fpr"), col="black", lty=9, lwd
    =2, add=TRUE)
legend(0.5, 0.3, c("Clinical Factors Model", "PAS Score"), lty=c(1,
    9), lwd=c(2,2))


#####################################################
### code chunk number 12: prelim.priorprob
#####################################################
library(car)
appy.si$logitpp <- car::logit(appy.si$priorprob)
lrm2.5 <- lrm(dx ~ car::logit(priorprob) + wbc + gender, data=appy.si
    )
# use glm() to get profile likelihood confidence intervals
m2.5.glm <- glm(dx ~ car::logit(priorprob) + wbc + gender, family="
    binomial", data=appy.si)
ci.pl <- confint(m2.5.glm)
ci.an <- confint.default(m2.5.glm)
m2.5.se <- sqrt(diag(lrm2.5$var))
m2.5.tab <- cbind(m2.5.glm$coefficients, m2.5.se)
c.m2.5 <-lrm2.5$stats["C"]
m2.5.aic <- round(AIC(lrm2.5),3)


#####################################################
### code chunk number 13: priorprob.mice
#####################################################
options(digits=3)
require(mice)
```

```
#creates 5 imputed datasets by default
imp <- mice(appy, seed=1234, printFlag=FALSE)
fit <- with(imp, glm(dx ~ car::logit(priorprob) + wbc + gender,
    family="binomial") )
pooled <- pool(fit)
m2.5pool <- summary(pooled)[,c("est", "se", "t", "df", "Pr(>|t|)", "
    lo 95", "hi 95")]
rownames(m2.5pool) <- c("$\\beta_{0}$", "$\\beta_{logit(priorprob)}$"
    , "$\\beta_{wbc}$",
                            "$\\beta_{gender}$")
print(xtable(m2.5pool, caption="Pooled Model Fit - Multiple
    Imputation"), table.placement="H",
      size="small", sanitize.rownames.function = function(x) {x})


#####################################################
### code chunk number 14: imputation.type
#####################################################
comp.table <- cbind(m2.5.tab, m2.5pool[, c("est", "se")])
colnames(comp.table) <- c("SI:coef", "SI:se", "MI:coef", "MI:se")
rownames(comp.table) <- c("$\\beta_{0}$", "$\\beta_{logit(priorprob)}
    $", "$\\beta_{wbc}$",
                            "$\\beta_{gender}$")
# rownames(comp.table) <- c("Intercept", "logit(priorprob)", "wbc", "
    gender")
print(xtable(comp.table, caption="Comparison of Coefficients: Single
    (Median) vs. Multiple Conditional Imputation"),
      table.placement="H",sanitize.rownames.function = function(x) {x
          })


#####################################################
```

```
### code chunk number 15: libBoot
#####################################################
require(boot)

## Use the 'boot' library - follows Davison and Hinkley
## Adapted from <http://www.statmethods.net/advstats/bootstrapping.
    html>
## Bootstrap 95% CI for regression coefficients

# function to obtain regression weights
bs <- function(formula, data, indices) {
  d <- data[indices,] # allows boot to select sample
  fit <- lrm(formula, data=d)
  return(coef(fit))
}
# bootstrapping with 1000 replications
rep <- 1000
results <- boot(data=appy.si, statistic=bs,
          R=rep, formula=dx~logitpp + wbc + gender)
r <- results$t
# get 95% confidence intervals
# print(results)
ci.int <- boot.ci(results, type="bca", index=1) # intercept
ci.priorprob <- boot.ci(results, type="bca", index=2) # priorprob
ci.wbc <- boot.ci(results, type="bca", index=3) # wbc
ci.gender <- boot.ci(results, type="bca", index=4) # gender
bs.ci <- rbind(ci.int$bca, ci.priorprob$bca, ci.wbc$bca, ci.gender$
    bca)[,4:5]
dimnames(bs.ci)[[2]] <- c("2.5%", "97.5%")
row.names(bs.ci) <- c("Intercept", "logit(priorprob)", "wbc", "gender
    ")
```

```
### Wald ci vs vs: profile likelihood vs  bootstrap CI's
ci.pl <- confint(m2.5.glm)
row.names(ci.pl) <- c("Intercept", "logit(priorprob)", "wbc", "gender
    ")


#####################################################
### code chunk number 16: bootstrapDist
#####################################################
# boot:::plot.boot
par(mfrow=c(2,2))
hist(r[,1], freq=FALSE, nclass=30, xlab="Intercept", main=NULL)
hist(r[,2], freq=FALSE, nclass=30, xlab="logit(priorprob)", main=NULL
    )
hist(r[,3], freq=FALSE, nclass=30, xlab="wbc", main=NULL)
hist(r[,4], freq=FALSE, nclass=30, xlab="gender", main=NULL)
par(mfrow=c(1,1))


#####################################################
### code chunk number 17: profileLik.ci
#####################################################
print(xtable(ci.pl, caption="Profile Likelihood Confidence Intervals"
    ), table.placement="H")


#####################################################
### code chunk number 18: bootstrap.ci
#####################################################
print(xtable(bs.ci, caption="Bootstrap Confidence Intervals"), table.
    placement="H")
```

```
#####################################################
### code chunk number 19: bootDetach
#####################################################
detach("package:boot")


#####################################################
### code chunk number 20: bsvalclinfact
#####################################################
set.seed(123)
options(digits=2)
# validate(m2.1, B=200)   #Need to explain the various indices
lrm2.1 <- update(lrm2.1, x=T, y=T)
cal2.1 <- calibrate(lrm2.1, B=1000)
plot(cal2.1) ##TODO Explain this plot


#####################################################
### code chunk number 21: bsvalgestalt
#####################################################
set.seed(321)
options(digits=2)
# validate(lrm2.5, B=200)   #Need to explain the various indices
lrm2.5 <- update(lrm2.5, x=TRUE, y=TRUE)
cal2.5 <- calibrate(lrm2.5, B=1000)
plot(cal2.5)


#####################################################
### code chunk number 22: chapter2.Rnw:550-552
#####################################################
```

```
lp <- lrm2.5$linear.predictors
boxplot(plogis(lp) ~ appy.si$dx, ylab="Predicted Probability", xlab="
    Appendicitis: 0=no, 1=yes")


##################################################
### code chunk number 23: chapter2.Rnw:568-577
##################################################
# model refitted using glm
m2.5glm <- glm(dx ~ car::logit(priorprob) + wbc + gender, family="
    binomial", data=appy.si)
m2.6glm <- glm(dx ~ car::logit(priorprob) + wbc + gender + doc,
    family="binomial", data=appy.si)
# anova(m2.5glm, m2.6glm, test="Chisq")
std.d <- residuals(m2.5glm, type="deviance")/sqrt(1 - lm.influence(m2
    .5glm)$hat)
# pred <- predict.glm(m2.5glm, type="response")
# plot(pred, std.d, xlab="Predicted Probability", ylab="Standardize
    Deviance Residuals", ylim=c(-2,2))
plot(std.d, xlab="Index", ylab="Standardized Deviance Residuals",
    ylim=c(-2,2))
abline(h=c(-2,2), lty="dotted")


##################################################
### code chunk number 24: chapter2.Rnw:584-587
##################################################
### TODO: le Cessie test
lC <- (resid(lrm2.5, 'gof'))
plC <- lC["P"]
```

```
#####################################################
### code chunk number 25: clinician.fixed
#####################################################
options(digits=3)
n.doc <- length(table(appy$doc))
min <- min(table(appy$doc))
max <- max(table(appy$doc))
```

A.3 Chapter 3 R Code

```
### R code from vignette source 'chapter3.Rnw'

###################################################
### code chunk number 1: chapter3.Rnw:37-50
###################################################
library(car)
require(lme4)
require(xtable)
load("../data/appy.si.Rdata")
m3.1 <- glmer(dx ~ car::logit(priorprob) + wbc + gender +  (car::
    logit(priorprob) | doc), family=binomial,
            data=appy.si)
m3.1tab <- xtable(summary(m3.1)@coefs)
rownames(m3.1tab) <- c("$\\beta_{0}$", "$\\beta_{logit(priorprob)}$",
     "$\\beta_{wbc}$",
                           "$\\beta_{gender}$")
xtable::caption(m3.1tab) <- "Random Intercept/Random Slope Model"
xtable::label(m3.1tab) <- "lme4RIRS"
print(m3.1tab, digits=3, caption.placement="bottom",  table.placement
   ="H",
      sanitize.rownames.function = function(x) {x})


###################################################
### code chunk number 2: chapter3.Rnw:57-68
###################################################
m3.2 <- glmer(dx ~ car::logit(priorprob) + wbc + gender + (1 | doc),
             family=binomial, data=appy.si)
m3.2.var <- VarCorr(m3.2)$doc[1]
m3.2.sd <- sqrt(VarCorr(m3.2)$doc[1])
m3.2tab <- xtable(summary(m3.2)@coefs)
```

```
rownames(m3.2tab) <- c("$\\beta_{0}$", "$\\beta_{logit(priorprob)}$",
    "$\\beta_{wbc}$",
                           "$\\beta_{gender}$")
xtable::caption(m3.2tab) <- "Random Intercept Model"
xtable::label(m3.2tab) <- "lme4RI"
print(m3.2tab, digits=3, caption.placement="bottom", table.placement
   ="H",
      sanitize.rownames.function = function(x) {x})


#####################################################
### code chunk number 3: lrtGLMM
#####################################################
lrt3 <- anova(m3.2, m3.1)
rownames(lrt3) <- c("Random Intercept", "Random Intercept/Slope")
print(xtable(lrt3, caption="Likelihood Ratio Test"), table.placement=
   "H")


#####################################################
### code chunk number 4: m3.AGQ
#####################################################
m3.3 <- update(m3.2, nAGQ=10)
m3.3se <- sqrt(diag(vcov(m3.3)))          #std errors
# vcov(m3.3)                              #the variance covariance
    matrix for the fixed effects
# fixef(m3.3)                             #reports the fixed-
    effect coeficients
m3.3beta <- getME(m3.3, "beta")           #fixed effects
    coefficients
m3.3.var <- VarCorr(m3.3)$doc[1]
m3.3.sd <- sqrt(VarCorr(m3.3)$doc[1])     #standard deviation of
```

```
    the random effect
#  mmME <- eval(formals(getME)$name)        #things that can be
    extracted


##################################################
### code chunk number 5: chapter3.Rnw:96-103
##################################################
m3.3tab <- xtable(summary(m3.3)@coefs)
rownames(m3.3tab) <- c("$\\beta_{0}$", "$\\beta_{logit(priorprob)}$",
    "$\\beta_{wbc}$",
                           "$\\beta_{gender}$")
xtable::caption(m3.3tab) <- "Random Intercepts Model (using Adaptive
    Gaussian Quadrature)"
xtable::label(m3.3tab) <- "lme4RIagq"
print(m3.3tab, digits=3, table.placement="H",
      caption.placement="bottom", sanitize.rownames.function =
          function(x) {x})


##################################################
### code chunk number 6: glmer.predict (eval = FALSE)
##################################################
## cmre.m3.3 <- ranef(m3.3, drop=TRUE)$doc
## # calculate the predicted probability for a boy with priorprob=10,
    wbc=8
## eta.ml <- crossprod(c(1, 10, 1, 8), fixef(m3.3))
## pred.ml <- binomial()$linkinv(eta.ml)
## ## prob(girl with priorprob=10, wbc=8)
## eta.fl <- crossprod(c(1, 10, 0, 8), fixef(m3.3))
## pred.fl <- binomial()$linkinv(eta.fl)
## # prob for a boy with priorprob=90, wbc=20
```

```
## eta.mh <- crossprod(c(1, 90, 1, 20), fixef(m3.3))
## pred.mh <- binomial()$linkinv(eta.mh)
## eta.girl <- crossprod(c(1, 90, 0, 20), fixef(m3.3))
## pred.fh <- binomial()$linkinv(eta.girl)


#####################################################
### code chunk number 7: chapter3.Rnw:130-147 (eval = FALSE)
#####################################################
## ## TODO: Learn how to extract a particular random effect (extract
    pieces of S4 objects)
## ranef(m3.3)          # what Bates calls the conditional mode of the
    random effect --> a list of data frames
## str(ranef(m3.3)$doc)       # a data.frame
## ranef(m3.3)$doc[,1]        # select the first column --> this is a
    vector
##
## (coef(m3.3))                    # the subject specific coefficients
## # linear predictor for a boy with a median wbc and priorprob
## eta.boy <- crossprod(c(1, 90, 1, 20), fixef(m3.3))
##
## # linear predictor for a girl with a median wbc and priorprob
## eta.girl <- crossprod(c(1, median(appy.si$priorprob), 0, median(
    appy.si$wbc)), fixef(m3.3))
## (eta.doc.boy <-  cmre + eta.boy)
## eta.doc.girl <- cmre + eta.girl
## (prob.boy <- binomial()$linkinv(eta.boy))
## (p.doc.boy <- binomial()$linkinv(eta.doc.boy))
## p.doc.girl <- binomial()$linkinv(eta.doc.girl)
## ### TODO: produce a side-by-side histograms (all doctors in study,
    boy and girl predicted probabilities.
```

A.4 Chapter 4 R Code

```
### R code from vignette source 'chapter4.Rnw'

###################################################
### code chunk number 1: chapt4.prelim
###################################################
require(cacheSweave)
setCacheDir("cache")
require(R2jags)
require(xtable)
require(car)    #for logit() function



###################################################
### code chunk number 2: load.chop.data
###################################################
load("../data/appyLB.RData")
# drop rows if any NA
appyLB <- appyLB[-which(apply(appyLB,1,function(x) any(is.na(x)))),]
appyLB$gender <- ifelse(appyLB$sex =="Female",0,1)        #convert to
    integer: male=1, female=0



###################################################
### code chunk number 3: load.hch.data
###################################################
load("../data/appy.si.Rdata")                                  # load
    the simple imputed data
a <- appy.si[, c("dx", "gender", "priorprob", "wbc", "doc")]    #
    select variables
a$gender <- ifelse(a[,"gender"]=="Female",0,1)
```

```
a$doc <- as.numeric(a$doc)


#####################################################
### code chunk number 4: m1a.chop
#####################################################
# The model
m1a.model = "
model {
# Likelihood
 for (i in 1:n) {
    appy[i] ~ dbin(p[i], 1)
    logit(p[i]) <- b.0 + b.wbc*wbc[i]+ b.gender*gender[i]
    }
"
m1a.priors = "
# independent normal priors
 b.0 ~ dnorm(0,0.0001)
 b.gender ~ dnorm(0,0.0001)
 b.wbc ~ dnorm(0,0.0001)
}
" #close quote for model string

# Write model to a file (piecewise, so parts can be re-used)
writeLines(c(m1a.model, m1a.priors) ,con="m1a.txt")
# Bundle data into a list
chop.data <- list(appy = appyLB$appy, wbc=appyLB$wbc, gender=appyLB$
    gender,  n = nrow(appyLB))
# Parameters to estimate
params <- c("b.0", "b.gender", "b.wbc")

# inits are chosen as radom deviates from glm() coefficient estimates
```

```
inits.m1 <- function() {
  list("b.0"=rnorm(1, mean=-3.7, sd=1), "b.wbc"=rnorm(1, mean=0.23,
     sd=1), "b.gender"=rnorm(1, mean=1.12, sd=1))
}

## Send to JAGS
load.module("glm")
m1a <- jags(data=chop.data, inits=inits.m1, parameters.to.save=params
    , model.file="m1a.txt",
                  n.chains=3, n.iter=30000, n.burnin=1000, n.thin=5,
                      DIC=TRUE, progress.bar="none")
########
# create data frame and use xtable
m1a.df <- as.data.frame((m1a$BUGSoutput$summary)[,c("mean", "sd", "
    2.5%", "50%", "97.5%")])
m1a.table <- xtable(m1a.df[1:nrow(m1a.df) - 1,], caption="CHOP data:
    Posterior Summary", digits=3)
m1a.mcmc <- as.mcmc(m1a) #convert to mcmc (coda) object
########

# plot(m1a.mcmc)
# xyplot(m1a.mcmc)       #show how well the chains are mixing
# densityplot(m1a.mcmc)     #the posterior density plots
# plot(m1a.mcmc[, c("b.gender", "b.wbc")])

## # Combine the chains
#HPDinterval(as.mcmc(rbind(m1a.mcmc[[1]],m1a.mcmc[[2]], m1a.mcmc
    [[3]])))[c("b.gender","b.wbc"),]
# m1a.combined <- as.mcmc(rbind(m1a.mcmc[[1]],m1a.mcmc[[2]], m1a.mcmc
    [[3]]))
# plot(density(m1a.combined[,"b.gender"]))
# Diagnostic plots using coda()
```

```
# autocorr.plot(m1a.combined[1:100,])

# Priors from CHOP model
m1a.mean <- unlist(m1a$BUGSoutput$mean)[c("b.gender", "b.wbc")]
                # become prior means
m1a.sd <- unlist(m1a$BUGSoutput$sd)[c("b.gender", "b.wbc")]
                    # becomes prior sd
m1a.precision <- 1/m1a.sd^2
                                                    # becomes prior
    precision
m1a.sd.cons <- m1a.sd*2
                                                        # stddev*2
m1a.precision.cons <- 1/m1a.sd.cons^2

### TODO Use these to automatically remember prior
# chop.prior.wbc     <- sprintf("b.wbc ~ dnorm(%s , %s)", m1a.mean["b.
    wbc"], m1a.precision["b.wbc"]))
chop.prior.wbc.cons <-sprintf("b.wbc ~ dnorm(%s , %s)", m1a.mean["b.
    wbc"], m1a.precision.cons["b.wbc"])


#####################################################
### code chunk number 5: chapter4.Rnw:128-130
#####################################################
rownames(m1a.table) <- c("$\\beta_{0}$", "$\\beta_{gender}$", "$\\
    beta_{wbc}$")
print(m1a.table, table.placement="H", caption.placement="bottom",
    sanitize.rownames.function = function(x) {x})


#####################################################
### code chunk number 6: Model.2a.diffusepriors
```

```
#####################################################
# The model
m2.model= "
model {
# Likelihood
 for (i in 1:n) {
    appy[i] ~ dbin(p[i], 1)
    logit(p[i]) <- b.0 +  b.logitpp*logitpp[i] + b.wbc*wbc[i] + b.
       gender*gender[i]
     }
"
m2a.priors= "
# Diffuse priors
 b.0 ~ dnorm(0,0.0001)
 b.logitpp ~ dnorm(0, 0.0001)
 b.wbc ~ dnorm(0,0.0001)
 b.gender ~ dnorm(0,0.0001)
"

m2.predictions= "
# posterior predictions
 or.gender <- exp( b.gender )
 or.wbc <- exp( b.wbc )
 or.logitpp <- exp( b.logitpp)
}
" # close quote for modelstring

# Write model to a file:
writeLines(c(m2.model, m2a.priors, m2.predictions), con="m2a.txt")

params.m2a <- c("b.0", "b.logitpp", "b.wbc", "b.gender", "or.logitpp"
    , "or.wbc", "or.gender")
```

```
# inits are chosen as radom deviates from glm() coefficient estimates
# glm(dx ~ logit(priorprob) + wbc + gender, family='binomial', data=
    appy.si)
inits.m2 <- function() {
  list("b.0"=rnorm(1, mean=-4.9, sd=1), "b.logitpp"=rnorm(1, mean
      =0.86), "b.wbc"=rnorm(1, mean=0.28, sd=1),
       "b.gender"=rnorm(1, mean=1.15))
}

m2.data <- list(appy = a$dx,logitpp=car::logit(a$priorprob), wbc=a$
    wbc,
                        gender=a$gender, n = nrow(a))
load.module("glm")
m2a <- jags(data=m2.data, inits=inits.m2, parameters.to.save=params.
    m2a, model.file="m2a.txt",
                n.iter=30000, n.burnin=1000, n.thin=5, progress.bar
                    ="none")
m2a.mcmc <- as.mcmc(m2a)
m2a.df <- as.data.frame((m2a$BUGSoutput$summary)[,c("mean", "sd", "
    2.5%", "50%", "97.5%")])
m2a.table <- xtable(m2a.df[c(1,3,4,2,7,8,6),], caption="HCH data:
    Posterior Summary", digits=3)


#######################################################
### code chunk number 7: m2a.summary
#######################################################
rownames(m2a.table) <- c("$\\beta_{0}$", "$\\beta_{logit(priorprob)}$
    ", "$\\beta_{wbc}$", "$\\beta_{gender}$",
                         "$OR_{logit(priorprob)}$", "$OR_{wbc}$", "$
                             OR_{gender}$")
```

```
print(m2a.table, table.placement="H", caption.placement="bottom",
    sanitize.rownames.function = function(x) {x})


#####################################################
### code chunk number 8: Model2b
#####################################################
# The model
m2b.priors= "
# Diffuse Priors for intercept and priorprob
 b.0 ~ dnorm(0,0.0001)
 b.logitpp ~ dnorm(0, 0.0001)
 b.gender ~ dnorm(0, 0.0001)
"
# Write model to a file:
writeLines(c(m2.model, m2b.priors, chop.prior.wbc.cons, m2.
    predictions), con="m2b.txt")

params.m2 <- c("b.0", "b.gender", "b.wbc", "b.logitpp", "or.gender",
    "or.wbc", "or.logitpp")

load.module("glm")
m2b <- jags(data=m2.data, inits=inits.m2, parameters.to.save=params.
    m2, model.file="m2b.txt",
                n.iter=30000, progress.bar="none")

m2b.mcmc <- as.mcmc(m2b)
m2b.df <- as.data.frame((m2b$BUGSoutput$summary)[,c("mean", "sd", "
    2.5%", "50%", "97.5%")])
m2b.table <- xtable(m2b.df[c(1,3,4,2,7,8,6),], caption="Posterior
    Summary - priors from CHOP", digits=3)
rownames(m2b.table) <- c("$\\beta_{0}$", "$\\beta_{logit(priorprob)}$
```

```
    ", "$\\beta_{wbc}$", "$\\beta_{gender}$",
                            "$OR_{logit(priorprob)}$", "$OR_{wbc}$",  "$
                                OR_{gender}$")


#######################################################
### code chunk number 9: chapter4.Rnw:237-238
#######################################################
print(m2b.table, table.placement="H", caption.placement="bottom",
    sanitize.rownames.function = function(x) {x})


#######################################################
### code chunk number 10: vI
#######################################################
m3.string= "
model {
# Likelihood
 for (i in 1:n) {
    appy[i] ~ dbin(p[i], 1)
    logit(p[i]) <-  a[doc[i]] + b.logitpp*logitpp[i] + b.wbc*wbc[i] +
         b.gender*gender[i]
    }
for (j in 1:ndoc){
 a[j] ~ dnorm(a.hat[j], tau.a)
 a.hat[j] <- mu.a
}

mu.a ~ dnorm(0, 0.0001)
tau.a <- pow(sigma.a, -2)
sigma.a ~ dunif(0, 100)
b.logitpp ~ dnorm(0, 0.0001)
```

```
b.gender ~ dnorm(0, 0.0001)
"

m.close = "}"

writeLines(c(m3.string, chop.prior.wbc.cons, m.close), con="m3.txt")


# Parameters to estimate
params <- c("a", "mu.a", "sigma.a", "b.logitpp", "b.wbc", "b.gender")

inits.m3 <- function() {
  list("a"=rnorm(23, mean=-5.5, sd=0.5), "mu.a"=rnorm(1, mean=-5.5,
      sd=0.5), "sigma.a"=rnorm(1, mean=1, sd=0.5),
        "b.logitpp"=rnorm(1, mean=1), "b.wbc"=rnorm(1, mean=0.3, sd=1)
            , "b.gender"=rnorm(1, mean=1.5))
}
load.module("glm")

# data for JAGS glmm
hch.glmm.data <- list(appy = a$dx, logitpp=car::logit(a$priorprob),
    wbc=a$wbc,
                          gender=a$gender, doc=a$doc, n = nrow(a), ndoc=
                              length(unique(a$doc)))

m3 <- jags(data=hch.glmm.data, inits=inits.m3, parameters.to.save=
    params, model.file="m3.txt",
                  n.iter=30000, n.burnin=1000, n.thin=5, progress.bar
                      ="none")
m3.mcmc <- as.mcmc(m3) #convert to coda object
m3.df <-as.data.frame((m3$BUGSoutput$summary)[,c("mean", "50%", "sd",
     "2.5%", "97.5%")])
```

```
m3.table <- xtable(m3.df[c("mu.a", "sigma.a", "b.logitpp", "b.wbc", "
    b.gender"),], caption="Random Coefficient Model",
                    label="HCH - GLMM with Random Coefficient", digits
                        =2)
save(m3, file= '../data/m3.Rdata')


##################################################
### code chunk number 11: postCoef
##################################################
plot(m3.mcmc[, c("mu.a", "b.gender", "b.logitpp", "b.wbc")])


##################################################
### code chunk number 12: sigmadoc
##################################################
densplot(m3.mcmc[,"sigma.a"], main="")


##################################################
### code chunk number 13: chapter4.Rnw:340-343
##################################################
rownames(m3.table) <- c("$\\mu_{\\alpha}$", "$\\sigma_{doc}$", "$\\
    beta_{logit(priorprob)}$", "$\\beta_{wbc}$",
                            "$\\beta_{gender}$")
print(m3.table, table.placement="H", caption.placement="bottom",
    sanitize.rownames.function = function(x) {x})


##################################################
### code chunk number 14: caterplot
##################################################
```

```
# Adapted from < http://users.aims.ac.za/~paulhewson/catplot.R>
catplot <- function(df, xl="Clinician Identifier (n=23)",xd=TRUE){
if (xd[1]==TRUE) xrange <- c(1: dim(df)[1]) else xrange <- xd
plot(xrange, df[,3], ylim=c(min(df), max(df)), xlab = xl,
  ylab = "Posterior Median with 95% Credible Intervals", main = NULL,
      pch = 16, col =  "red")
arrows(xrange, df[,3], xrange, df[,1], length = 0.01)
arrows(xrange, df[,3], xrange, df[,5], length = 0.01)
}
df <- summary(m3.mcmc)[[2]]
catplot(df[1:23,])
abline(h=df["mu.a", 3], lty="dotted")


#####################################################
### code chunk number 15: vIvSnc
#####################################################
m4.string= "
model {
# Likelihood
 for (i in 1:n) {
    appy[i] ~ dbin(p[i], 1)
    logit(p[i]) <-  a[doc[i]] + b[doc[i]]*logitpp[i] + b.wbc*wbc[i]
                    + b.gender*gender[i]
    }
for (j in 1:ndoc){
 a[j] ~ dnorm(a.hat[j], tau.a)
 b[j] ~ dnorm(b.hat[j], tau.b)
 a.hat[j] <- mu.a
 b.hat[j] <- mu.b
}
```

```
mu.a ~ dnorm(0, 0.0001)
mu.b ~ dnorm(0, 0.0001)
tau.a <- pow(sigma.a, -2)
tau.b <- pow(sigma.b, -2)
sigma.a ~ dunif(0, 100)
sigma.b ~ dunif(0,100)
b.logitpp ~ dnorm(0, 0.0001)
b.gender ~ dnorm(0, 0.0001)
"

writeLines(c(m4.string, chop.prior.wbc.cons, m.close), con="m4.txt")
# Parameters to estimate
params <- c("mu.a", "mu.b", "sigma.a", "sigma.b", "b.wbc", "b.gender"
    )

# data for JAGS glmm
hch.glmm.data <- list(appy = a$dx, logitpp=car::logit(a$priorprob),
    wbc=a$wbc, gender=a$gender,
                          doc=a$doc, n = nrow(a), ndoc=length(unique(a$
                              doc)))

load.module("glm")
m4 <- jags(data=hch.glmm.data, inits=NULL, parameters.to.save=params,
     model.file="m4.txt",
                    n.thin=5, n.burnin=1000, n.iter=30000, progress.bar
                        ="none")
m4.mcmc <- as.mcmc(m4)
m4.df <- as.data.frame((m4$BUGSoutput$summary)[,c("mean", "50%", "sd"
    , "2.5%", "97.5%")])
m4.table <- xtable(m4.df[c("mu.a", "mu.b", "b.wbc", "b.gender", "
    sigma.a", "sigma.b"),],
                           caption="Random Slope and Intercept",
```

```
                        label="HCH - Random coefficients,
                            Uncorrelated", digits=3)
# rownames(m4.table)
### FIXME: Formula for varying intercept/varying slope
rownames(m4.table) <- c("$\\mu_{\\alpha}$", "$\\mu_{\\beta}$",
                        "$\\beta_{wbc}$", "$\\beta_{gender}$",
                        "$\\sigma_{\\alpha[doc]}$", "$\\sigma_{\\beta
                            [doc]}$")


#####################################################
### code chunk number 16: chapter4.Rnw:451-452
#####################################################
print(m4.table, table.placement="H", caption.placement="bottom",
    sanitize.rownames.function = function(x) {x})


#####################################################
### code chunk number 17: vivsCorr
#####################################################
m5.string= "
model {
# Likelihood
 for (i in 1:n) {
    appy[i] ~ dbin(p[i], 1)
    logit(p[i]) <-  a[doc[i]] + b[doc[i]]*priorprob[i] + b.wbc*wbc[i]
                    + b.gender*gender[i]
    }
for (j in 1:ndoc){
    a[j] <- xi.a*B.raw[j,1]
    b[j] <- xi.b*B.raw[j,2]
    B.raw[j,1:2] ~ dmnorm (B.raw.hat[j,], Tau.B.raw[,])
```

```
    B.raw.hat[j,1] <- mu.a.raw
    B.raw.hat[j,2] <- mu.b.raw
  }
  mu.a <- xi.a*mu.a.raw
  mu.b <- xi.b*mu.b.raw
  mu.a.raw ~ dnorm (0, .0001)
  mu.b.raw ~ dnorm (0, .0001)

  xi.a ~ dunif (0, 100)
  xi.b ~ dunif (0, 100)

  Tau.B.raw[1:2,1:2] ~ dwish(W[,], df)
  df <- 3
  Sigma.B.raw[1:2,1:2] <- inverse(Tau.B.raw[,])
  sigma.a <- xi.a*sqrt(Sigma.B.raw[1,1])
  sigma.b <- xi.b*sqrt(Sigma.B.raw[2,2])
  rho <- Sigma.B.raw[1,2]/sqrt(Sigma.B.raw[1,1]*Sigma.B.raw[2,2])
  b.priorprob ~ dnorm(0, 0.0001)
  b.gender ~ dnorm(0, 0.0001)
"

writeLines(c(m5.string, chop.prior.wbc.cons, m.close), con="m4.txt")
# Parameters to monitor
params <- c("mu.a", "mu.b", "sigma.a", "sigma.b", "rho", "b.wbc", "b.
    gender")
# data for JAGS glmm
hch.glmm.data <- list(appy = a$dx, priorprob=logit(a$priorprob), wbc=
    a$wbc, gender=a$gender,
                          doc=a$doc, n = nrow(a), W=diag(2), ndoc=length(
                              unique(a$doc)))
load.module("glm")
m5 <- jags(data=hch.glmm.data, inits=NULL, parameters.to.save=params,
```

```
    model.file="m5.txt",
                n.thin=5, n.burnin=1000,  n.iter=80000, progress.bar
                   ="none")
m5
m5.mcmc <- as.mcmc(m5)
m5.df <- as.data.frame((m5$BUGSoutput$summary)[,c("mean", "50%", "sd"
    , "2.5%", "97.5%")])
m5.table <- xtable(m5.df[c("mu.a", "mu.b", "b.wbc", "b.gender", "
    sigma.a", "sigma.b", "rho"),],
                caption="Correlated Random Slope and Intercept",
                label="HCH - GLMM with Correlated Random
                   Coefficients", digits=2)


#####################################################
### code chunk number 18: chapter4.Rnw:516-521
#####################################################
rownames(m5.table) <- c("$\\mu_{\\alpha}$",  "$\\mu_{\\beta}$",
                        "$\\beta_{wbc}$", "$\\beta_{gender}$",
                        "$\\sigma_{\\alpha[doc]}$", "$\\sigma_{\\beta
                           [doc]}$", "$\\rho$")
print(m5.table, table.placement="H", caption.placement="bottom",
    sanitize.rownames.function = function(x) {x})



#####################################################
### code chunk number 19: chapter4.Rnw:544-557
#####################################################
# Adapted the pprobit.probs \pkg{LearnBayes}
blogit.probs <- function (X1, fit) #adapted from bprobit.probs{
    LearnBayes}
```

```
{
    d = dim(X1)
    n1 = d[1]
    md = dim(fit)
    m = md[1]
    m1 = array(0, c(m, n1))
    for (j in 1:n1) {
        m1[, j] = binomial()$linkinv(X1[j, ] %*% t(fit))  #note use
            of arm{invlogit}
    }
    return(m1)
}


#####################################################
### code chunk number 20: m2a.plot.setup
#####################################################
attach.jags(m3, overwrite=TRUE)
m3.beta <- cbind(mu.a, b.logitpp, b.wbc, b.gender)
detach.jags()
w <- seq(5,25)
X.mh <- cbind(1, logit(90), w, 1) # X matrix males with priorprob=90
    for wbc 5:25
X.fh <- cbind(1, logit(90), w, 0) # females with priorprob=90, wbc
    5:25
X.ml <- cbind(1, logit(10), w, 1) # male with priorprob=10, wbc 5:25
X.fl <- cbind(1, logit(10), w, 0) # female with priorprob=10, wbc
    5:25
X.me <- cbind(1, logit(50), w, 1) # male with priorprob=50, wbc 5:25
X.fe <- cbind(1, logit(50), w, 0) # female with priorpob=50, wbc 5:25
# compute the predicted probabilities
pred.mh <- blogit.probs(X.mh, m3.beta)
```

```
pred.fh <- blogit.probs(X.fh, m3.beta)
pred.ml <- blogit.probs(X.ml, m3.beta)
pred.fl <- blogit.probs(X.fl, m3.beta)
pred.me <- blogit.probs(X.me, m3.beta)
pred.fe <- blogit.probs(X.fe, m3.beta)


#####################################################
### code chunk number 21: m3Highprob
#####################################################
plot(w, apply(pred.mh, 2, quantile, 0.5), type="l",col="blue", ylim=c
    (0,1), lwd=3,
      xlab="White Blood Cell Count", ylab="Probability of Appendicitis
          ")
lines(w, apply(pred.mh, 2, quantile, 0.025), lty=2, col="blue", lwd
    =2)
lines(w, apply(pred.mh, 2, quantile, 0.975), lty=2, col="blue", lwd
    =2)
lines(w, apply(pred.fh, 2, quantile, 0.5), type="l", ylim=c(0,1), col
    ="pink", lwd=3)
lines(w, apply(pred.fh, 2, quantile, 0.025), lty=2, col="pink", lwd
    =2)
lines(w, apply(pred.fh, 2, quantile, 0.975), lty=2, col="pink", lwd
    =2)
legend(20, 0.5, c("Boys", "Girls"), lwd=c(3,3), col=c("blue", "pink")
    )


#####################################################
### code chunk number 22: m3LowProb
#####################################################
plot(w, apply(pred.ml, 2, quantile, 0.5), type="l",col="blue", ylim=c
```

```
    (0,1), lwd=3,
      xlab="White Blood Cell Count", ylab="Probability of Appendicitis
          ")
lines(w, apply(pred.ml, 2, quantile, 0.025), lty=2, col="blue", lwd
    =2)
lines(w, apply(pred.ml, 2, quantile, 0.975), lty=2, col="blue", lwd
    =2)
lines(w, apply(pred.fl, 2, quantile, 0.5), type="l", ylim=c(0,1), col
    ="pink", lwd=3)
lines(w, apply(pred.fl, 2, quantile, 0.025), lty=2, col="pink", lwd
    =2)
lines(w, apply(pred.fl, 2, quantile, 0.975), lty=2, col="pink", lwd
    =2)
legend(7.5, 0.8, c("Boys", "Girls"), lwd=c(3,3), col=c("blue", "pink"
    ))


#######################################################
### code chunk number 23: m3Midprob
#######################################################
plot(w, apply(pred.me, 2, quantile, 0.5), type="l",col="blue", ylim=c
    (0,1), lwd=3,
      xlab="White Blood Cell Count", ylab="Probability of Appendicitis
          ")
lines(w, apply(pred.me, 2, quantile, 0.025), lty=2, col="blue", lwd
    =2)
lines(w, apply(pred.me, 2, quantile, 0.975), lty=2, col="blue", lwd
    =2)
lines(w, apply(pred.fe, 2, quantile, 0.5), type="l", ylim=c(0,1), col
    ="pink", lwd=3)
lines(w, apply(pred.fe, 2, quantile, 0.025), lty=2, col="pink", lwd
    =2)
```

```
lines(w, apply(pred.fe, 2, quantile, 0.975), lty=2, col="pink", lwd
    =2)
legend(5, 1, c("Boys", "Girls"), col=c("blue", "pink"), lwd=c(3,3))
par(mfrow=c(1,1))
```

BIBLIOGRAPHY

Albert, J., *Bayesian computation with R.* New York: Springer, 2009.

Alvarado, A., "A practical score for the early diagnosis of acute appendicitis." *Ann Emerg Med*, vol. 15, no. 5, pp. 557–564, May 1986.

Bates, D., Maechler, M., and Bolker, B., *lme4: Linear mixed-effects models using S4 classes*, 2011, Accessed 26 Nov 2011. [Online]. Available: http://R-Forge.R-project.org/projects/lme4/

Bhatt, M., Joseph, L., Ducharme, F. M., Dougherty, G., and McGillivray, D., "Prospective validation of the pediatric appendicitis score in a Canadian pediatric emergency department." *Acad Emerg Med*, vol. 16, no. 7, pp. 591–596, Jul 2009.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H., and White, J. S., "Generalized linear mixed models: a practical guide for ecology and evolution," *Trends in Ecology & Evolution*, vol. 24, no. 3, pp. 127–135, Mar 2009.

Bundy, D. G., Byerley, J. S., Liles, E. A., Perrin, E. M., Katznelson, J., and Rice, H. E., "Does this child have appendicitis?" *JAMA*, vol. 298, pp. 438–451, Jul 2007.

Burkett, J., Personal communication, 2011.

Cohen, J., "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

Davison, A. C. and Hinkley, D. V., *Bootstrap Methods and Their Applications.* Cambridge: Cambridge University Press, 1997, iSBN 0-521-57391-2. [Online]. Available: http://statwww.epfl.ch/davison/BMA/

Escribá, A., Gamell, A. M., Fernández, Y., Quintillá, J. M., and Cubells, C. L., "Prospective validation of two systems of classification for the diagnosis of acute appendicitis." *Pediatr Emerg Care*, vol. 27, no. 3, pp. 165–169, Mar 2011.

Feinstein, A. R., "The inadequacy of binary models for the clinical reality of three-zone diagnostic decisions." *J Clin Epidemiol*, vol. 43, no. 1, pp. 109–113, 1990.

Gelman, A., *Bayesian Data Analysis*, 2nd ed. Boca Raton, Fla.: Chapman & Hall/CRC, 2004.

Gelman, A. and Hill, J., *Data analysis using regression and multilevel/hierarchical models*. Cambridge ; New York: Cambridge University Press, 2007. [Online]. Available: http://www.stat.columbia.edu/~gelman/arm/

Gill, C. J., Sabin, L., and Schmid, C. H., "Why clinicians are natural Bayesians," *BMJ*, vol. 330, no. 7499, pp. 1080–1083, 2005.

Goldman, R. D., Carter, S., Stephens, D., Antoon, R., Mounstephen, W., and Langer, J. C., "Prospective validation of the pediatric appendicitis score." *J Pediatr*, vol. 153, no. 2, pp. 278–282, Aug 2008.

Hagendorf, B. A., Clarke, J. R., and Burd, R. S., "The optimal initial management of children with suspected appendicitis: a decision analysis," *Journal of Pediatric Surgery*, vol. 39, no. 6, pp. 880–885, Jun 2004.

Harrell, Jr., F. E., *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis*. New York: Springer, 2001. [Online]. Available: http://biostat.mc.vanderbilt.edu/wiki/Main/RmS

Harrell, Jr., F. E., "UseR!2010. invited lecture: Information allergy," 2010, Accessed 26 Nov 2011. [Online]. Available: http://www.r-project.org/conferences/useR-2010/Invited/harrell_talk.pdf

Harrell, Jr., F. E., *Hmisc: Harrell Miscellaneous*, 2011, Accessed 26 Nov 2011. [Online]. Available: http://CRAN.R-project.org/package=Hmisc

Harrell, Jr., F. E., *rms: Regression Modeling Strategies*, 2011, Accessed 26 Nov 2011. [Online]. Available: http://biostat.mc.vanderbilt.edu/wiki/Main/Rrms

Kabakoff, R. I., *R in Action: Data Analysis and Graphics with R*. Manning, 2011. [Online]. Available: http://www.manning.com/kabacoff/

Kharbanda, A. B., Taylor, G. A., Fishman, S. J., and Bachur, R. G., "A clinical decision rule to identify children at low risk for appendicitis." *Pediatrics*, vol. 116, no. 3, pp. 709–716, Sep 2005.

Kuhn, M., "CRAN Task View: Reproducible Research," Accessed 15 Nov 2011. [Online]. Available: http://cran.r-project.org/web/views/ReproducibleResearch.html

le Cessie, S. and van Houwelingen, J. C., "A goodness-of-fit test for binary regression models, based on smoothing methods," *Biometrics*, vol. 47, pp. 1267–1282, 1991.

Leisch, F., "Sweave: Dynamic generation of statistical reports using literate data analysis," in *Compstat 2002 — Proceedings in Computational Statistics*, Härdle, W. and Rönz, B., Eds. Physica Verlag, Heidelberg, 2002, pp. 575–580, iSBN 3-7908-1517-9. [Online]. Available: http://www.stat.uni-muenchen.de/~leisch/Sweave

Lintula, H., Pesonen, E., Kokki, H., Vanamo, K., and Eskelinen, M., "A diagnostic score for children with suspected appendicitis." *Langenbecks Arch Surg*, vol. 390, no. 2, pp. 164–170, Apr 2005.

Mandeville, K., Pottker, T., Bulloch, B., and Liu, J., "Using appendicitis scores in the pediatric ED," *Am J Emerg Med*, vol. 29, pp. 972–977, Nov 2011.

Ntzoufras, I., *Bayesian modeling using WinBUGS*. Hoboken, N.J.: Wiley, 2009. [Online]. Available: http://stat-athens.aueb.gr/~jbn/winbugs_book/

Plummer, M., "Jags: A program for analysis of Bayesian graphical models using Gibbs sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, 2003. [Online]. Available: http://www.ci.tuwien.ac.at/Conferences/DSC-2003/

R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, Accessed 26 Nov 2011. [Online]. Available: http://www.R-project.org/

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, p. 77, 2011.

Rossini, A., Heiberger, R., Sparapani, R., Maechler, M., and Hornik, K., "Emacs Speaks Statistics: A multiplatform, multipackage development environment for statistical analysis," *Journal of Computational and Graphical Statistics*, vol. 13, no. 1, pp. 247–261, 2004.

Rufibach, K., "reporttools: R functions to generate latex tables of descriptive statistics," *Journal of Statistical Software, Code Snippets*, vol. 31, no. 1, pp. 1–7, 8 2009. [Online]. Available: http://www.jstatsoft.org/v31/c01

Samuel, M., "Pediatric appendicitis score." *J Pediatr Surg*, vol. 37, no. 6, pp. 877–881, Jun 2002.

Sarkar, D., *Lattice : multivariate data visualization with R*. New York: Springer, 2008. [Online]. Available: http://lmdvr.r-forge.r-project.org/figures/figures.html

Schneider, C., Kharbanda, A., and Bachur, R., "Evaluating appendicitis scoring systems using a prospective pediatric cohort," *Annals of Emergency Medicine*, vol. 49, no. 6, pp. 778–84, Jun 2007.

Senn, S. and Julious, S., "Measurement in clinical trials: A neglected issue for statisticians?" *Statistics in Medicine*, vol. 28, no. 26, pp. 3189–3209, 2009.

Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T., *ROCR: Visualizing the performance of scoring classifiers.*, 2009, Accessed 4 Dec 2011. [Online]. Available: http://CRAN.R-project.org/package=ROCR

Steyerberg, E. W., *Clinical prediction models : a practical approach to development, validation, and updating.* New York, NY: Springer, 2009. [Online]. Available: http://www.clinicalpredictionmodels.org/

Toolan, T. M. University of Rhode Island. "A guide to producing your thesis with latex." Accessed 14 Nov 2011. June 2008. [Online]. Available: http://www.ele.uri.edu/info/thesis/guide

van Buuren, S. and Groothuis-Oudshoorn, K., "Mice : Multivariate imputation by chained equations," *Journal Of Statistical Software*, vol. To appear., 2011, Accessed 14 Nov 2011. [Online]. Available: http://www.stefvanbuuren.nl/publications.html

van der Heijden, G. J., Rogier, A., Donders, T., Stijnen, T., and Moons, K. G., "Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example." *J Clin Epidemiol*, vol. 59, no. 10, pp. 1102–1109, Oct 2006.

Yen, K., Karpas, A., Pinkerton, H. J., and Gorelick, M. H., "Interexaminer reliability in physical examination of pediatric patients with abdominal pain." *Arch Pediatr Adolesc Med*, vol. 159, no. 4, pp. 373–376, Apr 2005.