PREDICTION OF RECURRENCE IN THIN MELANOMA USING TREES AND RANDOM FORESTS

Richard M. Reiter

A Thesis Submitted to the
University of North Carolina Wilmington in Partial Fulfillment
Of the Requirements for the Degree of
Master of Science

Department of Mathematics and Statistics

University of North Carolina Wilmington

2005

Approved by

Advisory Committee

_____ _____

_____
Chair

Accepted by

_____
Dean, Graduate School

TABLE OF CONTENTS

ABSTRACT

In this paper, we will try to predict the recurrence of melanoma in the subset of patients with a “thin melanoma,” defined by having a Breslow Thickness less than 1.00mm. The study used the 1610 patient data base from the Duke Melanoma Clinic. The methods used to try classifying these patients as to whether or not they will suffer a recurrence are based on the statistical processes of trees and random forests introduced by Dr. L. Breiman. Successful results were achieved only when all patients were studied including those with an advanced stage of disease, but who still qualified as patients with a thin melanoma by histological measurement. The variables collected on each patient and used in this study related to the Breslow Thickness, Clark level, initial stage, age, primary site of lesion, local pathological events, histological type, sex, race and previous immunotherapy. The factors other than the initial stage and extent of disease proved too weak to give any meaningful results when only patients with local disease were studied. The results using a single tree and a random forest were compared. A detailed discussion of trees and random forest with use in the R-2.01 CRAN software packages is included. Although the results of this study did not provide the answers for this specific set of melanoma patients, it is felt that the techniques and programs written would be applicable in a differently defined set of patients. The results could be an important step in determining which patients should receive therapy and then evaluating the results of the adjuvant therapy on a more properly defined group of patients.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Dargan Frierson of the Mathematics and Statistics Department of the University of North Carolina Wilmington for his excellent guidance helping me to complete my thesis. I appreciate the dedication and hard work of Dr. Susan Simmons and Dr. Gabriel Lugo serving as advisors for this thesis. I would also like to thank Dr. H. Seigler , Professor of Surgery and head of the Duke Melanoma Clinic, for sponsoring my application to use the D.U.M.C. patient data base and providing clinical guidance for the study. Finally I would acknowledge the expertise of Dr. L. Breiman and his staff for creating and sharing the mathematics and methods of trees and random forest.

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

LIST OF ITEMS