The Immunoepidemiology of Lung Cancer and Smoking: Lymphocyte Subsets.

Joan Peters Blackmon
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://repository.lsu.edu/gradschool_disstheses

Recommended Citation
https://repository.lsu.edu/gradschool_disstheses/5222

This Dissertation is brought to you for free and open access by the Graduate School at LSU Scholarly Repository. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Scholarly Repository. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
The immunoepidemiology of lung cancer and smoking: Lymphocyte subsets

Blackmon, Joan Peters, Ph.D.
The Louisiana State University and Agricultural and Mechanical Col., 1991
THE IMMUNOEPIDEMILOGY
OF LUNG CANCER AND SMOKING:
LYMPHOCYTE SUBSETS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Interdepartmental Program

in

Veterinary Medical Sciences

(Veterinary Microbiology and Parasitology Option)

by

Joan Peters Blackmon
B.S., The University of Pittsburgh, 1963
A.M. in Organic Chemistry, Duke University, 1965
M.S.P.H. in Epidemiology,
The University of North Carolina, Chapel Hill, 1971
December 1991
To

my parents,
Elinor and David Peters

my husband,
Bill,

and my children,
Heather, Tim, and Amanda

with Love
ACKNOWLEDGMENTS

I wish to thank my major professor Dr. Mark Newman and Dr. Kenneth Lo of Mary Bird Perkins Cancer Center for providing me the opportunity to conduct a study of this nature, and the Cancer Society of Greater Baton Rouge and the Louisiana Lung and Cancer Trust for providing funding. Thanks go also to my minor professor Dr. E. Barry Moser for guidance in the statistical analysis and the use of SAS, and the faithful members of my committee, Dr. William Todd, Dr. Thomas Gillis, and Dr. Barbara Shane. Dr. Grace Amborski, who retired before this effort was completed, is also remembered for her encouragement.

My appreciation also extends to Dr. Johannes Storz, Chairman of the Department of Veterinary Microbiology and Parasitology and the various members of the faculty of that department, who have been supportive and helpful throughout the project. Others to whom I am grateful are Ms. Marilyn Dietrich for her help with the flow cytometer, for drawing blood for me, and for doing differential counts on the samples, and Mr. Michael Kearney, Mr. James Roberts, and Mr. R.J. Sharman for guiding this neophyte through the maze of computer technology that I needed to master.

iii
I am grateful to my husband Bill, who has been my mainstay throughout the trials and tribulations of this graduate school process, for lovingly "picking up the slack" for me at home. The children made tangible contributions to this work beyond their patient endurance of Mom's preoccupation with school. Thanks go to Heather for entering questionnaire data into the computer and to Tim and Amanda for labelling thousands of FACS tubes for me.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xiv</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>REVIEW OF LITERATURE</td>
<td>5</td>
</tr>
<tr>
<td>Epidemiologic studies in</td>
<td>5</td>
</tr>
<tr>
<td>Louisiana</td>
<td></td>
</tr>
<tr>
<td>Smoking and lung cancer</td>
<td>6</td>
</tr>
<tr>
<td>Carcinogenic properties of tobacco</td>
<td>8</td>
</tr>
<tr>
<td>Smoking in Louisiana</td>
<td>11</td>
</tr>
<tr>
<td>Other factors in lung cancer</td>
<td>13</td>
</tr>
<tr>
<td>Genotype and lung cancer</td>
<td>16</td>
</tr>
<tr>
<td>Diet and cancer</td>
<td>22</td>
</tr>
<tr>
<td>Methods of obtaining information on diet</td>
<td>30</td>
</tr>
<tr>
<td>Tumor immunity</td>
<td>32</td>
</tr>
<tr>
<td>Smoking and the immune system</td>
<td>38</td>
</tr>
<tr>
<td>Lung cancer and leukocyte subsets</td>
<td>40</td>
</tr>
<tr>
<td>Flow cytometry</td>
<td>45</td>
</tr>
<tr>
<td>Flow cytometric lymphocyte subset profiles</td>
<td>53</td>
</tr>
<tr>
<td>Markers used in this study</td>
<td>54</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>70</td>
</tr>
<tr>
<td>Selection of study subjects</td>
<td>70</td>
</tr>
<tr>
<td>The questionnaire</td>
<td>72</td>
</tr>
<tr>
<td>Selection of lymphocyte markers</td>
<td>78</td>
</tr>
<tr>
<td>Subject kits</td>
<td>81</td>
</tr>
<tr>
<td>Phlebotomy and processing of blood specimens</td>
<td>82</td>
</tr>
<tr>
<td>Differential</td>
<td>84</td>
</tr>
<tr>
<td>Fluorescence labelling</td>
<td>84</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>85</td>
</tr>
<tr>
<td>A novel method for FACS data processing</td>
<td>87</td>
</tr>
<tr>
<td>The master data set</td>
<td>96</td>
</tr>
<tr>
<td>The lymphocyte profile</td>
<td>96</td>
</tr>
<tr>
<td>Isolation of single effects</td>
<td>100</td>
</tr>
<tr>
<td>Other tests</td>
<td>101</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Ternary diagrams</td>
<td>101</td>
</tr>
<tr>
<td>The leukocyte profile</td>
<td>103</td>
</tr>
<tr>
<td><strong>RESULTS</strong></td>
<td>105</td>
</tr>
<tr>
<td>Study subjects</td>
<td>105</td>
</tr>
<tr>
<td>Response to questionnaire</td>
<td>107</td>
</tr>
<tr>
<td>Sample-to-sample spillover</td>
<td>109</td>
</tr>
<tr>
<td>Performance of the new data analysis method</td>
<td>114</td>
</tr>
<tr>
<td>Lymphocyte subsets</td>
<td>117</td>
</tr>
<tr>
<td>The lymphocyte profile</td>
<td>123</td>
</tr>
<tr>
<td>Circadian variation</td>
<td>138</td>
</tr>
<tr>
<td>The T(month)-by-disease interaction</td>
<td>141</td>
</tr>
<tr>
<td>Circannual variation</td>
<td>150</td>
</tr>
<tr>
<td>The disease-by-vegetable interaction</td>
<td>154</td>
</tr>
<tr>
<td>The vegetable effect</td>
<td>154</td>
</tr>
<tr>
<td>Lung cancer effect</td>
<td>159</td>
</tr>
<tr>
<td>The age-by-smoking interaction</td>
<td>159</td>
</tr>
<tr>
<td>Consumption of alcoholic drinks</td>
<td>164</td>
</tr>
<tr>
<td>Consumption of fat</td>
<td>168</td>
</tr>
<tr>
<td>Other factors tested</td>
<td>171</td>
</tr>
<tr>
<td>Patients and controls on fat intake</td>
<td>171</td>
</tr>
<tr>
<td>Disease and exposure to substances or occupations</td>
<td>172</td>
</tr>
<tr>
<td>Leukocyte subsets</td>
<td>174</td>
</tr>
<tr>
<td><strong>DISCUSSION</strong></td>
<td>179</td>
</tr>
<tr>
<td>Study subjects</td>
<td>179</td>
</tr>
<tr>
<td>The questionnaire</td>
<td>181</td>
</tr>
<tr>
<td>Dietary data</td>
<td>182</td>
</tr>
<tr>
<td>The laboratory method</td>
<td>183</td>
</tr>
<tr>
<td>The new method for editing and processing the FACS data</td>
<td>185</td>
</tr>
<tr>
<td>A perspective on analysis</td>
<td>192</td>
</tr>
<tr>
<td>Subset balances</td>
<td>193</td>
</tr>
<tr>
<td>The lymphocyte profile model</td>
<td>198</td>
</tr>
<tr>
<td>Circadian variation</td>
<td>199</td>
</tr>
<tr>
<td>The circannual variation</td>
<td>202</td>
</tr>
<tr>
<td>T(month)-by-disease interaction</td>
<td>203</td>
</tr>
<tr>
<td>The disease-by-vegetable interaction</td>
<td>204</td>
</tr>
<tr>
<td>The disease effect</td>
<td>205</td>
</tr>
<tr>
<td>Age-by-smoking interaction</td>
<td>205</td>
</tr>
<tr>
<td>Consumption of alcoholic drinks</td>
<td>206</td>
</tr>
<tr>
<td>Consumption of fat</td>
<td>206</td>
</tr>
<tr>
<td>Other factors tested</td>
<td>208</td>
</tr>
<tr>
<td>Usefulness of FACS data for diagnosis</td>
<td>209</td>
</tr>
<tr>
<td>The leukocyte profile</td>
<td>209</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Disease and fat intake and occupational exposures</td>
<td>210</td>
</tr>
<tr>
<td>SUMMARY AND CONCLUSIONS</td>
<td>212</td>
</tr>
<tr>
<td>WORKS CITED</td>
<td>218</td>
</tr>
<tr>
<td>APPENDIX A THE QUESTIONNAIRE</td>
<td>241</td>
</tr>
<tr>
<td>APPENDIX B MISCELLANEOUS STUDY MATERIALS</td>
<td>259</td>
</tr>
<tr>
<td>APPENDIX C DATA PROCESSING PROGRAMS</td>
<td>267</td>
</tr>
<tr>
<td>APPENDIX D CODEBOOK FOR THE ANALYSIS DATABASE</td>
<td>279</td>
</tr>
<tr>
<td>VITA</td>
<td>317</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cluster Differentiation Antigens</td>
<td>52</td>
</tr>
<tr>
<td>2. Diet Analysis Options in Effect</td>
<td>75</td>
</tr>
<tr>
<td>3. Markers Utilized</td>
<td>79</td>
</tr>
<tr>
<td>4. Origin of Comparison Subjects</td>
<td>106</td>
</tr>
<tr>
<td>5. Paired Differences of Estimates by Two Methods</td>
<td>114</td>
</tr>
<tr>
<td>6. Variability of Estimates within Method</td>
<td>117</td>
</tr>
<tr>
<td>7. Statistics for Selected Subsets</td>
<td>118</td>
</tr>
<tr>
<td>8. Tests of Subset Equivalence Hypotheses</td>
<td>120</td>
</tr>
<tr>
<td>9. Subjects Deleted because of Extreme Values of (HLA-DR) - (CD22)</td>
<td>124</td>
</tr>
<tr>
<td>10. Final Model for the Lymphocyte Profile</td>
<td>132</td>
</tr>
<tr>
<td>11. Fat Intake by Disease</td>
<td>172</td>
</tr>
<tr>
<td>12. Occupational Exposures</td>
<td>173</td>
</tr>
<tr>
<td>13. Exposures to Carcinogenic Substances</td>
<td>174</td>
</tr>
<tr>
<td>14. Differential Leukocyte Counts</td>
<td>175</td>
</tr>
<tr>
<td>15. The Leukocyte Profile</td>
<td>177</td>
</tr>
</tbody>
</table>
## LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The Mary Bird Perkins Tumor Registry Jurisdiction</td>
<td>71</td>
</tr>
<tr>
<td>2.</td>
<td>Setting a Rectangular Window on Lymphocytes</td>
<td>88</td>
</tr>
<tr>
<td>3.</td>
<td>The Distribution of Leukocytes on Scatter and Fluorescence</td>
<td>91</td>
</tr>
<tr>
<td>4.</td>
<td>The Lymphocyte Editing Process</td>
<td>92</td>
</tr>
<tr>
<td>5.</td>
<td>A Selected Probability Sample of Lymphocytes</td>
<td>93</td>
</tr>
<tr>
<td>6.</td>
<td>NK Cells Exhibit Higher Sidescatter than Other Lymphocytes</td>
<td>95</td>
</tr>
<tr>
<td>7.</td>
<td>The Ternary Diagram</td>
<td>102</td>
</tr>
<tr>
<td>8.</td>
<td>Plotting a Ternary Diagram in Rectangular Coordinates</td>
<td>104</td>
</tr>
<tr>
<td>9.</td>
<td>Age Distribution of Patients and Controls</td>
<td>109</td>
</tr>
<tr>
<td>10.</td>
<td>Income Classification of Patients and Controls</td>
<td>110</td>
</tr>
<tr>
<td>11.</td>
<td>Parish of Residence for Patients and Controls</td>
<td>111</td>
</tr>
<tr>
<td>12.</td>
<td>Educational Status of Patients and Controls</td>
<td>112</td>
</tr>
<tr>
<td>13.</td>
<td>Characterization of Subjects on Disease and Smoking</td>
<td>113</td>
</tr>
<tr>
<td>14.</td>
<td>Subsets Chosen for First-stage Analysis</td>
<td>114</td>
</tr>
<tr>
<td>15.</td>
<td>Profile Adjusted for Age and Disease, by Smoking</td>
<td>125</td>
</tr>
<tr>
<td>16.</td>
<td>Means of Adjusted Profile Logratios by Month of Year</td>
<td>126</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>17. Means of Adjusted Profile Logratios by Hour of Day</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>18. The Adjusted Data Against Adjusted Month</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>19. Distribution of Subjects in the 8-Factor Model by Age</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>20. Distribution of Subjects in the 8-Factor Model by Income</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>21. Distribution of Subjects in the 8-Factor Model by Parish</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td>22. Distribution of Subjects in the 8-Factor Model by Education</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>23. Circadian Variation in the Lymphocyte Profile</td>
<td>139</td>
<td></td>
</tr>
<tr>
<td>24. The Circadian Variation of the Lymphocyte Profile Logratios</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>25. The Distribution of Subjects by Month Enrolled in Study</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>26. Ln(CD3+CD22): The Interaction between T(month) and Disease</td>
<td>143</td>
<td></td>
</tr>
<tr>
<td>27. Ln(CD56+CD22): The Interaction between T(month) and Disease</td>
<td>144</td>
<td></td>
</tr>
<tr>
<td>28. Ln(CD3+CD22) Adjusted for All Effects, Including Disease and Disease-by-vegetable, except T(month) and T(month)-by-disease</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>29. Ln(CD56+CD22) Adjusted for All Effects, Including Disease and Disease-by-vegetable, except T(month) and T(month)-by-disease</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>30. Maximum Separation of Patients' and Controls' Profiles</td>
<td>148</td>
<td></td>
</tr>
<tr>
<td>31. Profile Logratios of Patients and Controls at Months of Maximum Separation</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>32. Ln(CD3+CD22) Adjusted for All Effects except Month, Including the Estimated T(month)-by-disease Effect</td>
<td>151</td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>33</td>
<td>Ln(CD56:CD22) Adjusted for All Effects except Month, Including the Estimated T(month)-by-disease Effect</td>
<td>152</td>
</tr>
<tr>
<td>34</td>
<td>Circannual Variation in Lymphocyte Profile</td>
<td>153</td>
</tr>
<tr>
<td>35</td>
<td>The Disease-by-vegetable Interaction</td>
<td>155</td>
</tr>
<tr>
<td>36</td>
<td>The Disease-by-vegetable Interaction Effect on the Lymphocyte Profile</td>
<td>156</td>
</tr>
<tr>
<td>37</td>
<td>The Isolated Vegetable Effect</td>
<td>157</td>
</tr>
<tr>
<td>38</td>
<td>The Isolated Vegetable Effect, by Vegetable Intake Category</td>
<td>158</td>
</tr>
<tr>
<td>39</td>
<td>The Effect of Lung Cancer on Lymphocyte Profile</td>
<td>160</td>
</tr>
<tr>
<td>40</td>
<td>The Logratios by Smoking Category by Three Age Groups</td>
<td>161</td>
</tr>
<tr>
<td>41</td>
<td>The Logratios by Smoking Category by Five Age Groups</td>
<td>162</td>
</tr>
<tr>
<td>42</td>
<td>The Age-by-smoking Interaction on Lymphocyte Profiles</td>
<td>163</td>
</tr>
<tr>
<td>43</td>
<td>The Effect of Intake of Alcoholic Drinks on Profile Logratios</td>
<td>165</td>
</tr>
<tr>
<td>44</td>
<td>Linear Trend in Profile Logratios by Intake of Alcoholic Drinks in 3 Groups</td>
<td>166</td>
</tr>
<tr>
<td>45</td>
<td>Alcohol Intake and the Lymphocyte Profile</td>
<td>167</td>
</tr>
<tr>
<td>46</td>
<td>Adjusted Profile Logratio Means by Fat Intake</td>
<td>169</td>
</tr>
<tr>
<td>47</td>
<td>Fat Intake Effect on Lymphocyte Profile</td>
<td>170</td>
</tr>
<tr>
<td>48</td>
<td>Comparison of Bonferroni simultaneous C.I. to bivariate 95% Estimation Ellipse</td>
<td>188</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

AIDS acquired immune deficiency syndrome
amFITC Fluorescein Isothiocyanate-conjugated goat anti-mouse IgG
ATCC American Type Culture Collection
BD Becton Dickinson Immunocytometry Systems, Mountain View, CA 94039
CD cluster determinant
ConA concanavalin A
CPS Current Population Survey
CR cumulative risk
CTL cytotoxic T-lymphocytes
DF degrees of freedom
DMSO dimethyl sulfoxide
FACS fluorescence-activated cell sorter(s)
FH Ficoll-Hypaque density gradient
FITC fluorescein isothiocyanate
HHHQ Health Habits and History Questionnaire
IL2 interleukin 2
LAK lymphokine-activated killer
NCI National Cancer Institute
NHANES II Second National Health and Nutrition Examination Survey
NK natural killer cell(s)
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAH</td>
<td>polycyclic aromatic hydrocarbons</td>
</tr>
<tr>
<td>PBL</td>
<td>peripheral blood lymphocytes</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate-buffered saline</td>
</tr>
<tr>
<td>PBS/G</td>
<td>10% complement-inactivated goat serum in phosphate-buffered saline</td>
</tr>
<tr>
<td>PCC</td>
<td>Mary Bird Perkins Cancer Center</td>
</tr>
<tr>
<td>PE</td>
<td>phycoerythrin</td>
</tr>
<tr>
<td>PWM</td>
<td>pokeweed mitogen</td>
</tr>
<tr>
<td>RDA</td>
<td>recommended daily allowance(s)</td>
</tr>
<tr>
<td>rIL2</td>
<td>recombinant interleukin 2</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>SAS</td>
<td>Statistical Analysis System</td>
</tr>
<tr>
<td>SCE</td>
<td>sister chromatid exchange(s)</td>
</tr>
<tr>
<td>SSD</td>
<td>standardized squared distance of a point from the estimated mean</td>
</tr>
<tr>
<td>TIL</td>
<td>tumor infiltrating lymphocytes</td>
</tr>
</tbody>
</table>
The relative proportions of lymphocyte subsets in the peripheral blood constitute a measure of immune status. Results of research to establish whether or not individuals with lung cancer differ in their lymphocyte composition have conflicted due to failure to control for factors other than cancer that affect lymphocyte balance. This study was designed to quantitatively estimate the effects of variables such as disease (lung cancer), smoking, age, circadian and circannual rhythms, and dietary factors on lymphocyte subsets. Questionnaires were given to 64 white males with primary lung cancer and 219 cancer-free white males and a blood sample was obtained. Lymphocytes were labelled with subset-specific, fluorescence-tagged antibodies and analyzed on a fluorescence-activated cell sorter (FACS). A computer algorithm for processing the FACS data was invented and utilized to determine proportions positive.

The CD3-CD22-CD56 lymphocyte profile was modelled using bivariate loglinear multivariable regression. Disease, age, smoking, hour-of-day, month-of-year, weekly intake of vegetables, and percentages of daily calories from fat and from alcoholic drinks were significant main effects and there were significant age-by-smoking, disease-by-month, and disease-by-vegetable-intake interactions. Smokers' profiles did not mature with increasing age while
non-smokers showed an increase in natural killer cells of 10 percentage points from age quartile <51 years to quartile >68 years. Subjects with >28% of their calories in fat averaged 5% more CD3+ and 4% fewer CD22+ cells than those with <28% fat in their diets. Lung cancer patients had 1% fewer CD22+ and 1% more CD3+ lymphocytes than controls, had 11.7% HLA-DR+ lymphocytes that were not CD22+ compared to 6.6% for controls, and averaged 24% lymphocytes compared to 37% for controls. Their lymphocyte profiles were different from controls' in pattern of circannual variation and patients showed more extreme effect from a low-vegetable diet than controls. Patients reported a higher fat intake, and more occupational exposures to carcinogenic substances, but did not differ from controls in β-carotene consumption.
INTRODUCTION

"Cancer is one of the most serious health problems in Louisiana, second only to heart disease" (Task Force on Environmental Health, 1984). Cancer mortality rates for all males and for black females in Louisiana between 1968 and 1978 were well above the national average during the same period. In particular, lung cancer accounts for 85% of the difference between Louisiana's cancer mortality and the average rates for the United States as a whole (Fontham, et al. 1988). Mortality rates from lung cancer among white males were 28% higher than U.S. rates for the period 1968-77 (Task Force on Environmental Health, 1984) and lung cancer has been the leading cause of death among American men since 1950 (Loeb et al. 1984). Reasons for the higher risk in Louisiana are unknown but early epidemiologic studies focused on environmental factors. More recently, Fontham and Correa contended that lifestyle is the major modifier, citing an exceptionally high prevalence of smoking among southern Louisianians. They implicate inadequate diet, secondarily (Fontham et al. 1988).

Cancer is a complex disease that can affect most tissues and organs of the body and can be induced by
disparate agents including viruses, chemicals, and radiation. Mechanisms for tumor induction and progression are still not well defined, although in general they involve alteration of cell DNA which codes for molecules critical to the regulation of cell growth and division. Susceptibility factors, such as deficiencies in the immune system, have not been characterized either. Since experimental induction of cancers in humans would be unconscionable, the etiology of human cancers has, for the most part, been explored observationally through epidemiologic methods. Furthermore, cancers are relatively rare events, therefore most investigations have been of the relatively economical, retrospective, case-control type. Case-control studies have the limitation of not establishing the precedence of the putative cause. Such studies can, however, provide clues toward the identity of tumorigenic agents, conditions, or processes. Experimental induction of tumors in animals with suspected agents or under conditions of vulnerability then strengthens the hypothesis that they are carcinogenic for man.

Classical epidemiologic studies in man have seldom involved biologic assays, but neither have laboratory studies adequately attended to effect-modifying factors intrinsic to the individuality and uniqueness of each human being. Human beings are not comparable to one another in the same way that inbred strains of experimental animals
are, therefore one has to identify covariates relevant to the process being studied and adjust for these through design and/or with statistical methods. There has been a trend over the last decade or two to incorporate biophysical measures into epidemiologic methodology to obtain more quantitative information about etiologic processes of and mechanisms of resistance to disease. "Biochemical epidemiology" combines the advantages of both laboratory science and survey methods (Brown et al. 1989). Measurement of in vivo carcinogen load, nutritional status, and immunologic parameters have been applications of this approach. The current study focussed on (1) immune status, as reflected by lymphocyte subset profiles, (2) its modulation by smoking, diet, and life-style factors, and (3) its possible role in the etiology of lung cancer.

The hypothesis that was the basis for the work to be described below is that environmental exposures and life-style factors such as smoking and inadequate diet are both important components of lung cancer etiology and that either or both can modulate the immune system in such a way as to enhance susceptibility to lung cancer. Specific objectives included:

(1) identification of life-style and environmental factors which affect lymphocyte subset balance in the peripheral blood compartment, and
(2) identification of differences in leukocyte/lymphocyte subset profiles between lung cancer patients and controls, after controlling for confounding variables, which might

(a) be useful for early diagnosis, or

(b) contribute to the understanding of immunologic defenses against lung cancer.

To address this hypothesis and these objectives, a case-control study incorporating a questionnaire that addresses diet, smoking behavior, and environmental/occupational exposures, and a laboratory assay that provides a measure of immune competence was designed and conducted. Results of this study are the substance of this dissertation.
Epidemiologic studies in Louisiana.

Early studies suggested that shipbuilding, sugarcane farming, other agriculture, mining, commercial fishing, and certain jobs within the petroleum industry were occupations in Louisiana that correlated with higher risk of lung cancer. Welders, workers in oil fields, operators, boilermakers, and painters were specific jobs which were more frequently recorded for cases than for controls (Rothschild and Mulvey 1982; Gottlieb, et al. 1979; Gottlieb and Stedman 1979; Gottlieb 1980). Residence within a mile of chemical or petroleum industries or near lumber industries was also tentatively identified as conferring a higher risk (Gottlieb, Shear, and Seale 1982). Most of the initial studies were based on analysis of death-certificate information. While useful for hypothesis generation, such sources did not include information on smoking, diet, or other possible confounders and lacked historical information on places of residence and other occupations (Wong and Foliart 1989).

Subsequent studies that incorporated interview data did support higher risks associated with sugar-cane farming (increased risk among smokers only) and shipbuilding
(Rothschild and Mulvey, 1982). The risk in shipbuilding appears to be associated with exposure to asbestos and is also elevated among asbestos cement workers (Wong and Foliart 1989). Since shipbuilding and sugar cane farming are predominantly industries of the southern parishes, they correlate with the geographic distribution of lung cancer in Louisiana as well. The findings of five cohort studies of refinery and chemical workers, however, indicated that there was no significantly increased risk associated with employment in those industries (Wong and Foliart 1989). Since workers in those industries would presumably be exposed to higher doses of potential carcinogens than residents of areas surrounding the plants, the negative findings of these industrial studies also are evidence against hypotheses of higher risk among residents near refineries and chemical plants. Case-control studies with interview data have also failed to find an association of lung cancer with residence near industries (Wong and Foliart 1989).

**Smoking and lung cancer.**

Smoking is associated with a number of cancers including those of the lung, larynx, oral cavity, esophagus, bladder, kidney, pancreas, cervix, and stomach (Subar, Harlan, and Mattson 1990). Greater than 95% of primary lung cancers among men and more than 85% among women, are associated with smoking (Loeb et al., 1984). Of the most
common histologic types of lung cancer, small-cell and squamous-cell tumors are the most strongly correlated with smoking. In addition to their statistical association with smoking, small cell and squamous cell tumors also tend to occur in the upper lobes of the lungs, which are thought to be more heavily exposed during cigarette smoking than the lower lobes (Celikoglu et al. 1986). In one study of 2668 lung cancer cases, only 1.9% of the males were non-smokers and 13% of the females were non-smokers (Kabat, Geoffrey, and Wynder 1984). In a 10-year prospective study of 6,027 men, 45 years of age and older, none of the 830 men who had never smoked on a regular basis developed lung cancer. Among the other 5197 men who had smoked there were 121 lung cancer cases. A dose-response gradient in risk, based on number of cigarettes smoked per day, was demonstrated for each histologic type of lung cancer except for the poorly differentiated squamous-cell tumors and large-cell tumors (for the latter of which there were too few cases; Weiss et al. 1972). According to a recently reported study in Louisiana involving 1253 cases of lung cancer, 98 percent of both white and black male cases, 94 percent of black female cases, and 88 percent of white female cases were smokers or former smokers (Fontham, Chen and Correa 1988).

A Louisiana study was also the first to document an increased risk of lung cancer among the non-smoking spouses of smokers (Correa et al. 1983). The relative risk for
spouses of people with less than 40 pack-years of cigarette smoking (equivalent to one pack per day for 40 years) was 1.48 and that for those married to people with greater than 40 pack-years of exposure was 3.1. Several other studies in the U.S.A. and other countries have also concluded that there is a connection between passive smoking at home and in the workplace with the development of lung cancer (Hirayama 1987). However a number of studies have not found an increased risk from passive smoking. Comparability among these studies is not especially good due to differences and flaws in design. Evaluation of the combined results of nine such studies (meta-analysis), however, led to the conclusion of an overall odds ratio of 1.12 with a 95% confidence interval of 0.95-1.30, that is, a non-significantly elevated risk (Fleiss and Gross 1991). Despite the inconsistency of epidemiologic findings, there is physiologic evidence that exposure to the cigarette smoke of others is detrimental. Decrements in lung function, based on measurement of forced expiratory volumes and flow rates, were found among both children whose mothers smoked and adult, non-smoking spouses of smokers (Loeb et al. 1984).

**Carcinogenic properties of tobacco.**

Tobacco smoke is a complete carcinogen, that is, it contains both initiating and promoting compounds. It is commonly accepted that polycyclic aromatic hydrocarbons
(PAH) are the major initiator carcinogens in pyrolyzed tobacco although they constitute less than 0.003% weight of the smoke condensate or tar. They are relatively unreactive in themselves but are activated to strongly electrophilic intermediates by normal mammalian enzymes termed mixed function oxidases (isozymes of the cytochromes P-450; Gelboin 1980). These enzymes themselves are induced 2- to 135-fold by smoking (Conney 1982). The reactive diol epoxide intermediates then can bind covalently to DNA, usually to the 2-amino group of guanine (Sims et al. 1974; Jeffrey et al. 1977). Cell transformation and tumor initiation correlate with the formation and persistence of this adduct (Nakayama 1984; Stowers 1985; Autrup, 1982). Damaged DNA can be detected and repaired by mechanisms present in the affected cell. It is known, however, that smokers have subnormal capacity to repair damaged DNA (Setlow 1983). If the cell replicates before the damage is repaired, a mutation may occur. Such a mutation in genes which code for key proteins that regulate cell division and differentiation constitutes initiation.

Initiation is not sufficient to produce a tumor; at least one other event termed promotion must take place. When a synthetic mixture of 17 of the most commonly identified PAH, combined in proportions comparable to their natural occurrence in the tar, was applied to mouse skin at a dose representative of the amount contained in a
carcinogenic amount of tar, the PAH did not induce tumors. However, when this synthetic mixture was added to a tumorigenic dose of tar, the tumor yield was enhanced greater than two-fold (D. Hoffmann, et al. 1978). This indicated that constituents in the tar acted synergistically with the PAH, that is, they were promoters, and also that the PAH are responsible for a large fraction of the tumorigenicity of cigarette smoke condensate.

An epidemiologic study in Finland indicated that workers exposed to high concentrations of PAH experience a higher risk of lung cancer (RR=1.7; Rantanen 1983). The mutagenicity of the urine of these workers validated their high exposure. PAH adducts to DNA in blood lymphocytes as well as antibodies in peripheral blood to such adducts have been detected among persons occupationally exposed to PAH (Harris et al. 1985). The demonstration of DNA adducts in human tissues, which are similar to those known to occur during tumor induction in animals, strongly suggests that exposure to PAH occupationally or through smoking are causally related to the development of cancer in humans.

Cocarcinogenesis and promotion by constituents in tobacco smoke and the suppression of DNA-repair mechanisms by smoking might be as important or more important than initiation by constituents of tobacco smoke. The carcinogenic effects of other exposure to PAH or exposure to other carcinogens can be substantially amplified by
concurrent cigarette smoking. It is well documented that the risks of mesothelioma and of lung cancer among people exposed to asbestos are elevated for smokers compared to non-smokers. For non-smokers who worked with asbestos, the risk relative to that of men who neither smoked nor worked with asbestos was 5 times. Smokers who did not work with asbestos had a relative risk of 11, but those smokers who also worked with asbestos had 53 times the risk of unexposed non-smokers for lung cancer (Wagner 1984)! There is evidence in several studies that smoking acts synergistically with exposure to uranium in causing lung cancer among uranium miners (Loeb 1984), and smoking more than 30 grams of tobacco per day seems to interact with heavy alcohol intake to greatly increase the risk of laryngeal cancer (Doll and Peto 1981).

**Smoking in Louisiana.**

Since most lung cancer can be attributed to smoking, could the excess lung cancer in Louisiana be due to a relatively high prevalence of smoking among Louisianians? A National Health Interview Survey in 1979 found that 36.9% of men over age 17 and 28.2% of women reported themselves to be smokers and showed that prevalence of smoking had declined since 1965. In the subsequent nationwide Current Population Survey (CPS) of 1985, 30.6-32%\(^1\) of all male

\(^1\)Range indicates confidence interval around the survey-estimated prevalence.
respondents and 24.3-25.7% of the female respondents over
the age of 16 reported themselves to be regular smokers,
showing a further decline. Louisiana ranked 19th among the
fifty states in overall smoking prevalence for males
(33.1±3.3%) and 43rd for prevalence among females
(22.3±2.7%). The four-state region which contained
Louisiana, Texas, Arkansas, and Oklahoma reported 33.6±1.4%
of white men and 37.3±3.4% of black men as smokers. The
rates for white men were about 3% higher than the national
average and those for black men at the national average
(Marcus et al. 1989). Given these statistics, higher
prevalence of smoking would not account for a 28% excess of
lung cancer among white males in Louisiana.

Somewhat different figures were reported from a 1983
lifestyles survey among men over 30 years of age in
Louisiana: 34.7-38.7% (compared with the CPS average of
33.6%) of white men and 46.9-54% (CPS average = 37.3%) of
black men reported themselves to be current smokers (Correa
and Johnson 1983). The somewhat higher rates among white
males might have resulted from not including the younger age
groups in this survey, since smoking prevalence was highest
in the 30-49-yr age groups in the CPS. Correa’s rates for
black males, however, are so much higher than the regional
average for black males (30.5-44.1%) in the CPS that the
confidence intervals of the two estimates do not overlap.
This would lead one to wonder if one study or the other had
a biased population sample. Although Louisiana did not have an exceptionally high prevalence of smoking in the early 1980s, it is still possible, that smoking prevalence among white males in Louisiana was truly higher than the national prevalence in the late 1950s and the early 1960s, a time period which would be more relevant for induction of cancers which were diagnosed between 1968 and 1977.

Lung cancer rates were higher in southern parishes of Louisiana than in the northern part of the state. Comparing lifestyles of men within an area of high lung cancer mortality to those of men within an area of the state with half the rate or less, Correa found that higher percentages of both white and black men smoked, a higher percentage of smokers used non-filter cigarettes, more whites reported that their parents were smokers, and smokers had taken up the habit at an earlier age in the area where there was higher mortality (Correa and Johnson 1983). Thus particularly intense smoking behaviors in parts of the state may have contributed disproportionately to the state-wide excess of lung cancer.

**Other factors in lung cancer.**

Smoking can be considered a necessary, if not sufficient cause for perhaps 85-90% of male lung cancers and over 70% of lung cancer among females (Loeb, et al. 1984). The portion of risk attributable to smoking for Louisianians was estimated to be 90% (Fontham, Chen, and Correa 1988).
However since non-smokers can succumb to lung cancer but not all smokers do succumb, there must be factors other than smoking which contribute to the development of lung cancers. Lee and coworkers (Lee, Fry, and Forey 1990) demonstrated that the recent declines in age-specific lung cancer rates for younger men and women in England and Wales from 1950 onwards cannot be entirely explained by changes in the "cumulative constant tar cigarette consumption" during the same period. This discrepancy is also evident for emphysema and especially for chronic obstructive lung disease. While it is perhaps the most important single factor in the etiology of lung cancer, smoking does not account for all lung cancers.

Although much attention has been focused upon smoking as the major cause of lung cancer, very little is known about the etiology of the disease among non-smokers. Occupational exposures, which have been identified as increasing risks for lung cancer independently of smoking, are arsenic, bischloromethyl ether, chromium, ionizing radiation, mustard gas, nickel, and PAH. It is worthy of note that Correa found higher percentages of smokers in those industries thought to be associated with high rates of lung cancer in southern Louisiana, than in other occupations (Correa and Johnson 1983). Overall, however, occupational exposures are considered to account for only about 4% of all cancers (Correa and Johnson 1983).
Males appear to be more susceptible to lung cancer than females. In a retrospective study of 1939 squamous cell carcinoma, small cell carcinoma and adenocarcinoma patients, who were diagnosed in Canada between 1979 and 1986, men outnumbered women in the ratios of 5.4:1, 3.0:1, and 1.6:1 for each of the three types respectively. The excess of lung cancer among males relative to females could not be attributed either to the 7-15% higher prevalence of smoking among males or to the fact that males were likely to have smoked more cigarettes per day than females.

The relative incidence of the various histologic types of lung cancer varies from one geographic area to another but overall squamous-cell carcinoma has been the most common type, followed by either small-cell carcinoma or adenocarcinoma and finally large-cell carcinoma (Dodds, Davis and Polissar 1986). The prevalence of the strongly smoking-associated squamous-cell carcinoma has been decreasing in recent years, however, and that of adenocarcinoma has been increasing (Rothschild 1982).

Some cell types of lung cancer, adenocarcinoma for instance, are less strongly associated with smoking than others. In the Canadian study, 14% of patients with adenocarcinoma (3.8% for men and 28% for women) had never smoked, while <3% of those with squamous cell or small cell carcinomas had been lifetime non-smokers. Among men in this study, cell type was not related to duration or intensity of
smoking as expressed in pack-years, but among women, those with small cell tumors had experienced exposure in pack-years which was significantly higher than that of women with the other two histologic types (McDuffie, Klaassen, and Dosman 1990). Some studies have documented an increasing proportion of adenocarcinoma over time among diagnosed lung cancers (Dodds, Davis, and Polissar 1986). This might partially be due to the reduction in incidence of the more strongly smoking-related cancers as the prevalence of smoking declines in that population. Another group has observed that among cases of lung cancer first diagnosed at autopsy, the male-to-female ratio is lower (24:19) and non-smokers are more common (30% of 40 cases compared to 8% of 107 cases) than among cases diagnosed prior to death (McFarlane, Feinstein, and Wells 1986). These findings may indicate both that lung cancer among non-smokers is under diagnosed, and that important etiologic factors other than smoking are being overlooked. Maleness as a susceptibility factor, the relative increase in adenocarcinoma of the lung with respect to other cell types, and the incidence of lung cancers in non-smokers, all have eluded explanation.

**Genotype and lung cancer.**

An excess of lung cancer among relatives of lung cancer patients has been reported, even after correction for smoking (Tokuhata and Lilienfeld 1963). Lynch et al. (1986) observed that the cumulative risk (CR) for relatives of lung
cancer patients for developing any type of cancer was elevated over that of the general population and that the CR of getting non-smoking-related cancers was higher than that of getting smoking-associated cancers. Only relatives of those with adenocarcinoma of the lung (as opposed to other histologic types of lung cancer) had more lung cancer per se (Lynch et al. 1986). The relatives of probands who had multiple primary tumors had significantly more cancer than relatives of those with a single cancer.

The close relatives of lung cancer patients also have been observed to have higher rates of impaired lung function and chronic obstructive pulmonary disease than those of patients with non smoking-related cancers, even when never-smokers were excluded from the analysis and variables were included to adjust for the effects of smoking (Lynch and Lynch 1989). A heritable weakness in the pulmonary system which predisposes individuals to chronic obstructive lung disease, might also render them more susceptible to lung cancer.

A study of the first degree relatives (i.e. parents, full siblings and children) of 337 white lung cancer patients in Louisiana showed a greater overall risk for cancer compared to families of the probands' spouses. The risk remained significant after adjustment for age, sex, cigarette smoking, and occupational/industrial exposure. The same proportions of relatives of the probands and their
spouses were smokers, but the relatives of probands were 1.4 times as likely to smoke 2 or more packs per day and 1.3 times as likely to have accumulated 60 or more pack-years of exposure. Increased risk was found among families of probands both for lung cancer and for cancers of the larynx, brain and nervous system, bone, endocrine glands, ovary, kidney, bladder, esophagus and stomach and leukemias-lymphomas (Ooi et al. 1986). Furthermore the crude odds ratios for having multiple relatives with cancer increased: $\text{OR}_1 = 1.67$, $\text{OR}_2 = 2.16$, $\text{OR}_3 = 3.66$, and $\text{OR}_{4\text{&more}} = 5.04$ (Sellers et al. 1987).

A Canadian study showed increased risk among first-degree relatives of lung cancer patients for having 2 or more persons within the relationship affected with cancer of any type, for having a multigenerational pattern of cancer, and for having multiple primary cancers within individuals (McDuffie 1991). Most of the excess cancers among patients' relatives were smoking-related and interviews of siblings of patients and controls showed that 68% of patients' siblings smoked while only 47% of controls' did. Therefore in this study, family history is confounded with smoking and no attempt was made to partition the risk. A trend of increasing odds ratios for having multiple relatives with cancer, similar to that in the Louisiana study, was also found here.
Studies are also underway to assess the role of the p53 gene in carcinogenesis. P53 codes for a protein which regulates transcription in cell nuclei and is thought to restrain cell growth and division. Dysfunctional, mutated forms of this gene have been found to be associated with various forms of cancer, including lung cancer (Lewis 1991; Takahashi et al. 1989).

The frequency of sister chromatid exchanges (SCE) in cultured lymphocytes is also thought to be a heritable susceptibility marker for cancer. SCE were more frequent in the lymphocytes of lung cancer patients than they were among controls. When controls were divided by family history of cancer into high-risk and low-risk groups, the level of SCE among the high risk individuals was indistinguishable from that of patients. Both patients and high-risk controls responded with increased SCE when their lymphocytes were cultured with benzo-a-pyrene compared to low-risk controls (Lynch and Lynch 1989). High levels of SCE could be due to enhanced activation of carcinogens, defective repair mechanisms for carcinogen-damaged DNA, chromosomal instability, or combinations of the three.

Genetic differences in the ability to metabolize carcinogens, specifically in the balance of activation to detoxification pathways has been proposed as one reasonable biogenetic mechanism for variance in susceptibility to cancer (Harris et al. 1980). Of particular interest are the
cytochromes P450, a family of enzymes responsible for metabolizing xenobiotics. Their genes are polymorphic; individuals may vary in their ability to metabolize xenobiotics by as much as a factor of 100. Several studies have found that high inducibility of aryl hydrocarbon hydroxylase, the P450 enzyme which activates members of the PAH family, is more prevalent among lung cancer patients than among controls (Lynch and Lynch 1989). It has also been established that P450db1 is the specific human isozyme responsible for metabolizing the anti-hypertensive drug debrisoquine and that persons who are high metabolizers of debrisoquine have an enhanced risk for developing lung cancer. If these people are also occupationally exposed to carcinogens, the two factors interact synergistically, resulting in a substantially higher risk than that expected if the effects were additive (Caporaso et al. 1989). Thus there is specific biochemical evidence for genes which influence susceptibility to lung cancer by enhancing the metabolism (activation) of carcinogens, inhibiting DNA repair mechanisms, or disrupting normal regulation of cell growth and division.

While it is true that most lung cancer patients have been smokers, and that smokers have 25 times the risk of non-smokers for developing lung cancer (Subar, Harlan, and Mattson 1990), perhaps no more than 10-15% of smokers will develop lung cancer. What it is that is different about
those who smoke but do not get lung cancer is not known. Either there is an element of the causal chain of events which has not yet been identified, or there is a protective mechanism which most but not all smokers possess.

Campaigns to educate the public to the dangers of smoking and various programs have been established to help smokers break the habit. Among the younger age groups in several countries lung cancer rates are leveling off or even decreasing (Kurihara 1987). It is improbable, however, that the practice of smoking cigarettes will diminish to the point where it is of negligible public health concern (Pierce et al. 1989). Although prevalence of smoking has declined since 1965 when 51.1% of males and 33.3% of females reported themselves to be smokers, the proportion of smokers who smoke more than 25 cigarettes per day increased from 13.7% in 1965 to 22.4% in 1979 (Loeb et al. 1984). The heavier smokers tend not to relinquish the habit. Therefore it would be valuable to be able to discriminate the high-risk smoker from the low risk smoker for the purpose of earlier diagnosis as well as to provide additional motivation for quitting. In addition, the determinants, other than smoking, of risk for lung cancer, particularly adenocarcinoma, should be identified since rates of adenocarcinoma appear to be increasing (Kurihara 1987). Such an indicator of risk would presumably be a construct of multiple behavioral and physiologic traits.
**Diet and cancer.**

It has long been recognized that diet affects one's health. Current media debates of the effects of diet on health, are perhaps most prominently focussed on heart disease, but much consideration has been given to dietary factors which may influence resistance to cancer as well. **On the basis of many studies of the relationship of diet to susceptibility to cancer, it has been estimated that as much as 30% of all cancers are due to dietary factors (Doll and Peto 1981).** Particular foods and/or nutrients have been found to be associated with cancers of specific tissues, for example, cruciferous vegetables inversely with colon cancer, salted and smoked foods with stomach cancer, and alcoholic beverages with cancer of the larynx. Vitamins A and C have been most often identified as protective against lung cancer although results have not been consistent. Diets high in protective factors may be one reason that some smokers avoid lung cancer. If diet could also be linked to some physiologic element of the body's defense system, such as some parameter of immune status, the hypothesis of an important role for diet in cancer prevention would be strengthened and better defined.

"In a dozen case-control and cohort studies, high intake of fruits and vegetables containing carotenoids has been associated with a reduced risk of lung cancer," (Willet 1990). At first it was thought that vitamin A was
responsible for increased resistance to lung cancer. Some animal studies demonstrated that vitamin A could reverse metaplastic changes in cells and promote differentiation to maturity, protect animals from acquiring tumors, and inhibit the progression of already induced cancers under some conditions. Unfortunately, these findings were not entirely consistent.

Vitamin A is available to humans either as the preformed retinyl esters from animal sources or as carotenoids derived from green leafy vegetables or green and yellow vegetables and fruits. Certain of the carotenoids, specifically α-carotene, β-carotene, and cryptoxanthin, are metabolized to vitamin A, primarily in the intestine but perhaps also in the liver (Peto et al. 1981). A few early studies, in which plasma drawn before the diagnosis of lung cancer was analyzed for retinol levels, indicated a protective effect for plasma retinol, but these results were not verified by further studies. Instead it was a carotenoid which emerged as the protective factor and not preformed vitamin A (Ziegler et al. 1984). It was found that higher levels of plasma β-carotene were protective, but there was no difference in the pre-diagnostic levels of retinol or carotenoids other than β-carotene. That a carotenoid rather than retinol is apparently the protective factor is, in fact, more reasonable because plasma levels of retinols are tightly regulated by homeostatic mechanisms and
are not influenced much by increased intake. Levels of carotenoids can vary widely and directly reflect levels of intake.

Differences in intakes of fruits and vegetables among residents of northern and southern parishes of Louisiana were correlated with higher rates of lung cancer in the southern region of the state. This lifestyles study in Louisiana also found an inverse relationship between dietary carotene intake and squamous-cell and small-cell carcinomas, but not for adenocarcinoma. It was retinol intake which was inversely associated with adenocarcinoma of the lung (Correa et al. 1988). However a 1981-82 study in Los Angeles County involving 149 cases of adenocarcinoma in women provided evidence for a strongly protective effect of β-carotene against adenocarcinoma (Wu, et al. 1985).

The mechanisms through which β-carotene would act to protect an individual from neoplastic disease are not known. Physiologic and chemical effects are both possible. Carotenoids are known to be efficient quenchers of singlet oxygen and β-carotene, in particular, is a good scavenger for free radicals (Burton and Ingold 1984). β-Carotene protects mice from psoralen-induced phototoxicity by quenching intermediate active species (Giles, Wamer, and Kornhauser 1985). It was also found to reduce chromosomal breaks in the bone marrow of mice exposed to benzo-α-pyrene and mitomycin C (Raj and Katz 1985). This ability to
deactivate certain intermediate active species could be directly protective against cancer initiation. Chewers of the combination of betel nut, lime, and tobacco are at high risk for oral cancer and micronucleated cells are characteristic of their buccal mucosa. β-Carotene was shown to be effective in preventing the formation of such micronuclei among chewers in the Philippines, (Stich et al. 1984).

Enhancement of immune defenses by β-carotene has also been postulated. There was a rise in the number of CD4⁺ lymphocytes and total CD3⁺ lymphocytes but not of CD8⁺ lymphocytes per mm³ in human blood following daily, oral, 180-mg doses of β-carotene (Alexander, Newmark, and Miller 1985). Such changes in analogous subsets have also been observed in mice. Mice supplemented with vitamin A or β-carotene were also able to reject incompatible skin grafts more rapidly than controls and in other experiments, the growth of tumors was slowed with supplementation (Alexander, Newmark, and Miller 1985). Recent studies suggest that cigarette smoking itself lowers plasma levels of β-carotene independently of reported intake of green and yellow vegetables (Willet 1990). The association of plasma levels of β-carotene with lung cancer is therefore confounded with smoking. Risk of lung cancer, adjusted for smoking, among individuals with low intakes and low serum levels of vitamin A and carotenoids was 1.3 to 2.7 times
that of persons with high intakes and high serum levels. Among smokers the enhancement of the risk of lung cancer when there is inadequate intake of carotenoids/vitamin A could range from 1.3 to 8.1 times. (Subar, Harlan, and Mattson 1990).

β-Carotene and vitamin A may not be the only constituents of fruits and vegetables which are protective. Vitamin C intake has been found to be inversely related to some cancers, but the evidence with respect to lung cancer has been mixed. An inverse gradient was found in Louisiana between vitamin C consumption and squamous- and small-cell cancers (Fontham et al. 1988). The authors suggest that the effect might be significant among Louisianians, when it has not been found to be significant in other studies, because a national study has shown that dietary intakes of vitamin C are particularly low in Louisiana compared to the rest of the country. Fifteen percent of low-income Louisianians and 9% of higher income residents were found to be low or deficient in serum levels of vitamin C (Ten-State Nutrition Survey 1968-70). Dietary data from the Bogalusa Heart Study suggests that Louisiana children do not ingest the recommended daily allowance (RDA) of vitamins A and C (Correa et al. 1988). Of the children surveyed, 25-50% did not get the RDA of vitamin A and 15-36% reported intakes of less than 1/3 the RDA of vitamin C. Thus low intake of these important nutrients prevails from an early age.
Some studies have also found that folate might be protective against lung and cervical cancers. Low folate levels are associated with cervical dysplasia and supplementation with folate tends to reverse the dysplastic process (Butterworth et al. 1982). Organic nitrites, nitrous oxide, cyanates, and isocyanates in cigarette smoke destroy vitamin B$_{12}$ and folate, resulting in reduced levels of both vitamins in peripheral blood of smokers. Smokers with potentially premalignant bronchial squamous metaplasia had lower levels of both folate and vitamin B$_{12}$ than smokers with normal sputum cytology. In a preliminary, randomized, double-blind trial supplementation with 10 mg of folate and 500 µg of hydroxocoalbumin (vitamin B$_{12}$) for four months caused improvement in the degree of metaplasia among those so supplemented compared to the group who received the placebo (Heimburger et al. 1988). Vegetables are a major source of folate.

Another dietary association which has appeared in some studies of lung cancer is the inverse association of lung cancer with the consumption of dairy products and eggs. A gradient effect over quartiles of intake relative to the highest intake was observed although the increased risk ratio was significant only for the lowest two quartiles of intake (Wu et al. 1985).

A Norwegian group of investigators found an intriguing interaction between alcohol intake and vitamin A index in
estimating the risk of lung cancer. For those with low vitamin A intake, persons with a high alcohol intake had 3.7-fold higher odds for developing lung cancer compared to non-drinkers, while those who had a high vitamin A intake and a high alcohol intake had one-fifth the odds of non-drinkers. The finding is plausible on the basis of animal studies which showed that animals who were fed ethanol regularly maintained higher plasma and tissue levels of vitamin A but lower liver stores than control animals (Kvale, Bjelke, and Gart 1983). In a Japanese study, however, those who drank alcohol were found to have about 30% lower serum levels of β-carotene than non-drinkers (Shibata et al. 1989; Stryker et al. 1988). In a regression of serum β-carotene on age and on dummy variables for categories of intake of green and yellow vegetables, drinking, smoking, and city of residence, drinking had the highest estimated coefficient of the categorical variables.

If drinking does indeed affect the protective effect of β-carotene or vitamin A against cancer, failure to look for and to control for this relationship may have led to some of the inconsistencies in results in previous studies of vitamin A and lung cancer risk.

Smoking and drinking are independent behaviors although use of cigarettes is commonly associated with use of alcohol. Smokers are more likely to drink beer regularly and to consume larger quantities of all alcoholic beverages
on any particular occasion (Centers for Disease Control 1989). Some studies estimate that 7% of the adult population are alcoholics and over 80% of these smoke cigarettes as well (Difranza and Guerrera 1989). It was suggested above that alcohol intake might potentiate a protective effect of high vitamin A intakes and at the same time worsen the risk associated with inadequate intake. Although drinking is known to lower plasma levels of β-carotene, perhaps the plasma and tissue levels of vitamin A are enhanced with high levels of ethanol intake. This relationship deserves further investigation. Furthermore the effects of alcohol intake on the immune system have not been reported and might have relevance for the risk of cancer.

There is recent evidence from the Second National Health and Nutrition Examination Survey (NHANES II) that smokers do not consume as nutritious a diet as non-smokers. Smokers tend to eat fewer fruits, vegetables, and high-fiber grains and therefore less vitamin C, folate, fiber, and vitamin A than non-smokers and intake varies inversely with smoking intensity. Interestingly however, former smokers had eating patterns which were no different than those of never smokers (Subar, Harlan, and Mattson 1990). The inverse association of vitamin C intake with increasing smoking intensity was subsequently verified in data from two other epidemiologic studies. Ex-smokers had intakes
intermediate between current smokers and non-smokers (Gridley, McLaughlin, and Blot 1990). If it is true that former smokers revert to healthier eating patterns when they stop smoking, this would imply that smoking has effects upon appetite and that this is an important component of the mechanism by which smoking increases the chances of developing not only lung cancer, but other smoking-related cancers as well.

**Methods of obtaining information on diet.**

Dietary information is obtained from interviews or questionnaires completed by the subjects. The question of the validity of food-frequency instruments has received a good deal of attention (Chu et al. 1984; Willett et al. 1985). The issue of what should serve as the "gold standard" in judging the quality of a particular method is the major problem. Comparisons of food frequency data to intakes recorded in a diet diary or to the estimates made by a third person such as a spouse suffer from the likelihood that errors in the responses could be correlated to an unknown extent. For instance subjects and spouses may both over- or under-estimate the use of a certain food. More objective standards are those based on a biochemical or physiologic response, e.g., blood levels of a particular nutrient (Willett et al. 1983). Even these, however, do not necessarily fluctuate in proportion to the intake of foods containing that substance. Individuals vary in their
digestive and absorptive processes. Absorption of one nutrient sometimes depends upon what other foods are consumed with it. The blood levels of certain nutrients, like vitamin A, are regulated within narrow limits by homeostatic mechanisms and correlate with intake only at the extremes of deficiency or toxic surplus. Finally, genetic differences can affect the individual's usual blood level of a substance such as cholesterol.

Diet recall methods suffer from the inability of subjects to remember accurately just what he ate more than 1 or 2 days previously. Over that short interval, there is a good deal of intraindividual variation such that the last day's intake might not be representative of his usual intake (Sempos et al. 1985). A 7-day diet record might be a more accurate method but requires a substantial commitment on the part of the respondent and therefore reduces participation rates to the point where representativeness of the sample is questionable.

This problem of validity of dietary data assessment must also be considered in the context of the uncertainties in other, more-accepted strategies of measurement of physical characteristics. For example, single measurements of blood pressure are notoriously imprecise, yet we consider them to be valid and meaningful estimates. Furthermore, it is important to be clear about the objectives of measurement. The goal of a clinician is usually validity
for the individual. In population studies, however, the reliability of the assessment for each individual is not so important as that of the mean intake for each subgroup of individuals. Additionally, it is of little importance if the estimates are biased in one direction or the other, so long as the subgroups of interest can be ranked in categories of low, medium, and high intake (Block 1982).

Food frequency questionnaires have been found to be a reasonable if not a perfect method for gathering useful dietary data which has validity on the group level. They are easily self-administered and therefore practical and cost efficient methods for large studies. They yield estimates which correlate well with diet histories and diet records (Block 1982), and permit the useful grouping of subjects into categories by level of intake.

**Tumor immunity.**

The role of the immune system in defense against neoplastic disease is a complex issue which is still far from resolution. Both non-specific and specific immune responses probably participate (Ada 1981). Strongly antigenic tumors in mice can induce a specific cell-mediated immune response, but most tumors in humans are thought to be only weakly antigenic (Stutman 1975; Melief, Cornelis, and Schwartz 1975; Doherty, Knowles, and Wettstein 1984). Studies of immunosuppressed transplant patients or those with immune deficiencies show that only certain kinds of
tumors are increased. Specifically these are non-Hodgkins lymphomas, hepatocarcinoma, cervical carcinoma, soft tissue sarcomas, and skin cancers, all of which have been tentatively linked to viruses (Matas, Simmons, and Najarian 1975; Purtilo, et al. 1984). In contrast, the common cancers: lung, gastrointestinal, prostate, and breast are not increased. This suggests that in humans only neoplasia associated with viral infection would be combated by a specific, antigen-driven immune response (Kinlen 1985). However, culture of pleural-effusion lymphocytes with mitomycin-C-treated autologous tumor cells has been used successfully to generate cytotoxic T-cell effectors which were specific for the host's lung tumor, but inactive against allogeneic targets (Üchida and Moore 1985). This finding indicates that antigen-specific immune responses to lung neoplasia may be important but that they are either suppressed or difficult to measure with current assay systems.

Several leukocyte types have been implicated as defenders against malignant cells including macrophages, cytotoxic T-lymphocytes (CTL) with antigen specificity, natural killer cells (NK; also called "large granular lymphocytes"), and another lymphocyte called a lymphokine-activated killer (LAK; Rosenberg and Mule 1985) which may overlap the NK population. NK are lymphocytes which are defined by their morphology and their ability to lyse
cultured tumor cells (specifically cells of the human myeloid leukemia cell line K562) in a non-MHC-specific manner. In rats NK cells were shown to be the major effectors in limiting the development of metastatic lung tumors when the animals were challenged intravenously with tumor cells (Weissler et al. 1987). Unlike NK, LAK can lyse freshly isolated, but not cultured, tumor cells in vitro. The precursor of LAK is a null cell but is not a NK because anti-CD2 + C' destroys cells with NK activity but not cells which differentiate to LAK in response to IL2. The phenotype of the effector is CD3+CD8+ but like NK, its action is not MHC-restricted (Grimm et al. 1982). Regression in the tumor burdens of some humans treated with LAK and/or interleukin 2 (IL2) demonstrates that these components of the human immune system are able to inhibit and even reverse tumor progression (Rosenberg and Lotze 1986; Rosenberg et al. 1987).

There are numerous reports in the literature that present evidence of immune suppression among patients with cancer. Some immune responses seem to be correlated with disease progression and others not. Lymphoblastogenesis assays with mitogens concanavalin A (ConA) and poke weed mitogen (PWM) appear to correlate with prognosis for lung cancer patients (Watanabe et al. 1990). Assay of NK activity in peripheral blood lymphocytes (PBL) was less definitively correlated. Peripheral blood monocytes from 8
of 31 lung-cancer patients had reduced cytocidal activity toward cultured tumor cells compared to monocytes from healthy controls (Nakata et al. 1985). Patients' monocytes showed reduced ability to upgrade complement receptor expression and the degree of inhibition was proportional to the spread of disease. Although valuable for estimating extent of disease and therefore prognosis for survival, these assays are probably not relevant for susceptibility evaluation, early diagnosis, or treatment.

Tumor infiltrating lymphocytes (TIL) function differently from PBL in the autologous mixed lymphocyte reaction. PBL from lung cancer patients responded poorly, while pleural effusions of lung cancer patients contained enough responding cells to give results comparable to the blood of healthy donors (Uchida and Micksche 1982). TIL from patients with metastatic melanoma matured into tumor-specific cytotoxic lymphocytes (LAK ?) and not NK upon culture with recombinant IL2 (rIL2), whereas PBL required repeated culture with rIL2 and autologous tumor cells to develop tumor-specific effectors, and non-specific effectors were produced as well. TIL from patients with sarcomas, lung cancers, renal cell carcinomas and carcinomas of the head and neck were reported to have MHC-nonrestricted CTL activity rather than autologous tumor cell-specific CTL activity after culture with rIL2 (Itoh, Platsoucas, and Balch 1988). But TIL from patients with lung adenomas and
squamous cell and large-cell carcinomas of the lung, after culture with IL2 possessed both non-specific and specific antitumor cytolytic potential (Rabinowich et al. 1987).

Typically, TIL were found to be inactive when freshly isolated from the tumor but after culture with rIL2 demonstrated tumor-cell cytolytic capacity. One subset of TIL which was originally negative for CD3 and CD8, matured in the presence of rIL2 to a CD8+ T-cell. These tumor-derived cells had greater activity against autologous tumor cells than similarly cultured CD8+ cells from PBL or lymphocytes isolated from normal lung tissue (Rabinowich et al. 1987). This evidence, then, supports the thought that the host does mount a specific immune response to a tumor and that tumor-specific effector cells are either generated within the tumor or home there. However, the response is purportedly suppressed at the tumor site, perhaps by tumor-secreted factors. Since the tumor-specific effectors home to the tumor compartment and are not present in the peripheral blood, they would not be useful for diagnostic or predictive purposes unless attraction of specific subsets of cells to the tumor site produces an alteration of their prevalence in the accessible peripheral blood.

The levels of NK activity in peripheral blood (Uchida and Micksche 1981; Weissler, Nicod, and Toews 1987) and lung tissues of cancer patients have been investigated but without clear trends. The results obtained seemed to depend
upon methods by which the cells to be evaluated were isolated and purified. Studies of cells obtained by lavage (Berman et al. 1990; Pitchenik, Guffee, and Stein-Streilein 1987) and by extraction from minced lung tissue (Bordignon et al. 1982b; Deshazo et al. 1987; Weissler, Nicod, and Toews 1987) indicate that monocytes can elaborate suppressor factors which diminish natural killer activity in the lungs (Uchida and Micksche 1983; Weissler, Nicod, and Toews 1987). Since the effect of pulmonary macrophages can be abolished by indomethacin, Weissler postulated that prostaglandins or other products of the arachidonic acid cascade could be the inhibiting factor. The tumor itself might also produce suppressor factors (North and Awwad 1990; Nakata et al. 1985, Hakim 1988); suppression of responsiveness of NK correlates with stage of the disease (presumably tumor burden). A serum-borne factor from the blood of bronchogenic cancer patients blocked the activation of PBL LAK of both patients and of normal controls with IL2 in vitro, but even patients' lymphocytes could respond if human serum was replaced with fetal calf serum (Dunlap et al. 1990).

It would seem then that there is considerable evidence that there are immune defenses against neoplasia and that tumor-specific and non-specific cytotoxic lymphocytes exist and accumulate at the tumor site. There also seem to be mechanisms for up-regulating and down-regulating responses.
The establishment of neoplasia probably indicates a state of immune suppression mediated by tumor-secreted factors, factors secreted by host leukocytes responding to other stimuli, and in the case of lung cancer, by exposure to components of tobacco smoke. The complexity of the system and the impossibility of reproducing the host environment in vitro has made it difficult to delineate the operative mechanisms.

**Smoking and the immune system.**

Smoking perturbs the immune system. The active component is unknown; approximately 6000 different chemicals have been isolated from cigarette smoke. A recent review summarizes most of the known effects in animals and humans of both mainstream and environmental smoke (Johnson et al. 1990). Light-to-moderate smokers had 25% higher total leukocyte counts than non-smokers, but had normal to slightly elevated ratios of helper (CD4) to suppressor (CD8) subsets, while heavy smokers had a low CD4⁺/CD8⁺ ratio (Ginns et al. 1982a; Petitti and Kipp 1986). Leukocyte counts were increased in smokers, especially in heavy (>30 cigarettes per day) smokers who inhaled (Johnson et al. 1990), but the proportion of leukocytes which were lymphocytes was not different from that in non-smokers (Miller et al. 1982). No significant age or sex effects were found in the leukocyte parameters measured by Miller, but the total population was only 95 men and women who ranged in age from
22 to 72 and there may not have been sufficient numbers to find differences. Tollerud and coworkers (1989a) found reduced numbers and percentage of Leu-11a+ (labels Fc receptor) cells (presumably natural killer cells) in both smokers and ex-smokers. He also found increased numbers of all leukocytes, a slightly higher proportion of eosinophils, and a higher CD4+CD8+ ratio among smokers compared to non-smokers. The percentage of CD4+ cells tended to increase among smokers with the number of cigarettes smoked per day (Tollerud et al. 1989b).

Natural killer activity in smokers, expressed as the percent specific 51Cr released from target cells, was 13.7 +/- 1.6 compared to 29.0 +/- 3.0% in non-smoking controls (Ginns et al. 1985). Effector cells were peripheral blood lymphocytes isolated by Ficoll-Hypaque fractionation. The cause of this deficit could have been reduced numbers of NK among PBL or suppression by regulatory lymphocytes or monocytes. Metastasis in melanoma patients has been observed to be inversely associated with NK activity among PBL. Furthermore among persons who have had a melanoma excised, smokers have a higher incidence of blood-borne metastasis than non-smokers (Burton 1983a), suggesting that this was due to reduced NK activity in the blood of smokers.

Smokers average lower levels of IgG, IgA, and IgM and higher levels of IgE than non-smokers, but no dose-response relationship could be detected. Smokers responded as well
as non-smokers to vaccination with A₂ influenza but unprotected smokers were more susceptible to infection than non-smokers. In general, smokers were more susceptible to infections of the lungs and of the urinary tract than non-smokers (Johnson et al. 1990).

The effects of smoking on the immune system would appear to be reversible within a relatively short time. After cessation of smoking, NK activity, IgM and IgG levels, and leukocyte counts in 35 subjects returned to levels of non-smokers within 3 months, while the levels in 29 subjects who continued to smoke did not change (Hersey, Pendergast, and Edwards 1983). An earlier study (Miller et al. 1982) found that relative percentages of CD8⁺ T-cells and the CD4⁺+CD8⁺ ratio returned to normal among heavy smokers within 6 weeks of smoking cessation. These findings imply that any susceptibility factor, which is associated with the effects of smoking on the immune system, relates to one of the earlier steps in the carcinogenic process, since increased risk for developing cancer persists for a decade or more after one stops smoking.

**Lung cancer and leukocyte subsets.**

Lung cancer is a disease which is highly fatal, mostly because it is not diagnosed in the early, localized stage. What is needed is a practical, non-invasive test which can identify persons at high risk for, or in the earliest stages
of lung cancer. As monoclonal antibodies to leukocyte-subset-specific surface molecules have become available and more is learned about the function of lymphocyte subsets, interest has been generated in whether relative proportions of various subsets of lymphocytes in peripheral blood have a meaningful relationship to disease or to susceptibility for disease. The most well-known such link is the correlation between CD4\(^+\):CD8\(^+\), helper:suppressor ratios with clinical acquired immune deficiency syndrome (AIDS). Is there a lymphocyte profile which correlates with vulnerability to lung cancer or with early stages of the disease?

Studies relating prevalence of specific subsets of lymphocytes to malignant disease states date back to the 1970s. It is difficult to compare the results of these early studies, however, because (1) some were done by fluorescence microscopy and others by flow cytometry, (2) different antibodies were used to define the subset of interest, (3) the study subjects varied in characteristics from study to study, for example age, sex, smoking, treated vs untreated and stage of disease among patients, and (4) results were calculated as absolute counts/mm\(^3\) in some studies and as relative proportions in others. The difficulty with consideration of absolute counts/mm\(^3\) is that the variability of blood volume among individuals is introduced into the estimate. Counts may vary with state of
hydration, leukocyte proportion, and lymphocyte proportion. Each subset as a percentage of total lymphocytes would theoretically have much better among-subjects comparability.

The studies of the mid-1970s used the presence of the receptor for sheep erythrocytes (CD2) to define T-cells, whereas today CD3 is considered to be the pan-T-cell marker and CD3 CD2\(^+\) NK are not regarded as T-cells. A deficit in CD2\(^+\) lymphocytes was noted among cancer patients, including lung cancer patients, in several of these early studies (Dellon et al., 1975; Kaszubowski et al. 1980). Dellon also observed a trend toward decreased numbers of CD2\(^+\) lymphocytes with severity of disease (control, localized, regional, metastatic). The differences of means among groups were statistically significant at P < 0.05 for squamous cell, oat cell and undifferentiated carcinomas and it is likely that a test for linear trend would have proved significant for the adenocarcinoma group as well.

As more subset-specific antibodies became available, a deficiency in percentage of CD4\(^+\) helper cells and an increase in percentage of CD8\(^+\) suppressor/cytotoxic cells and major histocompatibility class II molecule-bearing cells (HLA-DR\(^+\)) among advanced cancer patients was reported (Dillman et al. 1984). Similarly decreased CD4\(^+\) cells and increased CD8\(^+\) and HLA-DR\(^+\) cells were found among CD2\(^+\) lymphocytes in patients with chronic myelocytic leukemia; the percentage of HLA-DR\(^+\) (activated) T-cells was highly
correlated with percent CD8+ (Velardi et al. 1984). Velardi also observed that almost 14% of CD3+ lymphocytes were neither CD4+ nor CD8+. In fact, these cells might have been low-density CD8+ and not observed as positive under fluorescence microscopy. Estimation of CD8+ lymphocytes by microscopy gave values which were consistently lower than estimates on the same samples by flow cytometry (Kreuzfelder et al. 1987).

Balch reported deficits in percentages of HNK-1+ (CD57+) cells, some of which are NK, among 22 lung cancer patients as well as patients with melanoma and cancers of the breast, head and neck, and colon (Balch et al. 1982). Balch controlled for the effects of age and sex in his study and found that HNK-1+ cells were more prevalent in the older age groups and among men compared to women.

Ginns and his associates were the first investigators to control for smoking in their analyses. Persons with metastatic lung tumors displayed reduced percentages of CD3+ and CD4+ T-cells and a reduced CD4+÷CD8+ ratio. Percentage of CD3+ and CD8+ lymphocytes tended to increase and CD4+ cells to decrease with intensity of smoking. After adjustment for smoking, the profiles of persons with squamous cell carcinoma (n=10) were not different from normal smokers and patients with adenocarcinoma (n=7) had decreased percentages of CD8+ cells and an increased CD4+÷CD8+ ratio (Ginns et al. 1982b). Ginns's report was based on only 17
cases of primary lung cancer (Ginns et al. 1982a), however, and there may not have been enough power in those numbers to detect other differences which did, in fact, exist. An attempt to correlate subset percentages with stage of disease also was not definitive, probably because of the small number of patients (21, including metastatic lung tumors). There were no linear trends with three-group comparisons (control, limited disease, advanced disease) whether stage, lymph node involvement, or tumor burden were the criteria (Ginns, Rogol, and Murphy 1983).

Other studies may have shown negative results because of size as well. Eleven patients with bronchogenic carcinoma were compared with 10 young non-smokers, 12 older non-smokers and 11 smokers. No differences were seen in OKT11+ T-cells (CD2, sheep erythrocyte receptor), CD4+, and CD8+ subsets of T-cells, NKH-1+ cells, the CD29+ subset of CD4+ cells, or the CD4+CD8+ ratio, but the CD45R+ subset of CD4+ cells was reduced in the cancer patients (Kratikanont et al. 1987). No adjustment for age, sex, time of sampling, or smoking levels was done. The mean age of cancer patients (62.8) was considerably older than that of the other groups: 30.9, 55.8, and 48.0, respectively.

Dellon suggests that low T-lymphocyte counts might antecede squamous cell carcinoma of the lung, thus being a susceptibility marker (Dellon et al. 1975). Or it could be that diminished numbers of lymphocytes, in consequence of
the disease, might precede other signs and symptoms of malignancy and allow earlier diagnosis. With the present repertoire of monoclonal antibodies specific for markers which define functional subsets of lymphocytes and the highly developed technology of flow cytometry (Hoffman et al. 1980), it is now feasible to define mononuclear cell profiles in humans more extensively and precisely than was possible in some of previous studies. Furthermore, we are aware that it is important to control or adjust for such factors as age, sex, smoking, and even time that the sample was drawn (Levi et al. 1985; Ritchie et al., 1983). Failure to take such factors into consideration probably explains many of the conflicting results in previous studies and inadequate sample size was responsible for inability to detect effects in some instances.

Flow cytometry.

The technology of flow cytometry continues to evolve as of this writing. It had its beginnings in work by Moldavan in 1934 and Gucker et al. in 1947, but it was Wallace Coulter in 1956 who developed the first useful instrument which could differentiate and count blood cells based on size (Landay, Ohlsson-Wilhelm and Giorgi 1990). The Herzenbergs at Stanford designed an instrument which could separate cells stained with fluorescence-labelled antibodies. Their 1972 version incorporated an argon laser and was subsequently developed by Becton Dickinson
Immunocytometry Systems (BD) into the 1986 FACS 440 used in this study (Landay, Ohlsson-Wilhelm and Giorgi 1990). Improved, more versatile, and more user-friendly fluorescence-activated cell sorters (FACS) continue to be developed.

Concise descriptions of the mechanics of cell sorting and analysis is given by Landay (Landay, Ohlsson-Wilhelm and Giorgi 1990), Jackson (Jackson and Warner 1986), and Haynes (Haynes 1988). Briefly, it is accomplished by passing a suspension of fluorochrome-labelled cells through a small orifice, such that the cells pass one at a time through the laser beam. Light is scattered at small angles which depend upon the size of the cell and at wide angles (90°) in a manner that is characteristic of the internal structure or granularity of the cell. If the cell bears one or more specific markers labelled by fluorescein- (green-) or phycoerythrin- (red-) tagged ligands (usually antibodies), light energy is absorbed as the cell passes through the laser beam and light is emitted at wavelengths characteristic of the fluorochromes used. The scattered light and fluorescence emissions are detected and measured using an optical system composed of selected light filters, photodiodes, and photomultiplier tubes. Signals are then amplified, digitized and stored by a computer for further processing.
An instrument with a single laser is able to measure up to four attributes of each cell: size, granularity, and the presence (or absence) and surface density of up to two cell-surface antigens. Forward scatter (size) and side scatter (granularity) can be used to discriminate lymphocytes from granulocytes, monocytes, and erythrocytes. Then with appropriate antibodies to cell-surface molecules which distinguish subsets of lymphocytes from one another in terms of their function, one can estimate the relative abundance of functional subsets of interest among lymphocytes from the peripheral blood.

Problems have arisen when flow cytometric data from one laboratory was compared to that of another (Ault 1988). Differences in instruments from different manufacturers and the difficulty of maintaining the same analytical environment from day to day using the same instrument induce variability. Handling of samples and methods of labelling also affect the results obtained (Ashmore, Shopp, and Edwards 1989; Green and Stelzer 1988; Landay and Muirhead 1989). Cell surface differentiation antigens in human blood were found to be stable for up to three days when the whole blood was stored at room temperature (Shield et al. 1983), but when stored at 4°C overnight, reactivity toward T11, OKT3, and OKT4 antibodies was decreased while levels of OKT8 remained the same (Weiblen, Debell and Valeri 1983). Bongers and Bertrams (1984) report good stability if
lymphocytes were isolated soon after drawing the blood and then stored in a lymphocyte stabilization medium (Park and Terasaki 1974) for up to 72 hours. Ashmore (Ashmore, Shopp, and Edwards 1989) report lower viability in cells isolated from blood which had been stored for 24 hr. at 4°C, a higher percentage of CD20+ B-cells, and lower percentages of OKT3+, OKT4+, and OKT8+ as well as lower CD8+CD11- and CD4+CD29+ subsets. Samples labelled, then fixed with paraformaldehyde\(^{2}\) were stable for some parameters for up to two weeks at 4°C if washed with Hank's balanced salt solution after fixation, but unwashed, fixed cells changed in forward scatter properties and diminished in fluorescence intensity over a 2-week period (Lal, Edison, and Chused 1988).

The time of day and the time of year in which the sample is drawn has to be taken into consideration. Diurnal variation in the total WBC and percentage of lymphocytes has been documented (Levi et al. 1985; Ritchie et al. 1983) but there was less variation in the percentages of subsets than in the cellular concentration (cells of each subset per mm\(^3\)) in the blood (Kidd and Vogt 1989). In fact in Ritchie's work, when results were expressed as percentage of total lymphocytes, only Leu-3a+ (CD4) and HNK-1+ (Leu-7, CD57) subsets varied significantly. Giorgi (1986) also

\(^{2}\)(CH\(_2\)O)\(_n\), a polymeric form of formaldehyde.
reported that percentages and numbers of CD4+ cells vary in a circadian manner, dropping between 8:00 a.m. and noon, then rising throughout the afternoon and sustaining their highest numbers throughout the night. CD4+\(\div\)CD8+ ratios, then, showed a parallel circadian variation. There was seasonal variation, as well (Abo et al. 1984; Levi et al. 1988). The numbers of subjects used and the numbers of time intervals at which measurements were taken in these studies were too small to adequately define seasonal patterns.

Delay in processing samples can result in lower viability of cells and lead to differential loss of subsets (Park and Terasaki 1974; Ashmore, Shopp, and Edwards 1989). Procedures utilized to reduce or eliminate extraneous cells from the sample can also lead to selective loss of subsets of interest. For instance separation of lymphocytes from other leukocytes on a Ficoll-Hypaque density gradient (FH) resulted in differentially lower estimates for CD8+ cells and in decreased stability of fixed, labelled cells compared to those from lysed, whole blood (Renzi and Ginns 1987). DePaoli and coworkers also reported selective loss of CD8+ cells using FH and increased percentages of large granular lymphocytes with the Leu-7+CD16+ phenotype (DePaoli et al. 1984). Still another group reported that FH separation resulted in lower estimates on all markers measured compared to measurements on lysed whole blood (Green and Stelzer 1988). Although gradient centrifugation is done to remove
monocytes and granulocytes, in fact there are residual cells of these types retained in the sample and the longer the blood is kept before processing, the greater the problem. It is thought that the lower estimates and increased variability with this method may be due to this contamination with unwanted cells (Green and Stelzer 1988).

Investigators disagree with respect to the methods for gating of data, that is, the selection of the events which are to be considered lymphocytes. One's results are highly dependent upon the base population selected for the denominator and how comparable that is from subject to subject (Vitale et al. 1987). The problem arises because monocytes and residual erythrocytes and platelets overlap the limits of the lymphocyte population in scatter properties. To minimize the number of non-lymphocytes in the denominator population a window is selected which includes most, but not all of available lymphocytes. But the various subsets of lymphocytes are not completely homogeneous in scatter properties either. NK cells are larger and more granular than most lymphocytes; B cells and CD4$^+$ T-cells are characteristically small. Estimates of NK cells, B cells, and of CD4$^+$ cells were proportionally biased depending upon whether a gate containing 95% of lymphocytes or one containing 98% of lymphocytes was chosen (Loken et al. 1990). Fleisher also demonstrated that natural killer cells, some T-cells including the majority of HLA-DR$^+$ T-
cells, and some B-cells are excluded from the tight window "usually" chosen to select lymphocytes (Fleisher, Marti, and Hagengruber 1988). He proposed monocyte depletion via treatment with nickel carbonyl before gradient centrifugation so that open gating (no gating) could be used. Recently systems of gating based on identifying lymphocytes by two-color distribution on a pair of antibodies which label lymphocytes, monocytes, and granulocytes differentially have been developed to help to define the best scatter window for lymphocytes (Loken et al. 1990).

Monoclonal antibodies with the same molecular specificity (see Table 1) often give different estimates of percentage positive, perhaps because of differing affinity of binding, or a difference in the epitope bound (Kootte et al. 1988). For instance CD4+CD8+ ratios were greater when the BD antibodies Leu-3 (CD4) and Leu-2 (CD8) were used than those obtained when the Coulter reagents T4 and T8 were employed (Schuerch et al. 1987; Henny et al. 1986b). Antibodies are not always as specific as first thought. They may not label 100% of the subset which they supposedly define and they may label small numbers of other subsets. For example, Leu-7 (CD57; Knapp et al. 1989a) was first thought to be a marker for NK (Lanier et al. 1983); it labels 40-60% of all large, granular lymphocytes (Weissler et al. 1987), but it was also found on a subset of CD8+
### TABLE 1

**CLUSTER DIFFERENTIATION ANTIGENS**

<table>
<thead>
<tr>
<th>CD Number</th>
<th>Clones</th>
<th>Target Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD2</td>
<td>OKT11, Leu-5b</td>
<td>T-cells, NK</td>
</tr>
<tr>
<td>CD3</td>
<td>OKT3, Leu-4</td>
<td>Mature T-cells</td>
</tr>
<tr>
<td>CD4</td>
<td>OKT4, Leu-3a</td>
<td>Helper T-cells</td>
</tr>
<tr>
<td>CD7</td>
<td>T3-3A1</td>
<td>T-cells, NK, B-cells(?)</td>
</tr>
<tr>
<td>CD8</td>
<td>OKT8, Leu-2</td>
<td>Cytotoxic/Suppressor T-cells, NK</td>
</tr>
<tr>
<td>CD11</td>
<td>Leu-15, OKM1</td>
<td>Suppressor subset of CD8+ T-cells, Monocytes and neutrophils</td>
</tr>
<tr>
<td>CD14</td>
<td>63D3, Leu-M3</td>
<td>Monocytes</td>
</tr>
<tr>
<td>CD16</td>
<td>Leu-11b</td>
<td>NK and neutrophils</td>
</tr>
<tr>
<td>CD22</td>
<td>Leu-14</td>
<td>B-cells</td>
</tr>
<tr>
<td>CD29</td>
<td>4B4</td>
<td>Memory T-cells</td>
</tr>
<tr>
<td>CD45r</td>
<td>2H4, Leu-18</td>
<td>Naive T-cells, NK, B-cells</td>
</tr>
<tr>
<td>CD56</td>
<td>NKH-1, Leu-19</td>
<td>NK</td>
</tr>
<tr>
<td>CD57</td>
<td>HNK-1, Leu-7</td>
<td>NK, T-cell subset</td>
</tr>
<tr>
<td>(none)</td>
<td>MMA</td>
<td>Neutrophils, monocytes, activated T-cells</td>
</tr>
</tbody>
</table>
cytotoxic T-cells, on a subset of B-cells, on brain cells, and curiously enough, on some small-cell lung tumor cell lines (CD56 or NKH-1 was also present on those small-cell tumor lines; Koros et al. 1986). Later it was found that NK function was restricted to those CD57+ cells which also carry the Fc receptor CD16 (Lanier et al. 1986). The NKH-1 antibody (Griffin et al. 1983) is thought to be more sensitive and specific than Leu-7 for cells with NK activity. Not all CD8+ cells carry CD3 either; CD8 is found on some NK-cells, which are CD3− (Ault 1988), but CD8 is at lower density on NK-cells than on T-cells (Perussia, Fanning, and Trinchieri 1983).

Choice of fluorochrome can affect the determination of proportion that are positive by 3 percentage points or more (Kootte et al. 1988). Whereas phycoerythrin (PE) is a good partner to use with fluorescein isothiocyanate (FITC) for double-labelling experiments, it must be kept in mind that the phycobiliprotein PE is a very large molecule (72–240 kd) which can in some cases interfere with antibody binding (Hoffman 1988).

**Flow cytometric lymphocyte subset profiles.**

Giorgi (1986) asserted that "standards for normal levels of the various lymphocyte subsets have not been established, so that each laboratory must generate its own reference range. "There is no 'gold standard' against which
flow cytometric data can be measured" (Kidd and Vogt 1989). Variations associated with gender, age, race, and smoking have been described. A decrease in the percentages of CD3+ cells and of CD8+ cells has been observed among elderly men compared to men under 40 years of age (Nagel, Chrest, and Adler 1981). The deficit in CD8 was also reflected in a higher CD4+CD8+ ratio among older men. The proliferating capacity of CD8+ cells and their cytolytic capacity seem to diminish in older subjects as well (Mariani, et al. 1990). Lymphocytes with the marker of naivete, CD45R, decrease with age and their reciprocal subset (CD29+) increase. The percentage of CD4+ cells with neither CD45R or CD29 increases also. The proportion of lymphocytes with a NK phenotype increases throughout life (Abo, Cooper, and Balch 1982).

Altering occur in leukocyte subset balance during viral or bacterial illness. Expansion of lymphocyte populations in the former or macrophage and granulocyte populations in the latter are expected. Chronic exposure to UV radiation such as in sunbathing has been shown to depress levels of circulating CD4+ cells (Burton et al. 1983b).

**Markers used in this study.**

For purposes of clarity, each reagent antibody should be referenced by the CD designation of its target antigen, the unique name assigned by the World Health Organization
WHO) Subcommittee on Immunology (Knapp et al. 1989), rather than by the common name determined by its origin. Various clones raised against particular cell-surface molecules, such as the OK series developed by Ortho Pharmaceutical Co. (Kung and Goldstein 1980) or the Leu- series developed by BD, received common names at development and before they were typed by WHO, but have a single CD designation which is specific for the cell-surface molecule to which they bind. They do not necessarily bind the same epitope, but do react with the same molecule. It is important to note, however, which particular clone is utilized in a specific study because of the small differences in estimated percentage positive which were noted above for different clones with the same CD specificity.

CD3 is the non-variant, T-cell-glycoprotein molecule which is associated with the receptor complex Ti for antigen. Although the bonding between CD3 and Ti is not covalent, the molecules co-modulate from the cell surface when either is precipitated by its antibody. Most CD3 antibodies induce proliferation of cultured PBL, and induce non-specific cytotoxicity in a test cytotoxic T-cell clone. Added to unseparated PBL, CD3 antibody enhances poke-weed-mitogen-induced production of IgM, however if B cells and T cells are first isolated then recombined, the effect of CD3 antibody is to down-regulate IgM production (Gorog, Batory, and Lanzavecchia 1987). The precise function of the CD3
molecule is unknown but it may be involved in signal transduction (Krensky, Lanier, and Engleman 1985). The lymphocytes which bear this molecule are a subset of cells bearing CD2, the receptor for sheep erythrocytes. Cells which bear CD2 but not CD3 are considered to be NK cells and conversely, most cells which have NK activity do not express the CD3 molecule.

CD7 might function as an Fc receptor for IgM, although some cell lines are positive for CD7, yet do not bind IgM. It is thought that the 3A1 antibody binds an epitope on the CD7 molecule other than the Fc receptor and that the latter is mutated in the cell lines which do not bind IgM. It is found on most, but not all T cells and some NK cells (and perhaps on a "small" subset of B cells; Sandrin et al. 1987). Modulation of the CD7 molecule causes a partial inhibition of the mixed lymphocyte reaction (Lazarovits et al. 1987). α-CD7 binds up to 85% (Haynes, Eisenbarth, and Fauci 1979; Haynes et al. 1980; Eisenbarth et al. 1980) of CD2⁺ lymphocytes (T cells and NK cells) in peripheral blood. Since this antibody was readily available, it was included in the panel to be compared with the CD3 antibody in the proportion of cells bound.

The CD4 molecule discriminates the helper/inducer subset of T cells and is also found at low density on monocytes. Helper cells are necessary accessory cells for antigen-specific immune responses and normally comprise
about 60% of T cells in peripheral blood. Antibodies reactive with the CD4 antigen may be specific for any one of 5 to 7 available epitopes. Certain individuals have been found whose lymphocytes are not reactive or have very low reactivity to OKT4, yet they react with another α-CD4, Leu-3, and apparently function normally (Taylor et al. 1987). CD4 appears to function as a receptor for the invariant or monomorphic part of the class II major histocompatibility antigen (MHC). Short peptide sequences have been identified which bind either to the β1 domain of HLA-DR or to the outer surface of the CD4 molecule and block the interaction of the two cells (Durandy and Fischer 1987). The CD4 molecule is essential, then, for antigen-presenting cells to activate helper cells and it also functions as the receptor for class II MHC antigen on target cells on the rare occasion when a CD4 lymphocyte differentiates into a cytotoxic effector.

α-CD45 designates antibodies which bind to "leukocyte common antigen," a 180-220kd glycoprotein molecule found on more than 95% of all leukocytes. A variant of this molecule is designated CD45R or CD45resticted. Cells which bind α-CD45R antibodies at high density have the characteristics of naive or less mature lymphocytes. When activated by phytohemagglutinin (PHA) they convert to expression of CD29 at high density and reduced expression of CD45R, and they do not revert when stimulation is discontinued (Sanders, Makgoba, and Shaw 1988). The addition of α-CD45R with PHA
enhances cell proliferation, probably by up-regulating the receptor for interleukin 2 (IL-2, Cobblod, Hale, and Waldmann 1987; Ledbetter et al. 1985).

The CD45R+ subset had previously been described among CD4+ T cells as suppression inducers since they induced CD8+ T cells to inhibit the production of immunoglobulin by pokeweed-mitogen-driven B cells (Morimoto et al 1985a), but could not provide help for antibody production. The CD29+ subset were considered the help inducers because they could enhance production of antibody in response to antigen stimulation. Further studies led to a reinterpretation, however. Upon stimulation with immobilized α-CD3, neo-natal and presumably naive T cells, which are high in CD45R and low in CD29, were capable of producing IL-2 upon stimulation, but not the repertoire of cytokines with which adult T cells responded (Ehlers and Smith 1991). Since IL-2 is all that was required to generate cytotoxic effector cells against allogeneic targets (Yamashita, Bullington, and Clement 1990), it is possible that the induction of suppressor cells requires only IL-2 as well, thus explaining the observed effects of CD45R+ T cells. The CD45R molecule is found on approximately 42% of unfractionated peripheral-blood T lymphocytes, on about the same fraction of CD4+ T cells, on 54% of CD8+ T cells, on 5-30% of monocytes, and on over 30% of B cells and null cells in the peripheral blood (Coulter Immunology product bulletin 1988).
Lymphocytes which express high densities of CD45R carry little CD29 and vice versa. CD29 was first used to identify a subset of CD4+ T cells. The function associated with this marker was the provision of help for PWM-driven immunoglobulin production by B cells, for which the subset was characterized as a helper-inducer population. CD4+ T cells bearing CD29 were also observed to act as inducers of activation among CD8+ cytotoxic T cells (Kalish, Morimoto, and Schlossman 1985).

The CD29bright subset proliferates sluggishly when stimulated by Concanavalin A or in the autologous mixed lymphocyte reaction AMLR, while its reciprocal subset, the CD45Rbright subset, responds well in AMLR (Morimoto et al. 1985b). CD4+CD45R+ T cells also are activated by immobilized anti-CD3 antibody and cytokines, whereas CD4+CD29+ cells are not (Wasik and Morimoto 1990). This is what one might expect from immature cells and differentiated cells, respectively. Later work led to the suggestion that CD29+ cells would be better described as memory cells since they respond to specific antigen to which they were sensitized in vivo. Consistent with this definition is the observation that they are rarely found in neonatal cord blood and increase in prevalence with age (Sanders, Makgoba, and Shaw 1988; DePaoli, Battistin, and Santini 1988). Since the proportion of antigen-specific cells increases with age, this might be one reason why the T cells of older people are
less responsive to non-specific stimuli like lectins. Weksler found that older people had the same number and percentage of T cells as the young but that only $\frac{1}{4}$ to $\frac{1}{2}$ of them could proliferate in response to lectins (PHA; Weksler 1980). The 4B4 (CD29) antigen is found on about 41% of peripheral blood T cells, on a similar proportion of CD4$^+$ T cells, on 43% of CD8$^+$ T cells, on 5-30% of B cells, and on over 30% of null cells, macrophages and thymic lymphocytes (Coulter Immunology product bulletin 1988).

Whereas CD29 and CD45R are expressed in high density on mutually exclusive subsets of CD4$^+$ T cells, there are some 16-30% of helper/inducer cells which do not express large quantities of either (Coulter product bulletins; Morimoto 1985a). The proportion of CD45R$^+$CD29$^-$ cells increases with age. Could these perhaps represent tolerized cells? When the cells identified by expression of CD29 or CD45R were defined as help-inducers or suppression-inducers, the inclusion of those antibodies in this study seemed more relevant. With the present understanding of CD29 and CD45R as markers of memory and naive cells, respectively, the meaning of any differences found between subgroups will not be as easily interpreted.

The subset of CD3$^+$ T cells which is the reciprocal of the CD4$^+$ population, bear CD8. Normal ratios of CD4 to CD8 T cells are approximately 3:2, but variation around that value is considerable. Immature thymocytes can carry both
CD4 and CD8 molecules, but mature PBL bear one or the other. CD8 cells function predominantly as cytotoxic effectors or suppressor cells, although CD8+ cells with helper function have been documented. The ratio of CD4 to CD8 is thought to reflect balance in immune status.

The CD8 antigen functions as the receptor for a monomorphemic epitope on Class I MHC molecules. Antigen-specific cytotoxic T lymphocytes can kill only cells which bear autologous Class I molecules. Antibody to CD3 can induce non-specific cytotoxicity in a CD8+ cell line. The reaction against a Class-I negative target cell line was inhibited by many anti-CD8 antibodies including many which did not bind epitopes which recognize Class I antigen. This would indicate more of a regulatory role than one of cell adhesion for the CD8 molecule (van Seventer et al. 1987).

Little is known about the mechanisms by which CD8+ suppressor cells are activated and function. The role of antigen is unclear since antigen-depleted CD4+ inducer cells were able to cause CD8+ responders to differentiate into antigen-specific suppressor cells and only CD3 and Class I MHC molecules on the inducer cells were required. The CD3 and CD8 molecules were essential on the responder cell (Krensky, Lanier, and Engleman 1987). Cross-linkage of CD3 with CD8 or CD4 molecules on a lymphocyte activates the cell. Furthermore cross-linkage of CD3 with either CD8 or CD4 mediated by either MHC Class I or MHC Class II molecules
in the thymus may determine whether a specific immature
CD4⁺CD8⁺ thymocyte develops into a helper (CD4⁺) or a
cytotoxic/suppressor (CD8⁺) lymphocyte (Emmrich 1987).

The FITC-conjugated Leu-2a was used for double-
labelling with PE-conjugated CD11b⁺CD18. CD11b/CD18 is the
receptor for the inactivated form of bound complement
molecule C3 (C3b) and its function on granulocytes is to
induce phagocytosis. It is curious that it identifies the
subset of CD8⁺ cells with suppressor function rather than
cytotoxic cells (Clement, Grossi, and Gartland 1984a). CD11b/CD18 belongs to a family of leukocyte function-
associated antigens (LFA) which all share a common 90kd β-
chain glycoprotein, designated CD18, and varying α-chains.
This family of molecules mediates intercellular adhesion
among leukocytes. From inhibition studies CD18 seems to be
the molecule which effects adhesion; antibodies to the α-
chains are less than 10% efficient at inhibiting adhesion
among polymorphonuclear leukocytes (PMN; Patarroyo and
Ansotegui 1987).

CD11b antibodies precipitate a glycoprotein of 155kd
found on monocytes and PMN, but only on CD8⁺ suppressor
lymphocytes and some NK cells (Zarling and Kung 1980;
Clement, Dagg, and Landay 1984b. The CD11b⁺ NK subset
probably also is dimly CD8⁺ which means that the only way to
distinguish them from CD8⁺CD11b⁺ suppressor T cells is
through the presence or absence of the CD3/Ti complex.
OKM1 is another antibody to CD11b. Cross-linking C3b receptors on monocytes and PMN enhances phagocytosis of IgG coated particles via Fc receptors and under certain circumstances, also allows the internalization of soluble ligands and C3bi-coated particles in the absence of IgG. The exposure of PMN to sheep erythrocyte ghosts coated with C3bi induced a respiratory burst (Fearon and Wong 1983).

W6/32 binds a monomorphic determinant on the Human Lymphocyte Antigens (HLA) A, B, and C, the MHC class I antigens. It not only precipitates these molecules from the PBL of all humans, it reacts with the PBL of all primates. In addition to lymphocytes W6/32 is found on PMN, monocytes, and eosinophils, but not on erythrocytes (Barnstable et al. 1978). The presence of the associated β2 microglobulin is essential to retain the conformation of the antigenic site, but W6/32 does not bind to the smaller molecule (Brodsky and Parham 1982; Parham, Barnstable and Bodmer 1979). W6/32 should label all living PBL giving an estimate of the quality of the sample.

One of the human MHC class II antigens is commonly designated HLA-DR. L243 was identified as an antibody to a monomorphic determinant on this highly polymorphic molecule (Lampson and Levy 1980). B cells and monocytes present antigen to T cells in the context of HLA-DR molecules. CD4 on helper/inducer T cells acts as a receptor for a
monomorphic determinant on the HLA-DR molecule while the CD3/Ti complex interacts with antigen and another portion of the HLA-DR molecule. HLA-DR can also be expressed by activated T cells as demonstrated in allogeneic mixed leukocyte culture (Evans et al. 1978). It will appear on CD8+ cells during cytomegalovirus infection and efforts have been made to correlate levels of HLA-DR on PBL and rejection of renal transplants (Henny et al. 1986a). In this study, the number of activated T cells can be estimated by the difference between the prevalence of HLA-DR+ lymphocytes and the number of B cells.

A surface marker which normally appears only on activated cells is the receptor for transferrin (CD71), the protein which transports iron in the bloodstream (Goding and Burns 1981). Resting, normal cells do not express this molecule and thus PBL should be negative for it.

Another marker of activation is the interleukin 2 receptor (IL-2R, CD25). Exposure to mitogens or antigen stimulates T cells to express the receptor for this growth regulating lymphokine. Increased numbers of PBL show this marker in the presence of an acute viral infection (Coulter Immunology product bulletin, 1986). Cross-linking of CD2 or CD3 is another stimulus which increases the expression of IL-2R. For a long time the presence of this molecule on T cells only was studied, but subsequently it was recognized that activated monocytes and activated B cells could also
bear IL-2R (Herrman et al. 1985; Olive et al. 1986). There is evidence that CD3+ NK also are responsive to IL-2 (Trinchieri 1986). There are actually two IL-2R's, a high-affinity receptor on a 75kd protein and a lower-affinity Tac antigen on a 55kd molecule which is highly glycosylated. The two molecules appear on the surface of the cell as a complex. The "anti-Tac" antibody binds to the low affinity receptor but blocks the binding of IL-2 to the high-affinity receptor thus blocking many IL-2-dependent immune responses (Waldmann, Goldmann, and Tsudo 1987).

The antibody 63D3 reacts with the CD14 molecule which appears in high density on monocytes. Granulocytes and B cells stain weakly and lymphocytes other than B cells do not react at all. Some granulocytic leukemias also react weakly (Ugolina et al 1980). The glycoprotein is also found in varying amounts in the blood as though it is a peripheral, easily shed molecule (Bazil, Horejsi, and Hilgert 1987). This antibody is useful for estimating the amount of monocytic contamination within the lymphocyte scatter window.

The antibody MMA reacts with an antigen found predominantly on cells of myelomonocytic origin, but which can be induced on T cells, particularly CD4+ helper/inducer cells, by culture with ConA. Depletion of MMA+ cells abrogated proliferative responses in the remaining cells to mitogens and antigen, but the addition of 10% MMA+ cells
restored proliferative capacity. The antigen is present on the bone marrow cells which are the precursors of granulopoietic colony-forming cells (Hanjan, Kearney, and Cooper 1982). MMA on a lymphocyte, then, probably indicates an activated T cell. In our hands, the MMA antibody in excess was lytic to monocytes and granulocytes.

Most peripheral blood B cells express a glycoprotein antigen, which exists as a heterodimer with glycoprotein chains of 130 and 140 kd (Ling, Maclennan, and Mason 1987). Cross-linking of this antigen does not activate B cells of itself, but will enhance activation by anti-Ig antibodies or by antigen, and the presence of CD22 is necessary for Ca\textsuperscript{2+} influx and activation to be induced by anti-\(\mu\) antibodies (Pezzutto, et al. 1987). CD22 is exclusively a B-cell antigen, i.e., is not expressed by T cells, monocytes, platelets, granulocytes, thymocytes, or activated T cells. It is characteristically strongly expressed on Hairy-Cell Leukemia cells, and is found on all non-Hodgkin's lymphomas of B-cell type which were tested (Schwarting, Stein, and Wang 1985).

There is a subpopulation of lymphocytes which are larger than other lymphocytes and contain cytoplasmic azurophilic granules. Among these are cells which have the ability to kill tumor cells, particularly those of the K562 line, without antigen- or MHC-restriction. They have been called "natural killer" cells (NK) and are thought to be an
important defense against neoplastic processes. While there are subsets of CD3+ cells which have natural killer activity, NK are generally considered to be CD3-. The genes which code for the T-cell antigen receptor T, are neither rearranged nor expressed on NK. Most if not all NK bear Fc receptors (CD16) and the CD11/CD18 LFA-1 complex and blocking either of these will inhibit NK function. Crosslinking CD2 on these cells enhances their effector capacity. While the above molecules seem to participate in NK function, they are not unique to NK but are expressed by other cytotoxic effectors as well. At least some NK are able to respond to IL-2, but it is not known that all NK respond (Hercend et al. 1985; Trinchieri 1986). Therefore there has been considerable confusion about the phenotype and properties of the NK.

NKH-1 (against CD56) is an antibody which is thought to be more specific for lymphocytes with natural killer activity than any other marker found thus far. CD56 is an isoform of the neural cellular adhesion molecule (N-CAM) and thus is found on neuroectodermal cells. NKH-1 has also been detected on human small-cell lung tumor lines and is found on a primitive invertebrate, the sea urchin. The latter suggests evolutionary conservation of the molecule (Koros et al. 1987). Addition of an IgM NKH-1 antibody and complement to lymphocytes removed essentially all NK activity (Hercend et al. 1985) as did FACS elimination of lymphocytes labelled
with the IgG1 NKH-1 antibody (Griffin et al. 1983) and with
the Leu-19 antibody to CD56 as well (Lanier et al. 1986b).
NKH-1 was found on 9 of 9 NK lines, one of which also bore
CD3 (Lanier et al. 1986b). In double-labelling experiments,
27% of NKH-1+ cells expressed CD8, 74% were CD2+, 80% were
CD11b+, 85% bore Leu-7 (CD57), while 2% or less expressed
CD3, CD4, HLA-DR, CD14 (monocytes), or CD20 (B-cells).
Addition of OKM1 antibody (α-CD11b) with complement
diminishes NK activity but does not eliminate it (Zarling
and Kung 1980). Therefore, while CD56+ cells seem to
include all of those lymphocytes with natural killer
activity, they are heterogeneous with respect to other
surface markers and it may be that not all CD56+ cells are
NK.

The consensus is that cells with natural-killer
activity bear a Fc receptor for IgG (CD16; Phillips and
Babcock 1983). CD16 is involved in cellular activation;
specifically bridging of the CD2 molecule to the CD16
molecule is required for enhancement of NK cytolysis.
Certain antibodies are capable of forming such a bridge and
hence activating the cell, or at higher temperatures (37°C),
co-modulating CD2 and CD16 from the surface (Anasetti et al.
1987). Antibodies which bind only CD2 or only CD16,
however, do not co-modulate the other molecule, so there is
no evidence for association of CD2 and CD16 such as that
which exists between CD3 and T₁. In addition, enhancement
or inhibition of NK function can be produced by antibodies reacting with either CD2 or CD16 individually depending upon the epitopes to which they bind (Uggla et al. 1987). CD16 is also found on neutrophils, but is not the same molecule as the Fc receptor on B lymphocytes. α-CD16 (Leu-11b) does not label monocytes (Perussia et al. 1984).

Susceptibility to lung cancer seems to have multiple components, which can interact with environmental, genetic, and behavioral factors. Furthermore all of these things impact immune defenses. In the study of disease etiology and particular factors, it is vital to measure and adjust for other factors known to be involved. Otherwise the results of studies done under varying conditions may be inconsistent, non-comparable, and of limited value. Accounting for known components of the variance in data also increases the probability of recognizing additional effects because the variance properly attributed to each factor comes out of the "error" variance. The statistical test for the significance of a factor is, of course, based on the ratio of the effect variance to the error variance.
MATERIALS AND METHODS

Selection of study subjects.

Lung cancer cases were volunteers from among white male patients who came to Mary Bird Perkins Cancer Center (PCC) for irradiation of their tumors between January 29, 1988 and January 4, 1990. Patients were considered eligible for the study if they had not received prior radiation treatment or chemotherapy for their disease and were judged by their doctors to be strong enough to participate.

Healthy white males of a similar age to the patients were needed for comparison subjects. A man was not eligible if he had ever had a cancer other than skin cancer. A two-to-one ratio of smokers to non-smokers was sought, such that smokers might be discriminated into a group more like patients and a group more similar to non-smokers. A 3:1 ratio of controls to patients was planned in order to optimize the efficiency of the study (Rothman 1986, 99). Volunteers were recruited from among spouses of female patients at PCC, from community and civic organizations within the 11-Parish area which comprises the Mary Bird Perkins Tumor Registry (Figure 1), and from among those informed of the study through a brochure (Appendix B) displayed at PCC.
Figure 1. The Mary Bird Perkins Tumor Registry jurisdiction.
Initial contact with an organization was made by mail. A brief description of the study and a copy of the questionnaire accompanied a request to address a regular meeting of the group for the purpose of recruiting volunteers. At a typical meeting, a short presentation about cancer in general and this study in particular was delivered using simple visual aids, then volunteers were invited to participate. A medical technologist was brought to each group so that the blood sample could be drawn immediately and with the least inconvenience to the volunteer. Participants signed the consent form, the blood was drawn, and then each was given a questionnaire to be completed at home, and an addressed, stamped envelope for its return.

The questionnaire.

The questionnaire for this study (Appendix A) was designed to assess variable personal attributes and behaviors, which are thought to impact the immune system and to affect one's susceptibility to lung cancer, such as ethnicity, socioeconomic status, geographic area of residence, familial history of cancer, occupational history, exposure to certain chemicals and metals, diet, and smoking practices. The majority of the questionnaire, in particular the dietary section, was an adaptation of the Health Habits and History Questionnaire (HHHQ), version 02, designed by Gladys Block of the National Cancer Institute (Block et al.
The software used to estimate nutrient intake was the 1989 version 2.2.

The questionnaire borrowed some life-style items (with permission) from a questionnaire designed by Dr. Marise Gottlieb specifically for case/control studies of cancer in Louisiana. Questions concerning the relevance of current dietary habits to those which were practiced some years ago, about medications, about recent illnesses and history of atopy or autoimmune disease, and about a second occupational history were the additions of this author. By using mostly items which had already been used successfully in questionnaires, the issue of pretesting was largely circumvented. The final version was, however, given to 4 individuals with characteristics similar to those of the study subjects, to be sure that there were no problems of clarity and to estimate the time that was required to complete it (45-60 minutes).

A subject was given 2-3 weeks to return his questionnaire before being reminded by telephone. Subjects who were delinquent were contacted at least twice. If a patient became too ill or died before completing his

1Available along with the companion nutrient analysis software for use in any study from Gladys Block, PhD., Epidemiologist, Applied Research Branch, Division of Cancer Prevention and Control, National Institutes of Health, Bethesda, MD 20892.

2It is commonly recognized that the development of cancer usually involves a period of latency of 5-20 years following the initial event.
questionnaire, his spouse or next of kin was asked to complete it.

Questionnaires were coded and keyed into files using the Statistical Analysis System\(^3\) (SAS) program editor. To verify data entry, the data files for each subject were printed and visually compared, byte by byte, with the coded columns on each questionnaire. Additionally the letter codes, which are card codes in the original HHHQ, were retained in the data stream to serve as markers for specific positions in the data records and used to check for misalignment of the data via a character left out of or an extraneous character inserted into the input stream. The appropriate food-frequency data were abstracted from the questionnaire file and formatted into 12 80-column records as required by the diet analysis software. Dietary analysis was conducted using the options in Table 2.

The Block dietary analysis software incorporates strategies for detecting possibly invalid data. The main criteria for flagging a subject were (1) too many items were skipped, (2) no serving size was used, (3) the number of different foods eaten daily was too small, (4) the frequency of eating a particular food was too high, or (5) too many items were coded with the same serving size. This software

\(^3\)SAS Institute Inc., SAS Circle, Box 8000, Cary, NC 27512-8000.
## TABLE 2

### DIET ANALYSIS OPTIONS IN EFFECT

<table>
<thead>
<tr>
<th>OPTION</th>
<th>INTERPRETATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addfats = Yes</td>
<td>Use information on fats used in cooking and added to vegetables.</td>
</tr>
<tr>
<td>Addfibs = No</td>
<td>Add fiber if response to 'raw foods' question.</td>
</tr>
<tr>
<td>Addsalt = Yes</td>
<td>Increases sodium according to response to 'add salt at the table' question.</td>
</tr>
<tr>
<td>Advise = No</td>
<td>Do not print advice for subject.</td>
</tr>
<tr>
<td>Bigfat = No</td>
<td>Allocates medium rather than large portion of butter added in cooking.</td>
</tr>
<tr>
<td>Carotfib = Yes</td>
<td>Produce an output file containing intakes of specific carotenoids and fiber.</td>
</tr>
<tr>
<td>Codec = Yes</td>
<td>Allows type-specific nutrient contributions from the kind of dry cereal consumed.</td>
</tr>
<tr>
<td>Colapsxl = No</td>
<td>Did not use extra-large portion sizes.</td>
</tr>
<tr>
<td>Cutoff = 4</td>
<td>Include nutrient values even of foods eaten very infrequently.</td>
</tr>
<tr>
<td>Darkques = No</td>
<td>Applies to question not asked in Version 2.</td>
</tr>
<tr>
<td>Dietques = Yes</td>
<td>Includes data on special-diet question.</td>
</tr>
<tr>
<td>Dropfood = 0</td>
<td>Allows exploration of effect of leaving 1 or more foods out of the analysis.</td>
</tr>
<tr>
<td>OPTION</td>
<td>INTERPRETATION</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Frtadj = Yes</td>
<td>Allows use of 'How many fruits/week?' to prevent overestimation of vitamin C, fiber, and carotenoids.</td>
</tr>
<tr>
<td>Htwtques = Yes</td>
<td>Utilizes the information on height, weight, age, and sex.</td>
</tr>
<tr>
<td>Illadj = No</td>
<td>Do not adjust portion size downward if subject lost 15 lbs. or more.</td>
</tr>
<tr>
<td>Keep = Original</td>
<td>No recalculation using all medium portion sizes.</td>
</tr>
<tr>
<td>Leanques = No</td>
<td>Applies to question not asked in Version 2.</td>
</tr>
<tr>
<td>Medonly = No</td>
<td>Changes all portion sizes to medium for all subjects.</td>
</tr>
<tr>
<td>Morefds = 0</td>
<td>No additional foods were added.</td>
</tr>
<tr>
<td>Moreinfo = Yes</td>
<td>Print out nutrients/1000 calories, nutrients/kg body wt., nutrients/g of solid food, and dietary variability indices.</td>
</tr>
<tr>
<td>Oldadj = No</td>
<td>Do not assume that oldest people chose too small a portion size.</td>
</tr>
<tr>
<td>Omit = 2</td>
<td>Written-in foods only considered if eaten more than once/week.</td>
</tr>
<tr>
<td>Pillques = Yes</td>
<td>The vitamin supplement question was asked.</td>
</tr>
<tr>
<td>Portions = Agesex</td>
<td>Portion sizes are age/sex-specific.</td>
</tr>
</tbody>
</table>
### TABLE 2 - Continued

<table>
<thead>
<tr>
<th>OPTION</th>
<th>INTERPRETATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recalc = No</td>
<td>Do not change all portion sizes to medium and recalculate for comparison.</td>
</tr>
<tr>
<td>Restadj = Yes</td>
<td>Include foods eaten in restaurants.</td>
</tr>
<tr>
<td>Someall = All</td>
<td>Analyze all files, rather than specific ID's.</td>
</tr>
<tr>
<td>Sources = No</td>
<td>Do not want contribution of each food to each nutrient calculated.</td>
</tr>
<tr>
<td>Versionq = 2</td>
<td>Using Version 2 of the questionnaire.</td>
</tr>
<tr>
<td>Study =</td>
<td>(These files were all from the same study.)</td>
</tr>
<tr>
<td>Subfruit = Yes</td>
<td>Avoids double counting of fruits mentioned in two places on questionnaire.</td>
</tr>
<tr>
<td>Todo = 99</td>
<td>Look for up to 99 subject files.</td>
</tr>
<tr>
<td>Toskip = 0</td>
<td>Process all respondent files.</td>
</tr>
<tr>
<td>Tunaques = No</td>
<td>Applies to question not asked in Version 2.</td>
</tr>
<tr>
<td>Vegadj = Yes</td>
<td>Allows use of 'How many vegetables/week?' to prevent overestimation of vitamins A and C and fiber.</td>
</tr>
<tr>
<td>X1 = No</td>
<td>Did not use extra-large portion sizes.</td>
</tr>
<tr>
<td>Yearcol = 0.01923</td>
<td>Used year as largest time-period option.</td>
</tr>
</tbody>
</table>
was utilized to determine the quality of the subjects' dietary data.

**Selection of Lymphocyte Markers.**

Markers to be used in this study were selected primarily on the basis of their relevance to immune status as it might relate to susceptibility to cancer. Most of the reagent antibodies were harvested from clones available from the American Type Culture Collection (Rockville, MD; ATCC); some were purchased from commercial sources as indicated below.

Most labelling was by the indirect method with fluorescein-isothiocyanate (FITC)-conjugated goat anti-mouse IgG as the second antibody. Only three determinations were done by direct double labelling with a FITC-conjugated antibody and a phycoerythrin-(PE-)conjugated antibody: CD4/CD45R, CD4/CD29, and CD8/CD11. Basic information about each of the antibodies used in this study was provided above and is summarized in Table 3. Antibodies to CD16 (Leu-11b, IgM), and CD22 (Leu-14, IgG1b) were purchased from Becton Dickinson (San Jose, CA) as were FITC-conjugated CD4 (Leu-3a, IgG1), FITC-conjugated CD8 (Leu-2a, IgG1), and PE-conjugated CD11b (Leu-15, IgG2a) antibodies. Coulter Immunology (Hialeah, FL) was the source for CD25 (IL-2R1, IgG2a), CD56 (NKH-1, IgG1), PE-conjugated CD45R (2H4, IgG1) and PE-conjugated CD29 (4B4, IgG1) antibodies. Clones which
### TABLE 3

MARKERS UTILIZED

<table>
<thead>
<tr>
<th>Marker</th>
<th>M.W.(kd)</th>
<th>Clone</th>
<th>Target Subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD3</td>
<td>26,20,16</td>
<td>OKT3</td>
<td>Mature T-cells</td>
</tr>
<tr>
<td>CD7</td>
<td>40</td>
<td>3A1</td>
<td>T-cells, NK</td>
</tr>
<tr>
<td>CD29</td>
<td>59</td>
<td>OKT4</td>
<td>Helper T-cells</td>
</tr>
<tr>
<td>CD45R</td>
<td>135</td>
<td>4B4</td>
<td>Memory subset of CD4+ T-cells</td>
</tr>
<tr>
<td>CD8</td>
<td>32,33</td>
<td>OKT8</td>
<td>Suppressor/cytotoxic T-cells, NK subset</td>
</tr>
<tr>
<td>CD11/CD18</td>
<td>155/95</td>
<td>Leu-15</td>
<td>Suppressor subset of CD8+ T-cells</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>43</td>
<td>W6/32</td>
</tr>
<tr>
<td>HLA-DR</td>
<td>34,28</td>
<td>L243</td>
<td>Monocytes, B-cells, activated T-cells</td>
</tr>
<tr>
<td>CD71</td>
<td>200(95)</td>
<td>OKT9</td>
<td>Dividing cells</td>
</tr>
<tr>
<td>CD25</td>
<td>55</td>
<td>IL-2R</td>
<td>Monocytes, B-cells and activated T-cells</td>
</tr>
<tr>
<td>CD14</td>
<td>55</td>
<td>63D3</td>
<td>Monocytes</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>MMA</td>
<td>Monocytes, Granulocytes, and activated T-cells</td>
</tr>
<tr>
<td>CD22</td>
<td>130,140</td>
<td>Leu-14</td>
<td>B-cells</td>
</tr>
<tr>
<td>CD56</td>
<td>220/135</td>
<td>NKH-1</td>
<td>NK</td>
</tr>
<tr>
<td>CD16</td>
<td>50-65</td>
<td>Leu-11b</td>
<td>NK, neutrophils</td>
</tr>
</tbody>
</table>

* OK antibodies originally developed by Ortho Pharmaceutical Co., Raritan, NJ.
produced the antibodies OKT3 (CD3, IgG₂a), T3-3A1 (CD7, IgG₁), OKT4 (IgG₂b), OKT8 (IgG₂a), OKM1 (CD11b, IgG₂b), W6/32 (IgG₂a), L243 (IgG₂a), OKT9 (IgG₁), 63D3 (CD14, IgG₁), and MMA (IgM) were obtained from ATCC.

The hybridoma cell lines from ATCC were cultured at 37°C in roller bottles to a maximum volume of 500 cc in DMEM, Iscove medium, or RPMI-1640 (GIBCO, Grand Island, NY) as specified by ATCC. The media were buffered with NaHCO₃, supplemented with 10% fetal bovine serum (Hyclone, Logan, UH), 2mM L-glutamine (GIBCO), 20 μM 2-mercaptoethanol (Sigma), and 50 μg/ml gentamicin sulfate (Schering Corp., Kenilworth, NJ). When viability dropped to <20%, the mixtures were centrifuged and the supernates passed through 45 micron filters. They were concentrated to a volume of 20-30 cc under 5-10 lbs. of nitrogen pressure in a Novacell 150™ with a membrane permeable to molecules of <100 kd and 0.1% sodium azide was added as a preservative. Each concentrate was then titrated using human leukocytes (or human tumor-cell lines HSB-2 and MT-2 for the transferrin-receptor antibody) to determine the optimal dilution, that is, that dilution at which a 50 μl aliquot added to 5-7.5 x 10⁵ leukocytes in 50 μl of PBS yielded the maximum fluorescence intensity. The antibodies purchased from commercial sources were reconstituted as recommended by the manufacturer and their concentrations were adjusted by dilution with PBS so that a 50 μl aliquot would contain the
appropriate concentration of antibody to label $5-7.5 \times 10^5$ cells. Small stocks of diluted antibodies were prepared before sample collection and stored at 4°C until needed.

FITC-conjugated goat anti-mouse IgG (amFITC), preadsorbed with human serum proteins to reduce non-specific binding to monocytes and B cells, was obtained from Sigma Chemical Company. The solution provided by Sigma was separated into 50-μl aliquots and stored frozen in the dark until used. It was diluted 1:100 with phosphate-buffered saline (PBS) at the time of labelling. Optimal fluorescence was obtained using 100 μl of diluted amFITC per $5-7.5 \times 10^5$ cells. Irrelevant mouse antibodies, of the IgG1, IgG2a, IgG2b, and IgM isotypes, used as isotype-specific, negative controls were also purchased from Sigma.

**Subject kits.**

The materials needed for the collection of data from each subject were pre-assembled and labelled with the 3-digit study number for that subject, beginning with 001, 002,..., and consecutively. The study number was prefaced with a P for patient or a C for control. Included in each kit were the questionnaire, an addressed, stamped, return envelope, a consent form (see Appendix B), a 3" x 5" card

---

Using a volume of 50 μl increased volume measurement and dispensation precision.

137 mM NaCl, 2.78 mM KCl, 7.89 mM Na$_2$HPO$_4$, 1.47 mM KH$_2$PO$_4$; pH=7.2-7.4.
for the subject's name and phone number (for follow-up on questionnaires), and all of the tubes and storage vials required for the collection, processing, and labelling of the blood sample. The kits for patients also included a medical information form (see Appendix B), which defined the histologic type, stage, and location of the subject's tumor and was to be completed by the patient's radiologist.

**Phlebotomy and processing of blood specimens.**

A licensed phlebotomist was employed to obtain the samples. Blood was drawn from an antecubital vein into two 10 cc, green-top vacutainers (BD) containing lithium heparin and stored at room temperature until processing. Processing was done on the same day that the samples were collected and was begun within 12 hours of collection in all cases.

Samples from 1 to up to a maximum of 12 subjects were processed at one time. The work-up was completed, to the point of fixing the labelled cells, within between 4 hours for one subject and 13 hours for 12 subjects. For each subject 4.3 cc of blood were aseptically removed from one heparinized vacutainer and 4.0 cc transferred to a borosilicate test tube for preparation for FACS analysis. From the blood remaining in the pipette, two slides were prepared for staining for a manual differential count and duplicate capillary tubes were filled for obtaining a microhematocrit determination. The remaining blood in the
vacutainers was centrifuged at 500 x g for 10 min.\(^6\) at room temperature (RT). The plasma, erythrocytes, and buffy coat cells were cryopreserved for future study.

The 4 ml sample which was reserved for FACS analysis was centrifuged. The buffy coat was transferred to a 15 cc, polystyrene centrifuge tube. Fourteen milliliters of freshly prepared lysing buffer\(^7\) were added and the contents of the tube mixed immediately by inversion until the mixture clarified (3-5 min.). The leukocyte suspension was centrifuged and the supernatant solution decanted. The pellet was dispersed in 1 cc of cold, isotonic PBS with a Pasteur pipette, another 13 cc of PBS were added, and the cell suspension mixed by inversion. The mixture was centrifuged and the pellet resuspended in enough refrigerated PBS with 10% complement-inactivated goat serum\(^8\) (PBS/G) to adjust the cell concentration to between 0.5 to 1 million cells per 50 µl aliquot (1.5 to 3 cc of PBS/G). A control containing cells exposed only to the goat amFITC second antibody was included in the panel for each subject to detect any residual non-specific binding.

\(^6\)Unless otherwise indicated, all centrifugation was done at 500 x g for 10 min.

\(^7\)100 ml of distilled H₂O added to 829 mg NH₄Cl, 109 mg KHCO₃, 3.7 mg Na₂H₂(CH₃CO₂)₄C₂N₂ (ethylenediaminetetraacetic acid, disodium salt) just prior to use.

\(^8\)The purpose of the goat serum was to block protein binding sites on the cells, especially any receptors for goat immunoglobulins.
**Differential.**

A manual differential was done on a stained slide for each subject. One hundred cells were classified as lymphocytes, monocytes, eosinophils, basophils or neutrophils. All differential determinations were carried out by the same ASCP medical technologist.

**Fluorescence labelling.**

 Appropriately labelled 12x75 mm Falcon tubes for each subject were set in an ice bath. A 50 µl aliquot of the leukocytes in PBS/G was pipetted into each tube. The first antibody (50 µl) was added to each of the samples to be indirectly labelled. Both antibodies, for example, CD4-FITC and CD29-PE, were added at the same time to those samples to be doubly labelled. The samples were vortexed, covered with aluminum foil, and allowed to incubate on ice for 1 hr. Then 3 cc of PBS were added to each tube and the mixtures were centrifuged in the cold. The cells which had been directly labelled were fixed at this time.

 For those samples indirectly labelled, the PBS wash was decanted, 100 µl of amFITC was added to the cell pellet, the tube contents were well mixed by vortexing, and the mixtures incubated on ice in the dark for 30 min. The cells were then diluted with 3 cc of cold PBS and the mixtures centrifuged. The PBS was decanted, the cells were resuspended in the residual drop (@ 50 µl) of supernate by
vortexing, and 50 μl of cold 1% paraformaldehyde was added to each of the tubes, on ice, with vortexing. For each subject the following negative controls were prepared and fixed: cells + amFITC only, and cells + irrelevant antibody of isotypes IgG1, IgG2a, IgG2b or IgM + amFITC. The samples were covered and stored in the refrigerator for 4-24 hours, then diluted with 100 μl of PBS. Again, they were stored at 4°C until analyzed on the cell sorter. In nearly all cases, the FACS data were collected on the day after labelling.

**Instrumentation.**

The Fluorescence-Activated Cell Sorter was a BD FACS-440 equipped with an argon laser used at 200 mWatt, with a 488 μm primary emission line. Green fluorescence was detected with the primary photomultiplier tube with a 530/30 band pass filter. Red fluorescence was detected with the secondary photomultiplier tube with a 585/42 band pass filter. All fluorescence signals and 90° scatter were log amplified. Sheath fluid was 0.2μ-filtered deionized water. The cell sorter interfaced with Consort 40 software (Becton Dickinson) on a Micro-VAX II workstation from Digital Equipment Corporation (Westminster, MA). Formaldehyde-fixed chicken erythrocytes were used to align the sorter on scatter and green fluorescence each day that samples were

---

9Two grams of paraformaldehyde were heated in 100 cc PBS at 70°C until dissolved. The solution was stored at 4°C. It was diluted 1:2 with cold PBS just prior to use.
run. A threshold was set on forward scatter such that all lymphocytes were retained but residual erythrocytes, platelets, and debris were excluded as completely as possible. The flow rate was adjusted to the fastest rate achievable without sacrificing resolution on forward and side scatter; usually 500-1000 cells per second. To check for instrument drift, the 2-dimensional scatter displays of the first sample and the last sample for each subject were overlaid. If substantial drift had occurred, that individual's samples were rerun. Where the subject had a normal lymphocyte:granulocyte ratio, 15000 events were collected, (except that 20000 events were collected for the CD22, CD56, CD4/CD29, CD4/CD45R, and CD8/CD11 samples because the proportions of cells expected to be positive for those markers were small). For those subjects with relative granulocytosis, some of whom had only 5-10% lymphocytes, 33-100% more events were collected with the goal of obtaining data on at least 2000 lymphocytes.

No electronic compensation\textsuperscript{10} was done on red-green doubly labelled samples. Compensation did not help to discriminate the positive from the negative subgroups and completely destroyed the normality of the distributions, which would have made it impossible to use Gaussian curve-

\textsuperscript{10}The cell sorter is equipped with an electronic compensation mechanism which is supposed to correct for spillover of the red emission into the green detector and vice versa.
fitting strategies to estimate the components of the mixtures. Data were stored on magnetic tape for later analysis.

**A novel method for FACS data processing.**

Forward- by side-scatter displays of the first and last samples from each subject were overlayed. Contours which enclosed all of those coordinates for which n events were recorded were drawn for n=3 and n=10 in order to define the leukocyte distributions. A rectangular window on forward and side scatter was set to enclose all of the cells with the scatter properties of lymphocytes, while excluding as many non-lymphocytic events as possible (Figure 2). The events within this window were selected for further editing.

The following paragraphs describe a new method which was invented to achieve percentage estimates on lymphocyte subsets from the raw data generated by the cell sorter.

For each event (cell), an integer ranging in value from 1 to 256 and corresponding to relative intensity or scatter angle of the light detected was recorded for each of the four parameters. List mode FACS data in this system are coded in unsigned binary integers (Dean et al. 1990). Analysis of the data with a multidimensional approach required conversion of the data to ASCII files so that they could be manipulated in SAS. Software which accomplished
Figure 2. Setting a rectangular window on lymphocytes.
this, FCSLST V01-03, was obtained from Dr. Robert F. Murphy. Short programs in BASIC were then written, which compressed the ASCII files to about half the length of the file output from FCSLST and truncated large files at 2500 events.

All processing to this point was conducted on a VAX 8200. The files were then transferred to an IBM 3090 mainframe for further editing and analysis.

A probability sample of events from the crudely edited data was selected for designation as the lymphocyte population in the following way. Based on the assumption that lymphocytes were at least approximately homogeneous in their scatter properties and that events were normally distributed in the 2 scatter channels (see the histogramic data display in Figure 3), a bivariate mean vector and covariance matrix for sample lymphocytes were estimated from the coordinates of the events within the rectangular edit window. The number of points used to estimate the mean vector and covariance matrix was large compared to the dimensionality (2), therefore the estimates were considered to be very close to the true parameter values and the standardized squared distance of any sample point from the mean of a bivariate normal distribution was considered to be

---

1Dr. Robert F. Murphy, Center for Fluorescence Research in Biomedical Science, Carnegie Mellon University, 4400 Fifth Ave., Pittsburgh, PA 15213.

12Program written by Mrs. Cindy Fort, Systems Analyst.
distributed as a $\chi^2$ with 2 degrees of freedom (DF; Johnson and Wichern 1988, 145). To decide if a particular event was likely to belong to the population of lymphocytes, the standardized squared distance of that point from the estimated mean (SSD)\(^{13}\) was compared to the 95th percentile $\chi^2_2 = 5.99$:

$$
\begin{bmatrix}
    x_{1i} - \overline{x}_1 \\
    x_{2i} - \overline{x}_2
\end{bmatrix}'
\begin{bmatrix}
    S_1^2 & S_1S_2 \\
    S_1S_2 & S_2^2
\end{bmatrix}
\begin{bmatrix}
    x_{1i} - \overline{x}_1 \\
    x_{2i} - \overline{x}_2
\end{bmatrix} = \chi^2_2
$$

If the SSD was less than the $\chi^2$, the event was considered to be a lymphocyte; otherwise it was rejected.

In order to compensate for the distortion of the estimates of the mean vector and covariance matrix by outliers, the parameters of the lymphocyte distribution were reestimated after the outliers were excluded and the revised estimates used to retest the events within the original rectangle. In subjects with granulocytosis, for whom outliers were relatively numerous compared to lymphocytes, the first cut was made at 90%, so that a better second estimate of the parameters was obtained. The editing process is diagrammed in Figure 4. It was programmed using SAS steps and procedures (Appendix C). An example of a selected lymphocyte sample is shown in Figure 5.

---

\(^{13}\)Standardized squared distance is analogous to the univariate $Z$ statistic squared: $[(x-\mu)/s]^2$. 
Figure 3. The distribution of leukocytes on scatter and fluorescence.

Determination of the proportion fluorescence positive was made in the following way. Fluorescence intensity, on a logarithmic scale, of events of each subset was assumed to be normally distributed (Figure 3), that is, the autofluorescence of the negative subpopulation was normally distributed as well as the brightness of the positive subpopulation(s). As mentioned previously, scatter properties were also considered to follow a Gaussian distribution. A Fortran algorithm for fitting Gaussian distributions to data (McLachlin and Basford 1988) was customized to the requirements of this study to estimate up
Figure 4. The lymphocyte editing process.
Figure 5. A selected probability sample of lymphocytes.
to 6 distributions in 2 dimensions and to accommodate data on up to 6000 events.

The program required that initial values for the mean vector, the covariance matrices, and the relative proportions of events in each subset be provided. In order to achieve data-based initial values for the parameters SAS Proc Fastclus\(^{14}\) was used to group events into 2-6 clusters (depending on the particular antibody) on the basis of the two fluorescence or fluorescence and scatter data on each event. The mean vector, covariance matrix, and proportion that a cluster of events were of total events were calculated for each cluster and appended to the crude data for input to the normal-mixtures program. The latter then provided final estimates of percentages of cells in each group for each subject.

Lymphocytes, especially NK, vary somewhat in their properties (Figure 6). Therefore side scatter was chosen as a second dimension for discrimination where only one fluorescent antibody marker was employed. In the situation where double-labelling was done, the analysis dimensions were red and green fluorescence and side scatter was ignored, since subsets within subsets of lymphocytes would not differ appreciably on side scatter.

\(^{14}\)This clustering method is based upon a nearest-cluster criterion for assignment of events and does not utilize available frequency distribution information as the normals-mixture approach does.
Figure 6. NK cells exhibit higher sidescatter than other lymphocytes.

Because the method was new and previously untried some care was taken to evaluate and assure the quality of the results. The output from each file was compared with scatter plots of the data and expected ranges for the percentages and fluorescence means and variances to insure that convergence of the normals-fitting program was to appropriate means and covariances and that the estimates of subset percentages were reasonable. Estimates with excessively large variances were not accepted to avoid assigning events to the population with the large variance when they did not belong there.
A subset of 25 subjects were selected using a table of random numbers. Their 63D3 (monocyte marker) files were analyzed by the new method to estimate the extent of contamination by monocytes in the probability sample of lymphocytes.

The performance of the new FACS data processing method was compared with the conventional method. The estimated percentages for CD3⁺, CD4⁺, CD8⁺_{bright}, and CD8⁺_{dim} for 27 subjects were compared with estimates for those subsets made using conventional analysis procedures.

**The master data set.**

Non-dietary data from the questionnaire data file, the nutrient estimates produced by the dietary analysis software, data from the patients' medical forms, hematocrit and differential counts, and the lymphocyte subset proportion estimates were combined into a master data file, coded in ASCII, for statistical analyses. The codebook for the master data set is Appendix D.

**The lymphocyte profile.**

The estimates of subset proportions were based upon the analysis of a large number of lymphocytes, therefore the B-cell, T-cell, and NK proportions can be considered to be continuous data and looked upon as elements of a lymphocyte composition. All of the information necessary to define a mixture of n components can be expressed in n-1 proportions,
\( p_i \), since \( p_n = 1 - p_1 - p_2 - \ldots - p_{(n-1)} \). A profile is then represented as \( n-1 \) variables of the form \( p_1 + p_n, p_2 + p_n, \ldots, p_{(n-1)} + p_n \) (Aitchison 1986, 26). The information which fully described the B-cell/T-cell/NK lymphocyte profile was summarized with two terms, the ratios of proportions CD3+CD22 (T+B) and CD56+CD22 (NK+B). A logarithmic transformation of the ratios CD3+CD22 and CD56+CD22 normalized their distributions and allowed the components of the profile to be simultaneously regressed upon a linear combination of putative effector variables. SAS Proc GLM with the MANOVA request accomplished the mathematics required to estimate the probability that any particular independent variable was not a determiner of the outcome profile.

Independent variables were added, one at a time, to the model, beginning with variables for which there was prior evidence of association. The first 3 variables that were incorporated were age, disease, and smoking, followed by terms summarizing the time of day when the sample was collected and the month of the year when the sample was collected.

"Rhythm" implies a movement with periodicity, cyclic variation. Therefore whatever variation that was due to time of day or time of year would be expected to display a cyclic and continuous pattern. Circadian rhythms have commonly been modelled with a cosine function (COSINOR
method; Reinberg and Smolensky 1983). Parameters $a_j$ are estimated for the nonlinear regression model:

$$Y(t_j) = a_0 + a_j \cos[(2\pi a_2 t_j + a_3), \text{ where}$$

$t_j =$ time (when blood drawn),

$a_0 =$ mesor (or mean),

$a_1 =$ amplitude,

$a_2 =$ period, and

$a_3 =$ acrophase (phase).

Software exists which will estimate all of the parameters $a_j$ given the data. In this case, however, the function of time had to be constructed so as to permit incorporation into a larger, linear model. Therefore if a cosine function was to be used, it was necessary to estimate a value for period and one for acrophase and assume those to be constant. The mesor then is incorporated into the intercept of the larger model and the amplitude is the estimated parameter of the model for that time function. A starting estimate of the period and one of the phase for each cosine term were easily determined from inspection of data plots. The phase was verified and estimated more precisely with the NLIN procedure in SAS (SAS Institute, Inc. 1988, 675-712).

Variables were screened for effect on the lymphocyte profile by regressing the residuals from the current model
on them (semi-partial plots\textsuperscript{15}). The assumption was made, at
this stage that there was no covariance between the
independent variable being tested and the other variables
already in the model. In the absence of prior definitive
data on patterns of circadian and circannual rhythms, the
data were used to decide upon transformations for hour-of-day
and month-of-year from longitudinal variables to cyclic
variables. If a variable was found to contribute toward the
residual variance, the residual means by categories of the
variable were plotted to determine the best modelling
strategy for the variable. Then it was entered into the
model. Each time that a variable was added to the model,
the model was reassessed by backward elimination of the
interaction terms from a full model with all 2-way and 3-way
Terms were eliminated in order of least significance,
starting with the 3-way interactions. With models
containing 7 or 8 main effects, 3-way interaction terms were
added in groups of 5 or 6 and removed at once if \( P(>F) > 0.50 \). Once all of the 3-way interactions had been tried
with the exclusion criterion of 0.50, those remaining in the
model were excluded in order of least significance in the

\textsuperscript{15}In partial regression, both the dependent variable and
the independent variable are adjusted for the other
variables in the model and the relationship of the residuals
examined. In semi-partial regression, the dependent
variable is adjusted for other variables in the model, but
the independent variable is assumed to have no covariance
with other independent variables.
usual manner. The criterion for remaining in the model was an approximate\(^{16}\) F statistic with a probability of 0.05 or less. After new variables were added to the model, variables previously excluded were retested against the new residual variance.

**Isolation of single effects.**

Adjustment of the data for a factor or a combination of factors was conducted in the following way. The data were modeled as a linear combination of factors \(F_i\):

\[
\text{Ln(CD3+CD22), Ln(CD56+CD22)} = \text{Mean}_0 + F_i + \ldots + F_{n-1} + F_n + R_n^{17} = \text{Predicted} + R_n
\]

Data adjusted for factors \(F_1 - F_n\) were simply

\[
[\text{Ln(CD3+CD22), Ln(CD56+CD22)}]_{\text{adjusted}} = \text{Mean}_0 + R_n
\]

The effect of \(F_n\) was estimated as the Predicted from the model which included \(F_n\) minus the Predicted from the model without \(F_n\).

\[
F_n = [\text{Mean}_0 + F_1 + \ldots + F_{n-1} + F_n] - [\text{Mean}_0 + F_1 + \ldots + F_{n-1}]
\]

\(^{16}\)Since division of a covariance matrix by a covariance matrix does not yield a scalar quantity, test statistics for multivariate analysis are constructed from the determinates or traces or eigenvalues of the hypothesis and error matrices and have nearly an F distribution (Barker and Barker 1984).

\(^{17}\)R represents the residual or "error" variance element.
To isolate the effect $F_B$ of a variable $B$ which was involved in an interaction,

$$\ln(\text{CD3+CD22}), \ln(\text{CD56+CD22}) = \text{Mean}_0 + F_A + F_B + F_{A\times B} + R,$$

the data were adjusted for the effect $F_A$ of $A$, then further adjusted for the effect $F_{A\times B}$, of $A\times B$ estimated from the full model as described above, and the doubly adjusted data examined by levels of $B$.

**Other tests.**

For correlations and tests involving a proportion, rather than a ratio of proportions, a logit transformation ($\ln[p/(1-p)]$) was used to normalize the distributions (Rothman 1986). The SAS procedures MEANS and CORR were used to conduct paired t-tests and to test correlations.

**Ternary diagrams.**

It was often useful to make a visual representation of a three-component mixture on a two-dimensional ternary diagram (Aitchison 1986), an equilateral triangle on which each apex represents 100% of a particular component and the opposite side is 0% of that component (Figure 7). The legs which extend from an apex are scaled linearly from 100% to 0% for that component. Any point which falls on the straight line which connects $X\%$ on one leg to $X\%$ on the other contains $X\%$ of that component. Therefore the percentage of a component can be read from the sides of the triangle by
Figure 7. The ternary diagram.

projecting a line, parallel to the 0% base from the point to either side. To plot the profile points using existing SAS software the percentage of T-cells was aligned with the Y axis, and the 100% B-cell apex was placed at (0,0), then the lymphocyte composition profile was plotted on a rectangular coordinate system using the following transformation:
\[ Y = \% \text{CD}3^+, \]
\[ X = 100 - (\% \text{CD}3^+)^2 - \% \text{CD}22^+. \]

The printing of the rectangular coordinates was suppressed and a triangular grid was overlaid for reference (Figure 8).

**The leukocyte profile**

The leukocyte profile was examined in manner similar to that used for the lymphocyte profile. Dependence on disease, age, smoking, month of year, and hour of day was tested. The composition was represented by four logratios of neutrophils, monocytes, eosinophils and basophils to lymphocytes:

\[ \ln(\text{NEUT}\div\text{LYM}), \ln(\text{MONO}\div\text{LYM}), \ln(\text{EOS}\div\text{LYM}), \text{ and } \ln(\text{BAS}\div\text{LYM}). \]

Where the count of one of the numerator subsets was 0, the count was set to 0.25, in order to preclude calculating \( \ln(0) \) as a datum.
Figure 8. Plotting a ternary diagram in rectangular coordinates.
RESULTS

Study subjects.

Since patients with small-cell (oat-cell) carcinomas were almost always treated with chemotherapy prior to any radiation treatment, this major cell-type was not represented in the patient population. Similar numbers of patients with the other three major histologic types of lung cancer were enrolled. Seventy new white male patients with primary lung cancer of the 3 histologic types eligible for the study entered treatment at PCC between January 29, 1988 and January 6, 1990. Of these 25 of 26 (96%) with squamous-cell carcinoma, 20 of 24 (83%) with large-cell carcinoma and 19 of 20 (95%) with adenocarcinoma\(^\text{18}\) volunteered to participate. There was no evidence of differences in smoking history among the three histologic groups. In fact of those with the least smoking-related histotype, adenocarcinoma, only 4 of 17 (23.5%) were ex-smokers of over a year, whereas 52% of those with squamous cell and 47% of those with large cell carcinoma reported having quit the

\(^{18}\) The proportion representing each histologic type should not be taken as representative of incidence for the geographic area which Mary Bird Cancer Center serves since radiation therapy is not necessarily appropriate for the same proportion of patients from each group.
habit more than 1 year ago, but the differences were statistically non-significant.

The origins of comparison subjects are detailed in Table 4. In all 219 comparison subjects were enrolled.

<table>
<thead>
<tr>
<th>Group</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Assn. of Retired Persons</td>
<td>26</td>
</tr>
<tr>
<td>Council on Aging</td>
<td>7</td>
</tr>
<tr>
<td>Kiwanis International</td>
<td>34</td>
</tr>
<tr>
<td>Lions Clubs</td>
<td>10</td>
</tr>
<tr>
<td>American Legion</td>
<td>61</td>
</tr>
<tr>
<td>Amer. Leg. Veterans Home, Jackson, LA</td>
<td>24</td>
</tr>
<tr>
<td>Veterans of Foreign Wars</td>
<td>10</td>
</tr>
<tr>
<td>Salvation Army</td>
<td>6</td>
</tr>
<tr>
<td>University Baptist Church</td>
<td>8</td>
</tr>
<tr>
<td>Knights of Columbus</td>
<td>3</td>
</tr>
<tr>
<td>Referred by third party</td>
<td>4</td>
</tr>
<tr>
<td>Brochure, other publicity</td>
<td>15</td>
</tr>
<tr>
<td>Acquaintances</td>
<td>11</td>
</tr>
</tbody>
</table>

219
Response to questionnaires.

Of the total number of subjects enrolled, 16 controls and 4 patients failed to return questionnaires even after repeated reminders.

The dietary data for five controls and one patient were considered unusable because of too many items skipped and failure to indicate portion sizes. In most cases, missing portion sizes were recoded to "medium" since examination of the questionnaires indicated that the respondents were otherwise diligent in completing them. No one was eliminated for eating too few or too many different foods on a daily basis, because their questionnaires indicated that they ate an unremarkable variety of food items overall and seemed merely to have difficulty in estimating how often they ate certain foods. Of those who returned questionnaires 244 (55 patients and 189 controls) had usable dietary data. Of these 147 were considered to have no problems, 22 had what were judged to be miscellaneous minor errors, 64 had marked too high a percentage of the items with a medium or a small portion size, and 11 indicated too many or too few foods daily.

Comparisons of patient and control groups on age, parish of residence, income per person in the household, and education are presented in Figures 9-12. The age range of patients was well represented in the control group. There were subjects among the controls, but not among patients,
who were in their fourth decade. A majority of the controls (54%) and nearly half of the patients (46%) reported residence in East Baton Rouge Parish. The controls who resided in East Feliciana Parish came from the War Veterans' home in Jackson and presumably did not live the majority of their lives there, but had resided in other parishes. When household income was calculated on the basis of the number of individuals who shared it, 85% of the patients reported less than $10,000 per year per person whereas only 45% of the controls fell into that category. Controls also reported more education than patients: 73% of the controls had at least finished high school compared to 50% of patients. Eighteen percent of the patients reported less than an eighth grade education compared to 4% of controls. Nevertheless there were representatives in the control group of every subgroup of patients with respect to income and education such that control for these measures of socioeconomic status was possible.

The subjects grouped according to disease and smoking characteristics as illustrated in Figure 13. No patient was a life-time non-smoker. One individual had smoked only a pipe but had that exposure for 45 years. He was not included in the analysis. Despite efforts to recruit smokers, only 63 of the comparison subjects were current smokers. Smokers in the older age groups proved least available.
Distribution by Age

Figure 9. Age distribution of patients and controls.

Sample-to-sample spillover.

CD3 samples were run in duplicate in order to get an estimate of sample-to-sample variability. However evidence
Figure 10. Income classification of patients and controls.

was found that in some cases not enough time had been allowed between running of sequential samples on the FACS and that the first CD3 sample in some subjects was
contaminated with residual control cells. A subsample of subjects (n=112) was analyzed. The paired differences for the proportion positive in the second sample minus that of the first averaged to 0.0152 (SE=0.0045) and was
Figure 12. Educational status of patients and controls. Significantly different from 0 at P<0.0011. If the paired differences were graphed, the largest group of individuals clustered around 0 and 10 individuals formed a broad tail on the positive end of the difference distribution. Without
Figure 13. Characterization of subjects on disease and smoking.

the 10 outliers, the averaged paired differences were not different from 0 and the estimate for the standard deviation in determinations from the two samples per 102 individuals was 2.7%. When the rest of the samples were considered separately, virtually identical results were obtained, with a mean difference of 0.015 caused by 12 individuals in the tail of the distribution. Since the first samples of at least 22 individuals were contaminated, the estimate from the second sample was used for analysis for everyone.
Only one other problem with spillover from the previous sample was observed. The HLA-DR samples were preceded by the antibody W6/32 to a common leukocyte antigen and 20 files were seen to be contaminated. By going back to the original data which contained all leukocytes, an estimate of the percentage of contamination could be made by determining the percentage of granulocytes which displayed green fluorescence above background. Corrections were applied to the HLA-DR estimates.

**Performance of the new data analysis method.**

An average of 2041 (±376) lymphocytes were selected and analyzed. Only two of 25 randomly selected subjects had greater than 1% contamination with monocytes as identified by the antibody 63D3 (CD14).

Occasionally extraneous events (doublets, for instance) occurred at a distance from the minor component on fluorescence or small numbers (<20) events grouped at 0 fluorescence and influenced estimation of the distributions. The normal-mixtures-fitting routine treated each event as though it belonged to one of the specified subsets and adjusted parameter estimates to accommodate it. Where difficulties arose, such obviously irrelevant events were edited out of the data so that estimates would converge to realistic values.

In some instances the curve-fitting routine seemed to 'detect' the presence of more subpopulations than were
theoretically there. For instance with the CD4\(^+\)CD29\(^+\) doubly-labelled cells, there seemed to be more than 4 distinguishable populations. In particular, there were small numbers of cells which labelled very brightly on red fluorescence (CD29-PE). In 11 cases, 5 groups were needed to fit the data and in 22 cases, 6 groups were needed. The proportions positive in both groups were summed to give a total estimate for CD29-bearing CD4\(^+\) lymphocytes.

Cells which labelled with HLA-DR formed a broad peak of low amplitude. A more satisfactory solution was obtained if two distributions were fitted to the positive cells in the majority of cases. This was taken as an indication that distinct subgroups of lymphocytes express the class II major histocompatibility antigen at different densities. However, for the purposes of this study, the proportions were combined as one positive population. In 81 cases, a single positive distribution fitted better than 2.

Beginning around subject 205, a small (1-2\%), third population of cells appeared in the samples labelled with CD56, the NK marker. These cells exhibited brighter fluorescence than those presumed to be the NK, but had the side-scatter characteristics of average lymphocytes. Their percentage was estimated and because of their side-scatter characteristics added to that of non-NK. The brightness of the NK was also diminished in these later samples.
## TABLE 5

PAIRED DIFFERENCES OF ESTIMATES BY TWO METHODS

<table>
<thead>
<tr>
<th>Marker</th>
<th>Mean Difference</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD3</td>
<td>-0.0044</td>
<td>0.024</td>
</tr>
<tr>
<td>CD4</td>
<td>0.0116</td>
<td>0.032</td>
</tr>
<tr>
<td>CD8\text{bright}</td>
<td>0.0168</td>
<td>0.028</td>
</tr>
<tr>
<td>CD8\text{dim}</td>
<td>-0.0282</td>
<td>0.028</td>
</tr>
</tbody>
</table>

The estimates of CD3⁺, CD4⁺, CD8⁺\text{bright}, and CD8⁺\text{dim} done by the conventional method for 27 subjects were compared with estimates for those subsets done by the new method. The paired differences (Table 5) were not different from 0, indicating that both methods gave the same estimates overall. The standard deviation on an individual difference, however, was 0.14 or 14 percentage points, which caused the standard error of the mean difference to be on the order of 3%. Therefore a mean paired difference of even 6% would not have been considered significant. In other words the power of this comparison of 27 subjects to detect a difference in estimated mean percentage positive of 6% was only 50%.

The standard deviations of estimates by both methods for the group of 27 subjects are presented in Table 6.
TABLE 6

VARIABILITY OF ESTIMATES WITHIN METHOD

<table>
<thead>
<tr>
<th>Marker</th>
<th>Std. Dev.</th>
<th>Marker</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD3</td>
<td>0.095</td>
<td>CD3</td>
<td>0.081</td>
</tr>
<tr>
<td>CD4</td>
<td>0.109</td>
<td>CD4</td>
<td>0.113</td>
</tr>
<tr>
<td>CD8_{bright}</td>
<td>0.084</td>
<td>CD8_{bright}</td>
<td>0.090</td>
</tr>
</tbody>
</table>

Although the variances of the estimates appear to be the same, the estimates from the old method were based on an average of 2990 (±1330) events, whereas the longer files were truncated in the new method and the estimates were based on 2041 cells on the average. Therefore the new method was somewhat more precise in estimating the percentage positive.

Lymphocyte subsets.

Of the 17 lymphocyte markers measured, the eight indicated in Figure 14 were selected for first stage analysis on the basis that (1) they are commonly estimated markers, (2) they are relevant to possible immune defenses against cancer, and (3) the positive and negative populations, with the exception of the 2H4 and 4B4 subsets of CD4-positive cells, were relatively easily discriminable,
Figure 14. Subsets chosen for first-stage analysis

Therefore more accurately estimable. As one can see from Figure 14, some markers define subsets of sets. For example, nearly all cells which carry CD3 also bear CD4 or CD8. Likewise most CD4+ lymphocytes are considered to carry either 2H4 or 4B4 although some 15-30% have been reported to carry neither (Morimoto 1985). T-cells, B-cells, and natural killer cells are the component subsets of lymphocytes. The only lymphocyte commonly recognized as carrying the Class II major histocompatibility marker HLA-DR is the B-cell, however activated T-cells and monocytes express HLA-DR as well. Therefore the following relationships between the proportions of cells having particular markers were expected to hold:
CD4\(^+\) + CD8\(^+\) = CD3\(^+\)
CD45R\(^+\) + CDw29\(^+\) < CD4\(^+\)
CD3\(^+\) + CD22\(^+\) + CD56\(^+\) = 1.00
HLA-DR\(^+\) >= CD22\(^+\)

These relationships were probed utilizing all subjects for whom there were data. The distributions of estimated proportions were normalized by employing a logit transformation, \(ln[p/(1-p)]\).

Relevant statistics for the above subsets and combinations of subsets are presented in Table 7. The mean proportions and confidence intervals were determined using the logit-transformed variables and back-transformed. The Pearson correlation coefficients \(R\) are given for appropriate comparisons among the transformed variables and \(Pr(>|R|)\) under the hypothesis that there is no correlation was 0.0001 for each case.

The paired differences corresponding to the relationships above were tested for equality to zero and the results are presented in Table 8. The sum of the proportions of cells CD3-, CD56-, and CD22-positive accounted for all lymphocytes, that is, they summed to 1.00 as would be expected if the 3 antibodies do indeed label mutually exclusive subsets of lymphocytes and each labels 100\% of its specific subset. The differences \((1 - P_{CD3} - P_{CD56} - P_{CD22})\) were not normally distributed, due predominantly to kurtosis (the distribution was symmetrical); the mean was
<table>
<thead>
<tr>
<th>MARKERS</th>
<th>N</th>
<th>MEAN</th>
<th>95% C.I.</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4+CD8&lt;sub&gt;i&lt;/sub&gt;</td>
<td>277</td>
<td>0.659</td>
<td>0.46-0.82</td>
<td></td>
</tr>
<tr>
<td>[ln{p/(1-p)}]</td>
<td></td>
<td>[0.661]</td>
<td>[-0.17-1.49]</td>
<td></td>
</tr>
<tr>
<td>CD3</td>
<td>275</td>
<td>0.709</td>
<td>0.48-0.86</td>
<td>0.71</td>
</tr>
<tr>
<td>[ln{p/(1-p)}]</td>
<td></td>
<td>[0.892]</td>
<td>[-0.07-1.85]</td>
<td></td>
</tr>
<tr>
<td>CD4+CD8&lt;sub&gt;i&lt;/sub&gt;+CD8&lt;sub&gt;2&lt;/sub&gt;</td>
<td>276</td>
<td>0.751</td>
<td>0.55-0.88</td>
<td>0.68</td>
</tr>
<tr>
<td>[ln{p/(1-p)}]</td>
<td></td>
<td>[1.105]</td>
<td>[0.18-2.03]</td>
<td></td>
</tr>
<tr>
<td>CD4</td>
<td>279</td>
<td>0.431</td>
<td>0.23-0.65</td>
<td>0.77</td>
</tr>
<tr>
<td>[ln{p/(1-p)}]</td>
<td></td>
<td>[-0.248]</td>
<td>[-1.19-0.60]</td>
<td></td>
</tr>
<tr>
<td>CD45R+CD29</td>
<td>223</td>
<td>0.356</td>
<td>0.17-0.60</td>
<td></td>
</tr>
<tr>
<td>[ln{p/(1-p)}]</td>
<td></td>
<td>[-0.592]</td>
<td>[-1.60-0.42]</td>
<td></td>
</tr>
<tr>
<td>CD3+CD56+CD22</td>
<td>276*</td>
<td>1.003</td>
<td>0.80-1.20</td>
<td></td>
</tr>
<tr>
<td>CD22</td>
<td>279</td>
<td>0.090</td>
<td>0.03-0.26</td>
<td>0.43</td>
</tr>
<tr>
<td>[ln{p/(1-p)}]</td>
<td></td>
<td>[-2.311]</td>
<td>[-3.55+1.07]</td>
<td></td>
</tr>
<tr>
<td>HLA-DR</td>
<td>279</td>
<td>0.182</td>
<td>0.09-0.34</td>
<td></td>
</tr>
<tr>
<td>[ln{p/(1-p)}]</td>
<td></td>
<td>[-1.504]</td>
<td>[-2.35-(-0.66)]</td>
<td></td>
</tr>
</tbody>
</table>

* Six subjects with values below 0.75 were excluded.
### TABLE 8

**TESTS OF SUBSET EQUIVALENCE HYPOTHESES**

| HYPOTHESIS                                | STUDENT'S t | P(>|t|) |
|-------------------------------------------|-------------|---------|
| CD3-(CD4+CD8_{1+2})=0                    | -10.330     | 0.0001  |
| CD3-(CD4+CD8_{1})=0                      | 10.01       | 0.0001  |
| CD4-(CD42H4+CD44B4)=0                    | 16.31       | 0.0001  |
| 1-(CD3+CD56+CD22)=0                      | 0.515       | 0.607   |
| (HLA-DR)-CD22=0*                         | 31.52       | 0.0001  |

* [logit(HLA-DR - CD22+0.08)-logit(0.08)=0]

0.003, standard error of the mean (SEM) was 0.0058 indicating that the estimate was precise, and the median of the differences was 0.001, indicating that the distribution was symmetrical and that the mean was an adequate measure of central tendency.

There were at least 2 subsets of lymphocytes which bore CD8, one at high density and another at lower density. The proportions of lymphocytes which were CD4⁺ or CD8⁺ (including both dim and bright cells with CD8) summed to a total greater than the proportion CD3-positive (mean paired
difference = -0.045, \( p(>|t|) < 0.0001 \). If the dim cells are considered to be CD3\(^-\) natural killer cells (Ault 1988) and only the brighter cells were summed with CD4 then the proportion (CD4\(^+\) + CD8\(_{\text{bright}}^+\)) was less than the proportion CD3\(^+\) (mean paired difference = 0.044, \( p(>|t|) < 0.0001 \)), implying that there were some T-cells present which had neither CD4 nor CD8\(_{\text{bright}}\). Likewise, the CD45R\(^+\) and CD29\(^+\) subsets of CD4-positive cells did not sum to the proportion estimated as CD4-positive by independent labelling (0.36 v. 0.44). The correlations of CD3\(^+\) with CD4\(^+\) + CD8\(_{\text{total}}^+\), CD4\(^+\) + CD8\(_{\text{bright}}^+\), and of CD4\(^+\) with CD4\(^+\)4B4\(^+\) + CD4\(^+\)2H4\(^+\) were high, (0.68, 0.71 and 0.77, respectively).

The logit transformation normalized the distributions of proportions positive for HLA-DR and CD22. It was necessary to transform the quantity (HLA-DR\(^+\) - CD22\(^+\)) by adding the constant 0.08 before logit transformation so that the difference was a positive number for all subjects. The distribution of this logit was normal except for 2 outliers whose logit(difference) were greater than 3 standard deviations below the mean and 9 individuals who clustered around a value 2.7 standard deviations above the mean (Table 9). It was assumed that the outliers were individuals for whom something had gone awry in the data collection or analysis process and they were deleted in this comparison. The equivalent of the hypothesis that HLA-DR\(^+\) - CD22\(^+\) = 0 for the transformed difference was \( H_0: \logit(\text{HLA-} \)
DR⁺ - CD22⁺ + 0.08) = logit(0.08) = -2.442. The mean of the transformed difference, even with outliers excluded, was -1.70 (t₁₆₈ = 31.52, P(>t) < 0.0001). The proportion of cells expressing HLA-DR, then, exceeded the estimated proportion of B-cells by an average of 7.6%. Furthermore ANOVA showed that the patients, with 11.7% more HLA-DR⁺ cells than B-cells, had a significantly greater excess of HLA-DR⁺ lymphocytes (F₁,₂₇₄ = 46.75, P < 0.0001) than controls (6.6% more HLA-DR⁺ cells than B-cells).

The lymphocyte profile.

The first variables to enter the regression model were disease, age, smoking, hour-of-day, and month-of-year. Modelling decisions with respect to the last 3 were data-based.

It was reported in the literature that the effects of smoking on the immune system resolved within less than a year after quitting. Profile logratio means for each of five categories of smoking status are shown in Figure 15. Those who had stopped smoking at least 1 year were very much like those who had never smoked, while those who had stopped within the year in which the sample was drawn, clustered with current smokers. Therefore subjects were dichotomized on smoking.

Estimation of and adjustment for circadian and circannual variation in the lymphocyte profile within the
TABLE 9

SUBJECTS DELETED BECAUSE OF EXTREME VALUES OF (HLA-DR) - (CD22)

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>HLA-DR</th>
<th>CD22</th>
<th>DIFFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>C099</td>
<td>0.451</td>
<td>0.083</td>
<td>0.368</td>
</tr>
<tr>
<td>P006</td>
<td>0.411</td>
<td>0.045</td>
<td>0.366</td>
</tr>
<tr>
<td>C256</td>
<td>0.431</td>
<td>0.110</td>
<td>0.321</td>
</tr>
<tr>
<td>P120</td>
<td>0.472</td>
<td>0.159</td>
<td>0.313</td>
</tr>
<tr>
<td>P005</td>
<td>0.321</td>
<td>0.011</td>
<td>0.310</td>
</tr>
<tr>
<td>P051</td>
<td>0.348</td>
<td>0.046</td>
<td>0.302</td>
</tr>
<tr>
<td>C184</td>
<td>0.346</td>
<td>0.047</td>
<td>0.299</td>
</tr>
<tr>
<td>C114</td>
<td>0.363</td>
<td>0.080</td>
<td>0.283</td>
</tr>
<tr>
<td>P092</td>
<td>0.387</td>
<td>0.105</td>
<td>0.282</td>
</tr>
<tr>
<td>C143</td>
<td>0.256</td>
<td>0.289</td>
<td>-0.033</td>
</tr>
<tr>
<td>C277</td>
<td>0.035</td>
<td>0.107</td>
<td>-0.072</td>
</tr>
</tbody>
</table>

constraints of a linear model required the transformation of longitudinal time units into cyclic variables. Inspection of plots of the profile logratios ln(CD3+CD22) and ln(CD56+CD22), adjusted for age, disease, smoking, and an age-by-smoking interaction (Figures 16 and 17), suggested that cosine functions of the time units could be utilized in constructing effective transformations. Therefore two cosine functions with different periods were used to model
SMOKING IN FIVE CATEGORIES

A = CURRENT SMOKER
B = QUIT < 1 YEAR AGO
C = QUIT 1-4 YEARS AGO
D = QUIT 5-20 YEARS AGO
E = QUIT >20 YEARS AGO
OR NEVER SMOKED

Figure 15. Profile adjusted for age and disease, by smoking.
Figure 16. Means of adjusted profile logratios by month of year.
Figure 17. Means of adjusted profile logratios by hour of day.
the effect of each of the time variables month and hour, and
the transformation consisted of a linear combination of the
two functions. The adjusted-logratio plots were used to
estimate periods for the circannual variation and those for
the circadian variation; these were assumed to be constant
throughout the year and day respectively. Proc NLIN in SAS
(SAS Institute, Inc. 1988, 675-712) was utilized to estimate
the acrophase for each cosine function and the acrophase was
also assumed to be constant over the day or year for the
component with the shorter period. The residuals of the
profile logratios adjusted for other factors in the model
were then regressed on the linear combination of the two
cosine functions and an interaction term for the two
functions to estimate the appropriate coefficients for the
best linear combination. The transformations used were

\[
T(\text{month}) = 0.545 \cos[(2\pi \times 12)(\text{month} - 5.17)] \\
+ 0.967 \cos[2\pi \times 4(\text{subyear} - 1.7)] \\
+ 3.91 \cos[(2\pi \times 12)(\text{month} - 5.17)] \cos[(2\pi \times 4)(\text{subyear} - 1.7)], \text{and} \\
T(\text{hour}) = 0.102 \cos[(2\pi \times 24)(\text{hour} - 8.3)] \\
+ 0.123 \cos[(2\pi \times 6)(\text{subday} - 2.45)].
\]

An interaction term did not contribute significantly to the
fit of residuals to cosine functions of hour, but seemed
important for the transformation of month.

Residuals from the 5-factor model including age,
disease, smoking, \(T(\text{month})\), and \(T(\text{hour})\) were regressed on
weekly vegetable intake (VEG) and VEG contributed to the
explanation of the residual variance at the 0.05 level. The distribution of VEG was examined and 2 outliers were detected at 72 per week and 97 per week. Since these were unlikely to reflect true consumption rates and it was undesirable for them to have undue influence on the contribution of VEG to the model, the values for these individuals were recoded to 55 per week, a value just above the highest value for the other subjects. Furthermore when one categorized VEG: 0-6, 7-14, 15-20, 21-27, 28-34, and >34 per week, a threshold effect was suggested with the break point at 14 per week. The dichotomous VEG contributed at a higher level of significance than continuous VEG toward explaining the variance of the profile adjusted for the other variables, therefore VEG was considered to be dichotomous.

Similarly FAT intake as percent of daily calories was an important determiner of lymphocyte profile and performed better as a dichotomous variable with the breakpoint at 28% of daily calories, than as continuous variable.

Finally alcohol (ALC), expressed as the percent of daily calories which come from alcoholic drinks, was found to be important to explain variation in the lymphocyte profile. Eight effects in all were found which accounted for significant portions of the variance of the profile: disease, age, smoking, hour-of-day, month-of-year, vegetable intake, fat intake, and alcohol intake.
As variables were added to the model the transformations of month and hour were reevaluated. When T(month) was left out and a model built containing the other 7 effects disease became a non-significant contributor and a very simple model resulted:

\[ \ln(CD3+CD22), \ln(CD56+CD22) = \text{AGE SMOKING VEGETABLES FAT ALCOHOL T(hour) AGE-by-SMOKING}. \]

Yet when T(month), modelled as the linear combination (of 2 cosine terms and their interaction) above was added, the model suddenly expanded to include 3 3-way interaction terms and 14 2-way interaction terms, beyond the 8 main effects. This suggested that month was being overmodeled. Examination of a partial-partial plot, that is, a plot of the data adjusted for the above model (excluding T(month) and disease) against month adjusted for the same variables (Figure 18), a transformation involving only a cosine of month with a 12-month period was suggested. Thus the transformation of month was revised to

\[ T(month) = \cos(2\pi \pm 12(\text{month}-4.93)). \]

When the residuals from the above model (excluding month and disease) were regressed on T(month), T(month) "explained" 36% of that variance of the residual \( \ln(CD3+CD22) \) by month and 72% of that variance of the residual \( \ln(CD56+CD22) \) by
Figure 18. The adjusted data against adjusted month.
month\textsuperscript{19} which could be accounted for by regression on categorical month. Using this revised transformation of month, a final 8-factor model was derived (Table 10). This model "explained" 28\% of the variance in $\ln(\text{CD3}+\text{CD22})$ and 43\% of the variance in $\ln(\text{CD56}+\text{CD22})$. If one adjusts for the degrees of freedom contained by the model (Neter, Wasserman, and Kutner 1985, 241),

$$R^2 = 1 - [(n-1)+(n-p)](1 - R^2)$$

$$= 1 - [233+223](1-R^2)$$

$$= 1.045(R^2) - 0.045,$$

the adjusted $R^2$'s are 25\% for $\ln(\text{CD3}+\text{CD22})$ and 40\% for $\ln(\text{CD56}+\text{CD22})$.

No subject who had missing data on any of the independent variables or on any of the components of the lymphocyte profile could be used in the analyses based on the 8-factor model. The model building used the data from 234 subjects. They did not differ appreciably from the

\textsuperscript{19}The denominator for determining the proportion of residual variance assignable to $T(\text{month})$ was the proportion of total residual variance that could be attributed to the means of the residuals by month. The proportion of the total variance which was the variance of the monthly mean residuals from the overall mean residual was determined by regressing the residuals on categorical month, i.e. a perfect fit of the model to the monthly means:

$$R_{\text{CD3/CD22}}, R_{\text{CD56/CD22}} = \text{Mean} + \text{Eff(mo)}_{1\text{ld.f.}} + \text{Resid}_B$$

$$R_{\text{CD3/CD22}}, R_{\text{CD56/CD22}} = \text{Mean} + T(\text{month}) + \text{Resid}_A + \text{Resid}_B$$

$$T(\text{month})/\text{Eff(mo)}_{1\text{ld.f.}} = \text{proportion of monthly variance attributable to } T(\text{month})$$.
TABLE 10

FINAL MODEL FOR THE LYMPHOCYTE PROFILE

TYPE III TESTS, P(>F)

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>Ln(CD3/CD22)</th>
<th>Ln(CD56/CD22)</th>
<th>MANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISEASE</td>
<td>0.0052</td>
<td>0.0253</td>
<td>0.0169</td>
</tr>
<tr>
<td>AGE</td>
<td>0.0009</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>SMOKING</td>
<td>0.7492</td>
<td>0.0237</td>
<td>0.0067</td>
</tr>
<tr>
<td>VEGETABLES</td>
<td>0.0060</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>FAT</td>
<td>0.0001</td>
<td>0.0175</td>
<td>0.0004</td>
</tr>
<tr>
<td>ALCOHOL</td>
<td>0.0003</td>
<td>0.0058</td>
<td>0.0013</td>
</tr>
<tr>
<td>MONTH</td>
<td>0.2912</td>
<td>0.0080</td>
<td>0.0219</td>
</tr>
<tr>
<td>HOUR</td>
<td>0.0004</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>DISEASE-by-VEGETABLES</td>
<td>0.0385</td>
<td>0.0069</td>
<td>0.0237</td>
</tr>
<tr>
<td>DISEASE-by-MONTH</td>
<td>0.0053</td>
<td>0.0029</td>
<td>0.0061</td>
</tr>
<tr>
<td>AGE-by-SMOKING</td>
<td>0.8543</td>
<td>0.0026</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.281 \quad 0.431 \]

* F_{11,222} statistics for Wilk's Lambda, Pillai's Trace, and Hotelling-Lawley Trace were all equal

**Uncorrected
composition of the larger group on age, income, parish of residence, or education (Figures 19-22).

Circadian variation.

A linear combination of two cosine functions with different periods was needed to model the circadian variation. The data spanned only 16 hours of the day so there was no information about variation between 11 p.m. and 7 a.m. The data suggested that there might be a post-prandial association; maxima occur near 8:30 a.m., 2:30 p.m., and 8:30 p.m. There did not seem to be any confounding of the hour-of-day term with other variables in the final model. Means of logratios adjusted for all other factors in the model were computed for the 3 maxima and 2 minima spanned by the data and the corresponding profiles were plotted in Figure 23. There is more variation from "maximum" to "minimum" than one would expect for a three-hour period. For instance, B-cells more than double, from 6.9% to 15.4% between 2:30 p.m. and 5:30 p.m.

The modeling of the circadian or hour-of-day effect on the lymphocyte profile as the linear combination of two cosine functions of hour of day performed well in accounting for the variance of the logratios with hour-of-day (Figure 24). When the profile logratios were adjusted for all effects except hour, T(hour) could claim 57% of the variance of the adjusted ln(CD3+CD22) means by hour and 81% of the
Figure 19. Distribution of subjects in the 8-factor model by age.
Figure 20. Distribution of subjects in the 8-factor model by income.
Figure 21. Distribution of subjects in the 8-factor model by parish.
Figure 22. Distribution of subjects in the 8-factor model by education.
CIRCADIAN RYTHYM

Adjusted for All Other Effects

Figure 23. Circadian variation in the lymphocyte profile.
Figure 24. The circadian variation of the lymphocyte profile logratios.
variance of the adjusted ln(CD56+CD22) means by hour\(^{20}\). Furthermore no interactions with T(hour) were significant when T(hour) was added to the model.

**The T(month)-by-disease interaction.**

Confounding arises in consideration of circannual variation in the lymphocyte profile and the difference in the lymphocyte profile between lung cancer patients and controls because subject composition with respect to disease varied with month of year (Figure 25). This might be the reason that the effect of disease was not apparent in the model which excluded month. Furthermore there is a significant T(month)-by-disease interaction effect, i.e. the patients' profile varies differently over the year from the controls' profile. The interaction term in the model can be visualized as a correction factor of sorts. The data means adjusted for all effects in the model except the interaction term for T(month)-by-disease are plotted by month in Figures 26 and 27. Notice the almost complementary distribution of adjusted logratio means of patients by month, compared to the T(month) incorporated into the model (see 'predicted' curves in Figures 32 and 33). The predicted (in Figures 26 and 27) are the predicted from the full model minus the effects of all of the components of the

\(^{20}\)Calculated in the same manner as the proportion attributable to T(month); see previous footnote.
Figure 25. The distribution of subjects by month enrolled in study.
Figure 26. \( \ln(T/B) \) MONTH BY DISEASE EFFECT: The interaction between T(month) and disease.
Figure 27. Ln(CD56/CD22): The interaction between T(month) and disease.
Figure 28. Ln(CD3/CD22) adjusted for all effects, including disease and disease-by-vegetable, except T(month) and T(month)-by-disease.
Figure 29. Ln(CD56/CD22) adjusted for all effects, including disease and disease-by-vegetable, except T(month) and T(month)-by-disease.
model except $T(\text{month})$-by-disease, in other words, $\text{Mean}_0 + F_{T(\text{month})\text{-by-disease}}$.

Without $T(\text{month})$ the disease-by-vegetable interaction is significant at only $p<0.0567$ and disease is significant as a main effect at $p<0.02$. When the disease-by-vegetable interaction is removed, the effect of disease becomes non-significant. But if one retains both the main effect of disease and the disease-by-vegetable interaction in the model without $T(\text{month})$ or the $T(\text{month})$-by-disease interaction and plots the data adjusted for this model by month (Figures 28 and 29) one definitely observes a different month effect for patients than for controls. Patients seem to experience maxima in September and minima around the first of April. It is somewhat difficult to identify trends in the patient data, however, because of the wide confidence intervals on the estimates.

Because the monthly variations of profiles for patients and controls are about $180^\circ$ (6 months) out of phase, one would expect that the difference between patients' and controls' profiles would be maximal around May and November. If one adjusts the data for all variables except month and disease, one can see in Figure 30 that the profiles of patients and controls are indeed widely separated. The profile logratios are plotted in Figure 31. There was only one patient in November who had all of the data to be
Figure 30. Maximum separation of patients's and controls's profiles.
**DISCRIMINATING DISEASE, ln(T/B) Means**

- P - Patients
- C - Controls
- 95% C. Limit
- Adjusted Means

**DISCRIMINATING DISEASE, ln(NK/B) Means**

- P - Patients
- C - Controls
- 95% C. Limit
- Adjusted Means

**Figure 31.** Profile logratios of patients and controls at months of maximum separation.
data to be included in the profile analysis, so there is no confidence interval on that profile.

**Circannual variation.**

If one isolates the effect of month by adjusting for all effects in the model except month and $T(\text{month})$-by-disease, then subtracting the effect of $T(\text{month})$-by-disease (estimated as the predicted from the full model minus the predicted from the model without the month-by-disease term), then plots the adjusted means by month, one sees the plots in Figures 32 and 33. The 'predicted' is that from regressing the adjusted data on $T(\text{month})$. Both logratios exhibit maxima in late April and minima in late October. The adjusted data for patients also fit the pattern seen in controls because of the adjustment for $T(\text{month})$-by-disease. It is difficult to graph profiles by month on a ternary diagram because there are so many points, but four points representing the maximum, the minimum, and a point midway between on either side of the maximum were graphed (Figure 34) to illustrate the cycle. The profiles represent the mean adjusted profiles for January and February, April and May, July and August, and October and November. The ternary plot also illustrates how deceptive and inadequate the logratio plots are for interpreting changes.

These data would suggest then that the lymphocyte profile does vary with month of year and that Caucasian men with lung cancer experience a cycle which is about 180° out
Figure 32. \( \ln(T/B) \) adjusted for all effects except month, including the estimated \( T(\text{month}) \)-by-disease effect.
Figure 33. $\ln(CD56/CD22)$ adjusted for all effects except month, including the estimated $T(\text{month})$-by-disease effect.
Figure 34. Circannual variation in lymphocyte profile.
of phase with that of Caucasian men who do not have lung cancer.

The disease-by-vegetable interaction.

Patients' and controls' lymphocyte profiles seem to respond to vegetable intake differently. The classical statistical interaction between 2 dichotomous variables is seen in Figure 35. The magnitude of the effect, on both logratios, of low vegetable intake seems to be much greater for patients than for controls. Figure 36 illustrates how that interaction is translated into the effect on the composite lymphocyte profile. Patients who say that they consume more than 14 vegetables per week have lymphocyte profiles very similar to those of the controls. However the profiles of those who have a low vegetable intake are affected more severely than those of controls. Patients with low intake have 21.7% NK compared to 17.2% for those with high intake, whereas the controls with low intake only have 19.8% NK. The difference in NK is evenly distributed among T-cells and B-cells among controls, whereas patients have predominantly a relative loss of B-cells.

The vegetable effect.

When the vegetable effect was isolated (Figure 37), it looked not much different from the control effect, which is what one would expect. When the adjusted logratio means were plotted against vegetable intake category, the
Figure 35. The disease-by-vegetable interaction.
Figure 36. The vegetable-by-disease interaction effect on the lymphocyte profile.
Figure 37. The isolated vegetable effect.
Figure 38. The isolated vegetable effect, by vegetable intake category.
threshold effect which caused vegetable consumption to be dichotomized can be seen (Figure 38). If the model without disease effects was fitted to controls only, the vegetable contribution was non-significant (Manova F statistics = 1.53, p<0.22). Therefore the vegetable effect was large only among patients.

Lung cancer effect.

The disease effect isolated from the effects of the two interactions is presented in Figure 39. Statistically the difference is not likely to have been found by chance (p<0.017), but it is questionable whether there is any meaningful or interpretable biological difference here.

The age-by-smoking interaction.

Throughout the modeling process the age-by-smoking interaction was an important effect. The importance of the interaction persists among controls alone, (as one would expect since the significance of the interaction in the full model among all subjects is calculated as a Type III sum-of-squares). The logratio plots (Figures 40 and 41) are not especially informative about the changes in the actual profile. If one groups subjects by age such that 1/4 fall into the "young" category, 1/4 fall in the "old" category and the other 50% form the "middle" group, as in Figure 40, the relationship of logratio to smoking category appears to be linear. If one categorizes smoking by 10-year age groups,
DISEASE EFFECT

Adjusted for Disease Interactions

Figure 39. The effect of lung cancer on lymphocyte profile.
Figure 40. The logratios by smoking category by three age groups.
Figure 41. The logratios by smoking category by five age groups.
Figure 42. The age-by-smoking interaction on lymphocyte profiles.
however, the linear relationship does not seem to hold at the higher age groups. There appears to be a plateau effect or even a decline in logratios at the oldest levels. Because the numbers are small in these uppermost age groups, however, it is not possible to draw any conclusions about the behavior of the upper tail. Note that there were no current smokers in the over 80 age group.

If one appraises the ternary plot in Figure 42, a dramatic difference is seen in the maturation of the lymphocyte profiles among smokers compared to non-smokers. As non-smokers age from less than 51 to older than 68, their percentage of NK rises from 14% to 24%. Their T-cells drop from 74% to 68% while B-cells are reduced from 11% to 7%. The profiles of smokers, however do not mature.

**Consumption of alcoholic drinks.**

The percentage of daily calories which come from alcoholic drinks seems to affect the lymphocyte profile. If one plots the profile logratios adjusted for the other variables in the model against alcoholic drink consumption in 5% increments (Figure 43), one can see a general linear trend, but the means estimated are not very precise since 74% of the subjects fall in the lowest two categories. If one categorizes alcoholic drink intake in 3 groups such that the 62 subjects who get over 5% of their daily calories from alcoholic drinks are divided evenly in the upper two categories, a linear trend is more apparent (Figure 44).
Figure 43. The effect of intake of alcoholic drinks on profile logratios.
Figure 44. Linear trend in profile logratios by intake of alcoholic drinks in 3 groups.
ALCOHOLIC DRINKS IN DIET

Figure 45. Alcohol intake and the lymphocyte profile.
The lymphocyte profiles of the three groups by low, moderate, and high intakes of calories from alcoholic drinks are plotted in Figure 45. The meaning of the changes seen here is unclear. Moderate drinkers have increased T-cells with roughly equal relative decreases in B-cells and NK. The heavy drinkers, however, have the same proportion of T-cells as the light- or non-drinkers, but increased NK at the expense of B-cells.

Consumption of fat.

The lymphocyte profile was affected by reported level of fat intake. There was an unmistakable threshold effect on the profile logratios, that occurred at 28% of daily calories in fat (Figure 46). The effect on the lymphocyte profile was not only highly significant (Manova F tests, p<0.0004), but substantial as well, there being a 5 percentage point rise in the proportion T-cells among those with a high-fat diet, accommodated by a 4% decrease in B-cells and 1% fewer NK (Figure 47).
Figure 46. Adjusted profile logratio means by fat intake.
Figure 47. Fat intake effect on lymphocyte profile.

H = >27% OF CALORIES
L = <28% OF CALORIES
Other factors tested.

Other variables were tested for influence on the lymphocyte profile by regressing the residuals from the 8-factor model on each of them individually. None of the following made a significant contribution toward explaining the variance of the residuals: parish of residence, income per person in household, education category, percent of daily calories in sweets, weekly fruit intake, weekly citrus intake, intake of green and yellow vegetables, total carotene intake, or \( \beta \)-carotene intake. Total vitamin C intake was significant at \( p<0.048 \). Variables representing the sum of occupational exposures to potentially carcinogenic materials and the sum of exposures of a year or more to occupations associated with higher incidence of lung cancer also failed to contribute toward explanation of the residual variance.

The effects of the vegetable- and fruit-related dietary items above were also tried on the residuals from a model which excluded the vegetable and vegetable-by-disease effects. None of them, including total vitamin C, made a significant contribution toward explaining the variance of the residuals.

Patients and controls on fat intake.

While modelling the lymphocyte profile, it was found that all 3-way interactions involving disease-by-fat had 0
DF. On investigation, it was discovered that only one patient out of the 48 in the model-building process reported a low fat intake, compared to 21 of 166 controls ($\chi^2 = 0.51$, likelihood ratio $\chi^2 = 0.025$). If fat intake was categorized by quartiles of intake among controls, patients were underrepresented in the lowest category and over-represented in the highest category ($p<0.005$, Table 11).

**TABLE 11**

<table>
<thead>
<tr>
<th></th>
<th>Fat Intake, % of Calories</th>
<th>Controls</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;32.6</td>
<td>46</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>32.6-38.1</td>
<td>47</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>38.1-41.8</td>
<td>46</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>&gt;41</td>
<td>47</td>
<td>23</td>
</tr>
</tbody>
</table>

$p(>\chi^2) < 0.005$

**Disease and exposure to substances or occupations.**

When subjects were dichotomized according to whether or not they had been exposed for a year or more to any of the substances associated with higher incidence of lung cancer
### TABLE 12

**OCCUPATIONAL EXPOSURES**

In your work, have you ever been exposed *for a year or more* to any of the following?

<table>
<thead>
<tr>
<th>Occupational Exposures</th>
<th>Industry/Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asbestos</td>
<td>Iron foundry</td>
</tr>
<tr>
<td>Radiation</td>
<td>Nickel smelting</td>
</tr>
<tr>
<td>Welding</td>
<td>Underground mining</td>
</tr>
<tr>
<td>Coal tar, soot, pitch, creosote, asphalt</td>
<td>Lumber industry or heavy wood dust</td>
</tr>
<tr>
<td>Mineral, cutting or lubricating oil</td>
<td>Rubber or cablemaking industry</td>
</tr>
<tr>
<td>Benzidine, β-naphthylamine</td>
<td>Chemical or plastics industry</td>
</tr>
<tr>
<td>Benzene</td>
<td>Pesticides, herbicides</td>
</tr>
<tr>
<td>Isopropyl oil</td>
<td>Mustard gas</td>
</tr>
<tr>
<td>Dyestuffs</td>
<td>Chromium</td>
</tr>
<tr>
<td>Arsenic</td>
<td>Cadmium, beryllium, vinyl chloride</td>
</tr>
</tbody>
</table>

(Table 12), patients were more likely to have reported exposure (35 of 48 patients compared with 86 of 186 controls; $p(>\chi^2) < 0.002$). When the number of exposures was
categorized by quartiles of controls (more than 50% of controls reported 0 exposures so that there were only 3 categories) and the patients' distribution across categories was examined, the patients were significantly differently distributed \( p < 0.004 \); Table 13. However, patients did not report having engaged in the occupations associated with higher incidence of lung cancer more often than controls.

**TABLE 13**

**EXPOSURES TO CARCINOGENIC SUBSTANCES**

<table>
<thead>
<tr>
<th>Number of Exposures</th>
<th>0</th>
<th>1</th>
<th>&gt;1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>100</td>
<td>44</td>
<td>42</td>
</tr>
<tr>
<td>Patients</td>
<td>13</td>
<td>16</td>
<td>19</td>
</tr>
</tbody>
</table>

\[ p(\chi^2) < 0.004 \]

**Leukocyte subsets.**

A differential leucocyte count of 100 cells was done for each subject. The mean percentages measured among the subjects are presented in Table 14, along with the expected ranges for adults (Wintrobe et al. 1974, 1794). The values
### TABLE 14

**DIFFERENTIAL LEUKOCYTE COUNTS**

#### NORMAL PERCENTAGES

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Mean</th>
<th>95% Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrophils</td>
<td>53</td>
<td>34.6 - 71.4</td>
</tr>
<tr>
<td>Lymphocytes</td>
<td>36</td>
<td>19.6 - 52.7</td>
</tr>
<tr>
<td>Monocytes</td>
<td>7.1</td>
<td>2.4 - 11.8</td>
</tr>
<tr>
<td>Eosinophils</td>
<td>3.2</td>
<td>0.0 - 7.8</td>
</tr>
<tr>
<td>Basophils</td>
<td>0.6</td>
<td>0.0 - 1.8</td>
</tr>
</tbody>
</table>

#### CONTROLS

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Mean</th>
<th>Std.Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymphocytes</td>
<td>37.24</td>
<td>0.78</td>
</tr>
<tr>
<td>Monocytes</td>
<td>1.87</td>
<td>0.49</td>
</tr>
<tr>
<td>Neutrophils</td>
<td>58.78</td>
<td>0.76</td>
</tr>
<tr>
<td>Basophils</td>
<td>0.97</td>
<td>0.49</td>
</tr>
<tr>
<td>Eosinophils</td>
<td>3.14</td>
<td>0.50</td>
</tr>
</tbody>
</table>

#### PATIENTS

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Mean</th>
<th>Std.Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymphocytes</td>
<td>24.45</td>
<td>1.39</td>
</tr>
<tr>
<td>Monocytes</td>
<td>2.43</td>
<td>0.27</td>
</tr>
<tr>
<td>Neutrophils</td>
<td>69.91</td>
<td>1.51</td>
</tr>
<tr>
<td>Basophils</td>
<td>0.38</td>
<td>0.09</td>
</tr>
<tr>
<td>Eosinophils</td>
<td>2.78</td>
<td>0.39</td>
</tr>
</tbody>
</table>
observed for monocytes are suspect. The mean percentage of monocytes over all of the subjects in the study was 1.6%, whereas the expected value was 7.1% (2.4-11.8%). No monocytes were observed among 100 cells for 85 of 286 individuals (30%) while if the expected percentage of monocytes were 7%, the probability of counting 100 cells without seeing at least one monocyte is only 0.0007 ([0.93]^{100}). The mean values found for the other 4 subsets agree with reported ranges.

A leukocyte profile was constructed of the logratios of neutrophils, monocytes, eosinophils, and basophils to lymphocytes in a manner analogous to the construction of the lymphocyte profiles. The leukocyte profile among these subjects was not influenced by smoking or by the hour of the day when the sample was drawn. The final quadrivariate model included disease, age, month of year, and the age-by-month interaction.

The eosinophils, basophils, and monocytes together comprised only about 4% of the leukocyte composition. In the above quadrivariate model there was only one non-zero eigenvalue for the quotient matrix \([SSCP]_{error}^{-1}[SSCP]_{model}\), indicating that the minor components of the leukocyte profile were not discriminated by the model. If the leukocyte profile was represented by the two logratios \(\ln(\text{neutrophils}+\text{lymphocytes}), \ln(\text{other}+\text{lymphocytes})\), the same independent variables were significant determiners and were
TABLE 15
THE LEUKOCYTE PROFILE

P(>F), Manova F Tests

<table>
<thead>
<tr>
<th></th>
<th>Quadrivariate Model</th>
<th>Bivariate Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISEASE</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>AGE</td>
<td>0.0157</td>
<td>0.0074</td>
</tr>
<tr>
<td>MONTH</td>
<td>0.047</td>
<td>0.0114</td>
</tr>
<tr>
<td>AGE-by-MONTH</td>
<td>0.040</td>
<td>0.009</td>
</tr>
</tbody>
</table>

more highly significant (Table 15). After adjustment for age and month in which the sample was drawn, the subjects had less than 1% difference in their mean percentages of monocytes, eosinophils and basophils, but patients had an average of 25.3% lymphocytes and 71.0% neutrophils, while the controls had 36.0% lymphocytes and 60.7% neutrophils. Nearly all of the disease effect, then, was reflected in the differing proportions of lymphocytes and neutrophils.

Adjustment for age, month, and age-by-month decreased the intergroup difference in the mean logratio of neutrophils to lymphocytes by 15%, while not reducing the intragroup variances at all. However it did remove confounding; an exaggerated patient-control difference due
to different age and month distributions among the two groups.
DISCUSSION

Study subjects.

The patient population in this study should not be considered to be representative of all Louisiana lung cancer patients with respect to histologic type since selection processes associated with treatment were operant. The relative prevalence of the major types of lung cancer varies geographically and with time, but most studies find that large-cell carcinoma is less common than adenocarcinoma, and adenocarcinoma less common than squamous cell carcinoma. In this study there were very similar numbers of the three types. The composition of the patients may reflect a greater use of radiation therapy among cases of large-cell carcinoma than among other types.

One would expect to find never-smokers or long-time ex-smokers among the men with the least smoking-related type of lung cancer, adenocarcinoma, but there were no life-time non-smokers at all among patients. If 95% of lung cancer among men is associated with having been a smoker, the probability of observing no non-smokers among 64 lung cancer patients is 0.0375. About 1 in 25 samples would contain all current smokers or ex-smokers.

21If 95% of lung cancer cases among men are associated with having smoked, the probability of observing no non-smokers among 64 patients is 0.0375. About 1 in 25 samples would contain all current smokers or ex-smokers.
patients is 0.0375 (0.95⁶⁴), which means that about one in 25 samples would contain all smokers or ex-smokers. The proportion of ex-smokers among the adenocarcinoma group was actually smaller than in other histologic types, but the difference in distribution was non-significant. Perhaps non-smokers who get carcinomas of the lung and especially those who get adenocarcinomas are less likely to receive radiation treatment for it.

Although all of the patients had smoked at some time, 6 of the patients (10.2%) reported that it had been 20 years or more since they had quit smoking. It is commonly believed among epidemiologists, that an ex-smoker's risk of lung cancer drops to that of a non-smoker after 15 years. This observation could be meaningful in terms of identifying alternative or auxiliary causes, however the in-depth analysis of the characteristics of these men was beyond the scope of the present phase of analysis.

The controls compare reasonably well with the patients in terms of demographic characteristics. Selection of volunteers from local civic groups like Lions and Kiwanis and even AARP tended to raise the average socioeconomic level of the controls. Each resource group tended to be different from the other in the socioeconomic characteristics and also in the age range represented, but this heterogeneity was unavoidable, since random-digit dialing or driver's license resourcing were not feasible for
this study. The veterans' organizations were the most balanced in terms of the characteristics for which similarity was desired among patients and controls, e.g. age, parish of residence, and socioeconomic indicators. It was difficult to find volunteers in their late 70's or 80's, however, and particularly difficult to find elderly smokers. Attempts to gain access to residence facilities for senior citizens for the purpose of recruiting volunteers were unsuccessful. The prevalence of smoking appears to decline as a birth-cohort\textsuperscript{22} ages, either through attrition of the habit or biological selection. It was fortunate that former smokers of at least one year were much like never-smokers in their lymphocyte profiles, or there would have been no "non-smokers" among patients. The empty cell would have created difficulties in the statistical analysis.

\textbf{The questionnaire.}

The questionnaire was relatively long. A large proportion of the time and effort required to complete it was due to the dietary portion, in which the food frequency questions on 99 requested estimates of both portion size and frequency. The overall response of the subjects was good given the length of the instrument. If subjects inadvertently skipped pages, these were sent to them for completion. A number of subjects placed check marks instead

\textsuperscript{22}A birth cohort is a group of people born in the same year or span of years.
of numbers in the food frequency boxes. Those also had to be mailed back for correction. Two patients could not read or write. For one patient, this problem was identified while the man was still at PCC and a nurse helped him to complete the questionnaire. The other patient became a 'non-responder'. It is unknown if any of the controls did not return the questionnaire because they were illiterate.

**Dietary data.**

The 3 dietary associations that were found with the lymphocyte profile suggest that the dietary data obtained in this study is meaningful, at least on a group level (Block 1982). If the food-frequency method used had been inadequate, one would expect reduced chances of finding differences among subgroups due to non-differential misclassification of subjects by nutrient intake (Rothman 1986, 87). The possibility exists, of course, that patients were more motivated to complete the questionnaire accurately than controls, but there is no reason to suspect that they would report higher consumption of any particular food than controls because of their diligence, or because of their disease status. When compared to controls on the overall quality of their dietary data, patients were no more likely than controls to have dietary data with none of the flaws flagged by the Block software. A major advantage of using the Block dietary questionnaire is that the dietary results
on this group of subjects will be comparable to the results of other studies which have used this instrument.

The laboratory method.

In deciding upon a method for isolating the leukocytes of interest, 3 options were considered. Many investigators prefer to work with lymphocyte populations isolated on density gradients, but some investigators have demonstrated that certain subsets, in particular subsets of CD8\(^+\) cells, can be lost (Renzi and Ginns 1987; Green and Stelzer 1988). Ficoll preparations are not quantitatively lymphocytes, either; substantial numbers of monocytes remain in the preparation. Monocytes are also the most difficult to exclude on the flow cytometer (Green and Stelzer 1988). Because of the fear of biasing their results due to selective loss of subsets on density gradients, many researchers do whole blood lysis. The problem with that approach is that considerable debris from the lysed erythrocytes remains behind. It is difficult to exclude such debris on the cytometer without also excluding small lymphocytes. By collecting the buffy coat, most of the erythrocytes are eliminated and clean samples can be obtained.

Sample to sample spillover occurred despite vigilance against it. Fluorescence histograms of each sample were monitored in the instrument's display window, and data acquisition by the computer was not begun until equilibrium
in the histogram was attained. The transport of a sample to the point where it passes through the laser beam involves passage through approximately 12 inches of plastic tubing, which can cause an effect much like hysteresis. It should be considered a flaw in the design of the instrument.

An unforeseen problem arose with the labelling of total leukocytes because with the Consort 40 software, one cannot avoid collecting data on granulocytes. In some individuals lymphocytes constituted only 5-6% of the total leukocytes. This meant that in order to collect data on even 2000 lymphocytes, a total of 40,000 events had to be recorded. This created a problem with data storage and handling, and resulted in collection of fewer lymphocytic events than were desired. Handling large data files was difficult given the computer-to-computer transfers which were required for data processing as well, and the longer files were truncated at 2500 sampled lymphocytes. Thus an average of only 2000 lymphocytes per sample were actually analyzed. The usual sample size in previous reports in the literature was between 5000 and 10000 lymphocytes. The analysis of relatively small samples was probably the reason that sample overlap constituted a serious problem. When a sample was truncated, the first 2500 recorded events were kept; it would have been preferable to have retained the last 2500 since the events carrying over from the previous sample
would have occurred much more frequently at the beginning of the current sample than at the end.

The new method for editing and processing the FACS data.

The only means of analyzing the cell sorter data which was available with the Consort 40 software, was to display a sample on the computer monitor in the two dimensions of scatter, visualize the leukocyte subpopulations by displaying contours (boundaries within which were coordinates which had been recorded at least n times), and draw an arbitrary polygon around the lymphocyte population with the mouse to select the lymphocytes. The sample could then be replayed on green fluorescence and side scatter, and the fluorescence-positive population isolated with the aid of the mouse in the same way. The computer would then provide the statistics for the selected population. If the positive and negative populations overlapped in fluorescence, the Consort 40 software provided a one-dimensional normal curve-fitting routine to assist in their separation.

The most immediate problem with using the available method was that it was extremely labor-intensive and time-consuming and therefore inadequate for processing a large number of files. It would have been impossible to monopolize the computer for the time required since it was needed for operating the FACS. Secondly the method was too subjective. If one were to reanalyze the same file without having
recorded the coordinates of the polygon used before, it was unlikely that one would select the same polygon. Visual estimation of the boundaries of the two-dimensional lymphocyte distribution is not accurate enough. Furthermore, one either had to create a new polygon for each sample from a subject, or ignore fluctuations in machine (stream) alignment which can cause the mean of the lymphocyte population on scatter to shift somewhat from sample to sample. An automated, objective method was needed which processed each sample in the same way, while being flexible enough to respond to variations in scatter properties among samples.

The conventional method for determining the percentage fluorescence positive was not entirely satisfactory either. There is no difficulty in discriminating the fluorescence-positive cells from negative populations when the two are completely separated but can be difficult when the difference of the mean fluorescence of the two is not at least 4 standard deviations. It has been standard practice to set a cut-off point at that channel above which only 5\% of the unlabelled control cells fall (Lanier et al. 1983). The number of cells of the sample which fell above this channel, less \( \frac{1}{19} \times \) the number of negative cells, are considered the positive population. No consideration was given to the possibility that some of the events recorded in the "positive" domain might not belong to the subset of interest, or to those situations where there could be more
than one positive subset having differing densities of marker expression. Another problem with this method was that positive cells which fell below the cutoff were counted as negative and therefore the proportion positive was variably biased depending upon the true proportion that was positive, and upon the proportion of the positive cells that lay below the cutoff channel. The 95th percentile for negative cells was rarely the 5th percentile for positive cells and the antigen density, which determines the mean brightness of positive cells, can vary from individual to individual, as well. Consider a hypothetical situation in which 50% of lymphocytes label with a marker antibody and the 95th percentile for control or negative cells is the 20th percentile for positive cells. Rounding to whole numbers, the above method would conclude that only 37% of the cells were positive! Furthermore it cannot be determined what percentage of positive cells fell below the 95th percentile for negative cells, unless one assumes normality of the distributions, finds the mode of the positive cells, assumes that the mode is the mean, and estimates the variance from the upper half of the positive distribution. If that approach were taken, however, fitting Gaussian curves to the fluorescence histogram would have been preferable. Indeed this has been another approach for estimating each population (Mann, Hand, and Braslawsky 1983). BD software includes software which will fit up to
five Gaussian curves to a fluorescence histogram, but it can be applied in only one dimension at a time.

\[
\mathbf{\mu} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}
\]

\[
\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 0.25 \end{bmatrix}
\]

**Figure 48.** Comparison of Bonferroni simultaneous C.I. to bivariate 95% estimation ellipse.

The current study not only took advantage of the normality assumption, but also used the advantage that data are recorded in multiple (in this case, four) dimensions. Considering multiple dimensions simultaneously, which is the essence of multivariate statistics, increases the precision and accuracy of estimation (Johnson and Wichern 1988). For instance, if one were to estimate a confidence interval for a variable in each of two dimensions separately but using a Bonferroni alpha+2 adjustment for making multiple (two)
interval estimates, then plot the confidence limits in two dimensions, the result would be a rectangular confidence area with dimensions of $\mu_i \pm 2.24 \sigma_i$ and having an overall confidence level of 95%. However if one considered both dimensions simultaneously and used the information contained in the covariance matrix, the boundary of constant statistical distance would be an ellipse with axes equal to $\mu_i \pm 2.448 \sigma_i$ for a joint confidence level of 95%. The confidence area so defined is 94% of that of the rectangle (Figure 48). Note that the 95% range of values of $x_2$ depends upon the value of $x_1$ and vice versa even though in this example the variables are orthogonal, that is, their covariance = 0. This arises from the joint improbability of both variables' having extreme values at the same time. Thus multivariate analytical techniques use more of the information in data such as that collected in this study and yield more precise and accurate estimates. The method invented for this study and described in "Materials and Methods," then, improved upon several aspects of the analysis of FACS data.

Estimation of the scatter mean of lymphocytes from the crudely edited cells worked well in most cases. Problems did arise in samples that were relatively lymphocytopenic in that a large percentage of contaminating, outlier cells compared to the actual lymphocytes were present and had undo influence in the estimation of the mean. In a few cases,
the first elliptical cut had to be made at 75% in order to exclude influential outliers.

The clustering technique for achieving starting values for the normal-curve-fitting routine which estimated the percentage of positive cells worked well. Using more clusters than actual subsets was helpful since the density of events with particular coordinates only indirectly, influences the clustering process; primarily the distance from the mean coordinates of each cluster is the criterion for assignment of an event to a cluster.

Because the group estimates on each parameter were as precise as those for the old method, and the new method sampled 33% fewer events, the new method seems to be actually more precise. Another way to visualize that is to realize that if as many events were sampled in the new method as in the old method the experimental variance component would have been reduced and the group estimates would have had smaller variances than those made by the old method. Whether the estimates from the new method differ from those produced by the old, is still an unanswered question, since the comparison of 27 subjects could show that they were different only if the mean paired difference was 6% or more!

The computerized method could be improved. No attempt was made in the data processing to assess residual contamination of the lymphocyte population by monocytes.
Roughly 8% of the subjects might have had more than 1% of monocytes in the selected elliptical confidence window, therefore biasing estimates of lymphocyte subsets in the negative direction. One could use the monocyte marker to flag those subjects with more than X% monocytes, for reediting. Alternatively, (or additionally) one could correct for the estimated number of monocytes in the selected lymphocyte window.

The latter procedure might be the preferred one, since tightening the window would begin to bias the estimation of NK because of their higher mean side scatter. Another source of variation in the overlap of lymphocyte and monocyte populations was day-to-day variation in degree of resolution on scatter which was achieved in aligning the instrument. While there did seem to be some interpersonal variation in the relative scatter properties of lymphocytes, monocytes, and granulocytes, observed among subjects run on the same day, the focus of the populations was directly affected by instrument alignment on a daily basis. The FACS 440 is not as easy to operate as, for instance, an ultraviolate spectrophotometer. Better results on these samples might have been obtained by a better-trained, more experienced operator.

For the last approximately 75 subjects, a small population (on the order of 1%) of lymphocytes apparently positive for the NKH-1 marker appeared at a higher mean
fluorescence than that of the majority of positive cells, and the mean fluorescence of the presumed NK declined slightly from what it had been in earlier samples. The brighter population was too small to be seen while collecting the data and became apparent only during the data processing. Because the NKH-1 antibody used was all of the same lot number and the aberration occurred during the last quartile of subjects, aging of the antibody solution was suspected to be the cause.

A perspective on analysis.

There were 2 possible approaches to the analysis of these data. The epidemiologic approach would have disease/no disease as the outcome of a function of multiple variables, including lymphocyte profiles, in a logistic regression model. The immunologic approach is to look at lymphocyte subset profile as the outcome variable and disease as one of the modifying characteristics. In this phase of the analysis, the immunologic approach was chosen.

One of the difficulties in evaluating differences in numbers and proportions of various leukocyte subsets is that the blood compartment is a closed system. Modifications in particular subset levels necessitate that other subsets are modified as well. Some researchers speak of "numbers" of cells with a particular phenotype (per mm$^3$ of blood); others prefer to operate in the context of relative proportions of
one cell type to another. To say that the number of T-cells is reduced in the blood of lung cancer patients (e.g. Ginns et al. 1982) without consideration of whether total lymphocytes are also reduced among such patients is of limited value. For instance, persons with granulocytosis could have an expanded blood volume and therefore the same number of T-cells in the total peripheral blood compartment as those with normal levels of granulocytes, yet the number of T-cells per cubic millimeter of blood would be reduced. Therefore the relative proportions of subsets are more meaningful. Even then a particular subset should not be taken out of the context of the total composition if possible; alterations in one subset do not occur without compensation among other subsets. A profile approach in treating compositional data is a more comprehensive and therefore more meaningful perspective in analysis.

**Subset balances.**

Estimates of proportions will be nearly normal in their mid-range, but since they are constrained to vary between 0 and 1, for low proportions or high proportions distributions of estimates become asymmetrical. Therefore, if one is to make meaningful estimates of central tendency and variation of estimates, a transformation is necessary to "normalize" the distributions. Either logit (log-odds) or arcsin transformations work very nicely; the logit transformation was chosen for these data merely because it was similar in
nature to the logratio transformations which were used for compositional analysis. Essentially it is the logratio representation of a two-component composition.

In considering both dimly and brightly CD8$^+$ lymphocytes their sum is greater than the total number of cells which are CD3$^+$. Although one group (Blue et al. 1985) found an average of 3% CD4$^+$CD8$^+$ cells in T-cells from normal donors, only activated, proliferating populations of T-cells have in general been found to contain meaningful numbers of cells expressing both CD4 and CD8 (Creemers 1987). In fact in Blue's work, cells were isolated by rosetting with sheep red blood cells and may have been activated in that process. Binding of the T11 (CD2) sheep erythrocyte receptor has been shown to activate thymocytes to express interleukin 2 receptors (IL-2R, Fox 1985). Thus overestimation of CD4$^+$ plus CD8$^+$ cells from counting doubly positive cells twice is an unlikely explanation for the excess.

The cells which express CD8 at low density are probably a subset of NK cells, which do not express CD3 (Perussia 1983; Ault 1988). There is also evidence that a small percentage of CD3$^+CD8^+$ cells are dimly CD8$^+$ (Wonigeit et al. 1988), which would mean that part of the population of lymphocytes which express CD8 at low density could account for the difference between CD3$^+$ and CD4$^+$ + CD8$^+$bright. Some extraneous CD4 reactivity could also be due to contaminating monocytes but this is unlikely; monocytes carry CD4 at low
density and only brightly labelled CD4\(^+\) cells were counted as positive. Also the sum of CD4\(^+\) and bright CD8\(^+\) is 5% less than CD3\(^+\), a difference too large to be accounted for by contaminating monocytes. Landay states that some CD3\(^+\) cells expressing \(\gamma/\delta\) T-cell receptor genes carry neither the CD4 nor the CD8 antigens (Landay 1990) which, along with putative CD8\(^+\)\(^{\text{dim}}\) T-cells, would help to explain why the sum of CD4\(^+\) and bright CD8\(^+\) cells was less than CD3\(^+\) cells and allow the dimly CD8\(^+\)-labelled NK cells to account for the excess in total CD8\(^+\) + CD4\(^+\).

The excess of HLA-DR\(^+\) cells over the number of B-cells presumably represents activated T-cells (Evans et al. 1978). Transplant recipients show increased expression of HLA-DR on CD8\(^+\) T-cells (Henny et al. 1986). Henny et al. found that expression of HLA-DR by CD8\(^+\) cells correlated with lower expression or steric hinderance of the CD8 epitope recognized by the Leu-2a anti-CD8 antibody, but did not affect the epitope recognized by OKT8. The lung cancer patients averaged a 5% greater excess of HLA-DR expression than controls. Since we have data on samples double-labelled with Leu-2a FITC and CD11 PE, the findings of Henny et al. could be tested among our subjects. If patients showed a greater difference in proportions of cells labelled with OKT8 and Leu-2a than controls and that difference correlates with higher HLA-DR\(^+\) - CD22\(^+\) percentage, there would be indirect evidence that more CD8\(^+\) T-cells in patients are
expressing HLA-DR than in controls. This might be postulated as evidence of an ongoing anti-tumor immune response.

For the purposes of this analysis, the 11 individuals in Table 10 were considered to have erroneous estimates for either HLA-DR or CD22 or both. For instance, subject C143 had an estimated 29% B-cells which is an unlikely percentage in a normal individual. However it is possible that these individuals did indeed differ from the rest of the population in terms of their HLA-DR+ and/or CD22+ lymphocytes.

The CD45R+ and CD29+ subsets of CD4-positive cells did not sum to the proportion estimated as CD4-positive by independent labelling (0.36 v. 0.44). These non-equivalencies could be due to a bias in the estimation of the subsets involved, but others have found that cells CD45R+ and CD29+ do not account for all CD4+ lymphocytes (Morimoto et al 1985a; Coulter product bulletins), that 16-32% of CD4+ T-cells express neither. Our results, 18% CD29' CD45R', agree with these estimates.

No analysis of CD45R+ and CD29+ subsets has yet been done aside from verifying that the 2 subsets do not account for all CD4+ lymphocytes. That the correlation between the sum of CD4+CD45R+ and CD4+CD29+ with CD4+ was as high as 0.77 could be interpreted to suggest that the CD4+CD45R'CD29+ cells are a relatively stable percentage of total CD4+
lymphocytes. Investigation of the subsets of CD4+ lymphocytes will be conducted from the perspective of the CD4+ profile, that is, the natural logs of CD29+ \div CD45R- CD29' and CD45R+ \div CD45R'CD29' will be used to summarize the information about CD4+ T-cells. In this way the uncharacterized subset(s) of CD4+ T-cells will not be ignored.

The proportions CD3+, CD22+, and CD56+ summed to 1.00. Although the distribution of the difference 1 - CD3+ - CD22+ - CD56+ was not normal, it was symmetrical in that the mean, median, and mode were nearly identical. The mean, then, was a meaningful quantity. In any case, the distribution of the mean itself, is likely to be normal (by the Central Limit Theorem). To test its difference from 0 using the paired t-test was, then, appropriate. That the profile sums to 1.00 implies that CD3, CD22, and CD56 label exclusive and complementary subsets. Sources differ as to the percentage of B-cells which carry CD22 and the percentage of NK which carry CD56, but nowhere has it been suggested that overlap of the three markers occurs. Furthermore no one has previously reported characterization of lymphocytes with this particular set of markers. Triple-labelling experiments with a dual-laser instrument and 3 fluorochromes could verify the exclusivity of the markers and their comprehensiveness.
The lymphocyte profile model.

Mathematical modelling was the method of choice for these data. Some of the variables of interest, especially the outcome variables, are continuous. Stratified analysis (contingency table analysis) necessitates the categorization of all of the variables (Kleinbaum, Kupper, and Morgenstern 1982, 321) and when one categorizes a continuous measure, information is almost always lost. Furthermore, when one is exploring many factors, observations rapidly become too thinly spread among categories and precision in estimation is lost (Rothman 1986). In addition, the purpose here was to examine the effect of study factors, including lung cancer, on the lymphocyte profile as a whole, rather than on each component separately, and multivariate, multivariable regression is the only available analysis strategy which can accommodate all of these features of the study variables.

The partitioning of the variance of the lymphocyte profile among study factors was more successful than the univariate logratio $R^2$s would indicate. Measurement of lymphocyte subsets involves substantial experimental error. Even duplicate CD3 determinations on aliquots of the same

---

23Researchers often refer to regression of one variable on several other variables as 'multivariate' whereas the better term is 'multivariable', reserving 'multivariate' to describe simultaneous regression of more than one outcome variable on one or more independent variables.
buffy-coat isolate had an estimated standard deviation of 2.7%. It was not possible to estimate any portion of this major source of variability since replication was not feasible, so it all remains in the total unexplained variation. Under these circumstances, it was remarkable that 25% of the variance of the CD3+CD22 logratio and 40% of the variance of the CD56+CD22 logratio were assigned to the independent variables studied.

Circadian variation.

There is little known about the variation of relative proportions of lymphocyte subsets over hour-of-day (Levi et al. 1985; Ritchie et al. 1983; Kidd and Vogt 1989; Giorgi 1986), or month-of-year (Abo et al. 1984; Levi et al. 1988). Subsets have been considered separately and not in the context of their membership in a composition. Insufficient numbers of subjects were assessed at time intervals too widely spaced to estimate the pattern of fluctuation, for example, the study of four individuals at 4-hour intervals to assess the circadian variation of T-, B-, and K-cells (Abo et al. 1981). Levi et al. (1985) measured 5 individuals, also at 4-hour intervals, then estimated the parameters (mesor, amplitude, and acrophase) for each 2 cosine functions with periods of 12 hrs. and 24 hrs. from a total of 30 data points, without properly accounting for the degrees of freedom in the model, and not taking into consideration that they were estimating 6 parameters from
measurements at 6 values of the independent variable time (over-parameterization). Finally other factors which affect the balance of lymphocyte subsets were not considered. Therefore there was no credible, previously proposed function which could be used in modeling and controlling for variation of the lymphocyte profile over time.

Abo and coworkers (1981) assessed the levels of T-cells, and immunoglobulin-bearing cells (B-cells) among 4 individuals at 4-hour intervals throughout the day. He found minima for both subsets at 8:00 a.m. and maxima at midnight. Those subjects whose blood was sampled between 5:00 and 6:00 p.m. in the current study had the lowest percentages of B-cells and the patterns of fluctuation in these data simply do not agree with theirs. That this study measured different individuals rather than the same individuals throughout the day may have had some influence on the results, although sex and race were controlled and the hourly logratio means were adjusted for 7 other profile modifiers including smoking, age, and disease status, thus removing much of the interpersonal variation. Since these were data-based transformations, degrees of freedom were "used" in estimating the parameters of the linear combination of cosine functions, but the consequences would be negligible in the context of a total of 233 DF. Of course the only way that the proposed hourly pattern can be supported or refuted is through testing with another data
set. Since there were no data for 8 nighttime hours, the
daytime pattern cannot be projected to include that third of
the day.

The magnitude of the circadian variation in the
lymphocyte profile was unexpected. During the one 3-hour
period, B-cells increased 9% (Figure 23). Examining the
profile logratio graphs, maxima are observed at 1.5 to 2
hours after what would be mealtimes, suggesting that there
might be a post-prandial effect. Studies of the same
individuals sampled at 2- or 3-hour intervals throughout the
day, on several days with mealtime schedules deliberately
modified on each day, would show whether or not lymphocyte
profiles were correlated with mealtimes. Lack of such
cycling during the sleeping hours would also support a post-
prandial influence.

The peripheral blood accounts for only a small
percentage of the total population of lymphocytes, the rest
being sequestered in various tissues and in the lymph.
Lymphocytes migrate through the blood, lymph, and the
tissues by means of surface molecules specific for ligands
on high endothelial cells which line venules (Chin,
Sackstein, and Cai 1991). It is plausible that the
digestive process induces the release of factors which cause
expression of subset-specific ligands on the venule
endothelium in gut-associated tissues such as the Peyer's
patches and removes lymphocytes from the peripheral
circulation for a period of time, then allows them to return. The up-regulation of such ligands on cultured high endothelial venule cells by the cytokines TNF-α, γ-interferon, and granulocyte-macrophage colony-stimulating factor has been demonstrated (Cai 1991). To my knowledge, there is no data available on circadian patterns in lymphocyte migration.

The linear combination of two cosine functions of hour, even with the simplification of fixed periods and acrophases, accounted for the major portion of the variation of logratio means with hour of day. The remaining variation might still be due a more complex relationship of lymphocyte profile to hour of day, but it might also represent confounding of some other, as yet undetermined variable with both hour and lymphocyte profile.

The circannual variation.

There was a good deal of covariance between month of the year in which the samples were drawn and the other variables in the model. The partial-partial plot (Figure 18), with adjustment for 7 variables, reveals a simpler pattern of profile variation with month of year than was seen at first. The cosine of month function with a 12-month period was moderately successful in modelling the residual variance of the lymphocyte profile with month, accounting for 36% of the CD3+CD22 logratio variance and 72% of the CD56+CD22 logratio variance. A better fit might have been
possible if the period and acrophase had not been fixed, but those constraints were necessary for incorporating a nonlinear function into a linear model. Also some of the remaining variance with month could be due to confounding with attributes of the subjects, which were either not measured in this study or were not discovered in the model-building process.

The variation of lymphocyte profile over the year as illustrated in Figure 34 is modest, especially when compared to the fluctuation of profiles throughout the day. It is possible that the variation is due, in part, to seasonal variation in exposure to pathogens, or to seasonal modulation in climate. If the latter were a factor, then one would expect patterns of variation to differ geographically, particularly with latitude of residence.

T(month)-by-disease interaction.

The nature of the variation in lymphocyte profile with month is better characterized for controls than it is for patients because of the relatively low numbers of patients, but these data indicate, at least, that patients' profiles do not fluctuate in the same pattern that controls' do. A cosine function about 180° out of phase with controls' fits the data, but would only be just one of multiple patterns which would fit data with as much variation as the patients'. One explanation is that the disease effect merely overwhelms the circannual effect on lymphocyte
profiles, but if one looks at Figures 28 and 29, there does seem to be a maximum for patients at about September and a minimum around March, which is nearly the reverse of the behavior of the controls's profiles. Actually, the data as plotted in Figures 28 and 29 suggest that patients's and controls's profiles are more nearly 120° or 4 months out of phase, but a variation in acrophase was not allowed for in the transformation of month used in this model.

The disease-by-vegetable interaction.

One can speculate about what the disease-by-vegetable interaction means. For instance, if substances in vegetables protect against initiating events, perhaps patients suffer more initiating events than controls given the same deficit in dietary protection and hence have a requirement for a greater expansion of NK. In a sense, then, patients as a group are more vulnerable to the negative effects of deficient diet. This could possibly be an effect which precedes the development of cancer in these people. Unfortunately the magnitude of the difference is probably not great enough for it to be used as a discriminator and it would only be useful for prediction among those people who happen to eat less than 14 vegetables per week in any case. The difference in lymphocyte profile between those who eat more than 14 vegetables per week and those who eat fewer is significant only among patients.
The disease effect.

The differences between lymphocyte profiles of lung cancer patients and controls are contained for the most part in the disease-by-T(month) and disease-by-vegetable-intake interactions. If one adjust for both of those effects, the remaining difference is small (Figure 39) and probably not physiologically meaningful.

Age-by-smoking interaction.

These results with respect to smoking and age agree with those of Abo, Cooper, and Balch (1982) in that the proportion of lymphocytes with a NK phenotype increases with age. They also agree with Tollerude's findings of lower percentages of NK in smokers (Tollerud et al. 1989). Tollerud also found reduced percentages of cells with the NK phenotype among ex-smokers, whereas this study did not; ex-smokers' profiles in this study "caught up" with those of never-smokers. The finding of lower levels of immunoglobulins G, A, and M in the blood of smokers (Johnson et al. 1990) would not be attributable to lower percentages of B-cells among smokers, because the differences in B-cell levels among smoking and non-smoking subjects in these data are quite small and probably not meaningful in a physiological sense. Smokers' profile logratios adjusted for all other effects in the model do not exhibit a significant linear age trend within the age range of these subjects (Manova F-tests: p<0.12).
Consumption of alcoholic drinks.

It is unlikely that the alcohol effect seen in these data is due to chance since it is highly significant (Manova p<0.0013), but whether it has implications for health status is unclear. The pattern of change in the lymphocyte profile seen in Figure 43 is not readily interpretable. Part of the difficulty may lie in the manner in which the lymphocyte profile is modelled; what does a linear relationship with the logarithms of ratios of proportions have to do with the behavior of the subset proportions themselves? With respect to the age effect on non-smokers, there was a clear transition from status "young" to "old" with the middle-aged group falling inbetween. There is no such "trend" in the profiles of consumers of alcohol, but then it must be remembered that in ternary diagram, we are not dealing with 2 dimensions, as we are accustomed to do, but with 3. The 3 dimensions also bear the constraint that they must sum to 1.00. It would have been helpful to have had CBC data to see if the individual subset concentrations are affected by consumption of alcoholic beverages.

Consumption of fat.

It is interesting to note that the lymphocyte profile is affected by intake level of fats, but even more intriguing is the unmistakable threshold effect (Figure 47). Not only is there a threshold effect, but it occurs at an intake level of about 28%. This is quite close to the 30%
recommended by the National Research Council (1989), and a number of health-related organizations, as the upper limit for percentage of calories in the diet from fats. The effect itself appears to be substantial, as well, having a magnitude of 5% in T-cells, about 4% in B-cells and 1% in NK (Figure 49). One estimate of the 5th percentile to 95th percentile range for T-cells is 59-81% (Giorgi 1986), therefore a 5% difference is relatively large. The particular subsets of T-cells which are increased would be of interest and might contribute to interpretation.

Another interesting finding with respect to fat intake, is that only one patient of 48 in the analysis fell into the low-fat intake category compared to 11% of the controls (p<0.036). The difference holds even if never-smokers are eliminated from the control group. The categorization of fat intake was based on the effect of fat intake on lymphocyte profile and entirely independent of patient/control status, therefore it is likely that the finding of higher reported fat intake among patients is real.

That only 11% of controls report <28% of daily calories in fat reflects dietary habits in Louisiana in general. A high intake of fats has been associated with cancers of the colon, breast and prostate, but never before associated with lung cancer. Dietary factors, however, are not independent of one another, and perhaps fat in this situation serves as a surrogate for low β-carotene intake or low fruit intake,
both of which have been associated with lung cancer. However when \(\beta\)-carotene intake was divided into quartiles and then the distribution of patients and controls among quartiles was compared, there was no evidence that patients differed from controls in intake of \(\beta\)-carotene intake. Nor did they differ from controls in weekly intake of fruit, weekly intake of citrus, total carotene intake or total vitamin C intake. It is also possible that patients have modified their diets since their diagnosis of cancer to include more fruits and vegetables but did not reduce their fat consumption.

Other factors tested.

It would not be expected that demographic factors would affect the lymphocyte profile except perhaps indirectly through dietary habits. Neither would a history of exposures to carcinogenic substances be expected to have affected the lymphocyte profile, although current exposures could have an effect. It was interesting that weekly vegetable intake was important in modelling lymphocyte profiles, but specific nutrients such as \(\beta\)-carotene were not, and intake of green and yellow vegetables, in particular, was not. The potential of total vitamin C to contribute has yet to be explored.

Vitamin C intake did not appear to be important in the earlier models of the lymphocyte profile, which illustrates that model building is a dynamic and in a sense an iterative
process. The remodelling of month of year underscores this fact, as well.

**Usefulness of FACS data for diagnosis.**

Unfortunately the contribution of measurement error to the variance of the profile estimation could not be quantitated. From the variance present in the replicate CD3 samples, it is thought that the experimental variance is large. The precision with which the lymphocyte profile for an individual can be determined would be enhanced if replicate samples were prepared and more lymphocytes analyzed per sample. The usefulness of FACS data for classifying individuals depends directly upon the variance of individual estimates. Adjustment for factors which contribute to interpersonal differences in FACS measures are helpful if they reduce intra-group differences to a greater extent than they do inter-group differences, but the precision with which one can estimate group means is irrelevant for prediction of group membership for an individual unless one can estimate the individual's profile with precision.

**The leukocyte profile.**

During the data gathering phase of the study, patients seemed to be relatively lymphocytopenic compared to controls. This was verified in the analysis and is the major difference in the leukocyte profiles of patients and
controls. Pulmonary bacterial infections are common among persons with lung cancer which could explain the relative granulocytosis. Smoking was not a significant determinant of the relative number of lymphocytes and neutrophils, which agrees with the observations of Miller et al. (1982). On the other hand, they reported no association with age whereas age was an important modifier in this study. The leukocyte profile was not explored as assiduously as was the lymphocyte profile, and other modifiers remain to be identified.

Adjustment for age and month of year in which the sample was collected reduced the difference between patient and control means by 15% while not reducing the intragroup variance at all. Still a control would have less than a 17% chance of having a neutrophil to lymphocyte ratio higher than the estimated adjusted mean for patients. It would also not be difficult to improve the individual estimate of the percentages of lymphocytes and neutrophils by counting more leukocytes, thereby reducing that component of the intragroup variance. Combining this indicator with similarly improbable measures could identify an undiagnosed patient.

**Disease and fat intake and occupational exposures.**

Preliminary evidence was found in this study that lung cancer patients were more likely to consume higher levels of fat than controls and were more likely to have reported
occupational exposures to potentially carcinogenic substances than controls. Evidence of the former association has not been found previously for lung cancer. Both associations will be explored in the context of potential confounders when the data is examined from the epidemiologic perspective. It is quite possible that the difference in occupational exposures might be explained by the differences in socioeconomic status between patients and controls. On the other hand, the differences in specific exposures could be quite meaningful in interpreting the high incidence of lung cancer in Louisiana.
SUMMARY AND CONCLUSIONS

The purpose of this study was to measure lymphocyte subsets among lung cancer patients and controls to determine whether (1) there were significant differences between the two groups after adjustment for other modifiers, and (2) to project whether those differences could be exploited, particularly in terms of earlier diagnosis of the disease. Realizing the complexity of the cancer phenomenon, wide-ranging, detailed data were gathered so that appropriate controls and adjustments could be made in the analysis.

In the process of the study a new method for automating the analysis of cell sorter data was invented. It removed a great deal of the subjectivity involved in other methods, and therefore should reduce intraanalyst bias and interanalyst variance. Certainly the method needs refinement, however the proof that it worked well for this study is that meaningful differences among subset profiles were found by other variables assessed. A comparison of subset estimates with determinations using a conventional method indicated that the new method gave estimates indistinguishable from those of the other method but was more precise than the old one. In addition to being more
objective, the method designed in this study was automated and therefore labor-saving as well.

The first phase of the analysis was confined to 8 of the 17 subsets or subset combinations actually measured. The analysis is ongoing and the results presented here, aside from the lymphocyte profile analysis, included only some general observations of the relationships among the subsets analyzed. The major focus of these results is the modelling of the lymphocyte profile, the relative balance of CD3+ T-cells, CD22+ B-cells, and CD56+ natural killer cells, as a log-linear function of variables recorded about the subjects in the study. By summarizing the 3-component profile with 2 variables and regressing these simultaneously on the independent variables, a holistic view of the effect of involved variables on the lymphocyte composition was achieved. With the help of ternary diagrams, one profile can be visually compared to another, as well. The effects of 8 factors on the profile have been estimated from these data and presented for verification or refutation in future research.

Previously reported results with respect to differences in relative proportions of lymphocyte subsets between persons with lung cancer and those without have been conflicting. It is likely that the disagreement has been due to failure to adjust for the many factors which influence lymphocyte balance. In the process of building
this model, it was demonstrated that disease effect could be
masked by confounding with another variable, that is, with
the month of year in which the sample was collected, also
with weekly vegetable intake. Differences between patients
and controls in lymphocyte profile were predominantly within
the disease-by-month and disease-by-vegetable-intake
interactions.

Although previous research has demonstrated that
lymphocyte subsets vary with circadian and circannual
rhythms, data to date has been insufficient to describe the
patterns of variation. With the data of the current study,
cosine functions with periods of 6 and 24 hours model a
large proportion of the variance of lymphocyte profile with
hour of the day. There appear to be maxima at 8:30 a.m.,
2:30 p.m. and 8:30 p.m. Similarly the profile seemed to
vary as the cosine of month with a period of 12 months and
maxima for the logratios in May. The profile of patients
appeared to be 6 months out of phase with that of controls.

The relationship of lymphocyte profile with age,
between the mid-forties and the early eighties appeared as
a consistent trend for non-smokers on the ternary plot
across 3 categories of age, with the most dramatic
difference being the increase of natural killer cells from
14% to 20% to 24% of the composition. Smokers retain the
lymphocyte profile of younger men which may not serve them
well as their bodies age. The effects of smoking seem to be
reversed in as little as one year after quitting, but are not instantaneously reversed.

Low vegetable intake in patients had a greater effect on their lymphocyte profiles than on the profiles of controls. There was a significant variation in lymphocyte profile with moderate (>5% of daily calories from alcoholic drinks) and heavy (>12%) drinking, but its meaning is an enigma at this point. The effect of a high-fat (>28% of daily calories) diet was surprisingly substantial, amounting to a shift of +5% in T-cells. It is interesting that the threshold effect of fat falls very near the upper limit recommended by specialists in nutrition for the daily dietary intake of fat. The dietary effects on components of the immune system observed in this study have not been reported previously and should offer fertile ground for further research.

Whether these findings concerning the lymphocyte profile could be useful in predicting a case of lung cancer prior to conventional diagnosis is doubtful. The difference between patients and controls is not large even after adjustment for other variables and the intragroup variance of the estimates remains substantial. A large component of that variance is the experimental error in the individual measurement. Until that can be reduced, the usefulness of FACS data for individual diagnosis will be limited.
Perhaps when the analysis is completed patients will be found to be different from controls on enough indices, including but not limited to FACS data, that a composite differentiation scale can be constructed. Potentially useful differences have been found in the leukocyte profile and in the proportion of cells HLA-DR\(^+\) which are not B-cells. Patients were found to average 11% fewer lymphocytes and 10% more neutrophils than controls and 5% more HLA-DR expression that was independent of that of B-cells. In addition, patients reported exposure to suspected carcinogens more frequently, and had less of a tendency than controls to eat low-fat diets.

In conclusion, patients did display differences in their CD3-CD56-CD22 profiles from those of comparable controls, but the differences found were mostly contained in the differing patterns of circannual variation and more extreme effect of low vegetable intake. Difference in lymphocyte profile, then, would be an unlikely discriminator for lung cancer. A major difference in lymphocyte-to-neutrophil balance was found, however, that has potential as a component for a screening battery and a role for the relative excess of HLA-DR expression among patients has not been excluded, if the adjustment for confounders does not decrease the intergroup difference more than the intragroup variability and if the experimental variability in
individual estimates of subset percentages could be controlled.
WORKS CITED


North, R.J. and M Awwad. 1990. Elimination of cycling CD4+ suppressor T cells with anti-mitotic drug release non-cycling CD8+ T cells to cause regression of an advanced lymphoma. *Immunology* 71:90-95.


APPENDIX A

The

QUESTIONNAIRE
HEALTH HABITS AND HISTORY QUESTIONNAIRE

This form asks you a variety of questions about your background, environment, and habits, which may affect or be related to your health. The information you provide will help scientists to understand more about the causes of disease.

This questionnaire will take about 40 minutes to complete. Please fill in the information requested, or place a check in the appropriate space. A few questions may be similar to ones you have answered before, but please do not skip any questions for this reason. If you are not sure about an answer, please estimate.

If you have any questions or would like help filling it out, please call Joan Blackmon at 388-1213. Please return this questionnaire by ___________________. We thank you for your time and your contribution to this research.

1x. In WHAT STATE or COUNTRY were you born? (Circle correct response.)

Alabama   Iowa   New Jersey   Vermont
Alaska     Kansas   New Mexico   Virginia
Arizona    Kentucky   New York   Washington
Arkansas   Louisiana   N. Carolina   W. Virginia
California MainE   N. Dakota   Wisconsin
Colorado   Maryland   Ohio   Wyoming
Connecticut Massachusetts   Oklahoma
Delaware   Michigan   Oregon   Puerto Rico
D. of C. Minnesota   Pennsylvania   Virgin Islands
Florida    Mississippi   Rhode Island   Guam
Georgia    Missouri   S. Carolina   Canada
Hawaii     Montana   S. Dakota   Cuba
Idaho      Nebraska   Tennessee   Mexico
Illinois   Nevada   Texas   Rest of World
Indiana    New Hampshire   Utah   Unknown

2x. With respect to your PRESENT home, in what PARISH do you reside? (Circle correct response.)

Ascension   E. Feliciana   Pointe Coupee   W. Baton Rouge
Assumption  Iberville   St. Helena   W. Feliciana
E. Baton Rouge Livingston   Tangipahoa

is your neighborhood (Check one.)

___ URBAN?
___ SUBURBAN?
___ RURAL?
___ OTHER?

what is your zip code?    __   __   __   __
PERSONAL INFORMATION, HABITS

1. When were you born? __________/______/______
   Month Day Year

2. How old are you? _____ years

3. Sex: 1 Male 2 Female

4. Race or ethnic background:
   1 __ White, not of Hispanic origin
   2 __ Black, not of Hispanic origin
   3 __ Hispanic
   4 __ American Indian/Alaskan native
   5 __ Asian
   6 __ Pacific Islander

5. Please circle the highest grade in school you have completed:
   1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16+

6. What is your marital status? 1 Single 3 Widowed
   2 Married 4 Divorced/Separated

7. How many times have you moved or changed residences in the last ten years? _____ times

3x. Is the total income, before taxes and other deductions, of your household (the TOTAL income of ALL members of the household) (Check one.)
   Income
   1 Under $10,000/yr?
   2 $10,000 - $19,999/yr?
   3 $20,000 - $29,999/yr?
   4 $30,000 - $39,999/yr?
   5 $40,000 - $49,999/yr?
   6 Or over $50,000/yr?
   9 I don't wish to answer.

4x. How many persons are there in your household? _____

5x. For how many of the last 30 years did you live in a
   RURAL AREA? __________ yrs
   SUBURBAN AREA? __________ yrs
   URBAN (DOWNTOWN) AREA? __________ yrs
6x. Have you ever lived within a mile of an industry which emitted DUST or ODORS or IRRITATING GASES?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
<td>Don't Know</td>
</tr>
</tbody>
</table>

If YES, tell me which plant you lived close to for the LONGEST period of time.

<table>
<thead>
<tr>
<th>KIND OF PLANT?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HOW LONG?</td>
<td>yrs</td>
</tr>
<tr>
<td>TYPE OF EMISSION?</td>
<td></td>
</tr>
</tbody>
</table>

(Choose one.)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dust</td>
<td>Odors</td>
<td>Irritants</td>
<td>Other</td>
</tr>
</tbody>
</table>

Next longest?

<table>
<thead>
<tr>
<th>KIND OF PLANT?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HOW LONG?</td>
<td>yrs</td>
</tr>
<tr>
<td>TYPE OF EMISSION?</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dust</td>
<td>Odors</td>
<td>Irritants</td>
<td>Other</td>
</tr>
</tbody>
</table>

Third longest?

<table>
<thead>
<tr>
<th>KIND OF PLANT?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HOW LONG?</td>
<td>yrs</td>
</tr>
<tr>
<td>TYPE OF EMISSION?</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dust</td>
<td>Odors</td>
<td>Irritants</td>
<td>Other</td>
</tr>
</tbody>
</table>

Fourth longest?

<table>
<thead>
<tr>
<th>KIND OF PLANT?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HOW LONG?</td>
<td>yrs</td>
</tr>
<tr>
<td>TYPE OF EMISSION?</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dust</td>
<td>Odors</td>
<td>Irritants</td>
<td>Other</td>
</tr>
</tbody>
</table>
8. Have you smoked at least 100 cigarettes in your entire life?  
   1. No  2. Yes  If Yes,
   - About how old were you when you first started smoking cigarettes fairly regularly?  
     ___ years old
   - On the average of the entire time you smoked, how many cigarettes did you smoke per day?  
     ___ cigarettes per day
   - Do you smoke cigarettes now?  
     1. No  2. Yes
     If No: How old were you when you stopped smoking?  ___ years old
     If Yes: On the average, about how many cigarettes a day do you smoke now?  ___ cigarettes

9. Have you ever smoked a pipe or cigars regularly?  
   1. No  2. Yes  If Yes,
   - For how many years?  ___ years
   - About how much?  ___ pipes or cigars per ___ (day or week)

If you have NEVER smoked, skip to question 1x

If you are a CURRENT or FORMER smoker

<table>
<thead>
<tr>
<th>Since you began smoking, how many years were there, altogether, when you did not smoke? ___</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many years did you smoke non-filter cigarettes? ___</td>
</tr>
<tr>
<td>How many years did you smoke filter cigarettes? ___</td>
</tr>
</tbody>
</table>

Thinking about how you USUALLY smoked, would you say that you (Check one.)

| ___ DIDN'T INHALE |
| ___ DREW SMOKE INTO MOUTH & THROAT ONLY |
| ___ INHALED DEEPLY INTO CHEST |
| ___ INHALED AT DIFFERENT LEVELS |
| ___ CAN'T DESCRIBE |

Usually, how long was the butt when you put the cigarette out? Was it (Check one.)

| ___ Was it more than 1/2 the cigarette? |
| ___ 1/3 to 1/2? |
| ___ 1/4 to 1/3? |
| ___ less than 1/4? |

When you were not drawing on it, did you usually

| ___ PUT THE LIGHTED CIGARETTE ON AN ASHTRAY? |
| ___ KEEP IT IN YOUR MOUTH? |
| ___ HOLD IT IN YOUR FINGERS? |
| ___ OTHER? |
7x. Did either of your parents (or the people who raised you) smoke?

- NO
- MOTHER (or mother-substitute)
- FATHER (or father-substitute)

8x. Have you lived with a spouse or roommate who smoked?

- NO
- SPOUSE
- ROOMMATE
- OTHER

9x. For how many years altogether did you live with a smoker? __________ yrs

10. During the past year, have you taken any vitamins or minerals?

1  —  No  2  —  Yes, fairly regularly  3  —  Yes, but not regularly  If Yes, ________ yrs

What do you take fairly regularly? # of PILLS per DAY, WEEK, etc.

<table>
<thead>
<tr>
<th>Multiple Vitamins</th>
<th># of PILLS per DAY, WEEK, etc.</th>
<th>How many milligrams or IUs per pill?</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-a-day type</td>
<td>pills per</td>
<td>IU per pill</td>
</tr>
<tr>
<td>Stress-tabs type</td>
<td>pills per</td>
<td>mg per pill</td>
</tr>
<tr>
<td>Therapeutic, Theragran type</td>
<td>pills per</td>
<td>IU per pill</td>
</tr>
</tbody>
</table>

Other Vitamins

- Vitamin A pills per IU per pill
- Vitamin C pills per mg per pill
- Vitamin E pills per IU per pill
- Calcium or dolomite pills per mg per pill

Other (What?):
1  —  Yeast  2  —  Selenium  3  —  Zinc  4  —  Iron  5  —  Beta-carotene
6  —  Cod liver oil  7  —  Other

Please list the brand of multiple vitamin/mineral you usually take:

FOR OFFICE USE

Q 10. mg or IU: 1 = 50-100  2 = 200-250  3 = 400-500  4 = 1000  5 = 5000  6 = 10,000  7 = 20,000-25,000  8 = 50,000  9 = Unk.
11. Are you on a special diet?
   1 ___No  2 ___Weight loss  3 ___For medical condition  4 ___Vegetarian  5 ___Low salt  
   6 ___Low cholesterol  7 ___Weight gain

12. How often do you eat the following foods from restaurants or fast food places?

<table>
<thead>
<tr>
<th>RESTAURANT FOOD</th>
<th>Almost every day</th>
<th>2-4 times a week</th>
<th>Once a week</th>
<th>1-3 times a month</th>
<th>1-5 times a year</th>
<th>1-6 times a year</th>
<th>Never, or less than once a year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fried chicken</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burgers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pizza</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese food</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexican food</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fried fish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other foods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

13. This section is about your usual eating habits. Thinking back over the past year, how often do you usually eat the foods listed on the next page?

First, check (/) whether your usual serving size is small, medium or large. (A small portion is about one-half the medium serving size shown, or less; a large portion is about one-and-a-half times as much, or more.)

Then, put a NUMBER in the most appropriate column to indicate HOW OFTEN, on the average, you eat the food. You may eat bananas twice a week (put a 2 in the "week" column). If you never eat the food, check "Rarely/Never." Please DO NOT SKIP foods. And please BE CAREFUL which column you put your answer in. It will make a big difference if you say "Hamburger once a day" when you mean "Hamburger once a week"!

Some items say "in season." Indicate how often you eat these just in the 2-3 month time when that food is in season. (Be careful about overestimating here.)

Please look at the example below. This person

1) eats a medium serving of cantaloupe once a week, in season.
2) has ½ grapefruit about twice a month.
3) has a small serving of sweet potatoes about 3 times a year.
4) has a large hamburger or cheeseburger or meat loaf about four times a week.
5) never eats winter squash.

EXAMPLE:

<table>
<thead>
<tr>
<th>Medium Serving Size</th>
<th>How often?</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>M</td>
</tr>
<tr>
<td>Cantaloupe (in season)</td>
<td>⅓ medium</td>
</tr>
<tr>
<td>Grapefruit</td>
<td>(⅓)</td>
</tr>
<tr>
<td>Sweet potatoes, yams</td>
<td>⅓ cup</td>
</tr>
<tr>
<td>Hamburger, cheeseburger, meat loaf</td>
<td>1 medium</td>
</tr>
<tr>
<td>Winter squash, baked squash</td>
<td>⅓ cup</td>
</tr>
</tbody>
</table>

Please go to next page.

For Office Use

On the following two pages, code the four characters for each food as follows:

- S-1 No
- M-2 Times
- L-3
- NS-9 NS-99

If respondent places a checkmark in the "How often" columns, do not impute "01", once. Instead, code "99", Not Stated. If respondent does not check a portion size, do not impute medium, but code "9".
### FRUITS & JUICES

<table>
<thead>
<tr>
<th>Item</th>
<th>Medium Serving</th>
<th>Your Serving Size</th>
<th>How often?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EXAMPLE</strong> - Apples, applesauce, pears</td>
<td>(1) or ¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apples, applesauce, pears</td>
<td>(1) or ¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bananas</td>
<td>1 medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peaches, apricots (canned, frozen or dried, whole year)</td>
<td>(1) or ¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peaches, apricots, nectarines (fresh, in season)</td>
<td>1 medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantaloupe (in season)</td>
<td>¼ medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watermelon (in season)</td>
<td>1 slice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strawberries (fresh, in season)</td>
<td>½ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oranges</td>
<td>1 medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orange juice or grapefruit juice</td>
<td>6 oz. glass</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grapefruit</td>
<td>(½)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang, fruit drinks</td>
<td>6 oz. glass</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other fruit juices, fortified fruit drinks</td>
<td>6 oz. glass</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any other fruit, including berries, fruit cocktail</td>
<td>¼ cup</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### VEGETABLES

<table>
<thead>
<tr>
<th>Item</th>
<th>Medium Serving</th>
<th>Your Serving Size</th>
<th>How often?</th>
</tr>
</thead>
<tbody>
<tr>
<td>String beans, green beans</td>
<td>½ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peas</td>
<td>½ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chili with beans</td>
<td>¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other beans such as baked beans, pinto, kidney beans, lima</td>
<td>¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corn</td>
<td>½ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter squash, baked squash</td>
<td>¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tomatoes, tomato juice</td>
<td>(1) or 6 oz.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red chili sauce, taco sauce, salsa picante</td>
<td>2 Tbsp. sauce</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broccoli</td>
<td>½ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cauliflower or brussel sprouts</td>
<td>½ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spinach (raw)</td>
<td>½ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spinach (cooked)</td>
<td>½ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mustard greens, turnip greens, collards</td>
<td>½ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cole slaw, cabbage, sauerkraut</td>
<td>½ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carrots, or mixed vegetables containing carrots</td>
<td>¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green salad</td>
<td>1 med. bowl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salad dressing, mayonnaise (including on sandwiches)</td>
<td>2 Tbsp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>French fries and fried potatoes</td>
<td>¾ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet potatoes, yams</td>
<td>¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other potatoes, including boiled, baked, potato salad</td>
<td>(1) or ¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rice</td>
<td>¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any other vegetable, including cooked onions, summer squash</td>
<td>¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Butter, margarine or other fat on vegetables, potatoes, etc.</td>
<td>2 pats</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### MEAT, FISH, POULTRY & MIXED DISHES

<table>
<thead>
<tr>
<th>Item</th>
<th>Medium Serving</th>
<th>Your Serving Size</th>
<th>How often?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamburger, cheeseburger, meat loaf</td>
<td>1 medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef—steaks, roasts</td>
<td>4 oz.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef stew or pot pie with carrots, other vegetables</td>
<td>1 cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver, including chicken livers</td>
<td>4 oz.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pork, including chops, roasts</td>
<td>2 chops or 4 oz.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fried chicken</td>
<td>2 sm. or 1 lg. piece</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicken or turkey, roasted, stewed or broiled</td>
<td>2 sm. or 1 lg. piece</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fried fish or fish sandwich</td>
<td>4 oz. or 1 sand.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuna fish, tuna salad, tuna casserole</td>
<td>¼ cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shell fish (shrimp, lobster, crab, oysters, etc.)</td>
<td>(3) ¼ cup or 3 oz.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other fish, broiled, baked</td>
<td>4 oz.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spaghetti, lasagna, other pasta with tomato sauce</td>
<td>1 cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pizza</td>
<td>2 slices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed dishes with cheese (such as macaroni and cheese)</td>
<td>1 cup</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### LUNCH ITEMS

<table>
<thead>
<tr>
<th>Item</th>
<th>Medium Serving</th>
<th>How often?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liverwurst</td>
<td>2 slices</td>
<td></td>
</tr>
<tr>
<td>Hot dogs</td>
<td>2 slices</td>
<td></td>
</tr>
<tr>
<td>Ham, lunch meats</td>
<td>2 slices</td>
<td></td>
</tr>
<tr>
<td>Vegetable soup, vegetable beef, minestrone, tomato soup</td>
<td>1 med. bowl</td>
<td></td>
</tr>
<tr>
<td>Other soups</td>
<td>1 med. bowl</td>
<td></td>
</tr>
</tbody>
</table>

### BREADS / SALTY SNACKS / SPREADS

<table>
<thead>
<tr>
<th>Item</th>
<th>Medium Serving</th>
<th>How often?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biscuits, muffins, burger rolls (incl. fast foods)</td>
<td>1 med. piece</td>
<td></td>
</tr>
<tr>
<td>White bread (including sandwiches), bagels, etc., crackers</td>
<td>2 slices, 3 crackers</td>
<td></td>
</tr>
<tr>
<td>Dark bread, including whole wheat, rye, pumpernickel</td>
<td>2 slices</td>
<td></td>
</tr>
<tr>
<td>Corn bread, corn muffins, corn tortillas</td>
<td>1 med. piece</td>
<td></td>
</tr>
<tr>
<td>Salty snacks (such as chips, popcorn)</td>
<td>2 handfuls</td>
<td></td>
</tr>
<tr>
<td>Peanuts, peanut butter</td>
<td>2 Tbsp.</td>
<td></td>
</tr>
<tr>
<td>Butter on bread or rolls</td>
<td>2 pats</td>
<td></td>
</tr>
<tr>
<td>Margarine on bread or rolls</td>
<td>2 pats</td>
<td></td>
</tr>
<tr>
<td>Gravies made with meat drippings, or white sauce</td>
<td>2 Tbsp.</td>
<td></td>
</tr>
</tbody>
</table>

### BREAKFAST FOODS

<table>
<thead>
<tr>
<th>Item</th>
<th>Medium Serving</th>
<th>How often?</th>
</tr>
</thead>
<tbody>
<tr>
<td>High fiber, bran or granola cereals, shredded wheat</td>
<td>1 med. bowl</td>
<td></td>
</tr>
<tr>
<td>Highly fortified cereals, such as Product 19, Total, or Most</td>
<td>1 med. bowl</td>
<td></td>
</tr>
<tr>
<td>Other cold cereals, such as Corn Flakes, Rice Kreespies</td>
<td>1 med. bowl</td>
<td></td>
</tr>
<tr>
<td>Cooked cereals</td>
<td>1 med. bowl</td>
<td></td>
</tr>
<tr>
<td>Sugar added to cereal</td>
<td>2 teaspn.</td>
<td></td>
</tr>
<tr>
<td>Eggs</td>
<td>1 egg = small. 2 eggs = medium.</td>
<td></td>
</tr>
<tr>
<td>Bacon</td>
<td>2 slices</td>
<td></td>
</tr>
<tr>
<td>Sausage</td>
<td>2 patties or links</td>
<td></td>
</tr>
</tbody>
</table>

### SWEETS

<table>
<thead>
<tr>
<th>Item</th>
<th>Medium Serving</th>
<th>How often?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ice cream</td>
<td>1 scoop</td>
<td></td>
</tr>
<tr>
<td>Doughnuts, cookies, cakes, pastry</td>
<td>1 pc. or 3 cookies</td>
<td></td>
</tr>
<tr>
<td>Pumpkin pie, sweet potato pie</td>
<td>1 med. slice</td>
<td></td>
</tr>
<tr>
<td>Other pies</td>
<td>1 med. slice</td>
<td></td>
</tr>
<tr>
<td>Chocolate candy</td>
<td>small bar, 1 oz.</td>
<td></td>
</tr>
<tr>
<td>Other candy, jelly, honey, brown sugar</td>
<td>3 pc. or 1 Tbsp.</td>
<td></td>
</tr>
</tbody>
</table>

### DAIRY PRODUCTS

<table>
<thead>
<tr>
<th>Item</th>
<th>Medium Serving</th>
<th>How often?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cottage cheese</td>
<td>1/2 cup</td>
<td></td>
</tr>
<tr>
<td>Other cheeses and cheese spreads</td>
<td>2 slices or 2 oz.</td>
<td></td>
</tr>
<tr>
<td>Flavored yogurt</td>
<td>1 cup</td>
<td></td>
</tr>
<tr>
<td>Whole milk and bevs. with whole milk (not incl. on cereal)</td>
<td>8 oz. glass</td>
<td></td>
</tr>
<tr>
<td>2% milk and bevs. with 2% milk (not incl. on cereal)</td>
<td>8 oz. glass</td>
<td></td>
</tr>
<tr>
<td>Skim milk, 1% milk or buttermilk (not incl. on cereal)</td>
<td>8 oz. glass</td>
<td></td>
</tr>
</tbody>
</table>

### BEVERAGES

<table>
<thead>
<tr>
<th>Item</th>
<th>Medium Serving</th>
<th>How often?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular soft drinks</td>
<td>12 oz. can or bottle</td>
<td></td>
</tr>
<tr>
<td>Diet soft drinks</td>
<td>12 oz. can or bottle</td>
<td></td>
</tr>
<tr>
<td>Beer</td>
<td>12 oz. can or bottle</td>
<td></td>
</tr>
<tr>
<td>Wine</td>
<td>1 med. glass</td>
<td></td>
</tr>
<tr>
<td>Liquor</td>
<td>1 shot</td>
<td></td>
</tr>
<tr>
<td>Decaffeinated coffee</td>
<td>1 med. cup</td>
<td></td>
</tr>
<tr>
<td>Coffee, not decaffeinated</td>
<td>1 med. cup</td>
<td></td>
</tr>
<tr>
<td>Tea (hot or iced)</td>
<td>1 med. cup</td>
<td></td>
</tr>
<tr>
<td>Lemon in tea</td>
<td>1 teaspn.</td>
<td></td>
</tr>
<tr>
<td>Non-dairy creamer in coffee or tea</td>
<td>1 Tbsp.</td>
<td></td>
</tr>
<tr>
<td>Milk in coffee or tea</td>
<td>1 Tbsp.</td>
<td></td>
</tr>
<tr>
<td>Cream (real) or Half-and-Half in coffee or tea</td>
<td>1 Tbsp.</td>
<td></td>
</tr>
<tr>
<td>Sugar in coffee or tea</td>
<td>2 teaspn.</td>
<td></td>
</tr>
<tr>
<td>Artificial sweetener in coffee or tea</td>
<td>1 packet</td>
<td></td>
</tr>
<tr>
<td>Glasses of water, not counting in coffee or tea</td>
<td>8 oz. glass</td>
<td></td>
</tr>
</tbody>
</table>
14. Think about your diet over the last year and the responses you have just made on this questionnaire. Are there any foods not mentioned which you ate at least once a week, even in small quantities, or ate frequently in a particular season? Consider other meats, breakfast foods, catsup, green chilies or jalapenos, avocado (guacamole), Mexican dishes, Chinese or other ethnic foods, other fruits or vegetables, as well as nutritional supplements (bran, etc.). Please take a look at the list at the bottom of the page.

<table>
<thead>
<tr>
<th>FOOD</th>
<th>Your Serving Size</th>
<th>How Often?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Day</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
</tr>
</tbody>
</table>

15. How often do you eat the skin on chicken?
   How often do you eat the fat on meat?
   How often do you add salt to your food?
   How often do you add pepper to your food?

16. How often do you use fat or oil in cooking?
   For example, in frying eggs, meat or vegetables? ______ times per day, week, month

17. What do you usually cook with?
   1  __ Don't know or don't cook  
   2  __ Soft margarine  
   3  __ Stick margarine  
   4  __ Butter  
   5  __ Oil  
   6  __ Lard, fatback, bacon fat  
   7  __ Pam or no oil

18. What kind of fat do you usually add to vegetables, potatoes, etc?
   1  __ Don't add fat  
   2  __ Soft margarine  
   3  __ Stick margarine  
   4  __ Butter  
   5  __ Half butter, half fatback, bacon fat

19. If you eat cold cereal, what kind do you eat most often?

20. Not counting salad or potatoes, about how many vegetables do you eat per day or per week?
   ______ per day, week

21. Not counting juices, how many fruits do you usually eat per day or per week?
   ______ per day, week

22. Have you gained or lost more than five pounds in the past year? (You may check more than one answer.)
   1  __ No  
   2  __ Lost 5-15 lbs.  
   3  __ Lost 16-25 lbs.  
   4  __ Lost more than 25 lbs.  
   5  __ Gained 5-15 lbs.  
   6  __ Gained 16-25 lbs.  
   7  __ Gained more than 25 lbs.

**OFFICE USE**

<table>
<thead>
<tr>
<th>Code</th>
<th>Amounts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DO YOU EAT THESE ONCE A WEEK?

<table>
<thead>
<tr>
<th>FOOD</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>veal, lamb</td>
<td>01</td>
<td>pancakes, waffles</td>
</tr>
<tr>
<td>beef</td>
<td>03</td>
<td>instant breakfast, oatmeal</td>
</tr>
<tr>
<td>mixed dish without chicken</td>
<td>05</td>
<td>pudding</td>
</tr>
<tr>
<td>Chinese dishes</td>
<td>06</td>
<td>milk shake</td>
</tr>
<tr>
<td>Mexican dishes</td>
<td>07</td>
<td>other dairy products</td>
</tr>
<tr>
<td>beef or ham</td>
<td>08</td>
<td>other dessert, sweet</td>
</tr>
<tr>
<td>potatoes or bean burritos</td>
<td>09</td>
<td>sour cream, dips</td>
</tr>
<tr>
<td>Polish or Italian sausage</td>
<td>10</td>
<td>diet salted dressing, cream</td>
</tr>
<tr>
<td>cream soups</td>
<td>11</td>
<td>cream soup</td>
</tr>
<tr>
<td>noodles</td>
<td>12</td>
<td>green chilies, jalapenos</td>
</tr>
<tr>
<td>onions</td>
<td>31</td>
<td>summer squash</td>
</tr>
<tr>
<td>asparagus</td>
<td>33</td>
<td>sweet greens</td>
</tr>
<tr>
<td>sweet red peppers</td>
<td>35</td>
<td>beans</td>
</tr>
<tr>
<td>apple</td>
<td>37</td>
<td>avocado, guacamole</td>
</tr>
<tr>
<td>pineapples</td>
<td>39</td>
<td>pineapple or pineapple juice</td>
</tr>
<tr>
<td>prunes or prune sauce</td>
<td>41</td>
<td>cranberry juice cocktail</td>
</tr>
<tr>
<td>grapes</td>
<td>43</td>
<td>margaritas</td>
</tr>
<tr>
<td>papayas</td>
<td>45</td>
<td>homemade or canned melon</td>
</tr>
<tr>
<td>nuts and seeds</td>
<td>47</td>
<td>raisins or lemon juice</td>
</tr>
<tr>
<td>bran</td>
<td>49</td>
<td>other vegetable juice</td>
</tr>
<tr>
<td>other fruit</td>
<td>51</td>
<td>other not mentioned here</td>
</tr>
</tbody>
</table>
10x. Think about the way you used to eat ABOUT 20 YEARS AGO. Was it very different from the way you have been eating in the past year?

- NO
- YES
- CAN'T SAY

11x. Please write MORE, LESS, or THE SAME to describe how much of the following foods you ate 20 years ago compared to now.

- Total calories?
- Charcoal-broiled meat?
- Butter and oil?
- Fruit?
- Vegetables?
- Vitamin Supplements?

12x. Do you take any prescription drugs as often as everyday or almost every day?

- NO
- YES
- OTHER

If YES, What is the most important or strongest medicine that you take? ____________________ (Name of drug or "Unknown")

What is the next most important? ____________________

13x. Have you had a cold, the flu, or another viral illness within the past three weeks?

- NO
- YES
- POSSIBLY
- OTHER

14x. Have you taken antibiotics (like penicillin, erythromycin, etc.) within the last three weeks?

- NO
- YES
- POSSIBLY
- OTHER
### MEDICAL INFORMATION

23. In the past five years, how many times have you been hospitalized?  
(if female, omit childbirths) ___ times

24. Have you ever had any of the following tests or treatments?  

<table>
<thead>
<tr>
<th>Test or Treatment</th>
<th>NO</th>
<th>YES</th>
<th>HOW MANY TIMES?</th>
<th>AGE AT FIRST TREATMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray treatments for acne, ringworm, enlarged tonsils, adenoids, thymus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment with radium, cobalt, or other radioactive isotopes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper GI series (x-ray of stomach after drinking white liquid)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower GI series (Barium enema)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

25. Have you ever been told by a doctor that you had any of the following conditions?  

<table>
<thead>
<tr>
<th>Condition</th>
<th>NO</th>
<th>YES</th>
<th>DON'T KNOW</th>
<th>DON'T KNOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart disease or angina</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart attack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High blood pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuberculosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic bronchitis or emphysema</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hay fever</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diverticulosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal/colon polyps</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic colitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thyroid condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidney disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bladder disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver cirrhosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stomach ulcers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other arthritis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteoporosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractured hip</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate trouble</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal Pap smear</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skin cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leukemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If yes to leukemia, skin, or other cancer, fill in below:

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Year 1st Diagnosed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

26. In the past year, have you had  

<table>
<thead>
<tr>
<th>Condition</th>
<th>NO</th>
<th>YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bleeding or sore gums</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bruise easily</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nosebleeds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty seeing in the dark</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequent or chronic fever</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequent constipation or hemorrhoids</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### FOR OFFICE USE

<table>
<thead>
<tr>
<th>Code</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Bladder</td>
</tr>
<tr>
<td>02</td>
<td>Bone</td>
</tr>
<tr>
<td>03</td>
<td>Brain</td>
</tr>
<tr>
<td>04</td>
<td>Breast</td>
</tr>
<tr>
<td>05</td>
<td>Cervix</td>
</tr>
<tr>
<td>06</td>
<td>Colon</td>
</tr>
<tr>
<td>07</td>
<td>Esophagus</td>
</tr>
<tr>
<td>08</td>
<td>Kidney</td>
</tr>
<tr>
<td>09</td>
<td>Liver</td>
</tr>
<tr>
<td>10</td>
<td>Leukemia</td>
</tr>
<tr>
<td>11</td>
<td>Lung, bronchus</td>
</tr>
<tr>
<td>12</td>
<td>Lymphoma, including Hodgkins</td>
</tr>
<tr>
<td>13</td>
<td>Mouth, oral</td>
</tr>
<tr>
<td>14</td>
<td>Nerves</td>
</tr>
<tr>
<td>15</td>
<td>Pancreas</td>
</tr>
<tr>
<td>16</td>
<td>Prostate</td>
</tr>
<tr>
<td>17</td>
<td>Rectum</td>
</tr>
<tr>
<td>18</td>
<td>Skin-Melanoma</td>
</tr>
<tr>
<td>19</td>
<td>Skin-Not melanoma (Basal or squamous)</td>
</tr>
<tr>
<td>20</td>
<td>Skin-Not specified</td>
</tr>
<tr>
<td>21</td>
<td>Stomach</td>
</tr>
<tr>
<td>22</td>
<td>Thymus</td>
</tr>
<tr>
<td>23</td>
<td>Testis</td>
</tr>
<tr>
<td>24</td>
<td>Other</td>
</tr>
</tbody>
</table>

---

Please go to next page.
15x. Has a doctor ever told you that you had:

- an **AUTOIMMUNE DISEASE**, where your body makes antibodies to your own tissues?

  - NO
  - YES
  - NOT SURE

If YES, __________ 

What kind? __________ 

(Name of disease or "Unknown")

<table>
<thead>
<tr>
<th>Question</th>
<th>NO</th>
<th>YES</th>
<th>NOT SURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>aplastic anemia?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>food allergy?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dust allergy?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pollen allergy?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>allergy to animal hair?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>allergy to insects?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>drug allergy?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other allergy?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### OCCUPATIONAL INFORMATION

27. What is your current employment status? Check the one that applies to the greatest percent of your time.

1  —  Employed  
2  —  Homemaker  
3  —  Retired  
4  —  Disabled, unable to work  
5  —  Unemployed  
6  —  Student  
7  —  Other

28. What has been your usual occupation or job — the one you have worked at the longest? (For example, carpenter, executive, salesman, foreman, waitress, truck driver)

Job/occupation: ____________________________

Years in this job: __________________________

In your work, did you spend more time 1  indoors  2  outdoors? (Please check one.)

29. In your work, have you ever been exposed for a year or more to any of the following?

<table>
<thead>
<tr>
<th>Substance</th>
<th>1</th>
<th>2</th>
<th>DON'T</th>
<th>1</th>
<th>2</th>
<th>DON'T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asbestos</td>
<td>NO</td>
<td>YES</td>
<td>KNOW</td>
<td>NO</td>
<td>YES</td>
<td>KNOW</td>
</tr>
<tr>
<td>Radiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Welding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coal tar, soot, pitch, creosote, asphalt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mineral, cutting or lubricating oil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benzidine, beta-naphthylamine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benzene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isopropyl oil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyesulfites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arsenic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron foundry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nickel smelting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underground mining</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lumber industry, or heavy wood dust</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubber or cablemaking industry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical or plastics industry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pesticides, herbicides</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mustard gas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cadmium, beryllium, vinyl chloride</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

16x. What was your second longest occupation or job?

(Job name or "None")

If you worked at a second occupation

How many years did you do that work? ____________ yrs

Was the work mostly INDOORS or OUT OF DOORS?

1  —  INDOORS 
2  —  OUT OF DOORS 
3  —  BOTH 
4  —  CAN'T SAY
17x. Have you ever worked for a year or more in the following industries:

<table>
<thead>
<tr>
<th>Industry</th>
<th>Code</th>
<th>Yes</th>
<th>No</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shipbuilding industry?</td>
<td>X174</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Construction industry?</td>
<td>X175</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Fishing industry?</td>
<td>X176</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Lumber, wood, furniture manufacturing, or paper industries?</td>
<td>X177</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Petrochemical industry?</td>
<td>X178</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Other chemical industries?</td>
<td>X179</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Metal refining, manufacturing, polishing, or plating indus-</td>
<td>X180</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>tries?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sugar cane farming?</td>
<td>X181</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Mining industry?</td>
<td>X182</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Insulation manufacturing or installation?</td>
<td>X183</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>
(Continuation of question 17x.)

asbestos manufacturing or manufacturing of products containing asbestos?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>YES</td>
<td>NOT SURE</td>
</tr>
</tbody>
</table>

cement manufacturing?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>YES</td>
<td>NOT SURE</td>
</tr>
</tbody>
</table>

demolition industry?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>YES</td>
<td>NOT SURE</td>
</tr>
</tbody>
</table>

cook?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>YES</td>
<td>NOT SURE</td>
</tr>
</tbody>
</table>

gas station attendant or auto mechanic?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>YES</td>
<td>NOT SURE</td>
</tr>
</tbody>
</table>

FAMILY HISTORY

30. Have any close relatives had cancer?  1  No  2  Yes

If YES, please fill this out for each blood relative who had cancer. Include your natural parents, sisters and brothers, daughters and sons, grandparents.

<table>
<thead>
<tr>
<th>One RELATIVE per line (Mother, son, etc.)</th>
<th>Circle one</th>
<th>If Alive, give age</th>
<th>If Dead, give age at death</th>
<th>Type of Cancer</th>
<th>Age at Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PLEASE GO TO NEXT PAGE
18x. In your work or hobbies, have you ever been exposed FOR A YEAR OR MORE to any of the following:

- petroleum refining?
  - NO
  - YES
  - NOT SURE

- smoke: wood, oil, sugar cane?
  - NO
  - YES
  - NOT SURE

- gasoline/motor oil?
  - NO
  - YES
  - NOT SURE

- x-rays, fluoroscopic equipment, or radio isotopes?
  - NO
  - YES
  - NOT SURE

- coke oven gases?
  - NO
  - YES
  - NOT SURE

OTHER HEALTH FACTORS

31. How tall are you? __ feet ___ inches  32. How much do you weigh? ____ pounds

33. What is the most you have ever weighed? ____ pounds

34. About how many times have you gone on a diet to lose weight?
   (1) Never  (2) 1-2  (3) 3-5  (4) 6-8  (5) 9-11  (6) 12 or more times

35. How many hours of sleep do you usually get at night?
   (1) 6 hours or less  (2) 7 hours  (3) 8 hours  (4) 9 hours or more

36. How often do you feel under stress which makes you tense or worried, or causes physical problems such as stomach or back trouble or headaches?
   (1) Every day  (2) Several times  (3) Several times  (4) Several times  (5) Rarely or never
   a week  a month  a year  never
37. Here is a list of active things that people do in their free time. How often do you do any of these things?

<table>
<thead>
<tr>
<th></th>
<th>MORETHAN</th>
<th>ABOUT</th>
<th>A FEW</th>
<th>A FEW</th>
<th>RARELY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ONCE A</td>
<td>WEEK</td>
<td>TIMES A</td>
<td>TIMES A</td>
<td>OR NEVER</td>
</tr>
<tr>
<td>Active sports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doing physical exercises</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jogging or running</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swimming or taking long walks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gardening, fishing, hunting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Something else</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

38. How many close friends do you have? (People that you feel at ease with, can talk to about private matters, and can call on for help.)

- None
- 1 or 2
- 3 to 5
- 6 to 9
- 10 or more

39. How often do you participate in the following groups or activities?

<table>
<thead>
<tr>
<th></th>
<th>MORETHAN</th>
<th>ABOUT</th>
<th>A FEW</th>
<th>A FEW</th>
<th>RARELY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ONCE A</td>
<td>WEEK</td>
<td>TIMES A</td>
<td>TIMES A</td>
<td>OR NEVER</td>
</tr>
<tr>
<td>Go to church or temple</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participate in group meetings or activities (such as clubs, PTA, professional, labor or service groups)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

19x. Today's date is ____________, 19__

Please take a moment to fill in any questions you may have skipped.

THANK YOU VERY MUCH for taking the time to fill out this information. The answers you have given will be very useful in interpreting the results of this study, and in helping to understand and control disease. Your participation is sincerely appreciated.

Reviewed by ____________________________
APPENDIX B

MISCELLANEOUS STUDY MATERIALS
CONSENT AND AUTHORIZATION FOR VENIPUNCTURE AND WITHDRAWAL OF BLOOD

LOUISIANA STATE UNIVERSITY
and
MARY BIRD PERKINS CANCER CENTER

I hereby consent to have a blood specimen of approximately 10 milliliters (1 tube) withdrawn from a vein in my arm, with a new, sterile needle, by a fully certified Medical Technologist. This blood sample will be used in the conduct of laboratory research investigations, and I have been informed that this is the best method of obtaining the blood sample for this purpose.

This procedure has been routinely used for many years in caring for patients and is generally considered by these investigators and others to be of little or no risk to me. There is some discomfort and the possibility of minor bruising and leakage of blood under the skin at the site of blood withdrawal. This usually heals quickly. I have been informed that any questions I may have concerning the procedures involved will be answered.

I understand that I shall be asked to supply information about aspects of my personal history and habits which are relevant to the health condition under study, but that I may choose not to answer a particular question or questions.

I understand that Mary Bird Perkins Cancer Center will protect my right to privacy and will not permit the research personnel to know my full name, or any information which would identify me. The results of the analysis of my blood sample and all questionnaire data will be coded, analyzed, and processed for presentation or publication in such a way that my identity will not be revealed.

I have agreed to do this of my own free will and without threat or coercion of any kind. I understand that I will not benefit medically from this procedure, but that it may provide helpful information for this research.

I have been informed that I may withdraw my consent and discontinue participation at any time without prejudice to myself or loss of access to care at this facility.

If I have questions at any time, I may call Dr. M.J. Newman (346-3329) or Mrs. Joan Blackmon (346-3353).

YOU ARE MAKING A DECISION WHETHER OR NOT TO PARTICIPATE IN THIS RESEARCH. YOUR SIGNATURE INDICATES THAT HAVING READ THE ABOVE INFORMATION, YOU HAVE DECIDED TO PARTICIPATE.

__________________________ Date
                      AM
                      PM Time

__________________________ Signature

__________________________ Signature of Investigator

__________________________ Signature of Witness
Medical Information Form for Lung Cancer Patients

Histologic Type:
- Oat Cell
- Squamous Cell
- Adenocarcinoma
- Large Cell
- Other (Mixed, Sarcoma, etc.)

Area Affected:
- Main Bronchus
- Right Lung or Left Lung
- Upper Lobe
- Middle Lobe
- Lower Lobe
- Other or not specified

Previous Therapy:
- Surgery
- Chemotherapy
  - Drug Prescribed
- None

Stage:
- Local
- Regional
- Disseminated

Physician
Co-Investigators:
Joan P. Blackmon, M.S.P.H.
Louisiana State U.
Mark J. Newman, Ph.D.
Louisiana State U.
Kenneth K. Lo, M.D., Ph.D.
Mary Bird Perkins
Cancer Center

VOLUNTEERS
are
NEEDED
to serve as
CANCER-FREE
comparison subjects.
BRIEF DESCRIPTION
OF THE LSU-MBPCC
LUNG CANCER STUDY

There are certain chemicals, called PAH's (Polycyclic Aromatic Hydrocarbons), which are present in cigarette smoke, in charbroiled meats and other foods, in petroleum products and elsewhere in the environment. We think that they may play a role in the development of certain cancers in humans, including lung cancer. We know that these chemicals can damage genes in cells by binding to the DNA molecules. Some people also make antibodies to their own damaged PAH-DNA and we would like to know if this plays any role in the likelihood that these people would develop lung cancer.

Another major focus of the study will be an evaluation of the involvement of different types of white blood cells in specific kinds of lung cancer. At LSU we have an intricate piece of machinery, called a Fluorescence-Activated Cell Sorter, which can identify and count many different types of white blood cells from a small sample of blood. Your white blood cells are the army of your immunological defense system and it is thought that imbalances in the proportions of the different types of cells may be involved in why people get lung cancer and in what type of lung cancer they get.

Although we are particularly interested in the above factors, it is also necessary to have information about other factors known or suspected to be involved with the causation of lung cancer. This allows us to identify combinations of characteristics which may predispose certain people to develop cancer. We obtain this information using a questionnaire. The questionnaire covers a wide range of descriptors, among which are dietary factors, smoking history and practice, work-related exposures, history of allergies and family history.

At this time, we are limiting the study to white males, because they have the highest rates of lung cancer. We will study men who come to MBP Cancer Center for radiation treatment of their lung tumors. The comparison group will consist of white, male volunteers who are cancer-free.

The above is only a very brief overview. If you have questions or would like to know more about this Study, please contact the Study Director Mrs. Joan Blackmon at LSU (388-1213).

We need
WHITE MEN
who have
NEVER HAD CANCER
(other than skin cancer)
who are
OVER 30 YEARS OF AGE
who are
SMOKERS, FORMER-SMOKERS,
or
NON-SMOKERS
and who are
RESIDENTS OF
E. Baton Rouge  W. Baton Rouge
E. Feliciana  W. Feliciana
Assumption  Ascension
Livingston  Tangipahoa
Iberville  St. Helen
or Pointe Coupee parishes.

Volunteers will need to come to Mary Bird Perkins Cancer Center to give a SMALL SAMPLE OF BLOOD (1 Tablespoonful) and to complete the QUESTIONNAIRE which will take 1 hour. EVENING and WEEKEND appointment times are planned.

Fill out the attached form and give it to the MBP Cancer Center receptionist or call Mrs. Joan Blackmon at 388-1213 to arrange an appointment.

263
Dear Madam,

Mary Bird Perkins Cancer Center is collaborating with Louisiana State University in a study through which we hope to identify factors involved in the causation of lung cancer. A brief description of the study is attached. The study group will consist of white males who come to the Cancer Center for treatment of their lung cancer. We also need a comparison group of white men who do not have lung cancer. Because cancer has touched your life, we hope that you will help us find volunteers to serve as healthy control subjects. Your blood relatives cannot be eligible because of the tendency for cancer to "run in families". But do you have a husband, in-law, neighbor, or friend, who is over 30 years of age, who has never had cancer (other than skin cancer), and who would be willing to come to the Cancer Center to 1. fill out a questionnaire which would take about an hour, and 2. give us a small sample of blood (about 1 tablespoonful)? We would be glad to accept as many gentlemen as are willing to participate.

If you can think of persons who would like to help us to identify ways to prevent lung cancer, please give each one a brochure which consists of a study description, a list of characteristics which control subjects must have, and directions for volunteering. These are available from the Center's receptionist. If you would prefer that we contact him(her) directly, please write the name(s) and phone number(s) on the attached blue form and give it to the receptionist; she will pass the information on to us.

Thank you for your help. We wish you success in your personal battle with cancer.

Sincerely,

Joan P. Blackmon, M.S.P.H.
Study Director
BRIEF DESCRIPTION OF THE
LSU LUNG CANCER STUDY

Co-Investigators:
Joan P. Blackmon, M.S.P.H.
Louisiana State University
Mark J. Newman, Ph.D.
Louisiana State University
Kenneth K. Lo, M.D., Ph.D.
Mary Bird Perkins Cancer Center

There are certain chemicals, called PAH's (Polycyclic Aromatic Hydrocarbons), which are present in cigarette smoke, in charbroiled meats and other foods, in petroleum products, and elsewhere in the environment. We think that they may play a role in the development of certain cancers in humans, including lung cancer. We know that these chemicals can damage genes in cells by binding to the DNA molecules. Some people also make antibodies to their own damaged PAH-DNA and we would like to know if this plays any role in the likelihood that these people would develop lung cancer.

Another major focus of the study will be an evaluation of the involvement of different types of white blood cells in specific kinds of lung cancer. At LSU we have an intricate piece of machinery, called a Fluorescence-Activated Cell Sorter, which can identify and count many different types of white blood cells from a small sample of blood. Your white blood cells are the army of your immunological defense system and it is thought that imbalances in the proportions of the different types of cells may be involved in why people get lung cancer and in what type of lung cancer they get.

Although we are particularly interested in the above factors, it is also necessary to have information about other factors known or suspected to be involved with the causation of lung cancer. This allows us to identify combinations of characteristics which may predispose certain people to develop cancer. The questionnaire covers a wide range of descriptors, among which are dietary factors, smoking history and practice, work-related exposures, history of allergies, and family history. For your own information, we can provide you with a printout of the dietary analysis, which will let you know how well your diet is supplying you with the major nutrients that you need.

The above is only a very brief overview. If you have questions or would like to know more about the Study, please contact the Study Director Mrs. Joan Blackmon at LSU (388-1213).
Dear Sir:

We know that this is a most difficult time for you and we would not bother you if there were an alternative. My colleagues and I, with the financial support of The Cancer Society of Greater Baton Rouge, hope to learn how to intervene to prevent so many people from developing lung cancer. Our research is designed to learn more about why people get cancer in their lungs and in particular, why some smokers are more likely to get it and others not. On the next page is a very brief description of the sorts of factors which we are studying.

The only way that we can study this disease is to obtain the help of people like you, who have a lung cancer. We are asking that you give us a small sample of blood (about 1 Tablespoonful) on which to do immunological analyses, and that you complete a detailed questionnaire, which will take about an hour. The information requested in the questionnaire is restricted to factors known to be or thought to be associated with lung cancer. Only a number, not your name, will be assigned to your blood analysis and questionnaire data so that all information that you give us will be confidential.

The blood must be obtained before you have your radiation treatments. If you would like to help with this research, please fill in the attached green form and give it to the appointments secretary when you make the appointment for your first treatment. Our medical technologist, Ms. Marilyn Dietrich, will be here at that time to obtain the blood specimen and to give you the questionnaire. It would be best if you completed the questionnaire at the Center and returned it to the receptionist before you leave, but if that is a problem, you are welcome to take it home and return it to the receptionist at your next visit. You may refuse to continue in the study at any time. If for whatever reason you feel that you cannot participate, indicate that on the green form, as well, but do please fill out the green form and turn it in at the reception desk before you leave today. If you have any questions about the study or the questionnaire, please call Mrs. Blackmon at 388-1213 or 346-3353.

We thank you for your help in this study.

Sincerely,

Joan P. Blackmon, M.S.P.H.
Study Director
THIS VERSION WAS DESIGNED FOR CD8 WHERE THERE ARE APPARENTLY AT LEAST TWO POSITIVE GROUPS. IT WILL SELECT TWO OF THREE OR FOUR CLUSTERS: THE ONE WITH THE HIGHEST FLUORESCENCE AND THE ONE WITH THE LOWEST. BIVARIATE (OF BY SS MEANS AND COVARIANCES ARE ESTIMATED FROM THE SELECTED CLUSTERS AND ALL EVENTS WITH FLUORESCENCE BETWEEN (THE LOWER MEAN + 2 STD. DEV.) AND (THE UPPER MEAN - 2 STD. DEV.) ARE ASSIGNED TO A THIRD CLUSTER. THE BIVARIATE MEANS, COVARIANCES AND POPULATIONS OF THOSE THREE CLUSTERS THEN ARE USED AS STARTER VALUES TO FIT A MIXTURE OF 3 BIVARIATE NORMAL CURVES TO THE DATA.

//SAST8 JOB (1940,60434,,75),'J BLACKMON',NOTIFY=VTJOAN,MSGCLASS=S,
// TIME=5,REGION=1024K
/*JOBPARM SHIFT=D
 // EXEC SAS,TIME=5
 //INC06309 DD DSN=VTJOAN.JBDAT8(JBC06309),DISP=SHR
 //INC06709 DD DSN=VTJOAN.JBDAT8(JBC06709),DISP=SHR
 //INC06809 DD DSN=VTJOAN.JBDAT8(JBC06809),DISP=SHR
 //INC06909 DD DSN=VTJOAN.JBDAT8(JBC06909),DISP=SHR
 //INC07009 DD DSN=VTJOAN.JBDAT8(JBC07009),DISP=SHR
 //INC07409 DD DSN=VTJOAN.JBDAT8(JBC07409),DISP=SHR
 //INF01308 DD DSN=VTJOAN.JBDAT8(JBP01308),DISP=SHR
 //INF01908 DD DSN=VTJOAN.JBDAT8(JBP01908),DISP=SHR
 //INF02008 DD DSN=VTJOAN.JBDAT8(JBP02008),DISP=SHR
 //OTC06309 DD DSN=VTJOAN.JBPROD8(JBC06309),DISP=SHR
 //OTC06709 DD DSN=VTJOAN.JBPROD8(JBC06709),DISP=SHR
 //OTC06809 DD DSN=VTJOAN.JBPROD8(JBC06809),DISP=SHR
 //OTC06909 DD DSN=VTJOAN.JBPROD8(JBC06909),DISP=SHR
 //OTC07009 DD DSN=VTJOAN.JBPROD8(JBC07009),DISP=SHR
 //OTC07409 DD DSN=VTJOAN.JBPROD8(JBC07409),DISP=SHR
 //OTP01308 DD DSN=VTJOAN.JBPROD8(JBP01308),DISP=SHR
 //OTP01908 DD DSN=VTJOAN.JBPROD8(JBP01908),DISP=SHR
 //OTP02008 DD DSN=VTJOAN.JBPROD8(JBP02008),DISP=SHR

******************************************************************************
/*This program takes crudely edited, decoded lymphocyte data
*and edits it on forward and side scatter. The mean vector
*covariance matrix for a bivariate normal distribution are
*estimated using PROC CORR. These are used to calculate a
*standardized, squared distance from the mean for each point
*in the crude data and events which fall outside a 95% confi-
dence ellipse are excluded. To eliminate the distortion of
*the parameter estimate caused by outliers in the crude data,
*the parameters are reestimated using the data from the first
*edit. Then the crude data are edited again based on a 90%
*confidence ellipse determined by the new estimate of the
*parameters.
*If the first edit of the data does not eliminate at least 4%
of the points, so that most of the outlier platelets, etc. are
*removed, the first edit is repeated using a 90% confidence
*ellipse, then the second edit is done as above.
*PROC FASTCLUS is used to estimate the number of events
*fluorescence-positive and to estimate the mean fluorescence
*and the standard deviation of the positive and negative popu-
*lations. An ASCII file containing the number of events retained
*after editing, the number of fluorescence parameters in the
*file (1 or 2), the fluorescence channels for each retained
*event, the mean vectors on fluorescence for the positive and
*negative populations and their standard deviations, and the
*estimated proportions for each population is constructed for
*further analysis.
*For each antibody a slightly different clustering procedure might
*be needed in that one may have to retain more than two clusters
*for the final estimate so that a minor positive population will
*be recognized. FOR T3 & T4, CLR=2 FOR LEU-14 & 3a1, CLR=4 OR 3
*FOR NKR-1, CLR=3 FOR T8, CLR=3
******************************************************************************
%MACRO DATAL;
OPTIONS REPLACE PS=60 LS=72;
DATA LYM (KEEP=P1 P2 P3 P4); /*CRUDE DATA*/
INFILE &DSN END=ENDD;
INPUT &1 TITLE $13.; /*READ FILENAME*/
SUBSTR(TITL1,1,2)='JB'; /*MAKE 1ST 2 CHAR'S 'JB'* /
SUBSTR(TITL2,TITL1); /*MAKE FILENAME GLOBAL VARIABLE*/
INDEX=SUBSTR(TITLE,5,3); /*MAKE FILENAME GLOBAL VARIABLE*/
IF SUBSTR(INDEX,1,1)='A' THEN SUBSTR(INDEX,1,1)='0';
DO Z=1 BY 1 WHILE (ENDD=0);
IF INDEX='014' OR INDEX='017' OR INDEX='043' THEN DO;
INPUT &1 P1 1-3 P2 4-6 P3 7-9 P4 10-12;
IF P1>50 THEN OUTPUT LYM;
ELSE DO;
INPUT &1 P1 1-3 P2 4-6 P3 7-9 P4 10-12;
OUTPUT LYM; /*READ IN FS, GF, RF,& SS */
END;
END;
DROP TITLE TITL1 TITL2 INDEX;
STOP;
RUN;
%MEND DATAL;

%MACRO COREDIT;
PROC CORR DATA=&SSET NOCORR NOPRINT COV OUTP=B(TYPE=COV);
TITLE;
VAR P1 P4; /*ESTIMATE BIVARIATE PARAM'RS*/
RUN;

DATA &CELLS (KEEP=P1 P2 P3 P4);
DO I=1 TO NPARM;
SET & POINT=I NOBS=NPARM; /*FETCH THE ESTIMATED PARAM'RS*/
SELECT (1);
WHEN (1) DO;
S1S1=P1;
S1S2=P4;
END;
WHEN (2) S2S2=P4;
WHEN (3) DO;
X1BAR=P1;
X2BAR=P4;
END;
WHEN (5) N=P1;
OTHERWISE;
END;
END;
DET=1/(S1S1*S2S2-S1S2*S1S2); /*CALCULATE THE ELEMENTS NEEDED*/
AIN=DET*S2S2; /*TO CALCULATE THE STANDARDIZED*/
BIN=DET*(-S1S2); /*DISTANCE FROM THE MEAN. */
CIN=DET*S1S1;
CHI95=5.99; /*THE CHISQ'S NEEDED DEPENDING */
CHI90=4.61; /*ON WHICH EXECUTION OF THIS */
CHI85=4.15; /*SUBROUTINE. */
CHI75=2.77; /*SUBROUTINE. */
CALL SYMPUT('MESG','NO'); /*INITIALIZE MACROVARIABLE*/
K=0;
DO I=1 TO NLYM;
SET LYM POINT=I NOBS=NLYM; /*EDIT DATA*/
X1=P1-X1BAR;
X2=P4-X2BAR;
TEST=(X1**2)*A1 + 2*X1*X2*B1 + (X2**2)*C1;
IF TEST<6CHI THEN DO;
K=K+1;
OUTPUT &CELLS;
END;
END;
IF (K/NLYM > 0.96) THEN CALL SYMPUT('MESG','YES'); /*IF > 96% KEPT THEN SET UP FOR */
/*REEDIT. */
CALL SYMPUT('KA', K);
STOP;
RUN;
%MEND COREDIT;

%MACRO PLOT;
OPTIONS PS=62 LS=125;
PROC PLOT DATA=LYM;
TITLE "CRUDE LYMPHOCYTE CUT OF &TTL";
PLOT P4*P1 / VAXIS=0 TO 120 BY 20 HAXIS=20 TO 210 BY 30;
RUN;
PROC PLOT DATA=EDTD;
TITLE "FINAL EDIT OF &TTL";
PLOT P4*P1 / VAXIS=0 TO 120 BY 20 HAXIS=20 TO 210 BY 30;
RUN;
%MEND PLOT;

%MACRO CLUS;
PROC STANDARD DATA=EDTD MEAN=0 STD=1 OUT=STAN;
TITLE; /*STANDARDIZE DATA TO MEAN=0 AND STD*/
VAR P2 &P; /*DEV = 1*/
RUN;
PROC FASTCLUS DATA=STAN MAXITER=0 MAXC=6 NOPRINT MEAN=MN;
TITLE; /*FIND SEEDS FOR FINAL CLUSTER*/
VAR P2 &P;
RUN;
PROC SORT DATA=MN;
BY _FREQ_;
RUN;
DATA RMN;
SET MN;
NUM=6 - (&CLR - 1);
IF _N_<NUM THEN DELETE; /*SELECT N CLUSTERS FOR SEEDS, N DE*/
/*PENDS ON WHICH ANTIBODY*/
PROC FASTCLUS DATA=STAN MAXC=6 CLR MAXITER=7 SEED=RMN NOPRINT
OUT=T3 DRIFT MEAN=OUT1;
VAR P2 &P; /*CLUSTER TO GET NUMBERS POSITIVE*/
/*AND NEGATIVE.*/
RUN;
PROC SORT DATA=OUT1; /*MAKE SURE THAT OUT1 IS IN ORDER*/
BY DESCENDING P2; /*OF DECREASING FLUORESCENCE*/
RUN;
DATA INTMED (KEEP=PR2 PR4 CLUSTER);
    SET T3;
    PR2=P2;
    PR4=P;
    RUN;

DATA MIX (KEEP=P2 &P CLUSTER);
    MERGE EDTD INTMED;
    RUN;

PROC PLOT DATA=MIX;
    TITLE "CLUSTERING &TTL ON GF AND SS, STD";
    PLOT P2*P=CLUSTER;
    RUN;

DATA NULL:
    DO K=1 TO OBS;
    SET OUT1 POINT=K NOBS=OBS;
    SELECT (K);
    WHEN (1) DO;
        CLSTR=CLUSTER;
        CALL SYMPUT('IND1',CLSTR);
    END;
    WHEN (OBS) DO;
        CLSTR=CLUSTER;
        CALL SYMPUT('INDNG',CLSTR);
    END;
    OTHERWISE;
    END;
    END;
    STOP;
RUN;

DATA CLUS1 CLUSNG;
    KEEP P2 &P CLUSTER;
    CLSTR1=IND1;
    CLUSTN=INDNG;
    SET MIX;
    IF CLUSTER=CLSTR1 THEN OUTPUT CLUS1;
    ELSE IF CLUSTER=CLUSTN THEN OUTPUT CLUSNG;
    ELSE;
    RUN;

PROC CORR DATA=CLUS1 NOCORR NOPRINT COV OUTP=B1(TYPE=COV);
    TITLE;
    VAR P2 &P; /*ESTIMATE P2 MEAN & VARIANCE*/
    RUN; /*FOR CLUSTER 1*/

PROC CORR DATA=CLUSNG NOCORR NOPRINT COV OUTP=BNG(TYPE=COV);
    TITLE;
    VAR P2 &P; /*ESTIMATE P2 MEAN & VARIANCE*/
    /*FOR NEGATIVE CLUSTER*/

DATA CLUS1 CLUS2 CLUSNG PLOTIT;
    KEEP P2 P4 CLSTR;
    DO I=1 TO NPARM;
    SET B1 POINT=I NOBS=NPARM;
    SELECT (I);
    WHEN (1) SF2=P2;
    WHEN (3) MNF2=P2;
    OTHERWISE;
    END;
    SDP2=SQRT(SF2);
    LOWER=MNF2 - 2*SDP2;
DO I=1 TO NPARM;
SET BNG POINT=I NOBS=NPARM;
SELECT (I);
WHEN (1) SPNG=P2;
WHEN (3) MNNG=P2;
OTHERWISE;
END;
END;
SDNG=SQR(SPNG);
UPPER=MNNG + 2*SDNG;

DO I=1 TO NLYM;
SET EDTD POINT=I NOBS=NLYM;
IF P2 LE UPPER THEN DO;
CLSTR=3;
OUTPUT CLUSNG;
END;
ELSE IF P2 > UPPER AND P2 < LOWER THEN DO;
CLSTR=2;
OUTPUT CLUS2;
END;
ELSE DO;
CLSTR=1;
OUTPUT CLUS1;
END;

OUTPUT PLOTIT;
END;
STOP;

PROC PLOT DATA=PLOTIT;
TITLE 'PLOT OF THE THREE CLUSTERS';
PLOT P2*P4 = CLSTR;

PROC CORR DATA=CLUS1 NOCORR NOPRINT COV OUTP=B1(TYPE-COV);
TITLE;
VAR P2 &P; /*ESTIMATE VAR-COVAR FOR */
RUN; /*CLUSTER 2 */

PROC CORR DATA=CLUS2 NOCORR NOPRINT COV OUTP=B2(TYPE-COV);
TITLE;
VAR P2 &P; /*ESTIMATE VAR-COVAR FOR */
RUN; /*CLUSTER 2 */

PROC CORR DATA=CLUSNG NOCORR NOPRINT COV OUTP=BNG(TYPE-COV);
TITLE;
VAR P2 &P; /*ESTIMATE VAR-COVAR FOR */
RUN; /*CLUSTER 2 */
%MEND CLUS;
%MACRO CREATDS;

DATA NULL ;
FILE OT&TITL;
NAME=SYMGET('TTL'); /*CREATE A FILE TO BE USED BY */
PARAM=2; /*NORFIT */
KAA=SYMGET('KA');
PUT NAME;
PUT KAA PARAM;
PAR=SYMGET('P');
DO I=1 TO NLYM;
SET EDTD POINT=I NOBS=NLYM;
IF PAR='P4' THEN PUT P2 P4 P1 P3;
ELSE IF PAR='P3' THEN PUT P2 P3 P1 P4;
ELSE;
END;
NG=3;
PUT NG;
NCOV=2;
PUT NCOV;
NSWH=2;
PUT NSWH;
DO J=1 TO NGRP;
  SET B1 POINT=J NOBS=NGRP;
  /*WRITE MEAN VECTOR FOR 1ST CLUS */
  /*B1 IS FROM PROC CORR ON 1ST CLUS*/
SELECT (J);
  END;
END;
DO J=1 TO NGRP;
  SET B2 POINT=J NOBS=NGRP;
  /*WRITE MEAN VECTOR FOR 2ND CLUS */
  /*B2 IS FROM PROC CORR ON 2ND CLUS*/
SELECT (J);
  WHEN (3) PUT P2 &P;
  OTHERWISE;
  END;
END;
DO J=1 TO NGRP;
  SET BNG POINT=J NOBS=NGRP;
  /*WRITE MEAN VECTOR FOR NEG CLUS */
  /*B3 IS FROM PROC CORR ON NEG CLUS*/
SELECT (J);
  WHEN (3) PUT P2 &P;
  OTHERWISE;
  END;
END;
DO J=1 TO NGRP;
  SET B1 POINT=J NOBS=NGRP;
  /*WRITE COVAR MATRIX FOR 1ST CLUS */
SELECT (J);
  WHEN (1) PUT P2 &P;
  WHEN (2) PUT P2 &P;
  OTHERWISE;
  END;
END;
DO J=1 TO NGRP;
  SET B2 POINT=J NOBS=NGRP;
  /*WRITE COVAR MATRIX FOR 2ND CLUS */
SELECT (J);
  WHEN (1) PUT P2 &P;
  WHEN (2) PUT P2 &P;
  OTHERWISE;
  END;
END;
DO J=1 TO NGRP;
  SET BNG POINT=J NOBS=NGRP;
  /*WRITE COVAR MATRIX FOR NEG CLUS */
SELECT (J);
  WHEN (1) PUT P2 &P;
  WHEN (2) PUT P2 &P;
  OTHERWISE;
  END;
END;
TOTPTS=0;
DO J=1 TO OBS1;
  SET B1 POINT=J NOBS=OBS1;
  SELECT (J);
  WHEN (5) N1=P2;
  OTHERWISE;
  END;
END;
DO J=1 TO OBS2;
  SET B2 POINT=J NOBS=OBS2;
  SELECT (J);
  WHEN (5) N2=P2;
  OTHERWISE;
  END;
END;
DO J=1 TO OBSNG;
  SET BNG POINT=J NOBS=OBSNG;
  SELECT (J);
  WHEN (5) NG=P2;
  OTHERWISE;
END;
END;
TOTPTS = N1 + N2 + NG;
PRP1 = N1/TOTPTS;
PRP2 = N2/TOTPTS;
PRPNG = NG/TOTPTS;
PUT PRP1 PRP2 PRPNG;
STOP;
RUN;
%MEND CREATDS;

%MEND OVERALL;

%MACRO OVERALL;
%DATA1;
%LET SETT=LYM;
%LET CELLS=EDITD;
%LET KIND=A %THEN %LET CHI=CHI95;
%ELSE %IF &KIND=B %THEN %LET CHI=CHI90;
%ELSE %IF &KIND=C %THEN %LET CHI=CHI85;
%ELSE %IF &KIND=D %THEN %LET CHI=CHI75;
%COREDIT;
%IF &MESS=YES %THEN %DO;
%LET CHI=CHI90;
%COREDIT;
%END;
%LET SETT=EDITD;
%LET CELLS=EDITD;
%LET CHI=CHI90;
%COREDIT;
%PLOT;
%CLUS;
%MEND CREATDS;

OPTIONS NODSNFRERR ERRORABEND;
DATA OUT;
  INPUT NAME $ X11 X12 X21 X22 S1S1A S1S2A S2S2A S1S1B S1S2B S2S2B
  PRP1 PRP2 TOTPTS;
  CARDS;
  RUN;
%LET CLR=3;
%LET P=P4;
%LET KIND=D;
%LET DSN=INP01308;
%OVERALL;
%LET KIND=A;
%LET DSN=INC06309;
%OVERALL;
%LET DSN=INC06709;
%OVERALL;
%LET DSN=INC06809;
%OVERALL;
%LET DSN=INC06909;
%OVERALL;
%LET DSN=INC07009;
%OVERALL;
%LET DSN=INC07409;
%OVERALL;
%LET DSN=INP01908;
%OVERALL;
%LET DSN=INP02008;
%OVERALL;
/*%LET DSN=INP02608;
%OVERALL; */
%LET DSN=INP04709;
%OVERALL;
//
//RUNIT PROC MEM=DUMMY
//GO  EXEC PGM=NORFIT4
//STEPLIB DD DSN=VTJOAN.FRT.LOAD,
//    DISP=SHR
//FT05F001 DD DUMMY
//FT06F001 DD SYSOUT=*  
//FT21F001 DD DSN=VTJOAN.JBPROD8(&MEM),DISP=SHR
//FT22F001 DD DSN=VTJOAN.JBFITS(&MEM),DISP=OLD
//FT23F001 DD DSN=VTJOAN.JBSUM8(&MEM),DISP=OLD
//RUNIT PEND
//******************************************************
  // EXEC RUNIT,MEM=JBC06309
  // EXEC RUNIT,MEM=JBC06709
  // EXEC RUNIT,MEM=JBC06809
  // EXEC RUNIT,MEM=JBC06909
  // EXEC RUNIT,MEM=JBC07009
  // EXEC RUNIT,MEM=JBC07409
  // EXEC RUNIT,MEM=JBP01308
  // EXEC RUNIT,MEM=JBP01908
  // EXEC RUNIT,MEM=JBP02008
This program creates a permanent SAS file containing the relevant results from the SUM files (for NKH-1 in this case.)

//SUMRY JOB (1940,60434,15), 'J BLACKMON', NOTIFY=VTJOAN, MSGCLASS=S,
//TIME=15, REGION=1024K
//*JOBPARM SHIFT=D
//*ROUTE PRINT LSUVAX.VTJOAN
//EXEC SAS, TIME=15
//INF00121 DD DSN=VTJOAN.JBSUM20(JBP00121),DISP=SHR
//ALL DD DSN=VTJOAN.JBSASOUT,DISP=OLD
//SYSIN DD *

OPTIONS PS=60 LS=132 NODSNFERR ERRORABEND;

%MACRO SET;

DATA PCT2;
INFILE DSN;
INPUT #1 NAME $8. #5 GMNGF1 GMNSS1 / GMNG GMNSSNG
#10 GVPOS1 / COVAR GVSSPOS1 #15 GVNEG / COVAR GVSSNEG
#20 GFPOS1 #25 EPPOS1 #30 NPOS1 NNEG
#34 ALOC1 ALOC2 #37 850 OVRL
#41 EMNGF1 EMNSS1 / EMNG EMNSSNG
#46 EVPOS1 / COVAR EVSSPOS1 #51 EVNEG / COVAR EVSSNEG;
TOTPTS = NPOS1 + NNEG;
FPFOS1 = NPOS1/TOTPTS;
SSTR=SUBSTR(NAME,4,3);
IF SUBSTR(SSTR,1,1)='X' OR SUBSTR(SSTR,1,1)='P' OR SUBSTR(SSTR,1,1)='A' OR SUBSTR(SSTR,1,1)='Y'
THEN SUBSTR(SSTR,1,1)='0';
DROP COVAR;
OUTPUT;
STOP;
RUN;

DATA ALL.NK2;
SET ALL.NK2 PCT2;
RUN;

%MEND SET;

%MACRO THREE;

DATA PCT3;
INFILE DSN;
INPUT #1 NAME $8. #5 GMNGF1 GMNSS1 / GMNGF2 GMNSS2 / GMNG GMNSSNG
#11 GVPOS1 / COVAR GVSSPOS1 #16 GVPOS2 / COVAR GVSSPOS2
#21 GVNEG / COVAR GVSSNEG #26 GFPOS1 GFPOS2 #32 EPPOS1 EPPOS2
#36 NPOS1 NPOS2 NNEG #40 ALOC1 ALOC2 ALOC3 #43 850 OVRL
#47 EMNGF1 EMNSS1 / EMNGF2 EMNSS2 / EMNG EMNSSNG
#53 EVPOS1 / COVAR EVSSPOS1 #58 EVPOS2 / COVAR EVSSPOS2
#63 EVNEG / COVAR EVSSNEG;
TOTPTS = NPOS1 + NPOS2 + NNEG;
FPFOS1 = NPOS1/TOTPTS;
FPFOS2 = NPOS2/TOTPTS;
SSTR=SUBSTR(NAME,4,3);
IF SUBSTR(SSTR,1,1)='X' OR SUBSTR(SSTR,1,1)='P'
OR SUBSTR(SSTR,1,1)='A' OR SUBSTR(SSTR,1,1)='Y'
THEN SUBSTR(SSTR,1,1)='0';
DROP COVAR;
OUTPUT;
STOP;
RUN;
DATA ALL.NK3;
  SET ALL.NK3 PCT3;
RUN;
%MEND THREE;
%MACRO FOUR;
DATA PCT4;
INFILE 6DSH;
INPUT #1 NAME $8. #5 GMNGF1 GMNSS1 / GMNGF2 GMNSS2 / GMNG1 GMNSSNG1 / GMNG2 GMNSSNG2 #12 GVPOS1 / COVAR GVSSPOS1 #17 GVPOS2 / COVAR GVSSPOS2 #22 GVNEG1 / COVAR GVSSNEG1 #27 GVNEG2 / GVSSNEG2 #32 GPPOS1 GPPOS2 GVNEG1 #38 EFPOS1 EFPOS2 EFNEG1 #42 NPOS1 NPOS2 NNEG1 NNEG2 #46 ALOC1 ALOC2 ALOC3 ALOC4 #49 050 OVRL $53 EMNGF1 EMNSS1 / EMNGF2 EMNSS2 / EMNG1 EMNSSNG1 / EMNG2 EMNSSNG2 #60 EVPOS1 / COVAR EVSSPOS1 #65 EVPOS2 / COVAR EVSSPOS2 #70 EVNEG1 / COVAR EVSSNEG1 #75 EVNEG2 / COVAR EVSSNEG2;
TOTPTS = NPOS1 + NPOS2 + NNEG1 + NNEG2;
FRPOS1 = NPOS1/TOTPTS;
FRPOS2 = NPOS2/TOTPTS;
FRNEG1 = NNEG1/TOTPTS;
SSTR=SUBSTR(NAME,4,3);
IF SUBSTR(SSTR,1,1)='X' OR SUBSTR(SSTR,1,1)='P' OR SUBSTR(SSTR,1,1)='Y' THEN SSTR=SUBSTR(SSTR,1,1)='0';
DROP COVAR;
OUTPUT;
STOP;
RUN;
DATA ALL.NK4;
  SET ALL.NK4 PCT4;
RUN;
%MEND FOUR;
DATA ALL.NK2;
  INPUT NAME $8. GMNGF1 GMNSS1 GMNGF2 GMNSS2 GMNG1 GMNSSNG GVPOS1 GVSSPOS1 GVNEG GVSSNEG GPPOS1 EFPOS1 NPOS1 NNEG ALOC1 ALOC2 OVRL EMNGF1 EMNSS1 EMNGF2 EMNSS2 EMNNG1 EMNSSNG1 EMNNG2 EMNSSNG2 EVPOS1 EVSSPOS1 EVNEG EVSSNEG EVPOS2 EVSSPOS2 EVNEG EVSSNEG SSTR $3.
CARDS;
RUN;
DATA ALL.NK3;
  INPUT NAME $8. GMNGF1 GMNSS1 GMNGF2 GMNSS2 GMNNG GMNSSNG GVPOS1 GVSSPOS1 GVPOS2 GVSSPOS2 GVNEG GVSSNEG GPPOS1 GPPOS2 EPPOS1 EFPOS2 NPOS1 NPOS2 NNEG ALOC1 ALOC2 ALOC3 OVRL EMNGF1 EMNSS1 EMNGF2 EMNSS2 EMNNG1 EMNSSNG1 EMNNG2 EMNSSNG2 EVPOS1 EVSSPOS1 EVPOS2 EVSSPOS2 EVNEG EVSSNEG SSTR $3.
TOTPTS FRPOS1 FRPOS2;
CARDS;
RUN;
DATA ALL.NK4;
INPUT NAME $8. GMNGF1 GMNSS1 GMNGF2 GMNSS2
     GMNG1 GMNSSNG1 GMNG2 GMNSSNG2
     GVP0S1 GVSSP0S1 GVP0S2 GVSSP0S2
     GVNEG1 GVSSNEG1 GVNEG2 GVSSNEG2
     GPPOS1 GPPOS2 GPNEG1 EPPOS1 EPPOS2 EPNEG1
     NP0S1 NP0S2 NNEG1 NNEG2 ALOC1 ALOC2 ALOC3 ALOC4
     OVRAL EMNGF1 EMNSS1 EMNGF2 EMNSS2
     EMNG1 EMNSSNG1 EMNG2 EMNSSNG2
     EVPOS1 EVSSPOS1 EVPOS2 EVSSPOS2
     EVNEG1 EVSSNEG1 EVNEG2 EVSSNEG2
     TOTPTS FRPOS1 FRPOS2 SSTR $3.;
CARDS;
RUN;

%LET DSN=INP00121;
%SET;
PROC SORT DATA=ALL.NK2;
   BY SSTR;
PROC PRINT DATA=ALL.NK2;
   VAR NAME GMNGF1 GMNSS1 GMNGF2 GMNSS2
       EMNGF1 EMNSS1 EMNGF2 EMNSS2
   OVRAL;
PROC SORT DATA=ALL.NK3;
   BY SSTR;
PROC PRINT DATA=ALL.NK3;
   VAR NAME GMNGF1 GMNGF2 GMNSS1 GMNSS2
       GPPOS1 EPPOS1 EPPOS2
       EMNGF1 EMNGF2 EMNSS1 EMNSS2
   OVRAL;
PROC SORT DATA=ALL.NK4;
   BY SSTR;
PROC PRINT DATA=ALL.NK4;
   VAR NAME GMNGF1 GMNGF2 GMNSS1 GMNSS2
       GPPOS1 GPPOS2 GPNEG1
       EPPOS1 EPPOS2 EPNEG1 EMNGF1 EMNGF2 EMNSS1 EMNSS2
   OVRAL;
//
APPENDIX D

CODEBOOK

for the Analysis Database

The analysis database consists of two ASCII-coded data files called QUESDI.DAT and FACSFILE.DAT.

The former contains questionnaire data, medical data for the patients which is descriptive of their cancers, and the results of the analysis of the dietary portion of the questionnaire. The codebook for QUESDI.DAT occupies pages 2-32 of this section.

FACSFILE.DAT contains the estimates of proportions of lymphocytes carrying the eight subset-specific molecules which were analyzed in this study as well as the estimated mean fluorescence of the positive and negative cell groups. Its codebook begins on page 33.
**FILE QUESDI.DAT**

(For the following codes, # indicates any integer from 1 to 9)

<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Subject ID</td>
<td>LCP### - Patient number</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LCC### - Control number</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>Birth state</td>
<td>(See codes below)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 or 99 - Missing</td>
</tr>
</tbody>
</table>

### STATE CODES

<table>
<thead>
<tr>
<th>State</th>
<th>Code</th>
<th>State</th>
<th>Code</th>
<th>State</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>AL</td>
<td>Maryland</td>
<td>MD</td>
<td>S. Carolina</td>
<td>SC</td>
</tr>
<tr>
<td>Alaska</td>
<td>AK</td>
<td>Massachusetts</td>
<td>MA</td>
<td>S. Dakota</td>
<td>SD</td>
</tr>
<tr>
<td>Arizona</td>
<td>AZ</td>
<td>Michigan</td>
<td>MI</td>
<td>Tennessee</td>
<td>TN</td>
</tr>
<tr>
<td>Arkansas</td>
<td>AR</td>
<td>Minnesota</td>
<td>MN</td>
<td>Texas</td>
<td>TX</td>
</tr>
<tr>
<td>California</td>
<td>CA</td>
<td>Mississippi</td>
<td>MS</td>
<td>Utah</td>
<td>UT</td>
</tr>
<tr>
<td>Colorado</td>
<td>CO</td>
<td>Missouri</td>
<td>MO</td>
<td>Vermont</td>
<td>VT</td>
</tr>
<tr>
<td>Connecticut</td>
<td>CT</td>
<td>Montana</td>
<td>MT</td>
<td>Virginia</td>
<td>VA</td>
</tr>
<tr>
<td>Delaware</td>
<td>DE</td>
<td>Nebraska</td>
<td>NE</td>
<td>Washington</td>
<td>WA</td>
</tr>
<tr>
<td>Dist.of Col.</td>
<td>DC</td>
<td>Nevada</td>
<td>NV</td>
<td>W. Virginia</td>
<td>WV</td>
</tr>
<tr>
<td>Florida</td>
<td>FL</td>
<td>New Hampshire</td>
<td>NH</td>
<td>Wisconsin</td>
<td>WI</td>
</tr>
<tr>
<td>Georgia</td>
<td>GA</td>
<td>New Jersey</td>
<td>NJ</td>
<td>Wyoming</td>
<td>WY</td>
</tr>
<tr>
<td>Hawaii</td>
<td>HI</td>
<td>New Mexico</td>
<td>NM</td>
<td>Puerto Rico</td>
<td>PR</td>
</tr>
<tr>
<td>Idaho</td>
<td>ID</td>
<td>New York</td>
<td>NY</td>
<td>Virgin Islds.</td>
<td>VI</td>
</tr>
<tr>
<td>Illinois</td>
<td>IL</td>
<td>N. Carolina</td>
<td>NC</td>
<td>Guam</td>
<td>GU</td>
</tr>
<tr>
<td>Indiana</td>
<td>IN</td>
<td>N. Dakota</td>
<td>ND</td>
<td>Canada</td>
<td>CN</td>
</tr>
<tr>
<td>Iowa</td>
<td>IA</td>
<td>Ohio</td>
<td>OH</td>
<td>Cuba</td>
<td>CU</td>
</tr>
<tr>
<td>Kansas</td>
<td>KS</td>
<td>Oklahoma</td>
<td>OK</td>
<td>Mexico</td>
<td>MX</td>
</tr>
<tr>
<td>Kentucky</td>
<td>KY</td>
<td>Oregon</td>
<td>OR</td>
<td>Rest of World</td>
<td>RW</td>
</tr>
<tr>
<td>Louisiana</td>
<td>LA</td>
<td>Pennsylvania</td>
<td>PA</td>
<td>Unknown</td>
<td>UN</td>
</tr>
<tr>
<td>Maine</td>
<td>ME</td>
<td>Rhode Island</td>
<td>RI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Record Column</td>
<td>Variable</td>
<td>Code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------</td>
<td>-------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 9</td>
<td>Parish of residence</td>
<td>(See codes below)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9 or 99 - Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 11</td>
<td>Neighborhood</td>
<td>1 - Urban</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 - Suburban</td>
<td>1 - Urban</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 - Rural</td>
<td>3 - Rural</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9 - Other/missing</td>
<td>9 - Other/missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 12</td>
<td>Zip code</td>
<td>######</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9 or 99999 - Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 17</td>
<td>Diet data quality</td>
<td>GOOD - Good: no errors or minor,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>correctable errors.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FIXD - &quot;Skipped&quot; assumed to mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;Never eat&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TUFU - Eat less than 6 foods daily</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MDSM - Same</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>portion size for all or almost all</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>' - Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parish</th>
<th>Code</th>
<th>Parish</th>
<th>Code</th>
<th>Parish</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ascension</td>
<td>AC</td>
<td>Iberville</td>
<td>IB</td>
<td>Tangipahoa</td>
<td>TA</td>
</tr>
<tr>
<td>Assumption</td>
<td>AS</td>
<td>Livingston</td>
<td>LI</td>
<td>W. Baton Rouge</td>
<td>WB</td>
</tr>
<tr>
<td>E. Baton Rouge</td>
<td>EB</td>
<td>Pointe Coupee</td>
<td>PC</td>
<td>W. Feliciana</td>
<td>WF</td>
</tr>
<tr>
<td>E. Feliciana</td>
<td>EF</td>
<td>St. Helena</td>
<td>SH</td>
<td>Other</td>
<td>OT</td>
</tr>
</tbody>
</table>

PARISH CODES

<p>| 1 21         | Birth month      | 01,02,...12 |
|              |                  | 99 - Missing |
| 1 23         | Birth day        | 01,02,...31 |
|              |                  | 99 - Missing |</p>
<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>Birth year</td>
<td>(1)890, (1)891, ... (1)960, 999 - Missing</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>Age in years</td>
<td>31, 32, ... 90, 9 or 99 - Missing</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>Sex</td>
<td>1 - Male, 9 - Missing</td>
</tr>
<tr>
<td>1</td>
<td>31</td>
<td>Race, ethnicity</td>
<td>1 - White, 2 - Black, 3 - Hispanic, 4 - Native American, 5 - Asian, 6 - Pacific Isl'dr, 9 - Missing</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
<td>Highest grade of schooling</td>
<td>01-16+, 99 - Missing</td>
</tr>
<tr>
<td>1</td>
<td>34</td>
<td>Marital Status</td>
<td>1 - Single, 2 - Married, 3 - Widowed, 4 - Divorced/ separated, 9 - Missing</td>
</tr>
<tr>
<td>1</td>
<td>35</td>
<td>How many times have you changed residence in last 10 years</td>
<td>01-98, 99 - Missing</td>
</tr>
<tr>
<td>1</td>
<td>37</td>
<td>Total household income</td>
<td>1 - Under $10,000/yr, 2 - $10,000-$9,999, 3 - $20,000-$29,999, 4 - $30,000-$39,999, 5 - $40,000-$49,999, 6 - Over $50,000, 7 - Don't wish to answer, 9 - Missing</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-----------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>1</td>
<td>38</td>
<td>Number of persons in household</td>
<td>1,2,...,98, 99 - Missing</td>
</tr>
</tbody>
</table>

Of the past 30 years, for how many years did you live in a

<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Area</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>Rural area</td>
<td>1 - 30, 9 or 99 - Missing</td>
</tr>
<tr>
<td>1</td>
<td>42</td>
<td>Suburban area</td>
<td>(As for 40)</td>
</tr>
<tr>
<td>1</td>
<td>44</td>
<td>Urban area</td>
<td>(As for 40)</td>
</tr>
<tr>
<td>1</td>
<td>46</td>
<td>Histologic type of lung cancer</td>
<td>1 - Oat cell</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Squamous cell</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - Adenocarcinoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 - Large cell</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 - Other (Mixed, sarcoma, etc.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>. - Control</td>
</tr>
<tr>
<td>1</td>
<td>47</td>
<td>Area of lung affected</td>
<td>1 - Main bronchus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Right lung (RL) upper lobe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - RL, middle lobe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 - RL, lower lobe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 - Left lung (LL), upper lobe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6 - LL, lower lobe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>. - Control</td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>Previous therapy</td>
<td>1 - Surgery</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Chemotherapy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>. - Control</td>
</tr>
<tr>
<td>1</td>
<td>49</td>
<td>Stage</td>
<td>1 - Local</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Regional</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - Disseminated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>. - Control</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Subject ID</td>
<td>(See record 1)</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-----------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>Lived within mile of industry emitting dust, odors, or irritants</td>
<td>0 - No, 1 - Yes, 9 - Missing</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>1st plant: Kind</td>
<td>(Up to 12 letters)</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>For how many years</td>
<td>01,02,...98</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>Type of emission</td>
<td>1 - Dust, 2 - Odors, 3 - Irritants, 4 - 1 + 3, 5 - 2 + 3, 6 - 1 + 2 + 3, 7 - 1 + 2, 9 - Other/Missing</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>2nd plant: Kind</td>
<td>(As for 8)</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>For how many years</td>
<td>(As for 20)</td>
</tr>
<tr>
<td>2</td>
<td>37</td>
<td>Type of emission</td>
<td>(As for 22)</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>3rd plant: Kind</td>
<td>(As for 8)</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>For how many years</td>
<td>(As for 20)</td>
</tr>
<tr>
<td>2</td>
<td>52</td>
<td>Type of emission</td>
<td>(As for 22)</td>
</tr>
<tr>
<td>2</td>
<td>53</td>
<td>4th plant: Kind</td>
<td>(As for 8)</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
<td>For how many years</td>
<td>(As for 20)</td>
</tr>
<tr>
<td>2</td>
<td>67</td>
<td>Type of emission</td>
<td>(As for 22)</td>
</tr>
<tr>
<td>2</td>
<td>68</td>
<td>Smoked 100 cigarettes over lifetime</td>
<td>1 - No, 2 - Yes, 9 - Missing</td>
</tr>
<tr>
<td>2</td>
<td>69</td>
<td>How old when began (in years)</td>
<td>01,02,...90, 9 or 99 - Missing/Non-smoker (M/N-s)</td>
</tr>
<tr>
<td>2</td>
<td>71</td>
<td>How many cigarettes per day on average</td>
<td>01,02,...98, 99 - M/N-s</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-----------------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>Smoke now</td>
<td>1 - No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - M/N-s</td>
</tr>
<tr>
<td>2</td>
<td>74</td>
<td>How old when quit (in years)</td>
<td>01,02, ...90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 - M/N-s or current smoker</td>
</tr>
<tr>
<td>2</td>
<td>76</td>
<td>How many cigarettes per day now</td>
<td>01,02, ...98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 or 9 - M/N-s</td>
</tr>
<tr>
<td>2</td>
<td>78</td>
<td>Ever smoked pipe or cigars regularly</td>
<td>1 - No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Yes</td>
</tr>
<tr>
<td>2</td>
<td>79</td>
<td>For how many years</td>
<td>01,02, ...90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 or 99 - M/N-s</td>
</tr>
<tr>
<td>2</td>
<td>81</td>
<td>How many per time</td>
<td>01,02, ...98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 or 99 - M/N-s</td>
</tr>
<tr>
<td>2</td>
<td>83</td>
<td>Time unit</td>
<td>1 - Day</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Week</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - M/N-s</td>
</tr>
<tr>
<td>2</td>
<td>84</td>
<td>Since beginning, how many years without smoking</td>
<td>01,02, ...90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 or 99 - M/N-s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>or Current smoker</td>
</tr>
<tr>
<td>2</td>
<td>86</td>
<td>Years smoking non-filter cigarettes</td>
<td>01,02, ...90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 or 99 - M/N-s</td>
</tr>
<tr>
<td>2</td>
<td>88</td>
<td>Years smoking filter cigarettes</td>
<td>01,02, ...90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 or 99 - M/N-s</td>
</tr>
<tr>
<td>2</td>
<td>90</td>
<td>How did you smoke</td>
<td>1 - Didn't inhale</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Into mouth &amp; throat only</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - Deeply into chest</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 - Different levels</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 - Can't describe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - M/N-s</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-------------------------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>2</td>
<td>91</td>
<td>How long was butt</td>
<td>1 - More than half</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - 1/3 to 1/2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - 1/4 to 1/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 - Less than 1/4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - M/N-s</td>
</tr>
<tr>
<td>2</td>
<td>92</td>
<td>Between puffs, where was the cigarette</td>
<td>1 - On an ashtray</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - In mouth</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - In fingers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 - Other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - M/N-s</td>
</tr>
<tr>
<td>2</td>
<td>93</td>
<td>Parent smoked</td>
<td>0 - No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 - Mother</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Father</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - Both</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Missing</td>
</tr>
<tr>
<td>2</td>
<td>94</td>
<td>Spouse or roommate</td>
<td>0 - No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 - Spouse</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Roommate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - Both</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Other/Missing</td>
</tr>
<tr>
<td>2</td>
<td>95</td>
<td>Total years lived with a smoker</td>
<td>01,02,...90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 or 99 - Missing</td>
</tr>
<tr>
<td>2</td>
<td>97</td>
<td>Vitamins or minerals during last year</td>
<td>1 - No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Yes, fairly regularly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - Yes, not regularly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Missing</td>
</tr>
<tr>
<td>2</td>
<td>98</td>
<td>One-a-day</td>
<td># - Number per time period</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 - Missing</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>Time period</td>
<td>1 - Day</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Week</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Missing</td>
</tr>
<tr>
<td>2</td>
<td>101</td>
<td>Stresstabs</td>
<td>(As for 98)</td>
</tr>
<tr>
<td>2</td>
<td>103</td>
<td>Time period</td>
<td>(As for 100)</td>
</tr>
<tr>
<td>2</td>
<td>104</td>
<td>Therapeutic/Theragran</td>
<td>(As for 98)</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>2</td>
<td>106</td>
<td>Time period</td>
<td>(As for 100)</td>
</tr>
<tr>
<td>2</td>
<td>107</td>
<td>Vitamin A: how many</td>
<td>01,02,...98, 99 - Missing</td>
</tr>
<tr>
<td>2</td>
<td>109</td>
<td>Time period</td>
<td>1 - Day, 2 - Week, 9 - Missing</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
<td>Mg or IU per pill</td>
<td>1 - 50-100, 2 - 200-250, 3 - 400-500, 4 - 1000, 5 - 5000, 6 - 10000, 7 - 20000-25000, 8 - 50000, 9 - Unknown/missing</td>
</tr>
<tr>
<td>2</td>
<td>111</td>
<td>Vitamin C</td>
<td>(As for 107)</td>
</tr>
<tr>
<td>2</td>
<td>113</td>
<td>Time period</td>
<td>(As for 109)</td>
</tr>
<tr>
<td>2</td>
<td>114</td>
<td>Mg or IU per pill</td>
<td>(As for 110)</td>
</tr>
<tr>
<td>2</td>
<td>115</td>
<td>Vitamin E</td>
<td>(As for 107)</td>
</tr>
<tr>
<td>2</td>
<td>117</td>
<td>Time period</td>
<td>(As for 109)</td>
</tr>
<tr>
<td>2</td>
<td>118</td>
<td>Mg or IU per pill</td>
<td>(As for 110)</td>
</tr>
<tr>
<td>2</td>
<td>119</td>
<td>Calcium/Dolomite</td>
<td>(As for 107)</td>
</tr>
<tr>
<td>2</td>
<td>121</td>
<td>Time period</td>
<td>(As for 109)</td>
</tr>
<tr>
<td>2</td>
<td>122</td>
<td>Mg or IU per pill</td>
<td>(As for 110)</td>
</tr>
<tr>
<td>2</td>
<td>123</td>
<td>Other vitamin or mineral supplement</td>
<td>1 - Yeast, 2 - Selenium, 3 - Zinc, 4 - Iron, 5 - β-Carotene, 6 - Cod liver oil, 7 - Other, 9 - Missing</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Subject ID</td>
<td>(See record 1)</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------</td>
<td>------</td>
</tr>
</tbody>
</table>
| 3      | 7      | Do you eat differently than you did 20 yrs ago | 0 - No  
|        |        |          | 1 - Yes  
|        |        |          | 9 - Can't say/ Missing |
| 3      | 8      | Total Calories | 0 - The same  
|        |        |          | 1 - Less  
|        |        |          | 2 - More |
| 3      | 9      | Char-broiled meat | (As for 8) |
| 3      | 10     | Butter & Oil | (As for 8) |
| 3      | 11     | Fruit | (As for 8) |
| 3      | 12     | Vegetables | (As for 8) |
| 3      | 13     | Vitamin supplements | (As for 8) |
| 3      | 14     | Take prescription drugs daily or almost daily | 0 - No  
|        |        |          | 1 - Yes  
|        |        |          | 9 - Missing |
| 3      | 15     | Most important or strongest kind | (Up to 12 letters)  
|        |        |          | 9 - Missing/None |
| 3      | 27     | Second most important | (As for 15) |
| 3      | 39     | Had cold, flu or other viral illness in last three weeks | 0 - No  
|        |        |          | 1 - Yes  
|        |        |          | 2 - Possibly  
|        |        |          | 9 - Missing |
| 3      | 40     | Taken antibiotics in last three weeks | (As for 39) |
| 3      | 41     | How many times hospitalized in past 5 years | 0,1,...8  
|        |        |          | 9 - Missing |

Have you ever had

<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
</table>
| 3      | 42     | X-ray treatments for acne, ringworm, thymus, adenoids | 1 - No  
|        |        |          | 2 - Yes  
|        |        |          | 9 - Missing |
| 3      | 43     | Number of times | 0,1,...8  
<p>|        |        |          | 9 - Missing |</p>
<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>44</td>
<td>Age at first treatment (in years)</td>
<td>01,02,...90 99 - Missing</td>
</tr>
<tr>
<td>3</td>
<td>46</td>
<td>Treatment with radioactive isotopes</td>
<td>(As for 42)</td>
</tr>
<tr>
<td>3</td>
<td>47</td>
<td>Number of times</td>
<td>(As for 43)</td>
</tr>
<tr>
<td>3</td>
<td>48</td>
<td>Age at first treatment</td>
<td>(As for 44)</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>Upper GI series</td>
<td>(As for 42)</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
<td>Number of times</td>
<td>(As for 43)</td>
</tr>
<tr>
<td>3</td>
<td>52</td>
<td>Age at first treatment</td>
<td>(As for 44)</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
<td>Lower GI series</td>
<td>(As for 42)</td>
</tr>
<tr>
<td>3</td>
<td>55</td>
<td>Number of times</td>
<td>(As for 43)</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td>Age at first treatment</td>
<td>(As for 44)</td>
</tr>
</tbody>
</table>

Has a doctor ever told you that you have

<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>58</td>
<td>Heart disease/angina</td>
<td>1 - No 2 - YES 9 - Missing/Don't know</td>
</tr>
<tr>
<td>3</td>
<td>59</td>
<td>Kidney disease</td>
<td>(As for 58)</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>Heart attack</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>61</td>
<td>Bladder disease</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>High blood pressure</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>63</td>
<td>Liver cirrhosis</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>Stroke</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
<td>Hepatitis</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>66</td>
<td>Tuberculosis</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>Stomach ulcers</td>
<td>&quot;</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>---------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>3</td>
<td>68</td>
<td>Chronic bronchitis /emphysema</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>69</td>
<td>Rheumatoid arthritis</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>70</td>
<td>Asthma</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>71</td>
<td>Other arthritis</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>Hay fever</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>73</td>
<td>Osteoporosis</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>74</td>
<td>Diverticulosis</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>Fractured hip</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>76</td>
<td>Rectal/colon polyps</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>77</td>
<td>Prostate trouble</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>78</td>
<td>Chronic colitis</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>79</td>
<td>Abnormal Pap smear</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>80</td>
<td>Diabetes</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>81</td>
<td>Skin cancer</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>82</td>
<td>Thyroid condition</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>83</td>
<td>Leukemia</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>84</td>
<td>Other cancer</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

First cancer:

3 85 Kind (See codes below)

3 87 Year first diagnosed (19)01,(19)02,
...(19)90
9 or 99 - Missing/no cancer

Second cancer:

3 89 Kind (As for 85)

3 91 Year first diagnosed (As for 87)
### Record | Column | Variable | Code
--- | --- | --- | ---
**In the past year, have you had**
3 | 93 | Bleeding or sore gums | 1 - No  
2 - Yes  
9 - Missing/Don't know
3 | 94 | Difficulty seeing in the dark | (As for 93)
3 | 95 | Bruise easily | "
3 | 96 | Frequent or chronic fever | "
3 | 97 | Nosebleeds | "
3 | 98 | Frequent constipation or hemorrhoids | "
**Has a doctor ever told you that you had**
4 | 7 | An autoimmune disease | 0 - No  
1 - Yes  
9 - Missing
4 | 8 | Kind | Up to 12 letters
4 | 20 | Aplastic anemia | 0 - No  
1 - Yes  
9 - Missing
4 | 21 | Food allergy | (As for 20)
4 | 22 | Dust allergy | "
4 | 23 | Pollen allergy | "
4 | 24 | Allergy to animal hair | "
4 | 25 | Allergy to insects | "
4 | 26 | Drug allergy | "
4 | 27 | Other allergy | "
<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>28</td>
<td>Current employment status</td>
<td>1 - Employed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Homemaker</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - Retired</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 - Disabled, unable to work</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 - Unemployed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6 - Student</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7 - Other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Missing</td>
</tr>
<tr>
<td>4</td>
<td>29</td>
<td>Job/occupation</td>
<td>Up to 12 letters</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>Years in this job</td>
<td>01, 02, ..., 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Missing</td>
</tr>
<tr>
<td>4</td>
<td>43</td>
<td>More time indoors or outdoors</td>
<td>1 - Indoors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Outdoors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Missing</td>
</tr>
</tbody>
</table>

In your work, have you been exposed for a year or more to any of the following:

<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>44</td>
<td>Asbestos</td>
<td>1 - No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Missing</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>Iron foundry</td>
<td>(As for 44)</td>
</tr>
<tr>
<td>4</td>
<td>46</td>
<td>Radiation</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>47</td>
<td>Nickel Smelting</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>Welding</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td>Underground mining</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>Coal tar, soot, pitch, creosote, asphalt</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>51</td>
<td>Lumber industry, or heavy wood dust</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>52</td>
<td>Mineral, cutting or lubricating oil</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>53</td>
<td>Rubber or cablemaking industry</td>
<td>&quot;</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4</td>
<td>54</td>
<td>Benzidine, beta-napthylamine</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>Chemical or plastics industry</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>56</td>
<td>Benzene</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>57</td>
<td>Pesticides, herbicides</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
<td>Isopropyl oil</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>59</td>
<td>Mustard gas</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>Dyestuffs</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>61</td>
<td>Chromium</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>62</td>
<td>Arsenic</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>63</td>
<td>Cadmium, beryllium, vinyl chloride</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cancer</th>
<th>Code</th>
<th>Cancer</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder</td>
<td>01</td>
<td>Mouth, Oral</td>
<td>13</td>
</tr>
<tr>
<td>Bone</td>
<td>02</td>
<td>Ovary</td>
<td>14</td>
</tr>
<tr>
<td>Brain</td>
<td>03</td>
<td>Pancreas</td>
<td>15</td>
</tr>
<tr>
<td>Breast</td>
<td>04</td>
<td>Prostate</td>
<td>16</td>
</tr>
<tr>
<td>Cervix</td>
<td>05</td>
<td>Rectum</td>
<td>17</td>
</tr>
<tr>
<td>Colon</td>
<td>06</td>
<td>Skin - Melanom</td>
<td>18</td>
</tr>
<tr>
<td>Esophagus</td>
<td>07</td>
<td>Skin - Not melanoma</td>
<td>18</td>
</tr>
<tr>
<td>Kidney</td>
<td>08</td>
<td>(Basal,squamous)</td>
<td>19</td>
</tr>
<tr>
<td>Liver</td>
<td>09</td>
<td>Skin - Not specif'd</td>
<td>20</td>
</tr>
<tr>
<td>Leukemia</td>
<td>10</td>
<td>Stomach</td>
<td>21</td>
</tr>
<tr>
<td>Lung</td>
<td>11</td>
<td>Thyroid</td>
<td>22</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>12</td>
<td>Uterus</td>
<td>23</td>
</tr>
<tr>
<td>(Inc. Hodgkins)</td>
<td></td>
<td>Other</td>
<td>24</td>
</tr>
</tbody>
</table>

CANCER CODES
<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>64</td>
<td>Second longest occupation</td>
<td>Up to 12 letters</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Missing</td>
</tr>
<tr>
<td>4</td>
<td>76</td>
<td>For how many years</td>
<td>## - Number</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 or 9 - Missing</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
<td>More time indoors or outdoors</td>
<td>(As for 43)</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>Subject ID</td>
<td>(See record 1)</td>
</tr>
</tbody>
</table>

Have you worked for a year or more in the following industries:

<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Industry</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>7</td>
<td>Shipbuilding</td>
<td>0 - No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 - Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Missing</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>Construction</td>
<td>(As for 7)</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>Fishing</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>Lumber, wood, furniture manufacturing or paper</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>Petrochemical</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>Other chemical</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>Metal refining, polishing, manufacturing or plating</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>Sugar cane farming</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>Mining</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>Insulation manufacturing or installation</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>Asbestos manufacturing or asbestos-containing products</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>Cement manufacturing</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>Demolition</td>
<td>&quot;</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>Cook</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>Gas station attendant or auto mechanic</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>Have any close relatives had cancer</td>
<td>1 - No, 2 - Yes, 9 - Missing</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>Number of relatives</td>
<td># - Number, 9 - Missing</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>Relation</td>
<td>1 - Mother, 2 - Father, 3 - Brother, 4 - Sister, 5 - Son, 6 - Daughter, 7 - Grandfather, 8 - Grandmother</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>Alive or dead</td>
<td>1 - Alive, 2 - Dead, 9 - Missing/No relatives with cancer (m/N-r)</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>Age if living or age at death</td>
<td>01, 02, ..., 98, 99 - M/n-r</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>Kind of cancer</td>
<td>(See codes above)</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>Age at diagnosis</td>
<td>(As for 26)</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>Relationship of 2nd case</td>
<td>(As for 24)</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
<td>Alive or dead</td>
<td>(As for 25)</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>Age if living or age at death</td>
<td>(As for 26)</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>Kind of cancer</td>
<td>(See codes above)</td>
</tr>
<tr>
<td>5</td>
<td>38</td>
<td>Age at diagnosis</td>
<td>(As for 26)</td>
</tr>
<tr>
<td>Record Column</td>
<td>Variable</td>
<td>Code</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>------</td>
<td></td>
</tr>
</tbody>
</table>
| 5 40          | Petroleum refining | 0 - No  
|               |          | 1 - Yes  
|               |          | 9 - Missing  |
| 5 41          | Smoke: wood, oil, sugar cane | (As for 40)  |
| 5 42          | Gasoline/motor oil | (As for 40)  |
| 5 43          | X-rays, fluoroscopic equipment, radioisotopes | "  |
| 5 44          | Coke oven gases | "  |

How tall are you

<table>
<thead>
<tr>
<th>Record Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
</table>
| 5 45          | Feet     | 4,5,...7  
|               |          | 9 - Missing  |
| 5 46          | Inches   | 01,02,...12  
|               |          | 99 - Missing  |
| 5 48          | How much do you weigh? | 100,101,...500  
|               |          | 999 - Missing  |
| 5 51          | Most you ever weighed | (As for 48)  |
| 5 54          | How many times you have dieted to lose weight | 1 - Never  
|               |          | 2 - 1-2  
|               |          | 3 - 3-5  
|               |          | 4 - 6-8  
|               |          | 5 - 9-11  
|               |          | 6 - 12 or more  |
| 5 55          | How many hours of sleep do you get at night | 1 - 6 or less  
|               |          | 2 - 7  
|               |          | 3 - 8  
<p>|               |          | 4 - 9 or more  |</p>
<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>56</td>
<td>How often do you feel significant stress</td>
<td>1 - Everyday</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Several times/ week</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - Several times/ month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 - Several times/ year</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 - Rarely/never (R/n)</td>
</tr>
<tr>
<td>5</td>
<td>57</td>
<td>Active sports</td>
<td>1 - &gt;1/week</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - About 1/week</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - Few times/ month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 - Few times/year</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 - R/n</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 - Missing</td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>Physical exercises</td>
<td>(As for 57)</td>
</tr>
<tr>
<td>5</td>
<td>59</td>
<td>Jogging or running</td>
<td>(As for 57)</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>Swimming/long walks</td>
<td>(As for 57)</td>
</tr>
<tr>
<td>5</td>
<td>61</td>
<td>Gardening, fishing, hunting</td>
<td>(As for 57)</td>
</tr>
<tr>
<td>5</td>
<td>62</td>
<td>Something else</td>
<td>(As for 57)</td>
</tr>
<tr>
<td>5</td>
<td>63</td>
<td>How many close friends do you have</td>
<td>1 - None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - 1 or 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 - 3 to 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 - 6 to 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 - 10 or more</td>
</tr>
<tr>
<td>5</td>
<td>64</td>
<td>How many relatives do you feel close to</td>
<td>(As for 63)</td>
</tr>
<tr>
<td>5</td>
<td>65</td>
<td>How many friends or relatives you see/talk</td>
<td>(As for 63)</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------</td>
<td>------</td>
</tr>
</tbody>
</table>
| 5      | 66     | Church or temple | 1 -> 1/week  
2 - 1/week  
3 - Few times/month  
4 - A few times/year  
5 - R/n |
| 5      | 67     | Clubs, PTA, professional, labor or service groups | (As for 66) |
| 5      | 68     | Month | 01,02,...,12  
99 - Missing |
| 5      | 70     | Day | 01,02,...,31 |
| 5      | 72     | Year | (198)8,(198)9, (199)0 |
| 5      | 73     | Version of diet questionnaire | 2 |
| 6      | 1      | Subject ID | (See record 1) |
| 6      | 7      | Line identifier | A |
| 6      | 8      | On special diet | 1 - No  
2 - Weight loss  
3 - For medical condition  
4 - Vegetarian  
5 - Low salt  
6 - Low cholesterol  
7 - Weight gain  
9 or 99 - Missing |
<p>| 6      | 9      | On second special diet | (As for 7) |
| 6      | 10     | No meaning | 9 |</p>
<table>
<thead>
<tr>
<th>Record Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 11</td>
<td>Amount of weight change in last year</td>
<td>1 - No</td>
</tr>
<tr>
<td>6 12</td>
<td>Desirable weight (DW) 100,101,...275</td>
<td></td>
</tr>
<tr>
<td>Female:</td>
<td>100 lbs. + 5 lbs. for each inch over 5 ft.</td>
<td></td>
</tr>
<tr>
<td>Male:</td>
<td>106 lbs. + 6 lbs. for each inch over 5 ft.</td>
<td></td>
</tr>
<tr>
<td>6 15</td>
<td>Lower bound for weight</td>
<td>DW X 0.95</td>
</tr>
<tr>
<td>6 18</td>
<td>Upper bound for weight</td>
<td>DW X 1.05</td>
</tr>
<tr>
<td>6 21</td>
<td>Grams solid food consumed daily</td>
<td>####</td>
</tr>
<tr>
<td>6 25</td>
<td>Daily carotene intake</td>
<td>####### micrograms (µg)</td>
</tr>
<tr>
<td>6 32</td>
<td>Retinol equivalents</td>
<td>####</td>
</tr>
<tr>
<td>6 36</td>
<td>Daily retinol intake</td>
<td>#### µg</td>
</tr>
<tr>
<td>6 40</td>
<td>Na/K ratio</td>
<td>.##</td>
</tr>
<tr>
<td>6 44</td>
<td>Linoleic/saturated fatty acids</td>
<td>.##</td>
</tr>
<tr>
<td>6 48</td>
<td>Percent of daily calories from fat</td>
<td>##.#</td>
</tr>
<tr>
<td>6 52</td>
<td>Percent of daily cal. from protein</td>
<td>##.#</td>
</tr>
<tr>
<td>6 56</td>
<td>Pct. of daily cal. from carbohydrate</td>
<td>##.#</td>
</tr>
<tr>
<td>6 60</td>
<td>Daily calories from sweets</td>
<td>######.#</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td>6</td>
<td>67</td>
<td>Pct. of daily cal. from sweets</td>
</tr>
<tr>
<td>6</td>
<td>71</td>
<td>Daily calories from alcoholic drinks</td>
</tr>
<tr>
<td>6</td>
<td>78</td>
<td>Pct. of daily cal. from alcoholic drinks</td>
</tr>
<tr>
<td>6</td>
<td>82</td>
<td>&quot;Other&quot; vitamin supplement</td>
</tr>
<tr>
<td>6</td>
<td>83</td>
<td>Supplemental Vit. A</td>
</tr>
<tr>
<td>6</td>
<td>90</td>
<td>Supplemental Vit. B</td>
</tr>
<tr>
<td>6</td>
<td>97</td>
<td>Supplemental Vit. C</td>
</tr>
<tr>
<td>6</td>
<td>104</td>
<td>Supplemental Vit. D</td>
</tr>
<tr>
<td>6</td>
<td>111</td>
<td>Supplemental Vit. E</td>
</tr>
<tr>
<td>6</td>
<td>118</td>
<td>Supplemental Iron</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>Subject ID</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>Line identifier</td>
</tr>
</tbody>
</table>

Daily nutrient estimates averaged over whole year:

<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>8</td>
<td>Total calories</td>
<td>####.#</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>Protein intake</td>
<td>####.# grams</td>
</tr>
<tr>
<td>7</td>
<td>22</td>
<td>Fat intake</td>
<td>####.# grams</td>
</tr>
<tr>
<td>7</td>
<td>29</td>
<td>Carbohydrate intake</td>
<td>####.# grams</td>
</tr>
<tr>
<td>7</td>
<td>36</td>
<td>Calcium intake</td>
<td>####.# mgs</td>
</tr>
<tr>
<td>7</td>
<td>43</td>
<td>Phosphorus intake</td>
<td>####.# mgs</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>Iron intake</td>
<td>####.# mgs</td>
</tr>
<tr>
<td>7</td>
<td>57</td>
<td>Sodium intake</td>
<td>####.# mgs</td>
</tr>
<tr>
<td>7</td>
<td>64</td>
<td>Potassium intake</td>
<td>####.# mgs</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>7</td>
<td>71</td>
<td>Vitamin A</td>
<td>#####. IU</td>
</tr>
<tr>
<td>7</td>
<td>78</td>
<td>Thiamin</td>
<td>#####. mgs</td>
</tr>
<tr>
<td>7</td>
<td>85</td>
<td>Riboflavin</td>
<td>#####. mgs</td>
</tr>
<tr>
<td>7</td>
<td>92</td>
<td>Niacin</td>
<td>#####. mgs</td>
</tr>
<tr>
<td>7</td>
<td>99</td>
<td>Vitamin C</td>
<td>#####. mgs</td>
</tr>
<tr>
<td>7</td>
<td>106</td>
<td>Saturated fat</td>
<td>#####. grams</td>
</tr>
<tr>
<td>7</td>
<td>113</td>
<td>Oleic acid</td>
<td>#####. grams</td>
</tr>
<tr>
<td>7</td>
<td>120</td>
<td>Linoleic acid</td>
<td>#####. grams</td>
</tr>
<tr>
<td>7</td>
<td>127</td>
<td>Cholesterol</td>
<td>#####. mgs</td>
</tr>
<tr>
<td>7</td>
<td>134</td>
<td>Dietary Fiber</td>
<td>#####. grams</td>
</tr>
</tbody>
</table>

*(See record 12 for breakdown of fiber from beans, grains and fruits and vegetables.)*

| 8      | 1      | Subject ID        | (See record 1) |
| 8      | 7      | Line identifier   | C              |

Error flags:

<p>| 8      | 8      | Number of times frequency not given for vitamin pill questions | # |
| 8      | 9      | Number of times nutrient amounts not given for vitamins | # |
| 8      | 10     | Number of food items skipped | ## |
| 8      | 12     | Number of foods where no portion size given | ## |
| 8      | 14     | Number of foods where no frequency given | ## |</p>
<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>16</td>
<td>Number of times large portion indicated</td>
<td>##</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>Number of times small portion indicated</td>
<td>##</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>Number of foods eaten (without beverages)</td>
<td>##</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>Average number of food servings eaten each day</td>
<td>##</td>
</tr>
<tr>
<td>8</td>
<td>24</td>
<td>Number of food time units out of range</td>
<td>##</td>
</tr>
<tr>
<td>8</td>
<td>26</td>
<td>Number of foods with frequency, time unit or serving size missing</td>
<td>##</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
<td>Number of food serving sizes out of range</td>
<td>##</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>Number of foods with frequency too high</td>
<td>##</td>
</tr>
<tr>
<td>8</td>
<td>31</td>
<td>Number where indicated once/time. If onces/(# of foods eaten) &gt; 0.7, probably unreliable data</td>
<td>##</td>
</tr>
<tr>
<td>8</td>
<td>33</td>
<td>Were recalculations done because subject was outlier</td>
<td>1 - Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Blank - No</td>
</tr>
<tr>
<td>8</td>
<td>34</td>
<td>Was subject below 10th percentile or above 90th percentile on total calories</td>
<td>1 - Below 10th</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 - Above 90th</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Blank - OK</td>
</tr>
<tr>
<td>8</td>
<td>35</td>
<td>Calcium from supplements</td>
<td>#####.# mgs</td>
</tr>
<tr>
<td>8</td>
<td>42</td>
<td>Total vitamin A</td>
<td>##### IU</td>
</tr>
<tr>
<td>8</td>
<td>47</td>
<td>Total vitamin C</td>
<td>##### IU</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>8</td>
<td>51</td>
<td>Weekly freq. of fruit</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>55</td>
<td>Weekly freq. of citrus fruit</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>59</td>
<td>Weekly freq. of all vegetables</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>63</td>
<td>Vegetable freq. excluding potatoes and rice</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>67</td>
<td>Weekly freq. of dark green &amp; deep yellow vegetables</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>71</td>
<td>Weekly freq. of tomatoes</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>75</td>
<td>Weekly freq. of carrots</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>79</td>
<td>Weekly freq. of salad</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>83</td>
<td>Weekly freq. of beef</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>87</td>
<td>Weekly freq. of fish and chicken</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>91</td>
<td>Weekly freq. of high-fiber cereals &amp; breads</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>95</td>
<td>Weekly freq. of alcohol</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>99</td>
<td>Number of eggs eaten per week</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>103</td>
<td>Weekly freq. of pork</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>107</td>
<td>Weekly freq. of hot dogs &amp; lunch meat</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>111</td>
<td>Weekly freq. of butter &amp; margarine</td>
<td>##.#</td>
</tr>
<tr>
<td>8</td>
<td>115</td>
<td>Weekly freq. of cheese except cottage cheese</td>
<td>##.#</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>8</td>
<td>119</td>
<td>Weekly freq. of whole milk</td>
<td>#.#</td>
</tr>
<tr>
<td>8</td>
<td>123</td>
<td>Weekly freq. of ice cream</td>
<td>#.#</td>
</tr>
<tr>
<td>8</td>
<td>127</td>
<td>Weekly freq. of sweets, soft drinks and desserts</td>
<td>#.#</td>
</tr>
<tr>
<td>8</td>
<td>131</td>
<td>Weekly freq. of fried fish &amp; fried chicken</td>
<td>#.#</td>
</tr>
<tr>
<td>8</td>
<td>135</td>
<td>No. of different fruits eaten at least once per week (1/wk)</td>
<td>01-12</td>
</tr>
<tr>
<td>8</td>
<td>137</td>
<td>No. of different fruits eaten at least once per month 1/mo</td>
<td>01,02,…12</td>
</tr>
<tr>
<td>8</td>
<td>139</td>
<td>No. of different vegetables eaten at least 1/wk</td>
<td>01,02,…20</td>
</tr>
<tr>
<td>8</td>
<td>141</td>
<td>No. of different vegetables eaten at least 1/mo</td>
<td>01,02,…20</td>
</tr>
<tr>
<td>8</td>
<td>143</td>
<td>No. of different meats, main dishes &amp; lunches eaten at least 1/wk</td>
<td>01,02,…18</td>
</tr>
<tr>
<td>8</td>
<td>145</td>
<td>No. of different meats, main dishes &amp; lunches eaten at least 1/mo</td>
<td>01,02,…18</td>
</tr>
<tr>
<td>8</td>
<td>147</td>
<td>No. of different breads, snacks and breakfasts eaten at least 1/wk</td>
<td>01,02,…17</td>
</tr>
<tr>
<td>8</td>
<td>149</td>
<td>No. of different breads, snacks and breakfasts eaten at least 1/mo</td>
<td>01,02,…17</td>
</tr>
</tbody>
</table>
The following are calculated excluding alcoholic beverages:

9  8  Total calories per day before alcoholic beverages (bab)  
    (See record 1)
9 15  Daily protein intake (bab)  
7 22  Daily fat intake (bab)  
9 29  Daily carbohydrate (bab)  
9 36  Daily calcium (bab)  
9 43  Daily phosphorus (bab)  
9 50  Daily iron (bab)  
9 57  Daily sodium (bab)  
9 64  Daily potassium (bab)
<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>71</td>
<td>Daily vitamin A (bab)</td>
<td>#.IU</td>
</tr>
<tr>
<td>9</td>
<td>78</td>
<td>Daily thiamin (bab)</td>
<td>#.mgs</td>
</tr>
<tr>
<td>9</td>
<td>85</td>
<td>Daily riboflavin (bab)</td>
<td>#.mgs</td>
</tr>
<tr>
<td>9</td>
<td>92</td>
<td>Daily niacin (bab)</td>
<td>#.mgs</td>
</tr>
<tr>
<td>9</td>
<td>99</td>
<td>Daily vitamin C (bab)</td>
<td>#.mgs</td>
</tr>
<tr>
<td>9</td>
<td>106</td>
<td>Daily saturated fat (bab)</td>
<td>#.grams</td>
</tr>
<tr>
<td>9</td>
<td>113</td>
<td>Daily oleic acid (bab)</td>
<td>#.grams</td>
</tr>
<tr>
<td>9</td>
<td>120</td>
<td>Daily linoleic acid (bab)</td>
<td>#.grams</td>
</tr>
<tr>
<td>9</td>
<td>127</td>
<td>Daily cholesterol (bab)</td>
<td>#.mgs</td>
</tr>
<tr>
<td>9</td>
<td>134</td>
<td>Daily dietary fiber (bab)</td>
<td>#.grams</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>Subject ID</td>
<td>(See record 1)</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>Line identifier</td>
<td>E</td>
</tr>
</tbody>
</table>

The following represent "seasonal" consumption:

<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>8</td>
<td>Total calories per day in summer (is)</td>
<td>#.</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>Daily protein intake (is)</td>
<td>#.grams</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>Daily fat intake (is)</td>
<td>#.grams</td>
</tr>
<tr>
<td>10</td>
<td>29</td>
<td>Daily carbohydrate (is)</td>
<td>#.grams</td>
</tr>
<tr>
<td>10</td>
<td>36</td>
<td>Daily calcium (is)</td>
<td>#.mgs</td>
</tr>
<tr>
<td>10</td>
<td>43</td>
<td>Daily phosphorus (is)</td>
<td>#.mgs</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>Daily iron (is)</td>
<td>#.mgs</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-----------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>10</td>
<td>57</td>
<td>Daily sodium (is)</td>
<td>#####.# mgs</td>
</tr>
<tr>
<td>10</td>
<td>64</td>
<td>Daily potassium (is)</td>
<td>#####.# mgs</td>
</tr>
<tr>
<td>10</td>
<td>71</td>
<td>Daily vitamin A (is)</td>
<td>#####.# IU</td>
</tr>
<tr>
<td>10</td>
<td>78</td>
<td>Daily thiamin (is)</td>
<td>#####.# mgs</td>
</tr>
<tr>
<td>10</td>
<td>85</td>
<td>Daily riboflavin (is)</td>
<td>#####.# mgs</td>
</tr>
<tr>
<td>10</td>
<td>92</td>
<td>Daily niacin (is)</td>
<td>#####.# mgs</td>
</tr>
<tr>
<td>10</td>
<td>99</td>
<td>Daily vitamin C (is)</td>
<td>#####.# mgs</td>
</tr>
<tr>
<td>10</td>
<td>106</td>
<td>Daily saturated fat (is)</td>
<td>#####.# grams</td>
</tr>
<tr>
<td>10</td>
<td>113</td>
<td>Daily oleic acid (is)</td>
<td>#####.# grams</td>
</tr>
<tr>
<td>10</td>
<td>120</td>
<td>Daily linoleic acid (is)</td>
<td>#####.# grams</td>
</tr>
<tr>
<td>10</td>
<td>127</td>
<td>Daily cholesterol (is)</td>
<td>#####.# mgs</td>
</tr>
<tr>
<td>10</td>
<td>134</td>
<td>Daily dietary fiber (is)</td>
<td>#####.# grams</td>
</tr>
</tbody>
</table>

11 1 Subject ID (See record 1)
11 7 Line identifier F

The following represent "out-of-season" consumption:

11 8 Total calories per day excluding summer or "out-of-season" (os)  
11 15 Daily protein intake (os)  
11 22 Daily fat intake (os)  
11 29 Daily carbohydrate (os)  
11 36 Daily calcium (os)  
11 43 Daily phosphorus (os)  

#####.

#####.

#####.

#####.

#####.

#####.

#####.

#####.

#####.
<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>50</td>
<td>Daily iron (os)</td>
<td>#.# mgs</td>
</tr>
<tr>
<td>11</td>
<td>57</td>
<td>Daily sodium (os)</td>
<td>#.# mgs</td>
</tr>
<tr>
<td>11</td>
<td>64</td>
<td>Daily potassium (os)</td>
<td>#.# mgs</td>
</tr>
<tr>
<td>11</td>
<td>71</td>
<td>Daily vitamin A (os)</td>
<td>#.# IU</td>
</tr>
<tr>
<td>11</td>
<td>78</td>
<td>Daily thiamin (os)</td>
<td>#.# mgs</td>
</tr>
<tr>
<td>11</td>
<td>85</td>
<td>Daily riboflavin (os)</td>
<td>#.# mgs</td>
</tr>
<tr>
<td>11</td>
<td>92</td>
<td>Daily niacin (os)</td>
<td>#.# mgs</td>
</tr>
<tr>
<td>11</td>
<td>99</td>
<td>Daily vitamin C (os)</td>
<td>#.# mgs</td>
</tr>
<tr>
<td>11</td>
<td>106</td>
<td>Daily saturated fat (os)</td>
<td>#.# grams</td>
</tr>
<tr>
<td>11</td>
<td>113</td>
<td>Daily oleic acid (os)</td>
<td>#.# grams</td>
</tr>
<tr>
<td>11</td>
<td>120</td>
<td>Daily linoleic acid (os)</td>
<td>#.# grams</td>
</tr>
<tr>
<td>11</td>
<td>127</td>
<td>Daily cholesterol (os)</td>
<td>#.# mgs</td>
</tr>
<tr>
<td>11</td>
<td>134</td>
<td>Daily dietary fiber (os)</td>
<td>#.# grams</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>Subject ID</td>
<td>(See record 1)</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>Line identifier</td>
<td>G</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>Dietary fiber - beans</td>
<td>#.# grams</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>Fiber - fruits &amp; vegetables</td>
<td>#.# grams</td>
</tr>
<tr>
<td>12</td>
<td>18</td>
<td>Fiber - grains</td>
<td>#.# grams</td>
</tr>
<tr>
<td>12</td>
<td>23</td>
<td>&quot;Nonlycopene carotenoids&quot; (i.e. excluding tomatoes)</td>
<td>#.# μg</td>
</tr>
<tr>
<td>12</td>
<td>28</td>
<td>Alpha-carotene</td>
<td>#.# μg</td>
</tr>
<tr>
<td>12</td>
<td>33</td>
<td>Alpha-carotene (is)</td>
<td>#.# μg</td>
</tr>
<tr>
<td>12</td>
<td>38</td>
<td>Alpha-carotene (os)</td>
<td>#.# μg</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>12</td>
<td>43</td>
<td>Beta-carotene</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>48</td>
<td>Beta-carotene (is)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>53</td>
<td>Beta-carotene (os)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>58</td>
<td>Lutein</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>63</td>
<td>Lutein (is)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>68</td>
<td>Lutein (os)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>73</td>
<td>Cryptoxanthin</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>78</td>
<td>Cryptoxanthin (is)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>83</td>
<td>Cryptoxanthin (os)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>88</td>
<td>Xanthins</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>93</td>
<td>Xanthins (is)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>98</td>
<td>Xanthins (os)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>103</td>
<td>Lycopene</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>108</td>
<td>Lycopene (is)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>113</td>
<td>Lycopene (os)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>118</td>
<td>Carotene from milk, cream, &amp; butter (0.27 of RE's)</td>
<td>###### µg</td>
</tr>
<tr>
<td>12</td>
<td>123</td>
<td>&quot;Carotene&quot; from eggs (Zero, no β-carotene in eggs, Beecher)</td>
<td>00000 µg</td>
</tr>
</tbody>
</table>

α- and β-carotenes and cryptoxanthin are the only carotenoids converted to Vitamin A, therefore their sum may be of interest.
FILE FACSFILE.DAT

<table>
<thead>
<tr>
<th>Record</th>
<th>Column</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Subject ID</td>
<td>LCP### - For patient number</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LCC### - For control number</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>B-cells: Mean of negative population (NP) on green fluorescence (GF)</td>
<td>###.###</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>B-cells: Mean of positive population (PP) on GF</td>
<td>###.###</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>B-cells: Proportion positive</td>
<td>0.###</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>HLA-DR: Mean of NP on GF (Two groups fitted)</td>
<td>###.###</td>
</tr>
<tr>
<td>1</td>
<td>36</td>
<td>HLA88-DR: Mean of PP on GF (Two groups fitted)</td>
<td>###.###</td>
</tr>
<tr>
<td>1</td>
<td>44</td>
<td>HLA-DR: Proportion in PP (Two groups fitted)</td>
<td>0.###</td>
</tr>
<tr>
<td>1</td>
<td>49</td>
<td>HLA-DR: Mean of NP on GF (Three groups fitted)</td>
<td>###.###</td>
</tr>
<tr>
<td>1</td>
<td>57</td>
<td>HLA-DR: Mean of higher PP (PP1) on GF (Three groups fitted)</td>
<td>###.###</td>
</tr>
<tr>
<td>1</td>
<td>65</td>
<td>HLA-DR: Mean of lower PP (PP2) on GF (Three groups fitted)</td>
<td>###.###</td>
</tr>
<tr>
<td>1</td>
<td>73</td>
<td>HLA-DR: Proportion in PP1 (Three groups fitted)</td>
<td>0.###</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>78</td>
<td>HLA-DR: Proportion in PP2 (Three groups fitted)</td>
<td>0.###</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Subject ID</td>
<td>(See record 1)</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>CD4-2H4 (green-red): Proportion with neither CD4 nor 2H4</td>
<td>0.###</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>CD4-2H4: Proportion with 2H4 but not CD4</td>
<td>0.###</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>CD4-2H4: Proportion with CD4 but not 2H4</td>
<td>0.###</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>CD4-2H4: Proportion with CD4 and 2H4</td>
<td>0.###</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>CD4-2H4: Mean on GF of population with neither CD4 nor 2H4</td>
<td>###.####</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>CD4-2H4: Mean on RF of population with neither CD4 nor 2H4</td>
<td>###.####</td>
</tr>
<tr>
<td>2</td>
<td>43</td>
<td>CD4-2H4: Mean on GF of population with 2H4 but not CD4</td>
<td>###.####</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>CD4-2H4: Mean on RF of population with 2H4 but not CD4</td>
<td>###.####</td>
</tr>
<tr>
<td>2</td>
<td>59</td>
<td>CD4-2H4: Mean on GF of population with CD4 but not 2H4</td>
<td>###.####</td>
</tr>
<tr>
<td>2</td>
<td>67</td>
<td>CD4-2H4: Mean on RF of population with CD4 but not 2H4</td>
<td>###.####</td>
</tr>
<tr>
<td>2</td>
<td>75</td>
<td>CD4-2H4: Mean on GF of population with both CD4 and 2H4</td>
<td>###.####</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-----------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>2</td>
<td>83</td>
<td>CD4-2H4: Mean on RF of population with both CD4 and 2H4</td>
<td>#.#.#.#.#</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Subject ID</td>
<td>(See record 1)</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>CD4-4B4 (green-red): Proportion with neither CD4 nor 4B4</td>
<td>0.#.#.</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>CD4-4B4: Proportion with 4B4 but not CD4</td>
<td>0.#.#.</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>CD4-4B4: Proportion with CD4 but not 4B4</td>
<td>0.#.#.</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>CD4-4B4: Proportion with CD4 and 4B4</td>
<td>0.#.#.</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>CD4-4B4: Mean on GF of population with neither CD4 nor 4B4</td>
<td>#.#.#.#.#</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>CD4-4B4: Mean on RF of population with neither CD4 nor 4B4</td>
<td>#.#.#.#.#</td>
</tr>
<tr>
<td>3</td>
<td>43</td>
<td>CD4-4B4: Mean on GF of population with 4B4 but not CD4</td>
<td>#.#.#.#.#</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
<td>CD4-4B4: Mean on RF of population with 4B4 but not CD4</td>
<td>#.#.#.#.#</td>
</tr>
<tr>
<td>3</td>
<td>59</td>
<td>CD4-4B4: Mean on GF of population with CD4 but not 4B4</td>
<td>#.#.#.#.#</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>CD4-4B4: Mean on RF of population with CD4 but not 4B4</td>
<td>#.#.#.#.#</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>CD4-4B4: Mean on GF of population with both CD4 and 4B4</td>
<td>#.#.#.#.#</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-----------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3</td>
<td>83</td>
<td>CD4-4B4: Mean on RF of population with both CD4 and 4B4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>Subject ID</td>
<td>(See record 1)</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>NKH-1: Mean of higher NP (NP1) on GF</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>NKH-1: Mean of lower NP (NP2) on GF</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>NKH-1: Mean of PP1 on GF</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>NKH-1: Mean of PP2 on GF</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>NKH-1: Proportion in NP1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>44</td>
<td>NKH-1: Proportion in PP1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td>NKH-1: Proportion in PP2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>54</td>
<td>NKH-1: Total proportion positive (PP1+PP2)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>59</td>
<td>CD3: Mean of NP on GF</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>67</td>
<td>CD3: Mean of PP on GF</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>75</td>
<td>CD3: Proportion in PP</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>Subject ID</td>
<td>(See record 1)</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>CD4: Mean of NP on GF</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>CD4: Mean of PP on GF</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>CD4: Proportion in PP</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>CD8: Mean of NP on GF</td>
<td></td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>CD8: Mean of higher PP on GF</td>
<td>###.###</td>
</tr>
<tr>
<td>5</td>
<td>44</td>
<td>CD8: Mean of lower PP on GF</td>
<td>###.###</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>CD8: Proportion in higher PP</td>
<td>0.###</td>
</tr>
<tr>
<td>5</td>
<td>57</td>
<td>CD8: Proportion in lower PP</td>
<td>0.###</td>
</tr>
<tr>
<td>5</td>
<td>62</td>
<td>The subject's number</td>
<td>###</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>Subject ID</td>
<td>(See record 1)</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>Source of control subject</td>
<td>(See codes below)</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>Hours between collection of blood and labelling</td>
<td>1,2,...,17</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>Month sample collected</td>
<td>01,02,...,12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>. - Missing</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>Day sample collected</td>
<td>01,02,...,31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>' ' - Missing</td>
</tr>
<tr>
<td>6</td>
<td>21</td>
<td>Year sample collected</td>
<td>(19)88,(19)89,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(19)90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>' ' - Missing</td>
</tr>
<tr>
<td>6</td>
<td>23</td>
<td>Month FACS done</td>
<td>01,02,...,12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>. - Missing</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>Day FACS done</td>
<td>01,02,...,31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>' ' - Missing</td>
</tr>
<tr>
<td>6</td>
<td>27</td>
<td>Year FACS done</td>
<td>(19)88,(19)89,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(19)90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>' ' - Missing</td>
</tr>
<tr>
<td>Record</td>
<td>Column</td>
<td>Variable</td>
<td>Code</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>6</td>
<td>29</td>
<td>Month questionnaire returned</td>
<td>01, 02, ..., 12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>. - Missing</td>
</tr>
<tr>
<td>6</td>
<td>31</td>
<td>Day questionnaire returned</td>
<td>01, 02, ..., 31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>' ' - Missing</td>
</tr>
<tr>
<td>6</td>
<td>33</td>
<td>Year questionnaire returned</td>
<td>(19) 88, (19) 89,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(19) 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>' ' - Missing</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>Hour of day blood drawn</td>
<td>07, 08, ..., 24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 - Missing</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td>Minutes after hour of blood draw</td>
<td>00, 01, ..., 59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 - Missing</td>
</tr>
<tr>
<td>6</td>
<td>39</td>
<td>Hematocrit, percent</td>
<td>20, 21, ..., 65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 - Missing</td>
</tr>
<tr>
<td>6</td>
<td>41</td>
<td>Neutrophils, percent</td>
<td>30, 31, ..., 98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 - Missing</td>
</tr>
<tr>
<td>6</td>
<td>43</td>
<td>Lymphocytes, percent</td>
<td>5, 6, ..., 65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 - Missing</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>Monocytes, percent</td>
<td>5, 6, ..., 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 - Missing</td>
</tr>
<tr>
<td>6</td>
<td>47</td>
<td>Eosinophils, percent</td>
<td>0, 1, ..., 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 - Missing</td>
</tr>
<tr>
<td>6</td>
<td>49</td>
<td>Basophils, percent</td>
<td>0, 1, ..., 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99 - Missing</td>
</tr>
<tr>
<td>Code</td>
<td>Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AARP</td>
<td>American Assn. of Retired Persons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COAGING</td>
<td>Council on Aging</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIWANIS</td>
<td>Kiwanis International</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIONS</td>
<td>Lions Clubs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMLEG</td>
<td>American Legion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMLEGJAK</td>
<td>Amer. Leg. Veterans Home, Jackson, LA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DUMAS</td>
<td>Referred by former Baton Rouge Mayor Woody Dumas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VFW</td>
<td>Veterans of Foreign Wars</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SALVARMY</td>
<td>Salvation Army</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERSFRND</td>
<td>Personal friend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHURCH</td>
<td>University Baptist Church</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KOFC</td>
<td>Knights of Columbus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUARD</td>
<td>School of Veterinary Medicine security guard</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CONTROL ORIGIN CODES
VITA

Joan Peters Blackmon was born in Pittsburgh, Pa. on August 23, 1942, the daughter of David Peters and Elinor Livelsberger Peters. She was educated in the public schools, then entered the University of Pittsburgh following her junior year in high school. As an undergraduate she received two National Science Foundation summer research fellowships and was elected to Phi Beta Kappa. She graduated *magna cum laude* with a B.S. degree in chemistry in April, 1963. Upon graduation she was awarded a research assistantship in organic chemistry at Duke University, where she earned the A.M. degree in 1965. She was employed by the Research Triangle Institute in Durham, N.C. for five years as a synthetic organic chemist. As a chemist, she was coauthor of 13 journal articles in professional journals.

In 1971, she earned the M.S.P.H. degree in epidemiology at the School of Public Health, University of North Carolina at Chapel Hill. She married William J. Blackmon and temporarily retired from career pursuits to bear and nurture three children: Heather, and twins Amanda and Timothy. A recipient of an Alumnae Federation Fellowship at Louisiana State University in 1984, she is currently a candidate for the Ph.D. degree in the Department of Veterinary Microbiology and Parasitology, School of Veterinary Medicine.
DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Joan Peters Blackmon

Major Field: Veterinary Medical Sciences

Title of Dissertation: The Immunoepidemiology of Lung Cancer and Smoking: Lymphocyte Subsets

Approved:

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Edgar Beny Moses

Daniel F. Church

William J. Todd

Thomas F. Gilles

Date of Examination:

October 21, 1991