

## Appendix A

### Models for Phase I: Classification

#### 0.5-year:

$\log(p/1-p) = 0.400 + 0.457 \cdot \text{Summary Stage Localized} + 0.322 \cdot \text{Summary Stage Regional} - 0.109 \cdot \text{Histology: } <8021 - 0.130 \cdot \text{CS site specific factor 1: Unknown} - 0.135 \cdot \text{Mets at DX-bone: Yes} - 0.136 \cdot \text{Age at diagnosis} - 0.169 \cdot \text{Mets at DX-brain: Yes} - 0.191 \cdot \text{Mets at DX-liver: Yes}$

#### 1-year:

$\log(p/1-p) = 0.170 + 0.311 \cdot \text{Summary stage: Localized} + 0.243 \cdot \text{Summary stage: Regional} + 0.203 \cdot \text{RX Summ Surg Prim Site: Yes} + 0.135 \cdot \text{Primary site: Upper lobe lung} + 0.129 \cdot \text{Histology: } >8083 - 0.109 \cdot \text{Mets at DX lung: Yes} - 0.112 \cdot \text{Histology: } <8021 - 0.121 \cdot \text{Histology: } 8042-8062 - 0.126 \cdot \text{Regional nodes examined} - 0.127 \cdot \text{Mets at DX-brain: Yes} - 0.136 \cdot \text{CS site specific factor 1: 20} - 0.150 \cdot \text{CS site specific factor 1: 30} - 0.151 \cdot \text{CS site specific factor 1: 40} - 0.159 \cdot \text{Mets at DX-liver: Yes} - 0.166 \cdot \text{CS site specific factor 1: 10} - 0.188 \cdot \text{CS site specific factor 1: Unknown}$

#### 1.5-year:

$\log(p/1-p) = 0.106 + 0.244 \cdot \text{RX Summ Surg Prim Site: Yes} + 0.196 \cdot \text{Summary stage: Localized} + 0.189 \cdot \text{Summary stage: Regional} + 0.139 \cdot \text{Histology: } >8083 + 0.129 \cdot \text{Primary site: Upper lobe lung} + 0.112 \cdot \text{Race: White} - 0.123 \cdot \text{Mets at DX liver: Yes} - 0.123 \cdot \text{CS site specific factor 2: Unknown} - 0.136 \cdot \text{CS site specific factor 1: 40} - 0.139 \cdot \text{Regional nodes examined} - 0.141 \cdot \text{CS site specific factor 1: 30} - 0.146 \cdot \text{CS site specific factor 1: 20} - 0.169 \cdot \text{CS site specific factor 1: Unknown} - 0.197 \cdot \text{CS site specific factor 1: 10}$

#### 2-year:

$\log(p/1-p) = 0.142 + 0.241 \cdot \text{RX Summ Surg Prim Site: Yes} + 0.165 \cdot \text{Summary stage: Regional} + 0.134 \cdot \text{Histology: } >8083 + 0.130 \cdot \text{Summary stage Localized} + 0.122 \cdot \text{Race: White} + 0.120 \cdot \text{Primary site: Upper lobe lung} - 0.115 \cdot \text{Mets at DX-liver: Yes} - 0.118 \cdot \text{Histology: } <8021 - 0.125 \cdot \text{Histology: } 8042-8062 - 0.139 \cdot \text{Regional nodes examined} - 0.150 \cdot \text{CS site specific factor 1: 40} - 0.154 \cdot \text{CS site specific factor 1: 30} - 0.156 \cdot \text{CS site specific factor 2: Unknown} - 0.174 \cdot \text{CS site specific factor 1: Unknown} - 0.179 \cdot \text{CS site specific factor 1: 20} - 0.211 \cdot \text{CS site specific factor 1: 10}$

#### 2.5-year:

$\log(p/1-p) = 0.132 + 0.241 \cdot \text{RX Summ Surg Prim Site: Yes} + 0.156 \cdot \text{Histology: } >8083 + 0.142 \cdot \text{Race: White} + 0.138 \cdot \text{Summary stage: Regional} + 0.107 \cdot \text{Primary Site: Upper lobe lung} - 0.114 \cdot \text{Histology: } 8042-8062 - 0.126 \cdot \text{Mets at DX-liver: Yes} - 0.147 \cdot \text{Regional nodes examined} - 0.169 \cdot \text{CS site specific factor 2: Unknown} - 0.183 \cdot \text{CS site specific factor 1: 40} - 0.189 \cdot \text{CS site specific factor 1: Unknown} - 0.192 \cdot \text{CS site specific factor 1: 30} - 0.215 \cdot \text{CS site specific factor 1: 20} - 0.262 \cdot \text{CS site specific factor 1: 10}$

#### 3-year:

$\log(p/1-p) = 0.152 + 0.190 \cdot \text{RX Summ Surg Prim Site: Yes} + 0.152 \cdot \text{Summary stage: Regional} + 0.151 \cdot \text{Histology: } >8083 + 0.121 \cdot \text{Race: White} + 0.109 \cdot \text{Summary stage: Localized} - 0.108 \cdot \text{Histology: } <8021 - 0.126 \cdot \text{Mets at DX liver: Yes} - 0.142 \cdot \text{Regional nodes examined} - 0.157 \cdot \text{CS site specific factor 1: Unknown} - 0.177 \cdot \text{CS site specific factor 2: Unknown} - 0.190 \cdot \text{CS site specific factor 1: 40} - 0.199 \cdot \text{CS site specific factor 1: 30} - 0.207 \cdot \text{CS site specific factor 1: 20} - 0.284 \cdot \text{CS site specific factor 1: 10}$

### Model for Phase II: Regression

$y = 1.462 + 7.072 * \text{Histology: } >8083 + 6.404 * \text{Histology: } 8021-8042 + 5.871 * \text{Histology: } 8063-8083 + 5.804 * \text{Histology: } 8042-8062 + 4.881 * \text{Histology: } <8021 + 4.705 * \text{Summary stage: Localized} + 2.591 * \text{Summary stage: Regional} + 1.908 * \text{Regional nodes examined} + 1.585 * \text{Diagnostic confirmation: Positive histology} + 1.456 * \text{Race: Other (American Indian, Asian/Pacific Islander, Unknown)} + 1.341 * \text{RX Summ Surg Prim Site: Yes} + 1.257 * \text{Diagnostic confirmation: Positive exfoliative cytology} + 1.015 * \text{Primary site: Upper lobe lung} - 1.020 * \text{Mets at DX-bone: Yes} - 1.176 * \text{Grade: Poorly differentiated (Grade III)} - 1.217 * \text{Mets at DX-brain: Yes} - 1.235 * \text{Age at diagnosis} - 1.829 * \text{Mets at DX-liver: Yes}$