Appendices – Table of Contents

Appendix 1. British Columbia Cancer Agency epidemiological questionnaire administered to lung cancer study participants

Appendix 2. Logistic regression Model 2 (article Table 2) expressed with beta coefficients and model constant (Stata statistical program printout)

```
. logit cancerno  ageC  educatC BMIC rellung pkyrC cigstat quittimeC
sexno##c.FEV1

Iteration 0:   log likelihood =  -537.9655
Iteration 1:   log likelihood = -499.51356
Iteration 2:   log likelihood = -472.93428
Iteration 3:   log likelihood = -472.73827
Iteration 4:   log likelihood =   -472.738
Iteration 5:   log likelihood =   -472.738

Logistic regression                                     Number of obs   =       2522
                                                        LR chi2(10)     =     130.45
                                                        Prob > chi2     =     0.0000
Log likelihood =   -472.738                             Pseudo R2       =     0.1212

------------------------------------------------------------------------------
    cancerno |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ageC |   .0454658   .0134467     3.38   0.001     .0191108    .0718209
     educatC |  -.0916623   .0630036    -1.45   0.146    -.2151471    .0318225
        BMIC |  -.0686361   .0238571    -2.88   0.004    -.1153951   -.0218771
     rellung |   .4200975   .2505529     1.68   0.094     -.070977    .9111721
       pkyrC |   .0088849    .004509     1.97   0.049     .0000474    .0177223
     cigstat |   .6632661   .2714387     2.44   0.015      .131256    1.195276
   quittimeC |  -.0175864   .0228021    -0.77   0.441    -.0622777    .0271048
     1.sexno |   2.285665   .7079188     3.23   0.001     .8981696     3.67316
        FEV1 |  -.0061832   .0067966    -0.91   0.363    -.0195042    .0071379
             |
sexno#c.FEV1 |
           1 |  -.0279219   .0086295    -3.24   0.001    -.0448355   -.0110083
             |
       _cons |  -3.534922   .7374864    -4.79   0.000    -4.980369   -2.089475
------------------------------------------------------------------------------
```

Abbreviations: FEV1, FEV1% or forced expiratory volume percent predicted in one second; FHxLCA, family history of lung cancer; PKYR, pack-years smoking history; SDIC, sputum DNA image cytometry.

[c], centered on mean value of continuous variables and the mode of ordinal variables (*Age* 58.2 years; *Education* centered on high school, *BMI* 26.9 kg/m2, *PKYR* 47.2, *Quit-time* 4.7 years).

Appendix 3. Logistic regression Model 3 (article Table 2) expressed with beta coefficients and model constant (Stata statistical program printout)

```
. logit Cancer  Age[c]  Education[c] BMI[c] FHxLCA PKYR[c] Smokestatus Quit-time[c]
sexno##c.FEV1 SDIC

Iteration 0:   log likelihood =  -537.9655
Iteration 1:   log likelihood = -497.54299
Iteration 2:   log likelihood = -470.27303
Iteration 3:   log likelihood = -470.05773
Iteration 4:   log likelihood = -470.05739
Iteration 5:   log likelihood = -470.05739

Logistic regression                                     Number of obs   =       2522
                                                        LR chi2(11)     =     135.82
                                                        Prob > chi2     =     0.0000
Log likelihood = -470.05739                             Pseudo R2       =     0.1262

------------------------------------------------------------------------------
      Cancer |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Age[c] |    .045605   .0135055     3.38   0.001     .0191347    .0720753
   Education[c] |  -.0950152   .0634379    -1.50   0.134    -.2193512    .0293208
         BMI[c] |  -.0698111   .0240301    -2.91   0.004    -.1169092   -.0227131
       FHxLCA |   .4522496   .2516854     1.80   0.072    -.0410447    .9455439
        PKYR[c] |    .008727   .0045573     1.91   0.056    -.0002052    .0176592
Smoke status |   .5968631   .2730776     2.19   0.029     .0616409    1.132085
   Quit-time[c] |   -.016746   .0227783    -0.74   0.462    -.0613907    .0278986
        1.Sex |   2.252948   .7174859     3.14   0.002     .8467018    3.659195
         FEV1 |  -.0064802   .0068931    -0.94   0.347    -.0199904       .00703
              |
 Sex*#c.FEV1 |
           1  |  -.0280037   .0087454    -3.20   0.001    -.0451443   -.0108631
              |
         SDIC |   .2641296   .1150227     2.30   0.022     .0386892    .4895699
        _cons |  -3.940088   .7671156    -5.14   0.000    -5.443607   -2.436569
------------------------------------------------------------------------------
```

Abbreviations: FEV1, FEV1% or forced expiratory volume percent predicted in one second; FHxLCA, family history of lung cancer; PKYR, pack-years smoking history; SDIC, sputum DNA image cytometry.

[c], centered on mean value of continuous variables and the mode of ordinal variables (*Age* 58.2 years; *Education* centered on high school, *BMI* 26.9 kg/m2, *PKYR* 47.2, *Quit-time* 4.7 years).

Appendix 4. Logistic regression model including Table 2 base Model 1 plus sputum DNA image cytometry, expressed with beta coefficients and model constant (Stata statistical program printout)

```
. logit Cancer  Age^c  Education^c BMI^c FHxLCA PKYR^c Smokestatus Quit-time^c Sex SDIC

Iteration 0:   log likelihood =  -537.9655
Iteration 1:   log likelihood = -500.34842
Iteration 2:   log likelihood = -488.19881
Iteration 3:   log likelihood = -488.18422
Iteration 4:   log likelihood = -488.18422

Logistic regression                                     Number of obs   =       2522
                                                        LR chi2(9)      =      99.56
                                                        Prob > chi2     =     0.0000
Log likelihood = -488.18422                             Pseudo R2       =     0.0925

------------------------------------------------------------------------------
    cancerno |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Age^c |    .061727   .0128882     4.79   0.000     .0364666    .0869874
  Education^c |  -.1312411   .0619426    -2.12   0.034    -.2526464   -.0098358
        BMI^c |  -.0799686    .024287    -3.29   0.001    -.1275702   -.0323669
      FHxLCA |   .4276221   .2469014     1.73   0.083    -.0562959      .91154
       PKYR^c |   .0146017   .0042418     3.44   0.001     .0062879    .0229156
 Smokestatus |    .592005   .2685174     2.20   0.027     .0657205     1.11829
  Quit-time^c |  -.0301089   .0225549    -1.33   0.182    -.0743157    .0140979
         Sex |   .1318689   .1892999     0.70   0.486    -.2391522    .5028899
        SDIC |   .2607657   .1128455     2.31   0.021     .0395925    .4819388
       _cons |  -4.521673   .4867886    -9.29   0.000    -5.475761   -3.567585
------------------------------------------------------------------------------
```

Abbreviations: FHxLCA, family history of lung cancer; PKYR, pack-years smoking history; SDIC, sputum DNA image cytometry.

[c], centered on mean value of continuous variables and the mode of ordinal variables (*Age* 58.2 years; *Education* centered on high school, *BMI* 26.9 kg/m2, *PKYR* 47.2, *Quit-time* 4.7 years).

Appendix 5. Cox proportional hazards model analogous to the logistic regression Model 3 (article Table 3)

```
. stcox Age^c  Education^c BMI^c FHxLCA PKYR^c Smokestatus Quit-time^c sexno##c.FEV1
SDIC

         failure _d:  cancerno == 1
   analysis time _t:  followup
                 id:  sid

Iteration 0:   log likelihood = -1050.0569
Iteration 1:   log likelihood = -1013.5353
Iteration 2:   log likelihood = -991.70087
Iteration 3:   log likelihood = -990.91627
Iteration 4:   log likelihood = -990.91476
Refining estimates:
Iteration 0:   log likelihood = -990.91476

Cox regression -- Breslow method for ties

No. of subjects =         2522                     Number of obs    =      2522
No. of failures =          139
Time at risk    =      6816460
                                                   LR chi2(11)      =    118.28
Log likelihood  =   -990.91476                     Prob > chi2      =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Age^c |    1.04621   .0132621     3.56   0.000     1.020537    1.072529
  Education^c |   .9299067   .0548556    -1.23   0.218     .8283744    1.043884
        BMI^c |   .9399262   .0210178    -2.77   0.006     .8996218    .9820363
      FHxLCA |   1.627494   .3744615     2.12   0.034     1.036743    2.554865
       PKYR^c |   1.008049    .003978     2.03   0.042     1.000282    1.015876
Smoke status |   1.385478   .3624822     1.25   0.213     .8296595    2.313658
  Quit-time^c |   .9890736   .0216441    -0.50   0.616      .947549    1.032418
     1.sexno |   7.087786   4.571627     3.04   0.002     2.002097    25.09205
        FEV1 |   .9956074   .0065769    -0.67   0.505        .9828    1.008582
             |
  Sex#c.FEV1 |
          1  |   .9761076   .0077003    -3.07   0.002     .9611313    .9913171
             |
        SDIC |   1.230358   .1322502     1.93   0.054     .9966358    1.518891
------------------------------------------------------------------------------
```

Abbreviations: FEV1, FEV1% or forced expiratory volume percent predicted in one second; FHxLCA, family history of lung cancer; PKYR, pack-years smoking history; SDIC, sputum DNA image cytometry.

[c], centered on mean value of continuous variables and the mode of ordinal variables (*Age* 58.2 years; *Education* centered on high school, *BMI* 26.9 kg/m2, *PKYR* 47.2, *Quit-time* 4.7 years).