# APPENDIX

Exposure assessment definitions and equations:

Jobs with “high” exposure were defined as those with at least 9% probability of exceeding PEL; this corresponded to the 50th percentile of non-zero probabilities assigned to all occupations among controls. Intermediate exposure groups were defined as “low” (θ = 0.1–2.9%) and “medium” (θ = 3.0–8.9%), which corresponded to less than the 25th and between the 25th and 50th percentile, respectively, and jobs with zero probability of exceedance were treated as unexposed. A second exposure metric, weighted duration of exposure, isdefined in equation (1):

|  |  |
| --- | --- |
| $$Weighted Duration\_{i}= \sum\_{k=1}^{K\_{i}}\left(PP\_{i,k} × D\_{i,k}\right)$$ | (1) |

where PP*i,k* is the predicted probability from the JEM for participant *i* during job number *k*, D*i,k* is the duration, and K*i* is the total number of jobs reported by participant *i* in the study. Weighted duration is analogous to cumulative exposure but uses probability instead of intensity. A third exposure metric utilized is the average probability of exposure weighted by duration defined by equation (2):

|  |  |
| --- | --- |
| $$Average Probability\_{i}= \frac{\sum\_{k=1}^{K\_{i}}\left(PP\_{i,k} × D\_{i,k}\right)}{\sum\_{k=1}^{K\_{i}}D\_{i,k}}$$ | (2) |

Exposure assessments were modified based on the participants’ reported tasks and materials handled. For example, if the JEM assigned no exposure, but the participant indicated they worked with materials known to be a source for PAHs, the exposure level for that job was modified based on *a priori* criteria below.

Supplemental Material: Algorithm for updating employee exposure based on materials handled and tasks performed on the job

In cases where the JEM indicates that, for some particular job, there is no exposure or risk of exposure but produces a conflict with the results of the supplemental questions: materials handled or tasks performed on the job, then the assigned score is updated accordingly. The general steps are outlined below:

1. JEM assigns score for job A
2. If the score for job A indicates no exposure or risk of exposure, then check supplemental question 1 (materials handled) and 2 (tasks performed)
3. If supplemental question 1 or 2 is YES for exposure, then update score for job A from 0 (None) to an imputed probability. If the score for job A is non-zero, then the score is not updated with the supplementary questions.

The algorithm updates the probability in step 3 using the probability 2.9%. This probability is the upper limit of the “low” exposure group described previously that defines the 25th percentile of non-zero probabilities assigned to all occupations among controls. The decision-making process was developed by the authors (DGL, JJS, and IB).

# SUPPLEMENTARY TABLES

Supplementary Table A1: Top 10 most common at risk industries (NAICS 3-digit code) based on Yes-No high-level of exposure classification using 3,514 unique occupations observed across all participants∆.

|  |  |  |
| --- | --- | --- |
| NAICS (3-digit code) | Code Description | Marginal percent |
| 722 | Food Service and Drinking Places  | 16.4 |
| 311  | Food Manufacturing  | 8.0 |
| 315  | Clothing Manufacturing  | 7.5 |
| 334  | Computer and Electronic Product Manufacturing  | 6.2 |
| 325 | Chemical Manufacturing | 5.4 |
| 322 | Paper Manufacturing | 4.7 |
| 524 | Insurance Carriers and Related Activities | 4.5 |
| 335 | Electrical Equipment, Appliance and Component Manufacturing | 3.5 |
| 339 | Miscellaneous Manufacturing | 3.4 |
| 236 | Construction of Buildings | 2.8 |

 Analysis for exposure at high level (estimated probability of exposure above 0.2 mg/m3 of coal tar pitch volatiles) is θ ≥ 9% in at least one job

∆ To obtain these marginal percentages we used the 3,514 unique occupations, which were unique identifiers that combined industry (NAICS) and occupation (SOC) codes. Only unique occupations that had non-zero probabilities for exposure at high levels (*n* = 597, 17%) were retained and their respective industry codes were truncated to the 3-digit NAICS code. For example, 72211-353031 is the combined code for Restaurant (Industry) and Waitress (Occupation) and 72211-352012 is the combined code for Restaurant (Industry) and Cook (Occupation), which are both identified as exposed by the JEM, are part of the industry code 722. Marginal percentages of at risk industries were then calculated for all 3-digit NAICS codes.

Supplementary Table A2: Exposure assessment to PAHs and breast cancer risk stratified by hormone receptor statusⱡ

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|   |   |   | ER/PR+ |  | ER/PR- |   |  Receptor |
| Exposure Assessment | Controls (%) |   | Cases (%) |   | OR | 95% CI |   | Cases (%) |   | OR | 95% CI |   | effect |
| Ever-Never: Any level |  |  |  |  |  |  |  |  |  |  |  |  |
| Never | 454 (39.8) |  | 262 (32.1) |  | ----- |  |  |  | 47 (29.9) |  | ----- |  |  |  |  |
| Ever | 687 (60.2) |  | 555 (67.9) |  | **1.27** | **1.04** | **1.56** |  | 110 (70.1) |  | 1.40 | 0.96 | 2.04 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  | p > 0.7 |
| Ever-Never: At maximum level† |  |  |  |  |  |  |  |  |  |  |  |  |
| Never | 454 (39.8) |  | 262 (32.1) |  | ----- |  |  |  | 47 (29.9) |  | ----- |  |  |  |  |
| Maximum level at low\* | 107 (09.4) |  | 69 (08.4) |  | 1.04 | 0.74 | 1.48 |  | 13 (08.3) |  | 1.03 | 0.53 | 2.00 |  |  |
| Maximum level at medium¶ | 178 (15.6) |  | 122 (14.9) |  | 1.15 | 0.86 | 1.53 |  | 30 (19.1) |  | 1.55 | 0.94 | 2.56 |  |  |
| Maximum level at high | 402 (35.2) |  | 364 (44.6) |  | **1.40** | **1.12** | **1.75** |  | 67 (42.7) |  | 1.44 | 0.95 | 2.17 |  |  |
|  |  |  |  |  | ptrend < 0.01 |  |  |  | ptrend = 0.06 |  | p > 0.9 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Duration (years) of exposure at any level |  |  |  |  |  |  |  |  |  |  |  |  |
| None (0) | 454 (39.7) |  | 262 (32.1) |  | ----- |  |  |  | 47 (29.9) |  | ----- |  |  |  |  |
| Short (0.1–4.2) | 229 (20.1) |  | 173 (21.2) |  | **1.35** | **1.04** | **1.75** |  | 35 (22.3) |  | 1.47 | 0.92 | 2.38 |  |  |
| Moderate (4.3–13.0) | 230 (20.2) |  | 190 (23.3) |  | **1.29** | **1.01** | **1.67** |  | 44 (28.0) |  | **1.66** | **1.05** | **2.62** |  |  |
| Long (13.1–82.2) | 228 (20.0) |  | 192 (23.4) |  | 1.17 | 0.90 | 1.53 |  | 31 (19.8) |  | 1.08 | 0.65 | 1.78 |  |  |
|  |  |  |  |  | ptrend > 0.1 |  |  |  | ptrend > 0.4 |  | p > 0.9 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Duration (years) of exposure at medium¶ or high levels  |  |  |  |  |  |  |  |  |  |  |  |  |
| None (0) | 454 (39.7) |  | 262 (32.1) |  | ----- |  |  |  | 47 (29.9) |  | ----- |  |  |  |  |
| Ever: Maximum at low level∆ | 107 (09.4) |  | 69 (08.4) |  | 1.04 | 0.74 | 1.48 |  | 13 (08.3) |  | 1.03 | 0.53 | 2.01 |  |  |
| Short (0.1–2.7) | 194 (17.0) |  | 153 (18.7) |  | **1.38** | **1.05** | **1.80** |  | 27 (17.2) |  | 1.30 | 0.77 | 2.17 |  |  |
| Moderate (2.8–9.0) | 196 (17.2) |  | 147 (18.0) |  | 1.21 | 0.92 | 1.60 |  | 32 (20.4) |  | 1.57 | 0.96 | 2.58 |  |  |
| Long (9.1–80.8) | 190 (16.7) |  | 186 (22.8) |  | **1.37** | **1.04** | **1.80** |  | 38 (24.2) |  | 1.55 | 0.96 | 2.52 |  |  |
|  |  |  |  |  | ptrend = 0.03 |  |  |  | ptrend = 0.04 |  | p > 0.4 |
| Duration (years) of exposure at the high level |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| None (0) | 454 (39.8) |  | 262 (32.1) |  | ----- |  |  |  | 47 (29.9) |  | ----- |  |  |  |  |
| Ever: Maximum at low\* or medium¶ levels◊ | 285 (25.0) |  | 191 (23.4) |  | 1.11 | 0.86 | 1.42 |  | 43 (27.4) |  | 1.35 | 0.86 | 2.11 |  |  |
| Short (0.1–2.3) | 134 (11.7) |  | 120 (14.7) |  | **1.58** | **1.17** | **2.14** |  | 21 (13.4) |  | 1.47 | 0.83 | 2.58 |  |  |
| Moderate (2.4–7.4) | 134 (11.7) |  | 99 (12.1) |  | 1.17 | 0.85 | 1.61 |  | 21 (13.4) |  | 1.48 | 0.84 | 2.60 |  |  |
| Long (7.5–74.1) | 134 (11.7) |   | 145 (17.7) |  | **1.44** | **1.07** | **1.94** |  | 25 (15.9) |  | 1.37 | 0.79 | 2.38 |  |  |
|  |  |  |  |  | ptrend = 0.02 |  |  |  | ptrend > 0.1 |  | p > 0.9 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Weighted Duration (Years) – Equation (1) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| None (0) | 454 (39.7) |  | 262 (32.1) |  | ----- |  |  |  | 47 (29.9) |  | ----- |  |  |  |  |
| Short (0.1–0.4) | 229 (20.1) |  | 165 (20.2) |  | 1.21 | 0.94 | 1.57 |  | 42 (26.8) |  | **1.67** | **1.06** | **2.63** |  |  |
| Moderate (0.5–1.7) | 229 (20.1) |  | 181 (22.2) |  | 1.27 | 0.98 | 1.65 |  | 27 (17.2) |  | 1.06 | 0.64 | 1.77 |  |  |
| Long (1.8–55.1) | 229 (20.1) |  | 209 (25.6) |  | **1.34** | **1.04** | **1.73** |  | 41 (26.1) |  | 1.46 | 0.91 | 2.34 |   |   |
|  |  |  |  |  | ptrend < 0.02 |  |  |  | ptrend > 0.2 |  | p > 0.7 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Average Probability – Equation (2) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| None (0) | 454 (39.7) |  | 262 (32.1) |  | ----- |  |  |  | 47 (29.9) |  | ----- |  |  |  |  |
| Low (0.01–0.02) | 229 (20.1) |  | 157 (19.2) |  | 1.17 | 0.90 | 1.52 |  | 37 (23.6) |  | 1.46 | 0.92 | 2.34 |  |  |
| Medium (0.03–0.07) | 229 (20.1) |  | 188 (23.0) |  | **1.35** | **1.04** | **1.74** |  | 39 (24.8) |  | 1.58 | 0.99 | 2.52 |  |  |
| High (0.08–0.88) | 229 (20.1) |  | 210 (25.7) |  | **1.31** | **1.01** | **1.69** |  | 34 (21.7) |  | 1.15 | 0.70 | 1.88 |   |   |
|  |  |  |  |  | ptrend < 0.02 |  |  |  | ptrend > 0.3 |  | p > 0.5 |

ⱡ Adjusted for age, centre, education, ethnicity, smoking (pack-years). All ptrend values are calculated by treating ordinal categories as continuous values

† Maximum level classification, regardless of duration, is the maximum exposure level to which the participant was exposed across all occupations.

\* Analysis for exposure at low level (estimated probability of exposure above 0.2 mg/m3 of coal tar pitch volatiles) is θ = (0.1 – 2.9%) in at least one job

¶ Analysis for exposure at medium level (estimated probability of exposure above 0.2 mg/m3 of coal tar pitch volatiles) is θ = (3.0 – 8.9%) in at least one job

 Analysis for exposure at high level (estimated probability of exposure above 0.2 mg/m3 of coal tar pitch volatiles) is θ ≥ 9% in at least one job

∆ To ensure referent group are truly unexposed, a nuisance variable was created for the low-exposed group where value = 1, if highest duration at low level exposure, else 0

◊ To ensure referent group are truly unexposed, a nuisance variable was created for the low/medium-exposed group where value = 1, if highest duration at low or medium level exposure, else 0