

## **SUPPORTING ONLINE MATERIAL**

### **MATERIALS AND METHODS**

#### **Study Design**

We carried out a cross-sectional survey of 250 workers exposed to benzene and 140 unexposed controls to evaluate the impact of low levels of benzene exposure on hematologic, cytogenetic and molecular endpoints. Exposed workers were enrolled from two shoe manufacturing factories. Unexposed controls were selected from three clothes manufacturing factories in the same region of Tianjin, China. Controls were frequency-matched by sex and age to exposed workers. The study was approved by Institutional Review Boards at the U.S. National Cancer Institute and the Chinese Academy of Preventive Medicine. Participation was voluntary, and written informed consent was obtained. The participation rate was approximately 95%. Blood samples were collected from 88 workers in June 2000 during the first year of the study and from the remaining workers (plus repeat samples from 28 subjects enrolled in the first year) in May and June 2001. The same personnel using essentially the same methods carried out the study in both years.

Subjects were administered a questionnaire requesting information on occupational history, environmental exposures, medical history and current medications, and past and current tobacco and alcohol use. Subjects provided a 29 ml peripheral blood sample, a buccal cell mouth rinse sample, and underwent a physical exam. Demographic characteristics of study subjects are shown in Table S1 by exposure categories based on the arithmetic mean of their benzene air levels in the month before phlebotomy.

## **Laboratory analyses**

### Blood Counts

Blood samples were delivered to the lab within 6 hours of being collected, the Complete Blood Count and differential were analyzed by a Beckman-Coulter® T540 blood counter, and the major lymphocyte subsets were analyzed by a Becton Dickinson FACSCalibur™ flow cytometer (Software: SimulSET v3.1). Coefficients of variation for all cell counts were < 10%.

### Methods Used to Culture Progenitor Cells

A portion of the hematopoietic stem and progenitor cells circulate in the blood stream in dynamic equilibrium with the stem cell pools in the bone marrow. These stem and progenitor cells can be cultured in colony-forming assays in semi-solid media containing appropriate growth factors. During the culture period, the progenitor cells establish individual colonies and terminally differentiated cells, such as lymphocytes, die out. The individual colonies can be classified microscopically according to the progenitor cell type. Colonies arising from the most primitive, early progenitor cells are called colony-forming-unit–granulocyte, erythroid, macrophage, megakaryocyte (CFU-GEMM) because the progenitors can give rise to any of these mature cells. Colonies derived from more committed progenitor cells that give rise to reticulocytes and erythrocytes are called burst-forming unit–erythroid (BFU-E), whereas those that give rise to granulocytes and macrophages are called colony-forming unit–granulocyte-macrophage (CFU-GM).

We applied these colony-forming assays to peripheral blood mononuclear cells from 29 benzene-exposed workers and 24 matched controls, which allowed us to examine the dose-dependent effects of benzene on different types of progenitor cell colony formation. Hematopoietic progenitor cells from the peripheral blood were cultured in semi-solid media as follows: 10 ml of heparinized blood was diluted with 2%

fetal bovine serum in Iscove's Modified Dulbecco's Medium (IMDM) and Ficoll-Paque centrifugation performed. The mononuclear cell fraction was isolated, washed and counted. The washed cells were mixed with growth factor-containing methylcellulose media (according to the protocol provided by StemCell Technologies Inc) with and without erythropoietin (EPO) and plated to a final concentration of 300,000 cells per 1.1 ml petri dish. The cells were then cultured for 10-12 days at 37°C in a humidified incubator under 5% CO<sub>2</sub>. In the medium containing EPO colonies of BFU-E, CFU-GM and CFU-GEMM are formed. In the medium without EPO only CFU-GM colonies are formed. The number of each colony type was scored in 6 petri dishes (3 in +EPO; 3 in - EPO), providing data for each subject on the number of colonies formed per 100,000 mononuclear cells plated from the formula:

$$\text{Colonies per 100,000 cells} = \frac{(\text{Mean of 3 dishes}) \times (1.1 \text{ ml / dish})}{(\text{Cell concentration/ml}) / 100,000 \text{ cells/ml}}$$

CFU-GM colonies in –EPO medium are almost twice that found with +EPO medium. The exact reasons for this are unknown, but this phenomenon has been observed in other laboratories and is probably the result of competition for growth factors by BFU-E and CFU-GEMM colonies in the presence of EPO.

#### Genotyping of SNPs in *MPO*, *NQO1* and *CYP2E1*

Genotyping of the four SNPs was performed by real-time PCR on an ABI 7900HT sequence detection system as described on the SNP500 website (<http://snp500cancer.nci.nih.gov>) (S1). Approximately 10% of samples were re-genotyped and there was 99.4% concordance between replicate analyses.

#### **Exposure assessment**

Exposure assessment for study factories has been described in detail previously (S2). The two factories with benzene present were near Tianjin, China and were

selected based on having a stable workforce, a sufficient number of workers in the work force that had been in the particular factory for at least five years, and little or no task rotation. The larger factory was a shoe manufacturing facility characterized by a high level of mechanization with local exhaust ventilation present. A low level of mechanization and no local exhaust ventilation present characterized the smaller facility. Exposure to benzene in both facilities occurred through inhalation of vapors from the use of benzene-containing glues. Benzene contents in these glues ranged from 0.6% to 34%. None of the subjects in either factory used personal protective equipment.

Individual benzene and toluene exposure was monitored by wearing 3M™ organic vapor monitors. Personal full-shift air monitoring took place every 1-2 months over a 16-month period in the larger factory which had lower benzene levels and 5 times in the factory with higher exposures, resulting in the successful collection of 2783 workplace samples. Analyses of the monitors for benzene and toluene were conducted by Gas Chromatography with a Flame Ionization Detector (GC-FID). In addition, exposure to benzene was measured 3 times in control factories and 3 times in the homes of about 85% of study subjects. However, no benzene could be detected in any of these samples. This result was confirmed in duplicate monitors analyzed at a U.S. laboratory, which had somewhat lower limits of detection (LOD) for benzene (0.04 ppm) and toluene (0.16 ppm). In addition, preliminary analyses of dermal exposure data indicated that this route of exposure did not contribute substantially to the total benzene doses received in this population (S2). Post-shift urine samples were collected during the week before phlebotomy and analyzed for benzene by Gas Chromatography with Mass Spectrometry (GC-MS) as previously described (S3). Lifetime cumulative benzene exposure levels were estimated using work history and factory records on production process, ventilation, glue use and shoe production.

Average exposures were calculated for the last month and last year by taking the arithmetic mean of the subject's individual measurements in the last month and year, respectively. All concentrations below the LOD [ $n = 370$  (13.3%), and  $n = 238$  (8.6%) for benzene and toluene, respectively] were replaced with values equal to the LOD divided by the square root of 2 (S4) (LOD = 0.2 ppm and 0.3 ppm for benzene and toluene, respectively). As the individual estimates of benzene exposure in the last month before phlebotomy were based on an average of 2 measurements per subject, we performed additional analyses to study the robustness of the exposure categorization based on these individual estimates. For these analyses, subjects were assigned to their respective job-groups based on the job and the location (e.g. room). Subsequently, a mixed-effect model (S5) was employed to characterize exposure in each job-title, as indicated by the following relationship:

$$Y_{hij} = \mu_h + \chi_{hi} + \varepsilon_{hij},$$

where,

$Y_{hij}$  = the natural logarithm of the exposure concentration measured on the  $j^{\text{th}}$  day of the  $i^{\text{th}}$  worker of the  $h^{\text{th}}$  job title,

$\mu_h$  = the mean (logged) exposure of the  $h^{\text{th}}$  job title,

$\chi_{hi}$  = random effect of the  $i^{\text{th}}$  worker in the  $h^{\text{th}}$  job title,

$\varepsilon_{hij}$  = random error of the  $j^{\text{th}}$  measurement from the  $i^{\text{th}}$  worker in the  $h^{\text{th}}$  job title.

Restricted maximum likelihood estimation (REML) was performed.

Maximum likelihood estimates of each subject's individual mean exposure was

determined as  $\hat{\mu}_{hi} = \exp\left(\hat{\mu}_h + \hat{\chi}_{hi} + \frac{\hat{\sigma}_\varepsilon^2}{2}\right)$ , where  $\hat{\mu}_{hi}$  is the estimated mean exposure of

the  $i^{\text{th}}$  worker in the  $h^{\text{th}}$  job title,  $\hat{\mu}_h$  is the REML estimate of  $\mu_h$ ,  $\hat{\chi}_{hi}$  is the predicted

random effect of the  $i^{\text{th}}$  worker in the  $h^{\text{th}}$  job title, and  $\hat{\sigma}_\varepsilon^2$  is the REML estimate of the

variance of the random error term  $\varepsilon_{hij}$ . The mean sample size for a given job title was 25 personal measurements. Correlation between these maximum likelihood estimates and the individual average exposures was high ( $r = 0.97$ ) and did not change the observed associations with hematological end-points.

The potential for co-exposures to other organic hydrocarbons was investigated as previously described (S2). There were a total of 11 different hydrocarbons detected in the selected samples from the larger factory, of which 9 (e.g. benzene, toluene, pentane, ethyl benzene, hexane, m-xylene, p-xylene, 1,1,1-trichloroethane, and heptane) were detected in over 25% of the samples. In the smaller shoe factory, 16 different hydrocarbons were detected in the selected samples, of which 8 could be detected in over 25% of the samples (e.g. benzene, toluene, pentane, hexane, methyl ethyl ketone, ethyl acetate, acetone and heptane). Exposure levels to the above mentioned hydrocarbons other than benzene and toluene were low in both factories (generally below 5 ppm) and were less than 5% of the current ACGIH (American Conference of Governmental Industrial Hygienists) Threshold Limit Value (TLV). None of the individual hydrocarbon measurements exceeded the respective TLV, except those for benzene. Although the presence of toluene was moderately correlated with benzene ( $r = 0.44$ ) in the shoe-manufacturing workplaces, other solvents were unlikely to have confounded the relationships observed here between benzene and peripheral blood cell counts because these solvents: a) are not considered hematotoxic; b) were present at relatively low levels; and c) did not correlate with benzene exposure.

All study subjects exposed to benzene at  $< 1$  ppm were employed in the larger factory, where they worked on average 8 hours a day, 6 days a week. Potential bias caused by other solvent exposures among these workers was explored further by evaluating a subgroup of 30 workers located in a relatively isolated part of the factory where shoe materials were cut and no direct solvent exposures were present. Exposure

to benzene in this group was low at  $0.29 \pm 0.15$  (mean  $\pm$  SD) and  $0.36 \pm 0.31$  ppm, in the last month and previous year, respectively. Toluene exposure levels were  $0.67 \pm 0.84$ , and  $1.42 \pm 1.96$  ppm in the last month and previous year, respectively. There was negligible exposure to other solvents measured (i.e., detectable levels were present for only pentane and hexane and concentrations were 0.5, and 0.3 ppm, respectively). We further explored if this group could be regarded as a truly low exposed group with average exposures below 1 ppm. None of the benzene measurements collected in the last month for this group was above 1 ppm (Figure S1) and only 25 out of the 337 (7.4%) measurements in the last year were above 1 ppm. Excursions above 1 ppm were minimal ranging from 1.01 to 1.74 ppm with an average of 1.25 ppm. This further corroborates that this sub-group is a truly low-exposed group with average exposures below 1 ppm even when the slight increases in previous months during the winter were accounted for.

### **Statistical Analysis**

Unadjusted summary measures are presented for all endpoints. Linear regression using the natural logarithm (ln) of each endpoint was used to test for differences between workers exposed to <1 ppm benzene in the month prior to phlebotomy and all controls. All analyses were carried out using SAS version 8.0 software (SAS Institute, Cary, North Carolina, USA). Evaluation of potential confounders showed that age (continuous variable), sex, current cigarette smoking status (yes/no), current alcohol consumption (yes/no), recent infections (flu or respiratory infections in the previous month), and BMI (Body Mass Index) were associated with one or more endpoints and therefore all results shown in Table 1, and in the text were adjusted for these variables. Linear trend analyses between continuous endpoints and ln benzene

air level shown in Table 1 and the text were further adjusted for ln toluene air level, which can competitively inhibit benzene's metabolism. Benzene and toluene air levels were based on the arithmetic mean of an average of two measurements per subject collected during the month prior to phlebotomy. Analyses of endpoints in workers exposed to <1 ppm versus controls were not adjusted for toluene, which was collinear with benzene. Linear trend analyses showed that toluene did not decrease cell counts. Findings were negligibly changed with adjustment for other potential confounders, including current medications, medical history and environmental exposures. The same model was used to test the influence of each genotype on WBC in benzene-exposed workers (tables S2, S3). Generalized Estimating Equations (GEE) analysis was used to adjust for the repeat measurements in all models (S6).

Risk [odds ratio with 95% confidence intervals (CI)] of having a WBC count <4,000/ $\mu$ l for each benzene exposure category vs. controls was analyzed by unconditional logistic regression, with adjustment for age, sex, current cigarette smoking status, current alcohol consumption, recent infections, and BMI. Trend analysis was carried out using ln benzene and was not adjusted for toluene, as it was highly correlated with benzene among subjects with a WBC count <4,000/ $\mu$ l.

The 53 subjects analyzed for colony progenitor cells shown in Figure 1 were selected to reflect a wide range of benzene exposure and to be comparable with controls with regard to age, sex and current smoking status. Nineteen workers exposed to benzene had exposures of <10 ppm (mean  $\pm$  SD =  $2.6 \pm 2.6$  ppm) and 10 had exposures  $\geq 10$  ppm ( $24.2 \pm 10.6$  ppm) in the month before phlebotomy. The trend between ln benzene and each endpoint was tested by linear regression for WBC and granulocyte counts, negative binomial regression for CFU-GM and BFU-E, and unconditional logistic regression for CFU-GEMM. All models were adjusted for age and

sex, and additionally for current smoking, current alcohol use, recent infections and BMI if they were statistically significant in the model (i.e., WBC analysis was adjusted for current alcohol use). Results were not adjusted for toluene exposure as it was highly correlated with benzene among this relatively small number of subjects. Analysis of the proportional decrease in white blood cells and granulocytes versus the decrease in colony formation from progenitor cells (Figure 1 legend) was calculated by dividing the subject's cell or colony count by the average count among the controls for that specific end-point. This resets the average response among the controls to 1 for all assays and makes the response among the exposed directly proportional to the response among the controls. As the proportional decrease in the different end-points did not describe a normal distribution, the ln of the ratio was used. In these analyses, the ln of the proportional decrease for each colony count was compared to the ln of the proportional decrease in WBC and granulocytes within exposure category using a paired T-test. Additionally, the non-transformed proportional decreases were compared using the non-parametric paired rank test (Wilcoxon test). Results from these analyses were similar to the results obtained from the paired T-test using the ln of the proportional decrease.

## **SOM TEXT**

### **ADDITIONAL RESULTS**

#### **Further analysis of WBC counts**

We have previously shown that a diagnosis of benzene poisoning (i.e., hematotoxicity) in China, which is a compensable condition based primarily on having a total WBC count  $<4,000/\mu\text{l}$  measured repeatedly over several months, is associated with subsequent development of hematological malignancies (S7, 8). We therefore explored the relationship between benzene exposure and WBC counts dichotomized into  $<$  or  $\geq$

4,000/ $\mu$ l, based on the single measurement available on our study subjects. There was a 3.16 (95% CI: 1.05-9.52), 3.18 (95% CI: 1.06–9.51) and 6.23 (95% CI: 1.76-22.09) fold risk for having a WBC count < 4000/ $\mu$ l for workers exposed to <1 ppm, 1-<10 ppm, and  $\geq$  10 ppm, vs. controls ( $p_{\text{trend}} = 0.0016$ ). Aksoy and Erdem (S9) reported that pancytopenia among Turkish workers exposed to benzene was associated with the development of leukemia. Yin et al. (S8) reported that workers in China with a history of chronic benzene poisoning (severity unknown) were at higher risk of developing leukemia. Rothman et al. (S7) reported that workers in Shanghai with a history of benzene poisoning (severity unknown) were at greater risk of hematologic malignancies and related disorders. Given the relationship between persistent hematotoxicity from benzene and future risk of hematologic malignancy, even the relatively mild hematologic effects observed among workers in the current study could raise concerns about future health risks.

### **Influence of genetic polymorphisms on benzene hematotoxicity**

We explored whether or not a sub-group of workers in the shoe-manufacturing workplaces was more susceptible to the hematotoxic effects of occupational benzene exposure by studying genetic polymorphisms in cytochrome P450 (CYP2E1), myeloperoxidase (MPO) and NAD(P)H:quinone oxidoreductase (NQO1) using validated TaqMan-based genotyping. CYP2E1 catalyzes the first step in benzene metabolism producing benzene oxide which can spontaneously rearrange to phenol (S10). Phenol is then converted to hydroquinone also by CYP2E1. The conversion of phenol and hydroquinone by MPO to toxic quinones and free radicals has been proposed to play a key role in the production of benzene-induced hematotoxicity and leukemia (S11). The G polymorphism at position -463 in the *MPO* gene promoter is associated with normal expression, and therefore potentially an elevated risk of benzene toxicity, compared to

individuals who have the A polymorphism at the same position, which has been associated with reduced expression. A recent study in China showed that the risk of benzene poisoning in those with the GG genotype was greater than in those with the GA + AA genotype (S12), but another study on a different population of Chinese workers showed no such association (S13). In the present study we found that the GG genotype was associated with a greater decline in WBC counts ( $p=0.04$ ) (Table S2), especially in the low exposed workers ( $<1$  ppm) ( $p=0.016$ ). No associations were found with the *CYP2E1* -1053C>T (rs2031920) variant in any category (Table S2). The impact of this mutant allele on CYP2E1 activity is unclear at present as functional studies conflict as to whether the variant raises or lowers CYP2E1 expression (S14).

NQO1 protects the blood and marrow against the toxic quinones and free radicals formed from the phenolic metabolites of benzene (S7, 15). Previously, we have demonstrated an association between the inactivating variant *NQO1* 609C>T allele and risk of benzene poisoning (S7). In the present study, no effect of the variant *NQO1* 609C>T allele on WBC counts was found (Table S2), but the *NQO1* 465CT genotype was associated with a decline in WBC counts in benzene-exposed workers ( $p=0.014$ ) (Table S2). No influence was found of the *NQO1* 465CT genotype on WBC counts in unexposed controls. Evidence of an interaction between benzene exposure and combined *MPO* -463GG and *NQO1* 465CT genotypes, which identify a subgroup of workers who have relatively high MPO capacity and potentially reduced NQO1 function, was observed (Table S3).

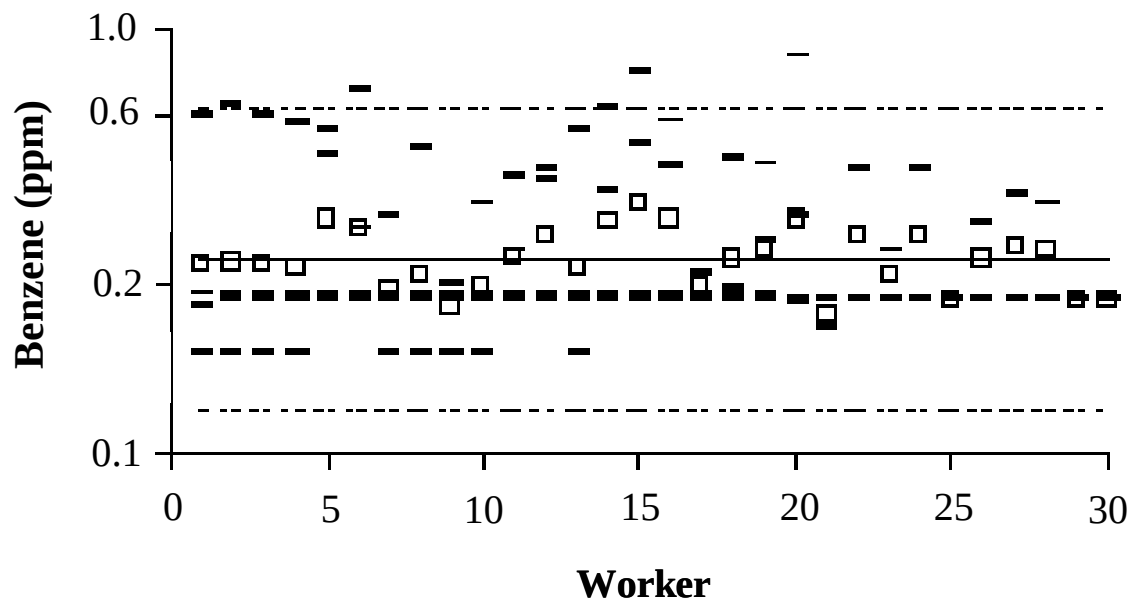
Wacholder et al. (S16) have raised concerns about false positive findings in studies of common polymorphisms and noted that the probability of a statistically significant observation being a false positive is higher in small studies, particularly when studying genetic variants with a relatively low prior probability of being truly associated with the event under study. Even though our study is relatively large compared to

previous genetic studies of workers exposed to benzene, it requires replication in even larger populations.

## REFERENCES

- S1. B. R. Packer *et al.*, *Nucleic Acids Res* **32 Database issue**, D528-32 (2004).
- S2. R. Vermeulen *et al.*, *Ann Occup Hyg* **48**, 105-16 (2004).
- S3. S. Waidyanatha *et al.*, *Carcinogenesis* **22**, 279 (2001).
- S4. R. W. Hornung, L. D. Reed, *Appl Occup Environ Hyg* **5**, 46-51 (1990).
- S5. R. C. Littell, G. A. Milliken, W. W. Stroup, R. D. Wolfinger, *SAS Institute Inc.* (1996).
- S6. S. L. Zeger, K. Y. Liang, *Biometrics* **42**, 121-30 (1986).
- S7. N. Rothman *et al.*, *Cancer Res* **57**, 2839-42 (1997).
- S8. S. N. Yin *et al.*, *Br J Ind Med* **44**, 124-8 (1987).
- S9. M. Aksoy, S. Erdem, *Blood* **52**, 285-92 (1978).
- S10. D. Ross, *Eur J Haematol Suppl* **60**, 111-8 (1996).
- S11. M. T. Smith, J. W. Yager, K. L. Steinmetz, D. A. Eastmond, *Environ Health Perspect* **82**, 23-9 (1989).
- S12. J. N. Xu *et al.*, *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi* **21**, 86-9 (2003).
- S13. J. Wan *et al.*, *Environ Health Perspect* **110**, 1213-8 (2002).
- S14. H. M. Bolt, P. H. Roos, R. Thier, *Int Arch Occup Environ Health* **76**, 174-85 (2003).
- S15. A. K. Bauer *et al.*, *Cancer Res* **63**, 929-35 (2003).
- S16. S. Wacholder, S. Chanock, M. Garcia-Closas, L. El Ghormli, N. Rothman, *J Natl Cancer Inst* **96**, 434-42 (2004).

**Figure S1. Individual exposure measurements (dashes), estimated individual mean values (squares), and group mean estimate (solid line) of benzene exposure in the last month for a sub-group of workers with minimal co-exposures (n=30). The two dashed lines represent the approximate 90% confidence intervals of the group mean estimate.**



**Table S1. Demographic characteristics of study subjects. \***

|                       | <b>Controls<br/>(n=140)</b> | <b>&lt;1 ppm<br/>(n=109)</b> | <b>1-&lt;10 ppm<br/>(n=110)</b> | <b>≥ 10 ppm §<br/>(n=31)</b> |
|-----------------------|-----------------------------|------------------------------|---------------------------------|------------------------------|
| Sex †                 |                             |                              |                                 |                              |
| Male                  | 52 (37)                     | 37 (34)                      | 39 (35)                         | 10 (32)                      |
| Female                | 88 (63)                     | 72 (66)                      | 71 (65)                         | 21 (68)                      |
| Current alcohol use † |                             |                              |                                 |                              |
| Yes                   | 43 (31)                     | 23 (21)                      | 34 (31)                         | 10 (32)                      |
| No                    | 97 (69)                     | 86 (79)                      | 76 (69)                         | 21 (68)                      |
| Recent infection †    |                             |                              |                                 |                              |
| Yes                   | 16 (11)                     | 10 (9)                       | 5 (5)                           | 3 (10)                       |
| No                    | 124 (89)                    | 99 (91)                      | 105 (95)                        | 28 (90)                      |
| Current smoking †     |                             |                              |                                 |                              |
| Yes                   | 39 (28)                     | 20 (18)                      | 25 (23)                         | 7 (23)                       |
| No                    | 101 (72)                    | 89 (82)                      | 85 (77)                         | 24 (77)                      |
| Age ‡                 | 30.34 ± 8.69                | 28.42 ± 7.84                 | 29.27 ± 8.20                    | 34.81 ± 8.09                 |
| Body Mass Index ‡♦    | 22.46 ± 3.93                | 22.14 ± 3.30                 | 22.74 ± 3.28                    | 22.43 ± 2.81                 |

\* For 28 subjects studied in both the first (2000) and second (2001) year of the study, data presented is from information collected in 2000.

§ Exposure categories based on arithmetic mean of an average of two measurements per subject collected during the month before phlebotomy.

† Number (percent)

‡ Mean ± standard deviation

♦ Two controls missing data for Body Mass Index.

**Table S2. Effect on white blood cell (WBC) counts of single nucleotide polymorphisms in *MPO*, *NQO1* and *CYP2E1* in unexposed controls and benzene-exposed subjects.\***

| Genotype                                       |       | N   | WBC §       | P-value | P <sub>trend</sub> |
|------------------------------------------------|-------|-----|-------------|---------|--------------------|
| <b><i>MPO</i> -463G&gt;A (rs2333227) †</b>     |       |     |             |         |                    |
| Control                                        | AA/AG | 38  | 6360 ± 1590 | ref.    |                    |
|                                                | GG    | 102 | 6530 ± 1760 | 0.83    |                    |
| Exposed                                        | AA/AG | 52  | 5870 ± 1340 | ref.    |                    |
|                                                | GG    | 223 | 5410 ± 1340 | 0.04    |                    |
| <b><i>NQO1</i> 465C&gt;T (rs4986998) ‡</b>     |       |     |             |         |                    |
| Control                                        | CC    | 132 | 6500 ± 1750 | ref.    |                    |
|                                                | CT    | 7   | 6100 ± 866  | 0.46    |                    |
| Exposed                                        | CC    | 257 | 5550 ± 1360 | ref.    |                    |
|                                                | CT    | 18  | 4790 ± 903  | 0.014   |                    |
| <b><i>NQO1</i> 609C&gt;T (rs1800566) ◆</b>     |       |     |             |         |                    |
| Control                                        | CC    | 47  | 6370 ± 1810 | ref.    |                    |
|                                                | CT    | 55  | 6370 ± 1510 | 0.55    |                    |
|                                                | TT    | 38  | 6800 ± 1860 | 0.26    | 0.26               |
| Exposed                                        | CC    | 67  | 5320 ± 1170 | ref.    |                    |
|                                                | CT    | 135 | 5570 ± 1340 | 0.34    |                    |
|                                                | TT    | 74  | 5450 ± 1480 | 0.63    | 0.63               |
| <b><i>CYP2E1</i> -1053C&gt;T (rs2031920) ◆</b> |       |     |             |         |                    |
| Control                                        | CC    | 77  | 6640 ± 1580 | ref.    |                    |
|                                                | CT    | 52  | 6370 ± 1980 | 0.42    |                    |
|                                                | TT    | 10  | 6090 ± 1130 | 0.36    | 0.27               |
| Exposed                                        | CC    | 182 | 5560 ± 1430 | ref.    |                    |
|                                                | CT    | 86  | 5400 ± 1220 | 0.26    |                    |
|                                                | TT    | 9   | 4880 ± 595  | 0.23    | 0.15               |

\* Samples were obtained from 28 exposed subjects in both years (2000 and 2001) and are treated as independent observations in summary data shown, distributing into benzene category based on exposure level in the year that the blood sample was collected. Statistical analyses were adjusted for repeated measures by generalized estimating equations (S6). Models were adjusted for age, sex, current smoking, current alcohol drinking, body mass index (BMI), recent infections, and in exposed workers in air benzene exposure and in air toluene exposure in the month before phlebotomy. There are two controls without BMI data and they are excluded from the statistical analysis.

§ Unadjusted total white blood cell (WBC) count as mean ± standard deviation.

† There are 415 observations on 387 unique subjects.

‡ There are 414 observations on 386 unique subjects.

◆ There are 416 observations on 388 unique subjects.

**Table S3. Effect on white blood cell (WBC) counts of the *MPO* -463GG and *NQO1* 465CT genotypes in combination in controls and workers exposed to benzene. \***

|                  | <i>MPO</i> — <i>NQO1</i> genotype | N   | WBC §       | <i>P</i> -value | <i>P</i> <sub>trend</sub> |
|------------------|-----------------------------------|-----|-------------|-----------------|---------------------------|
| <b>Control</b>   | AA/AG—CC                          | 37  | 6380 ± 1610 | ref.            |                           |
|                  | (AA/AG—CT) or (GG—CC)             | 96  | 6530 ± 1800 | 0.83            |                           |
|                  | GG—CT                             | 6   | 6200 ± 903  | 0.63            | 0.94                      |
| <b>Exposed †</b> | AA/AG—CC                          | 46  | 5990 ± 1350 | ref.            |                           |
|                  | (AA/AG—CT) or (GG—CC)             | 215 | 5450 ± 1340 | 0.028           |                           |
|                  | GG—CT                             | 12  | 4690 ± 947  | 0.006           | 0.004                     |

\* There are 412 observations on 384 unique subjects. Samples were obtained from 28 exposed subjects in both years (2000 and 2001) and are treated as independent observations in summary data shown, distributing into benzene category based on exposure level in the year that the blood sample was collected. Statistical analyses were adjusted for repeated measures by generalized estimating equations (S6). Models were adjusted for age, sex, current smoking, current alcohol drinking, body mass index (BMI), recent infections, and in exposed workers in air benzene exposure and in air toluene exposure in the month before phlebotomy. Trend tests carried out using a combined *MPO*—*NQO1* genotype variable coded as 1-3 (with 1 = neither at risk genotype, 2 = one or the other at risk genotype, and 3 = both at risk genotypes) with adjustment as above. A test for interaction ( $p = 0.03$ ) was carried out on all subjects with benzene exposure as a categorical variable, the combined *MPO*—*NQO1* genotype, a cross-product term between them, and adjustment for age, sex, current smoking, current alcohol drinking, body mass index (BMI) and recent infections. There are two controls without BMI data that are excluded from the statistical analysis.

§ Unadjusted total white blood cell (WBC) count as mean ± standard deviation.

† For AA/AG—CC genotype: there are 43 unique individuals, with repeat observations on 3. For AA/AG—CT or GG—CC genotypes: there are 191 unique individuals, with repeat observations on 24. For GG—CT genotype: there are 11 unique individuals and a repeat observation on one.