**Supplemental Material**

**Hospital-Based Case-Control Study of MDS Subtypes and Benzene Exposure in Shanghai**

This supplemental material contains more detail on the background, methods and statistical analysis used in the main paper. It also contains results of other biologically-motivated alternative groupings of different MDS subtypes.

**Background**

This current study is part of the Shanghai Health Study which was conceived with three arms designed to accomplish the following: (1) evaluate the relation between benzene exposure and specific lymphohematopoietic cancers via the case-control study design; (2), determine pathologic precursors to benzene-induced disease (mode of action) through disease progression studies; and, (3) conduct molecular epidemiology studies to determine the level of exposure at which blood counts are affected. This study falls under the case-control arm and assesses the benzene-related risk for MDS and its subtypes.

Direct examination of the bone marrow is required to distinguish MDS from aplastic anemia (AA) and, in some cases, leukemias or myeloproliferative diseases. Prior to the 1970s the differential diagnosis of MDS was severely hampered by the absence of a non-surgical biopsy method for direct analysis of bone marrow (Jamshidi et al. 1971). Thus, the recognition of MDS as a discrete disease entity emerged slowly over the past century with vague and imprecise terminology often used to describe cases of MDS, examples of which included “aplastic anemia,” “preleukemia,” “sub-acute leukemia,” and “atypical leukemia”, among others. All of these are now regarded as distinct disease entities on the basis of morphologic, clinical and molecular data. During the first half of the 20th century, this ambiguity contributed to a lack of understanding of the basic nature of bone marrow failure occurring in individuals chronically exposed to benzene. These were commonly described as pancytopenia, aplastic anemia, granulocytopenia, thrombocytopenia or anemia, while the concept of the myelodysplastic syndromes remained obscure. Given the benefit of hindsight, many cases of bone marrow failure/MDS were likely misclassified, and subsequently reported as AA. Thus, it is not surprising that the epidemiology of MDS is only just emerging, as sufficiently-sized populations need to be tracked over time and the disease diagnosed reliably as was accomplished in this study.

**Methods**

*Case classification*

All MDS cases with a diagnosis that differed when classified according to the 2001 or 2008 WHO criteria were re-evaluated using clinical records and pathology reports using the 2008 WHO criteria. In addition, all cases of refractory anemia with excess blasts, or RAEB) were re-evaluated to: (a) determine if a preceding diagnostic entity could be identified, since RAEB is an advanced form of MDS, and (b) classify RAEBs into subtypes I and II, the latter being a more severe/advanced form of MDS. Results of the reclassification by WHO classification guidelines are summarized in Table S1. The reevaluation resulted in a decrease in the number of Myelodyspastic Syndrome-Unspecified (MDS-U) and Refractory Cytopenia with Unilineage Dysplasia (RCUD) cases, and a slight increase in the number of RCMD cases.

**Table S1.**  MDS cases by WHO sub-type

|  |  |  |  |
| --- | --- | --- | --- |
| **MDS subtype** | **WHO 2001  (n)** | **WHO 2008  (n)** | **Updated diagnosis (n)** |
| Refractory cytopenia with multilineage dysplasia (RCMD) | 419 | 428 | 433 |
| Refractory anemia with excess blasts (RAEB)a | 105 | 105 | 103 |
| Refractory anemia (RA) | 37 | 37 | 38 |
| MDS-unspecified (MDS-U) | -- | 29 | 5 |
| Refractory cytopenia with unilineage dysplasia (RCUD) | -- | 24 | 16 |
| Refractory anemia with ringed sideroblasts (RARS) | 6 | 6 | 7 |
| MDS associated with isolated del(5q) (MDS w/ 5q-) | -- | 2 | 2 |
| Total MDS | 567 | 631 | 604 |
| Not MDS (for 2001 includes cytopenia and HDIES)b | 64 |  | 27 |

a For RAEB, there were 38 subtype I, and 63 with subtype II

b Hematopoietic dysplasia of indeterminate etiologic significance

In addition to examining the diagnostic subtypes, there is also a biologic (or an historic) rationale for grouping specific MDS subtypes as follows: (a) refractory anemia (RA) + refractory anemia with ringed sideroblasts (RARS) are both “low-risk” refractory anemias, i.e. not in a stage prone to advancing to AML; (b) RA + RARS + refractory anemia with excess blasts (RAEB) represents all refractory anemias, both low risk and high risk; and (c) refractory cytopenia with multi-lineage dysplasia (RCMD) + MDS-Unspecified (MDS-U), both of which involve pathologies of multiple lineage. In addition, we considered whether a “source diagnosis” (i.e., a previous MDS diagnosis that differed from the final diagnosis) was present that indicated progression from one subtype to another. Table S2 shows these groupings that were included in the analytical scheme.

**Table S2.**  Biologically-Based MDS Groupings

|  |  |  |  |
| --- | --- | --- | --- |
| **MDS Grouping** | **Definition** | **No. of  Cases (%)** | **No. of Controls (%)** |
| All inclusive RA | RA (39) + RA with Ringed Sideroblasts (RARS) (7) + all RAEB with a source diagnosisa of RA (40) or RARS (1) | 87 (14.4) | 170 (14.2) |
| Non-advanced RA | RA (39) + RARS (7) | 46 (7.6) | 91 (7.6) |
| Advanced RA | RAEBs w/ source diagnosis of RA (40) or RARS (1) | 41 (6.8) | 79 (6.6) |
| All inclusive multi-lineage MDS | RCMD (433) + MDS-Unspecified (MDS-U) (5) + all RAEB with source at diagnosis = RCMD or MDS-U (32) | 470 (77.8) | 929 (77.8) |
| Advanced RCMD | all RAEBs with source diagnosis of RCMD | 32 (5.3) | 62 (5.2) |
| Very advanced RCMD | all RAEB-II’s with source diagnosis of RCMD | 19 (3.1) | 37 (3.1) |

a A “source diagnosis” consisted of a previous MDS diagnosis which differed from the final diagnosis, implying progression from one subtype to another.

*Statistical Analysis*

The study team also employed other extensive quality assurance checks and preliminary analyses, e.g., checking that data on study subjects included all relevant data fields and assessing the number of cases and/or controls that needed to be excluded. Study data were assessed in several frequency tables to assess (a) the validity of data on potential confounders (e.g. by conducting nonsense checks) and (b) to ensure that the data were consistent with what is known about this cohort (date cross-checks, smoking rates by gender).

Phase 1, bivariable crude analysis (see Table S-2) identified important demographic, non-benzene exposure and lifestyle variables to be used in the second phase multivariable analyses which examined the independent effect of benzene exposure and other environmental factors on MDS. Odds ratios (ORs) with 95 percent confidence intervals (95%CI) and probability values (*p*-values) were computed for each variable. Additionally, Phase 1 permitted the determination of the most statistically efficient and biologically consistent manner in which to parameterize these variables for Phase 2. Subjects with missing values for a specific variable were excluded from the analyses.

Phase 2 entailed statistical modeling via conditional multivariable logistic regression allowing for a variable case-control ratio. The matching variables, age and gender, were considered strata. Case/control status (where cases were defined as either MDS or an MDS subgroup) was included as the dependent variable, while benzene and potential confounders were modelled as independent predictor variables. Parsimonious models were identified using the PROC LOGISTIC procedure in SAS version 9.3 (Cary NC), using a forward-stepwise procedure to qualify a variable’s entry and retention into the final (sub)group-specific multivariable models. The best-fitting models were selected by monitoring the value Akaike’s Information Criterion (AIC), while entering and removing variables using a *p*-value criterion of ≤0.10 for entry and retention of specific variables. Benzene EGs were forced into the model unless convergence could not be attained. In those situations, a binary ever/never benzene (any benzene exposure) covariate was substituted. ORs and 95%CIs for the semi-quantitative EGs and the selected covariates were computed. Individuals with missing data for the qualifying covariates were excluded from the models.

**Results**

**All MDS Subtypes (combined)**

**Table S3**. Results of bivariable unadjusted (crude) analyses on all (combined) subtypes of MDS, with odds ratios (OR) and 95 percent confidence intervals (95% CI).

|  |  |  |  |
| --- | --- | --- | --- |
| **Exposures / Variables** | **No. Cases** | **No. Controls** | **Crude Odds Ratio (95% CI)** |
| Benzene average exposure, ppm:  EG0: Unexposed  EG1: <0.3  EG2: 0.3-2.9  EG3: 3.0-11.9  EG4: 12.0 - | 531  22  14  11  26 | 1126  37  16  3  11 | [reference]  1.23 (0.71, 2.11)  1.99 (0.96, 4.13)  7.81 (2.17, 28.1)  4.81 (2.38, 9.74) |
| BMI (continuous)\* | 298 | 606 | 0.97 (0.95, 1.00) |
| Education:  College/post-grad  Middle/high school  None or primary | 96  314  192 | 286  645  258 | [reference]  1.51 (1.14, 2.00)  2.70 (1.94, 3.77) |
|  |  |  |  |
| Farm residence | 290 | 405 | 1.98 (1.59, 2.47) |
| Diesel fuel | 7 | 23 | 0.56 (0.23, 1.33) |
| Metal machining (job) | 25 | 76 | 0.61 (0.38, 0.98) |
| Anti-TB medication | 29 | 44 | 1.32 (0.82, 2.12) |
| Diabetes | 39 | 142 | 0.49 (0.33, 0.72) |
| Welding (job) | 11 | 34 | 0.64 (0.32, 1.27) |
| Smoking  Crops, raised  Livestock, raised  Fertilizer (occupational)  Herbicides (occupational)  Insecticides (occupation)  Organophosphate (occ)  Metals (occupational) | 191  154  75  91  27  32  68  31 | 375  219  91  120  19  39  90  81 | 1.00 (0.76, 1.33)  1.69 (1.30, 2.21)  1.88 (1.32, 2.67)  1.70 (1.24, 2.33)  3.03 (1.64, 5.57)  1.70 (1.03, 2.81)  1.68 (1.18, 2.41)  0.71 (0.46, 1.11) |

\*Number of cases and controls represent number of patients at or above the BMI median value (22.49).

**All-inclusive multi-lineage MDS**

This grouping includes RCMD and MDS-U as primary diagnoses that involve multiple lineages, and all RAEBs with a source diagnosis of RCMD. The overall exposure-response pattern was similar to that of aggregate MDS and particularly RCMD, since 92% of these cases are RCMD. Given that the exposure-response was stronger for RCMD alone, the addition of other forms of multi-lineage dysplasia to the analysis did not generate any additional material findings.

**Table S4.** Results of multivariate modeling for all-inclusive multi-lineage MDS, crude and adjusted odds ratios and 95 percent confidence intervals

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Exposure** | **No. Cases** | **No. Controls** |  | **Odds Ratio** | **95% CI** |  | **Odds Ratio** | **95% CI** |
|  | *Crude* | |  | *Adjusted* | |
| Benzene exposure, ppm:  Unexposed  <.3  0.3-2.9  3.0-12.0  12.0 - | 414  15  12  9  20 | 875  29  12  3  10 |  | [reference]  1.08  2.30  6.58  4.08 | -  0.57-2.03  1.02-5.22  1.76-24.5  1.91-8.73 |  | [reference]  1.53  2.61  4.59  3.40 | -  0.78-2.97  1.08-6.28  1.21-17.5  1.53-7.56 |
| Farm residence | 225 | 315 |  | 1.93 | 1.51-2.46 |  | 1.78 | 1.37-2.31 |
| Herbicides | 23 | 16 |  | 3.10 | 1.59-6.05 |  | 2.14 | 1.05-4.35 |
| Metal machining | 16 | 59 |  | 0.49 | 0.27-0.88 |  | 0.50 | 0.27-0.93 |
| Diabetes | 31 | 106 |  | 0.53 | 0.34-0.81 |  | 0.57 | 0.36-0.89 |
| Anti-tuberculosis medication | 24 | 31 |  | 1.54 | 0.90-2.64 |  | 1.89 | 1.05-3.39 |

**RCUD**

Results for the total unilineage dysplasia --a combined category of refractory anemia (RA, 39 cases) and refractory cytopenia with unilineage dysplasia (RCUD) other than RA (15 cases)--are presented in Table S4. The multivariable model did not converge for the EG3 and EG4 parameters since there were only seven cases and no controls in the two EGs combined, necessitating the need to further aggregate the EGs into just two EGs (<.3 ppm and 0.3 ppm and above). The aggregated exposure group of 0.3 ppm and above showed an adjusted OR exceeding 15.0 accompanied by a wide 95% CI (1.40-178). When examined further by subtype, the suggestive risk for total unilineage dysplasia did not result from the more common diagnosis of RA (Table S5), as there were only 2 cases and 3 controls exposed above background, precluding a meaningful analysis. Farm residence showed a strong positive statistical relationship with total RCUD, OR = 16.1; 95% CI 3.21-80.2.

**Table S5.** Results of multivariable modeling for total RCUD

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Exposure** | **No. Cases** | **No. Controls** |  | **Odds Ratio** | **95% CI** |  | **Odds Ratio** | **95% CI** |
|  | *Crude* | |  | *Adjusted* | |
| Benzene exposure, ppm:  Unexposed  <.3  0.3- | 45  1  8 | 104  2  1 |  | [reference]  1.00  16.00 | -  0.09-11.0  2.00-128 |  | [reference]  1.34  15.7 | -  0.09-20.1  1.40-178 |
| Farm Residence | 36 | 37 |  | 10.1 | 2.99-34.0 |  | 16.1 | 3.21-80.2 |
| Insecticides | 3 | 7 |  | 0.78 | 0.19-3.22 |  | 0.06 | 0.01-0.83 |

**RA**

Regarding RA, the further breakdown of non-advanced and advanced stages did not provide a sufficient number of study participants to facilitate further analysis. The results of multivariable modeling for all-inclusive RA likewise did not provide useful information.

**Table S6.** Results of multivariable modeling for RA

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Exposure** | **No. Cases** | **No. Controls** |  | **Odds Ratio** | **95% CI** |  | **Odds Ratio** | **95% CI** |
|  | Crude | |  | Adjusted | |
| Benzene exposure  Unexposed  >0 | 37  2 | 74  3 |  | [reference]  1.33 | -  0.22-7.98 |  | [reference]  2.15 | -  0.24-19.1 |
| Farm Residence | 24 | 23 |  | 10.7 | 2.42-47.0 |  | 21.1 | 2.61-171 |
| Insecticides | 2 | 5 |  | 0.72 | 0.14-3.74 |  | 0.05 | .002-1.02 |

**Reference**

Jamshidi K, Windshiti HE, Swaim WR. 1971. A new bone marrow biopsy needle is described. Scand J Haematol 8:69-71.