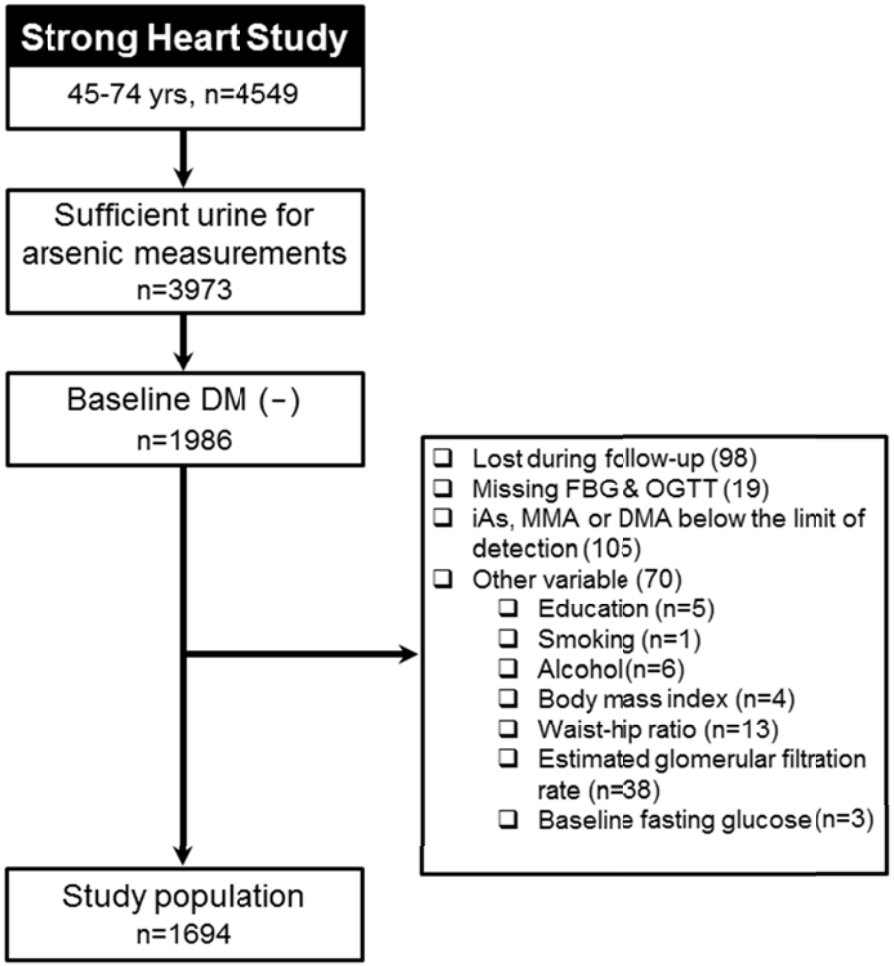
SUPPLEMENTARY DATA

Supplementary Figure 1. **Flow chart of participant selection.**



iAs, inorganic arsenic; DMA, dymethylarsinate; FBG, fasting blood glucose; MMA, monomethylarsonate; OGTT, oral glucose tolerance test.

Supplementary Figure 2. The mixture of generalized gamma distributions summarized the cumulative incidence of both diabetes and death in participants with highest and lowest quartile of arsenic exposure.

Brief Methods

The main competing event in our study was death. A total of 97 women and 128 men without developing diabetes died between 1989-1991 and December 31, 1999. Uncensored observations correspond to either the time when a participant developed diabetes or to the time when a participant died during follow-up. If π is the proportion of the total population of participants who developed diabetes and 1– π is the complementary proportion of patients who died without developing diabetes, we used a mixture according to π and 1- π of two generalized gamma distributions to model time to diabetes development and time to death.[1] We used the 3-parameter generalized gamma distribution (β for location, σ for scale, and **κ** for shape) for this analysis because of its flexibility to accommodate various hazard patterns. We modeled time to diabetes development with a generalized gamma distribution with density *f(t)*, and the times to death with another generalized gamma distribution with density *g(t)*. Hence, if T denotes the time to either diabetes development or death, the proportion with T < t is given by:

$$\Pr(T < t) = \Pr(T < t, event = diabetes) + Pr\ (T < t, event = death) = \ \pi\ [1 - F(t)] + (1 - \pi)\ [1 - G(t)]$$

where F and G were the survival functions corresponding to the *f* and *g* densities, respectively. $\pi\ [1 - F(t)]$ is the cumulative incidence of diabetes and $(1 - \pi)\ [1 - G(t)]$ is the cumulative incidence of death. The detailed analysis procedure has been described.[2] Publicly available algorithms at the Johns Hopkins STATPEI website (www.statepi.jhsph.edu) facilitated the development of the maximal likelihood function to fit these types of mixture models.

Competing risk graph using generalized gamma model – Sum of iAs and methylated As

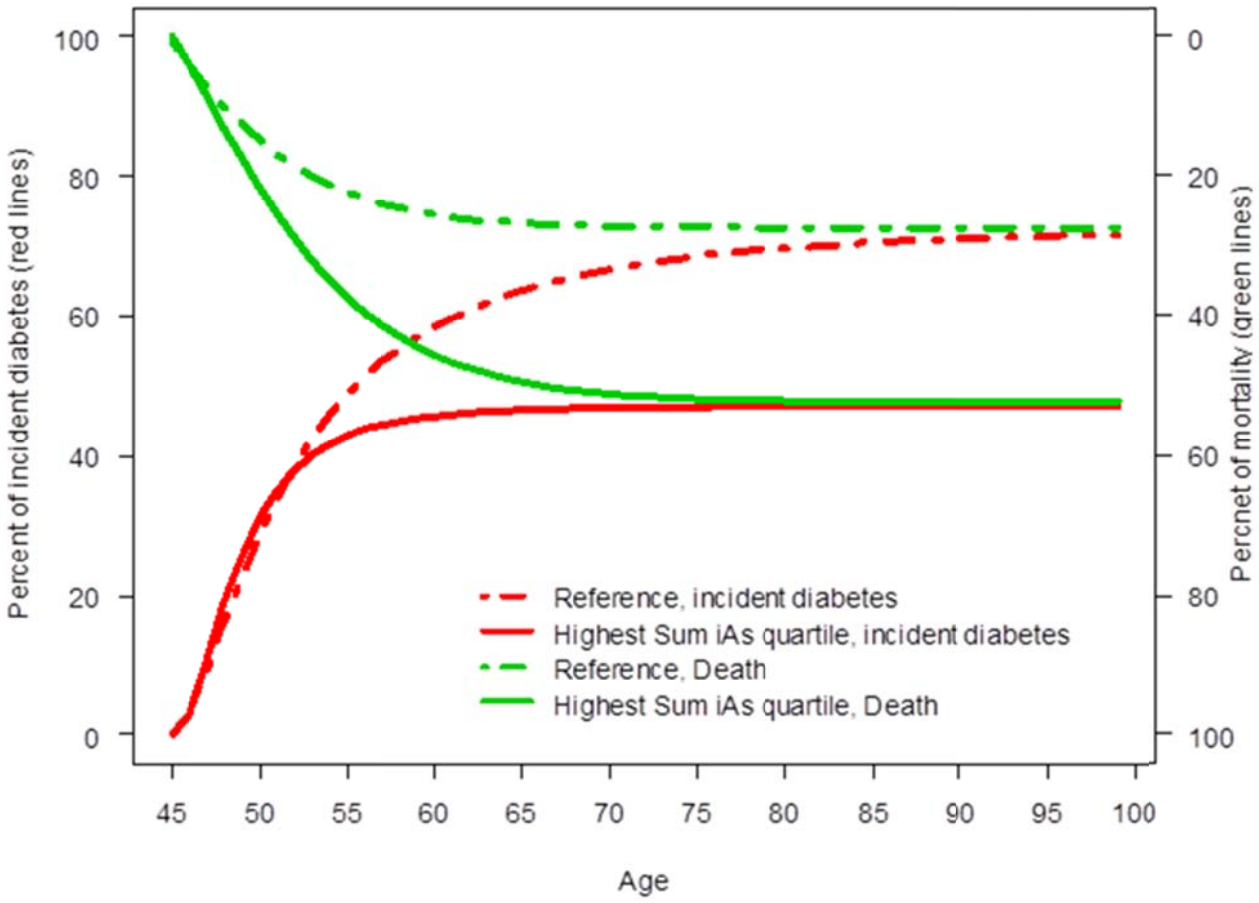
Full Saturated Model		**Incident diabetes**			**Death**			
	%	β_1	σ_1	κ_1	β_2	σ_2	κ_2	AIC
Reference group (Quartile 1)	0.727	2.60	0.95	0.0001	2.56	0.87	1	5506.978
Arsenic (Quartile 2)	0.645	2.36	0.92	0.0001	2.65	0.60	1	
Arsenic (Quartile 3)	0.809	2.55	1.00	0.0001	1.90	0.64	1	
Arsenic (Quartile 4)	0.471	2.05	0.72	0.0001	2.83	0.80	1	

Reference:

1. Cox C, Chu H, Schneider MF, Munoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in medicine*. 2007;26:4352-4374

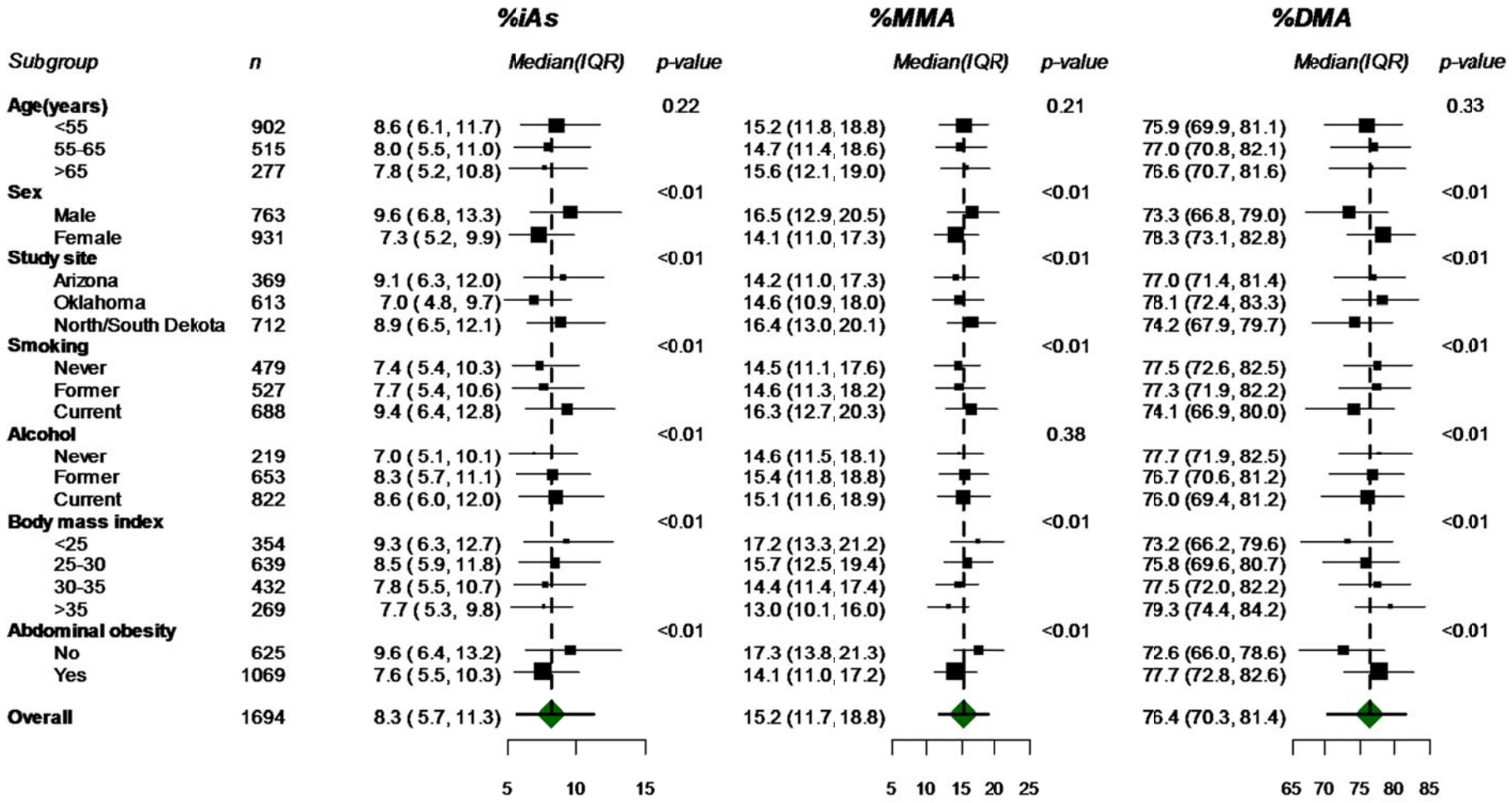
2. Checkley W, Brower RG, Munoz A, Investigators NIHARDSN. Inference for mutually exclusive competing events through a mixture of generalized gamma distributions. *Epidemiology*. 2010;21:557-565

SUPPLEMENTARY DATA



SUPPLEMENTARY DATA

Supplementary Figure 3. **Median (IQR) of arsenic metabolism biomarkers by participant characteristics.** The squares reflect the median for each arsenic metabolism biomarker and are proportional to sample size. The horizontal lines represent the interquartile range. The dashed vertical lines represent the overall median.



Supplementary Figure 4. Hazard ratios for incident diabetes by urine arsenic concentrations. Solid lines represent adjusted hazard ratios based on restricted quadratic splines for the log-transformed sum of inorganic and methylated arsenic species, with knots at the 10th, 50th, and 90th percentiles (3.8, 8.8, and 21.7 μg/g creatinine, respectively). The dotted lines represent upper and lower 95% CIs. The reference was set at the 10th percentile of the arsenic distribution (3.8 μg/g creatinine).