Published in final edited form as:

Cancer. 2014 October 15; 120(20): 3262–3263. doi:10.1002/cncr.28868.

Response to Limitations in the Imputation Strategy to Handle Missing Nativity Data in SEER

Renke Zhou, M.P.H., MS^{1,3}, Jane R. Montealegre, Ph.D.^{2,3}, E. Susan Amirian, Ph.D.^{3,4}, and Michael E. Scheurer, Ph.D., M.P.H.^{3,4}

¹Division of Biostatistics, The University of Texas School of Public Health, Houston, Texas

²Division of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas School of Public Health, Houston, Texas

³Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas

⁴Department of Pediatrics, Baylor College of Medicine, Houston, Texas

We appreciate the comments from Pineda et al., and agree with their statement that the distribution of missing birthplace is not at random and is related to vital status. While they are correct in noting a large discrepancy in the proportion of U.S.- and foreign-born individuals when comparing the imputed data to cases whose birthplace was unknown in 2006 but became known in 2012, we would like to comment on two issues: 1) when applying the exclusion criteria specified in our original analysis, the proportion of cases with updated nativity is small, representing 11% of cases with missing birthplace and < 5% of total cases; and 2) when applying the exclusion criteria, the nativity distribution of cases with updated nativity is quite different from what they present.

We requested a custom data linkage of Hispanic cases diagnosed from 1995–2004 who were reported in 2006 and 2012. For cervix cancer, there were 98 cases for whom nativity was unknown in 2006 but became known in 2012, representing 6% of 1,636 cases with missing nativity and < 2% of total cases. Among these, 53% were found to be U.S.-born and 47% were foreign-born, a distribution overall not different from the 1:1 ratio expected if death certificate reporting is independent of birthplace. For prostate cancer, nativity became known for 1,229 cases, < 10% of cases with missing data and < 5% of total cases. Similarly, the observed distribution reflected independence between reporting and birthplace (53% U.S.-born). Only for colorectal cancer was the observed distribution different from what may be expected by chance, with U.S.-borns accounting for 66% of cases with updated data.

Corresponding author: Michael E. Scheurer, One Baylor Plaza, MS: BCM 305, Houston, Texas 77030, Tel: 713-798-5547; Fax: 713-798-8711, scheurer@bcm.edu.

Department of Pediatrics and Dan L. Duncan Cancer Center Baylor College of Medicine, Houston, TX

Financial Disclosures: The authors do not have any financial disclosures.

Funding Disclosures: This research is partly supported by a grant from the National Institutes of Health (Grant P01CA082710, PI: M. Follen). JRM and RZ are supported by a UTHealth Innovation for Cancer Prevention Research Postdoctoral (JRM)/Predoctoral (RZ) Fellowship (The University of Texas School of Public Health Cancer Prevention and Research Institute of Texas grant # RP101503). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Cancer Prevention and Research Institute of Texas.

However, these 671 cases still represent 11% of cases with missing data and only 4% of total cases.

We did not observe significant changes in the survival curves presented in our original analyses with the addition of the updated data (affecting < 5% of cases). While birthplace will eventually (and inevitably, due to death certificate reporting) become known for all cases, there is no reason to suspect that the data will follow the same distribution as the cases for whom birthplace became known (11%). Given the need for timely analyses of the public SEER dataset, we argue that multiple imputation of missing birthplace is an improved approach over listwise deletion, the usual method for handling missing birthplace data in SEER which would exclude > 30% of the original sample.