

## Supplementary Material

### **Factors associated with self-reported anal cancer screening history in men who have sex with men**

*Joseph T. Hicks<sup>A,I</sup>, Lu-Yu Hwang<sup>B</sup>, Sarah Baraniuk<sup>C</sup>, Margaret White<sup>D</sup>, Elizabeth Y. Chiao<sup>E</sup>, Nkechi Onwuka<sup>F</sup>, Michael W. Ross<sup>G</sup> and Alan G. Nyitray<sup>H</sup>*

<sup>A</sup>Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, 501 D.W. Brooks Drive, Athens, GA 30602, USA.

<sup>B</sup>Center for Infectious Diseases, Division of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas School of Public Health at Houston, 1200 Pressler Street, Houston, TX 77030, USA.

<sup>C</sup>Coordinating Center for Clinical Trials, Department of Biostatistics, The University of Texas School of Public Health at Houston, 1200 Pressler Street, Houston, TX 77030, USA.

<sup>D</sup>Gordon E. Crofoot MD, PA, 3701 Kirby Drive, Suite 1230, Houston, TX 77098, USA.

<sup>E</sup>Department of Internal Medicine, Baylor College of Medicine, 6620 Main Street, Houston, TX 77030, USA.

<sup>F</sup>Division of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas School of Public Health at Houston, 1200 Pressler Street, Houston, TX 77030, USA.

<sup>G</sup>University of Minnesota, Program in Human Sexuality, 1300 South 2nd Street, Suite 180, Minneapolis, MN 55454, USA.

<sup>H</sup>Department of Psychiatry and Behavioral Medicine, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA.

<sup>I</sup>Corresponding author. Email: joseph.hicks@uga.edu

### *Data ascertainment*

Participants' anal cancer screening history, anal cancer awareness and demographic information were gathered in three ways. First, previous digital anorectal examination (DARE) history was obtained as part of eligibility assessment and study enrollment confirmation. Second, a written pre-test questionnaire was administered with questions regarding demographic information. In this pre-test, participants were also asked whether they had ever received anal cancer screening. Third, after education and self- (or partner) exam, participants completed a computer-assisted self-interview (CASI) (Qualtrics, Provo, Utah) that included questions on health, hygiene, and sexual behaviors. During the CASI, participants were also asked about their self-perceived level of knowledge about anal cancer "before today," their anxiety about anal cancer, whether they have a friend or family member with anal cancer, and whether they are aware that there is a vaccine that prevents HPV infection.

### *Logistic Regression*

The primary outcome of this analysis was the presence or absence of any lifetime anal cancer screening, defined as either DARE or anal Pap cytology at any point in the participant's life. Purposeful model selection was used to identify important covariates for the multiple logistic regression model. Demographic covariates that were assessed included age (continuous), race (black, white, other), Hispanic ethnicity (yes, no), education level (less than high school graduate, high school graduate/some college, college graduate and beyond), relationship status (single, partnered, widowed/divorced), health insurance coverage (yes, no), and body mass index (normal/underweight, overweight, obese). In addition, variables concerning the participants' awareness and knowledge of anal cancer and human papillomavirus were also assessed. These included how much the participant reported he knew about anal cancer before participating (none, little, some/a lot), how much the participant reported worrying about anal cancer (none, little, some, a lot), if the participant said he knew any friends or family that had been diagnosed with anal cancer (yes, no), and whether the patient was aware of the HPV vaccine (yes, no).

Covariates with important clinical significance or  $p$ -value  $< 0.25$  on univariable analysis were entered into the initial fitting of the multiple logistic regression model. During the selection process, a likelihood ratio test  $p$ -value threshold of 0.1 was used for a covariate's retention in the model. This relaxed threshold was used to allow the inclusion of variables that, though not statistically significant by the conventional 0.05 standard, may still affect the model estimates. However, when considering statistical significance in the final model, an alpha level of 0.05 was used. Once a variable was removed, the model was re-fit and the significance of each remaining variable was re-evaluated. Once all statistically important variables were identified, variables originally excluded based on univariable analysis were included into the model one at a time to re-assess their importance. Linearity of age was assessed by testing the statistical significance of a model with a cubic spline with three knots. Age was found to uphold the linearity assumption ( $p = 0.26$ ). The overall fit of the final model was examined with the Hosmer-Lemeshow test and indicated a good fit ( $p = 0.34$ ). All statistical analyses were performed using SAS software, Version 9.4 of the SAS System for Windows (Copyright © 2002-2012 SAS Institute Inc., Cary, NC, USA).