Supplementary Data

APPENDIX

COUNTRIES AND TERRITORIES WITH A TRAVEL ADVISORY ISSUED BY THE CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC) DUE TO REPORTED LOCAL MOSQUITO-BORNE TRANSMISSION OF ZIKA VIRUS THROUGH MARCH 1, 2016

American Samoa
Aruba
Barbados
Bolivia
Bonaire
Brazil
Cape Verde
Colombia
Costa Rica
Curaçao
Dominican Republic
Ecuador
El Salvador
French Guiana
Guadeloupe
Guatemala
Guyana
Haiti
Honduras

Jamaica
Marshall Islands
Martinique
Mexico
New Caledonia
Nicaragua
Panama
Paraguay
Puerto Rico
Saint Martin
Saint Vincent and the Grenadines
Samoa
Sint Maarten
Suriname
Tonga
Trinidad and Tobago
US Virgin Islands
Venezuela

Respondent-level checks for inclusion in sample

To help ensure the validity of findings, a series of respondent-level validation checks were done before data from each respondent was included in the sample.

Eliminating Duplicates and Gaming-Respondents. There are not generally concerns about respondent-level validity from randomized or probability-based samples. Respondents are not self-referred and thus have few ways to game the system and complete more surveys. However, there may be concern about such gaming from respondents sourced from opt-in panels, such as the supplementary sample used in this study. We used tactics during the survey recruiting, invitation, and completion process to address this.

First, the panel companies selected to provide sources for the study regularly take steps to validate respondents – that is, to ensure that they are who they say they are and that they are unable to complete more than one survey. The first step in this process is a double opt-in, meaning that to become a panel member a respondent needs to register at the website and then click on a confirmation link sent to their email address. The panel companies also take addi-tional steps including eliminating overused IP addresses, matching IP address to the respondent's self-reported lo-cation, and de-duplicating email addresses. The invitation for the survey is then matched to the email address so that there cannot be more than one invitation to a given email address. Finally, once respondents complete a survey, SSRS uses a cookie to block respondents from completing additional surveys from the same device. This last step was done for opt-in panel respondents and for respondents recruited through random sampling.

Eliminating Unqualified Respondents. SSRS did a review of open-ended answers to confirm recruiting qualifications. A few respondents answered the screener in a way that indicated they qualified for the study, but open-ended responses to subsequent questions revealed that they did not. For example, respondents who said they did not protect themselves because their travel took place 20 years prior clearly were not answering about travel that took place in the timeframe that is covered by the poll. Data from these respondents were not included in the sample.

<u>Eliminating Respondents Who Provide Responses That May Not Be Valid</u>. SSRS did additional checks for respondents who were apparently providing invalid responses to questions. Two checks included:

(1) Review of straightline responses: Respondents who answer a sequence of questions with the same response code are considered straightliners and can be indicative of poor-quality responses [Zhang C, Conrad FG. Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Surv Res Methods*. 2014;8(2):127-35]. The risk for this type of respondent behavior is highest where a relatively long series of questions with identical response options is asked, particularly when presented in a grid format. This was most likely to happen in the series of questions that asked for the various causes and symptoms of Zika virus. In deciding to remove cases, caution needs to be taken to ensure that data are not removed without cause – meaning the plausibility of seemingly-straightlined responses needs to be considered. To that end, cases were removed if: (1) a respondent stated she or he knew something about Zika virus; and (2) responded "no" to all possible causes and symptoms of Zika virus asked at the beginning of the poll. This combination, especially when explicitly offered a Don't Know option, seemed wholly unlikely and thus such respondents are likely to have provided poor quality responses. Data from these respondents were not included in the sample.

(2) Review of time for completion: The Zika poll required respondents – specifically those who said they knew at least something about the Zika virus – to answer approximately 50 questions. The mean length of taking the poll was estimated at 18 minutes. The most extreme respondents (1%) answered the poll in less than 5 minutes. Data from these respondents were not included in the sample.

Weighting parameter details
Details on the parameter variables used for weighting are below.
Demographic Variable Parameters:

- Gender (Male, Female) by Age (18-29, 30-44, 45-59, 60+)
- Race/Ethnicity (White/Non-Hispanic, Black/Non-Hispanic, Other/Non-Hispanic, Hispanic, 2+ Races/Non-Hispanic)
- Census Region (Northeast, Midwest, South, West) by Metropolitan Status (Metro/Non-Metro)
- Education (Less than High School, High School, Some College, Bachelor or higher)
- Household Income (Under $25K, $25K-$49,999, $50K-$74,999, $75K-$99,999, $100K-$149,999, $150K and over)

Online Survey Participation Variable Parameters:

- Watch TV (< 3 hours/day, 3+ hours/day)
- Internet for Personal Use (< 10 hours/week, 10+ hours/week)
- Express Political/Community Opinions Online (Less than once a month or more often, Not at all)
- Usually Try New Products (Statement describes me not at all/somewhat, a lot/completely)

Comparisons to RDD sample. As added measures of data validity and reasonableness, the demographic profile and destinations of the weighted Zika poll sample was compared to that of an independent, random sample of people screened concurrently through random digit dial (RDD). The purpose of the comparison was to test whether the sampling method in the Zika poll (combining two probability samples with non-probability sources) and mode of data-collection (online) yielded results that differ meaningfully from a RDD, full-probability sample.

The independent sample comes from the SSRS Omnibus, a telephone poll conducted using RDD over a five-day period involving interviews with 1,000 respondents reached by cell phones and landline telephones. Interviews on the SSRS Omnibus poll were conducted in English and Spanish. The data were weighted to match the sample of respondents to known population parameters based on the most recent supplement of U.S. Census Bureau's Current Population Survey (March 2016), as well as estimates of phone use (i.e., cell phone only, landline only, or both phone types) based on the most recent estimates from the CDC's National Health Interview Survey (June- Dec 2015).

Over three waves of the SSRS Omnibus poll, respondents were asked identical screening question as respondents to the Zika poll. In total, 277 respondents answered the questions in a way parallel to those qualified for the Zika poll. These "RDD respondents" were then compared to the Zika poll respondents with respect to demographics and travel destinations (focused on locations where at least 5% of Zika poll respondents had gone). Both data sets were weighted as the purpose was to identify the degree to which 'final' results would be similar from the two methodologies. Results of these comparisons are shown in the table on the following page.

	Zika Poll Respondents (%)	RDD Respondents (%)
Demographics		
Gender		
Female	50	48
Male	50	52
Age		
18-29	24	25
30-49	37	38
50-64	26	26
65 or older?	13	12
Race/ethnicity		
White	55	50
Hispanic	27	34
Black	9	8
Education		
4-year college education or more	42	41
Less than 4-year college education	58	59
Household income		
Less than $50,000	25	34
$50,000 to $100,000	29	2
$100,000 or more	46	35
Destinations Traveled		
Mexico	44	50
Miami	39	31
Puerto Rico	9	6
Dominican Republic	7	7
Jamaica	7	6
US Virgin Islands	7	4
St Martin/St Maarten	7	2
Aruba	6	2
Costa Rica	6	3
Cayman Islands	5	2
Bahamas	10	2

As far as demographics, the Zika poll sample and RDD sample yielded similar results. In both, about half of the respondents were female (50% of Zika poll respondents; 48% of RDD respondents), and the samples had very similar age profiles. Further, just over 40% of each set of respondents had a 4-year college education or more (42%; 41%). As we would predict for a sample of travelers to destinations in largely Spanish-speaking areas, both samples were less white (55%; 50%) and more Hispanic (27%; 34%) than the general population, which is 64% white and 16% Hispanic. The Zika poll respondents reported a somewhat higher household annual income than the RDD respondents (46% at $100,000 or more; 35% for the RDD respondents), though we note that the percentages for RDD respondents are slightly deflated by a 10% item nonresponse to the income question, compared with less than 1% item non-response for Zika poll respondents. (Telephone polls frequently have this level of item non-response for income [Yan T, Curtin R, Jans M. Trends in income nonresponse over two decades. *J Off Stat.* 2010;20(1):145-64].)

As far as destinations traveled, both samples shared the most common destinations: Mexico (44% of Zika poll respondents; 50% of RDD respondents) and Miami (39%;

31%). In addition, there were roughly similar shares of visitors to less frequently visited destinations in both samples, including Puerto Rico (9% Zika poll respondents; 6% of RDD respondents), the Dominican Republic (7%; 7%), Jamaica (7%; 6%), U.S. Virgin Islands (7%; 4%), St. Martin/St. Maarten (7%; 2%), Aruba (6%; 2%), Costa Rica (6%; 3%) and the Cayman Islands (5%; 2%). The Bahamas (10%; 2%) was also a less frequently visited des-

tination in both samples, though it was a more common destination for the Zika poll respondents.

Given that results are similar across these two methods, this gives a measure of confidence in the Zika poll methodology and greater confidence that the results are representative of the target population with respect to demographic characteristics and destinations traveled.