Appendix 1 (as supplied by the authors): Supplementary material

Random selection of controls

Using software R version 1.1.453, random numbers were generated between 1 and the maximum number of malaria tests conducted for the year, such that five controls could be selected for each case within that specific year. This number corresponded to the laboratory accession number from the Calgary Laboratory Services (CLS) database which was used to identify the participant. Even though, statistical power does not increase significantly by increasing the case to control ratio above four or five(1-3) since minimal resources were required to extract data by increasing the ratio from 4:1 to 5:1, it was decided to follow a 5:1 matching. It was also decided that 5 to 1 matching will increase the sample size thus accounting for the reduction in power that would have resulted from missing and incomplete data.

The rationale for cases and controls

Given the objective of the study, malaria positive individuals presented themselves as the outcome of interest. Controls were selected from the CLS database from among those who have undergone a malaria test to ensure that both the cases and controls got tested for malaria by the same diagnostic methods and that controls only included those who did not have malaria within the specificity of the diagnostic tests used. This was to limit the information bias due to incorrect outcome ascertainment. No measures were taken to mitigate selection bias that arose due to selection of controls that presented themselves to a care facility only in this study. Controls and cases were matched by year only to eliminate any variation due to immigration/emigration rates and other factors that impact travel by year. Controls were not matched by age and gender as the effect of each on likelihood of acquiring malaria was sought from the multivariable analysis.

Epidemiological risk factors

Our primary exposure of interest, VFR was defined as those who travelled primarily for the purpose of visiting friends and relatives and those who had a difference in risk for malaria exposure due to a history of recent travel to a malaria endemic region, disregarding the differences in ethnicity and immigration status of this population (4).*VFR* was a category under Reason for Visit variable in the Malaria History Form (MHF) and other categories included *Tourists, Business, New Immigrant* and *Visitor*. The two main other purposes of visit categories were *Tourist* travelers and *Business* travelers which included those who self-declared themselves as visiting a malaria endemic region for pleasure (without visiting family or friends) or for work-related purposes respectively. Information regarding age and gender was obtained from the CLS database. Age was a categorical variable of *Child* or *Adult*, where participants who are under 16 years of age were defined as a child. Gender was a categorical variable of *Female* and *Male*. Continent visited was derived from the country indicated in the *Travel Destination* section in MHF and categorized as either visiting *Africa* or *Other* continent for the statistical analyses. Africa was

used as a primary comparator to other continents due to past evidence suggesting imported malaria mainly resulted from travel to the African continent (5-7). Pre-travel advice sought was also a categorical variable of *Yes* or *No* which was collected from the MHF questionnaire that was completed before malaria testing. Malaria prophylaxis taken was a *Yes* or *N*o categorical variable and information for this was also collected from the MHF. Participants were categorized under *Yes* if they indicated that they have taken any one or more of the malaria prophylaxes listed under *Malaria Prophylaxis Taken* in the MHF and '*No'* if they indicated '*None'*. Duration of stay was derived from information provided in the MHF under *Departure Date* and *Date of Arrival in Canada*. This was calculated in days and later categories as less than or equal to two weeks of stay (*≤ 2 weeks*) or longer than 2 weeks (*>2 weeks).* First three digits of the zip code obtained from the CLS database, complemented by MHFs was used to determine the Ward of residence in Calgary and to develop an incidence map of Malaria cases in Calgary according to 2016 census population data.

Statistical analysis

For descriptive analysis, chi-square tests with continuity correction were used for categorical variables whereas ANOVA with the assumption of equal variance was used for continuous variables. This included all the cases and controls (N = 1348) selected for this study. Incidences of malaria map were drawn according to population demographics from the census in 2016 and using geographical boundaries of Wards from Open Calgary (8). For the multivariable logistic regression, numerous variables were examined including the effects of age, VFR status, region traveled to, pre-travel advice, duration and gender on testing positive for malaria. Evidence for effect measure modification between gender and VFR and continent travel to and VFR on a multiplicative scale was analyzed by including interaction variables in multivariate regressions. A chi-square test was used to analyze the differences in taking prophylaxis, taking pre-travel advice and duration of stay between VFR travelers and other travelers. Missing data were not examined further using imputations methods or sensitivity analysis.

Data access and cleaning methods

Investigators only had access data extracted from the CLS database regarding all malaria tests conducted in the CLS during the study period. They were stored and cleaned using Microsoft Excel (Version 1808) to remove repeated tests for individuals occurring within three months. Individuals from the CLS database was linked to individuals from the MHF using MIL numbers.

Sample size calculation

$$N = \frac{r+1}{r}\frac{(\bar{p})(1-\bar{p})\left(Z_{\beta}+Z_{\frac{\alpha}{2}}\right)^2}{(p_1-p_2)^2}$$ [Eq 1] (9) Where;

N	⇨	Number of cases	?
r	⇨	Ratio of controls to cases	5

$\bar{p}$	⇨	Average proportion of exposed	0.56 (Table 1)
Z_{β}	⇨	Standard deviation for power with 80%	0.84
$Z_{\frac{\alpha}{2}}$	⇨	Standard deviation for a significance level at alpha 0.05	1.96
p_1	⇨	Proportion of exposed in cases	0.61 (Table 1)
p_2	⇨	Proportion of exposed in controls	0.53 (Table 1)

Using the above equation and the proportion of exposed cases and controls from our study sample (Table 1), the number of cases required to detect an effect measure that is supported by 80% statistical power is 362 and the sample size required for this study is 1810.

References

1. Rothman K. Modern Epidemiology Little Brown and Company. Boston, Toronto. 1986.
2. Hennekens C, Buring J. Epidemiology in Medicine Little. Brown and Company Boston, MA. 1987.
3. Hennessy S, Bilker WB, Berlin JA, Strom BL. Factors influencing the optimal control-to-case ratio in matched case-control studies. American journal of epidemiology. 1999;149(2):195-7.
4. Barnett ED, MacPherson DW, Stauffer WM, Loutan L, Hatz CF, Matteelli A, et al. The visiting friends or relatives traveler in the 21st century: time for a new definition. J Travel Med. 2010;17(3):163-70.
5. Matteelli A, Carvalho AC, Bigoni S. Visiting relatives and friends (VFR), pregnant, and other vulnerable travelers. Infect Dis Clin North Am. 2012;26(3):625-35.
6. Legros F, Bouchaud O, Ancelle T, Arnaud A, Cojean S, Le Bras J, et al. Risk factors for imported fatal Plasmodium falciparum malaria, France, 1996-2003. Emerg Infect Dis. 2007;13(6):883.
7. Hendel-Paterson B, Swanson SJ. Pediatric travelers visiting friends and relatives (VFR) abroad: illnesses, barriers and pre-travel recommendations. Travel Med Infect Dis. 2011;9(4):192-203.
8. Ward Boundaries [Internet]. 2018 [cited August 2018]. Available from: https://data.calgary.ca/Government/Ward-Boundaries/r9vx-mhnf.
9. Charan J, Biswas T. How to calculate sample size for different study designs in medical research? Indian journal of psychological medicine. 2013;35(2):121.

Appendix to: Marasinghe DH, Cheaveau J, Meatherall B, et al. Risk of malaria associated with travel to malaria-endemic areas to visit friends and relatives: a population-based case–control study. *CMAJ Open* 2020. DOI:10.9778/cmajo.20190070.