

Original Article

Using social media listening and data mining to understand travellers' perspectives on travel disease risks and vaccine-related attitudes and behaviours

Catherine Bravo, PhD^{1,*}, Valérie Bosch Castells, MSc¹, Susann Zietek-Gutsch, MBA¹, Pierre-Antoine Bodin, MSc², Cliona Molony, PhD³, and Markus Frühwein, MD⁴

¹Sanofi Pasteur, Lyon, France, ²Pharma & Healthcare, Berger, Roland Berger, Paris, France, ³Sanofi US Services, Inc., Cambridge, MA, USA and ⁴Dr Frühwein & Partner, Munich, Germany

*To whom correspondence should be addressed. Email: Catherine.Bravo@sanofi.com

Submitted 6 July 2021; Revised 14 January 2022; Editorial Decision 17 January 2022; Accepted 17 January 2022

Abstract

Background: Travellers can access online information to research and plan their expeditions/excursions, and seek travel-related health information. We explored German travellers' attitude and behaviour toward vaccination, and their travel-related health information seeking activities.

Methods: We used two approaches: web 'scraping' of comments on German travel-related sites and an online survey. 'Scraping' of travel-related sites was undertaken using keywords/synonyms to identify vaccine- and disease-related posts. The raw unstructured text extracted from online comments was converted to a structured dataset using Natural Language Processing Techniques. Traveller personas were defined using K-means based on the online survey results, with cluster (i.e. persona) descriptions made from the most discriminant features in a distinguished set of observations. The web-scraped profiles were mapped to the personas identified. Travel and vaccine-related behaviours were described for each persona.

Results: We identified ~2.6 million comments; ~880 k were unique and mentioned ~280 k unique trips by ~65 k unique profiles. Most comments were on destinations in Europe (37%), Africa (21%), Southeast Asia (12%) and the Middle East (11%). Eight personas were identified: 'middle-class family woman', 'young woman travelling with partner', 'female globe-trotter', 'upper-class active man', 'single male traveller', 'retired traveller', 'young backpacker', and 'visiting friends and relatives'. Purpose of travel was leisure in 82–94% of profiles, except the 'visiting friends and relatives' persona. Malaria and rabies were the most commented diseases with 12.7 k and 6.6 k comments, respectively. The 'middle-class family woman' and the 'upper-class active man' personas were the most active in online conversations regarding endemic disease and vaccine-related topics, representing 40% and 19% of comments, respectively. Vaccination rates were 54%–71% across the traveller personas in the online survey. Reasons for vaccination reluctance included perception of low risk to disease exposure (21%), price (14%), fear of side effects (12%) and number of vaccines (11%).

Conclusions: The information collated on German traveller personas and behaviours toward vaccinations should help guide counselling by healthcare professionals.

Key words: Big Data, vaccines, web scraping, health-seeking behaviours, traveller personas, visiting friends and relatives, social media listening

Introduction

Over the past 20 years, international travel has sharply increased, exposing a large number of travellers to personal risks.^{1–4} Vaccine-preventable diseases currently play and will continue to play a significant role in travel-acquired illnesses, and their impact on the health and well-being of travellers is not limited to ‘exotic’ diseases in emerging economy destinations.⁵ Yet despite the benefits of vaccinations, uptake of recommended vaccines among at-risk travellers is highly variable^{6,7}; vaccine coverage rates for vaccine-preventable diseases need to be improved among such travellers.

The coronavirus 2019 (COVID-19) pandemic has deeply impacted the travel industry and highlighted the risk of travellers importing infectious diseases back to their country of origin, with unprecedented burden on healthcare systems and catastrophic social and economic consequences. This pandemic has only emphasized the importance of ensuring traveller acceptance of necessary vaccines so as to provide protection during their travels and reduce the risk of exportation and importation of infectious disease.

There have been a myriad of studies assessing traveller knowledge, attitudes and health-seeking practices.^{8–12} In addition, internet forums and websites provide huge amounts of data that may be used to supplement traditional questionnaire-based surveys, even from travellers who may not seek advice from healthcare professionals. However, information is lacking on the different types of travellers seeking travel-related health information. Travellers are often characterized by their purpose of visit such as personal or business/professional.¹³ Although backpackers and business travellers are usually well characterized, it would be of interest to bring more granularity on leisure travellers, who are usually considered as a single category but may have a wide range of attitudes and behaviours associated with their travel plans and profiles. Attempts to further characterize travellers may help determine factors that motivate those seeking travel-related health advice/information including vaccination recommendations and may better guide counselling by healthcare professionals.

The internet is also a crucial resource of real-time information^{14–16}; it is the most comprehensive accessible archive of electronic health-related information. It also serves as a communication and distribution channel for services and products suppliers and provides a platform for individuals or groups to express their opinions, concerns, and desires.^{17,18} As such, travellers have easy access to a wealth of information to enable them to research and plan their expeditions and/or excursions, and a platform from which to seek or discuss travel-related health information. The proportion for travellers who seek pretravel health advice is highly variable (23–86%),^{8,19–21} and of those who did so 23–73% consulted their general practitioner, 9–39% report using the internet, and 6–22% asked family and friends.^{8,19–21} The variation in the proportion seeking pretravel health advice between studies is likely attributable to a number of factors including country of origin, destination and reason for travel.

We undertook this study to better understand German travellers’ attitudes and behaviours towards vaccination, and their travel-related health information seeking activities using two approaches: web ‘scraping’ of comments posted on various

German travel-related sites, and an online survey. We also used this information to profile different types of German travellers seeking travel-related health information. The intention of our study was to identify the profiles of travellers that are present on blogs and website forums to better understand the factors that shape travellers’ decisions to get vaccinated and how healthcare professionals may improve awareness of disease risk targeting the appropriate traveller population with the appropriate information through the most suitable channel or media.

Methodology

Web scraping

‘Scraping’ of social media websites was undertaken between 18 and 21 January 2019 using a list of 64 German keywords to identify relevant comments and posts ([Appendix 1, Supplementary Table S1](#)). We also included German synonyms of the keywords and each keyword could be mapped into one or more of the five categories: vaccine-related, disease-related, vaccine name (brand names), disease (name) and destination related. We selected the most popular travel website forums in Germany based on daily traffic to these domains for scraping. The websites chosen (monthly traffic to these domains at the time of our analysis stated in parenthesis) for scraping were Holiday-Check.de (14 000 k), TripAdvisor.de (18 700 k), Lonelyplanet.de (48.7 k), Stefan-Loose.de (14.1 k), GeoForum.de (1300 k) and [ReiseFrage.net](#) (14.2 k). We created structured features linked to words identified a priori to characterize the travellers from web scraping; these included word or word patterns linked to vaccines, socio-demographics, sentiments, travel destination, prices, interests, destination country, destination continent, main accommodation, health, purpose of travel and travel duration. The raw data were extracted from the forums of these websites as unstructured text using open source libraries from Beautiful-Soup Python. We applied Natural Language Processing (NLP) techniques from open source tools, including NLP modules such as nltk, spacy, word2number, parsedatetime, dateparser, dateutil and performed routines such as Named Entity Recognition and keywords mapping using the predefined selection ([Appendix 1, Supplementary Table S1](#)). We converted raw data into the structured datasets, adding new structured features to the data, for example, onegram or bigrams. Unknown/unavailable data were matched based on the ‘nearest neighbour’ approach, a commonly used filling missing information technique in data science.²²

Online survey

A self-administered online survey characterizing the responders’ profile, their previous travel and their behaviours towards health issues associated with their travel, which consisted of 35 questions ([Appendix 2, Excel file](#)), was made available to German adults aged ≥ 18 years through advertisement on Facebook and Twitter (i.e. the two largest media platforms so as to be representative of the population using social media) between 25 January 2019 and 11 February 2019. The survey was proposed to 20 000 individuals, with an expected acceptance rate of 5%, to reach around 1000 participants. Participation was completely

voluntary and done anonymously, with no incentives to complete the survey. Only complete surveys were included in our analysis.

Traveller personas

Traveller personas were defined using K-means based on the results of the online survey.²³ K-means is an unsupervised machine learning approach used to define groupings (clusters) of observations based on features (variables) of those observations. It is the most popular partitioning clustering method.²⁴ The optimal number of clusters K was selected using the Elbow method to identify the balance point between the quality of the clustering analysis and the complexity of the model.²⁵ The smallest number of clusters accounting for the greatest variation in the data was selected. We considered K from 4 to 20. Categorical features were encoded by one-hot encoding so that the Euclidean distance could be calculated considering both numerical and categorical features. Clusters were described based on the most discriminant features, by comparing the distribution of the unique sample profiles across all clusters and all levels of the feature. In each cluster, features that showed a differentiating sample distribution (e.g. over represented) were picked as discriminant features, and these features were used to help name the personas for lay-interpretation.

The centroid (a defining feature that distinguishes clusters from each other) of each cluster was then identified, and projected over the web-scraped profiles. This was done after searching and extracting the same features from the forum comments as used for the clustering of the online survey profiles. The travel and vaccine-related behaviours were qualitatively described for each persona using both web-scraped and online survey data.

Sentiment analysis

To understand travellers' emotions towards infectious diseases endemic at their destination of interest, we performed a sentiment analysis on each disease using the Tone Analyzer from IBM Watson,²⁶ a cloud-based platform that uses deep learning to generate characteristics from the web-scraped text such as entities, keywords, categories and sentiment. The Tone Analyzer software measures emotions present in written text and has been used in various topics including healthcare.^{27,28} The software detects the different emotions/feelings expressed in the text and generates a score between 0 and 1 for each of the following five standard sentiments: anger, confidence, fear, joy and anxiety. Scores approaching 1 indicate an emotion strongly expressed and scores approaching 0, no emotion expressed. Each comment or post from social media could have >1 related sentiment. We calculated the average score of each sentiment by infectious disease and traveller persona.

Ethical statement

This study was conducted in line with General Data Protection Regulation rules and did not require ethical approval as it relied on publically available information or survey data collected anonymously via the internet where the participants could not be directly or indirectly identified, and there was no direct interaction with the participants.

Table 1. Demographic characteristics of web-scraped profiles and online survey respondents

Characteristic	Web-scraped profiles (n=65 k), %(n)	Online survey (n=1028), %(n)
Sex		
Male	27% (~17 k)	48% (488)
Female	31% (~20 k)	53% (540)
Unknown	42% (~27 k)	
Age, years		
< 25	11% (~7 k)	7% (72)
25–34	8% (~5 k)	23% (234)
35–44	10% (~7 k)	21% (212)
45–54	10% (~7 k)	19% (192)
55–64	8% (~5 k)	20% (206)
>65	2% (~1 k)	11% (112)
Unknown	51% (~33 k)	
Relationship status		
Single	2% (~1 k)	23% (235)
Couple	36% (~23 k)	69% (713)
Divorced	10% (~7 k)	8% (80)
Married	6% (~4 k)	
Unknown	46% (~30 k)	
Children		
No	6% (~4 k)	46% (472)
Yes	17% (~11 k)	54% (556)
Unknown	78% (~50 k)	
Occupation		
Unemployed		20% (201)
Student		5% (55)
Self-employed		7% (72)
Employed		68% (700)
Planned travel duration		
< 10 days		15% (150)
10 days to 3 weeks		70% (722)
> 3 weeks		15% (156)
Frequency of travel to destinations outside the European Union and North America		
≤ 1 annually		56% (547)
≥ 2 annually		44% (434)

Results

Web scraping

We identified ~2.6 million comments through web scraping. Duplicate comments (same verbatim) and user profiles (same identifiers) were screened and removed, leaving ~880 k unique comments that mentioned ~280 k unique trips (a unique trip was defined as 1 unique user on a topic) by ~65 k unique profiles. Of these, ~210 k comments specifically mentioned endemic diseases and vaccination. The demographic characteristics of the web-scraped profiles are summarized in Table 1.

The destinations commented on were most frequently in Europe (320 k comments; 37%) and Africa (180 k comments; 21%), followed by Southeast Asia (100 k comment; 12%) and the Middle East (92 k; 11%). The destinations with fewest comments were in South America (5 k comments; 1%) and

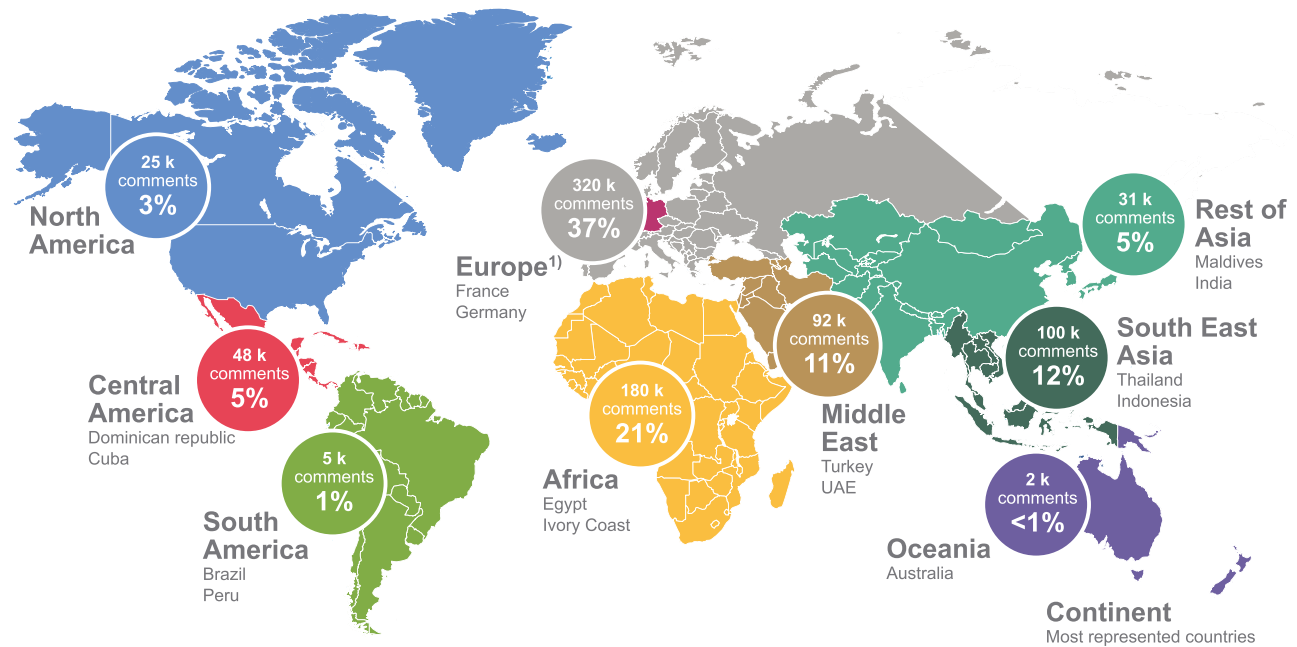


Figure 1. Web-scraped comments by destination; ‘scraping’ of social media websites was undertaken between 18 and 21 January 2019. Data shown are the number and percentage of overall comments

Oceania (2 k comments; <1%) (Figure 1). To optimize the data description based on travel destination, the countries were regrouped into seven zones: North America, Central and South America, Europe, Africa (excluding North Africa), North Africa and the Middle East, Asia and Oceania.

Online survey

Overall, 1028 questionnaires were completed with relevant responses. The characteristics of the respondents are summarised in Table 1. Survey respondents mainly travelled for leisure (86%; $n=889$), with a length of stay of 10 days to 3 weeks (70%; $n=724$). Respondents chose 93 countries for their last travel, and the two most popular destinations outside the European Union and North America were Asia (36%; $n=365$) and Africa (30%; $n=306$), with Thailand (14%; $n=141$) and Egypt (9%; $n=97$) the two most frequent destinations.

Excluding Europe and North America (not classified as exotic areas) and Oceania (underrepresented in the online survey), there was good concordance in the top 25 countries identified by web scraping with those identified in the online survey (Appendix 3, Supplementary Figure S1).

Traveller personas

The clustering approach produced eight distinct personas based on the most discriminant features for each type: ‘middle-class family woman’, ‘young woman travelling with partner’, ‘female globe-trotter’, ‘upper-class active man’, ‘single male traveller’, ‘retired traveller’, ‘young backpacker’, ‘visiting friends and relatives’. The purpose of travel was identified as leisure in most (82–94%) of the personas (highest for ‘single male traveller’ and lowest for ‘young backpacker’), except for the ‘visiting friends

and relatives’ persona where the main purpose was for a visit (86%).

Here, we describe in detail the ‘middle-class family woman’ persona as an example; the remainder is presented in detail in Appendix 4.

‘Middle-class family woman’

The ‘middle-class family woman’ persona was composed of 19% of the online survey responders and 30% of the web-scraped unique travellers; 50% were aged 35–54 years and most were in active employment (76%) and had an income in the range €30–100 000 (65%). Travel was mostly for leisure (93%) with her partner and/or family (1–2 children) for 10–21 days (70%) with hotel stays (89%) (Appendix 1, Supplementary Table S2). This persona tended to plan in advance (Figure 2) and had a slight preference for using travel agencies. Booking travel, planning vaccination and receipt of vaccination were all completed on average 3 months before departure. Overall, 60% received vaccinations before their last trip; this persona mostly went to the specialist physician for vaccination and tended to get their children vaccinated more frequently than themselves. They perceived a high level of risk to endemic diseases in Asia, but low risk in Sub-Saharan Africa, and were generally not sensitive to vaccine price. This persona showed more than average worrying behaviour online and often mentioned their partner and children in comments.

Main endemic diseases mentioned in online conversations (web-scraped comments)

Malaria and rabies were the two most commented diseases, mentioned in 12.7 k and 6.6 k web comments, respectively (Table 2).

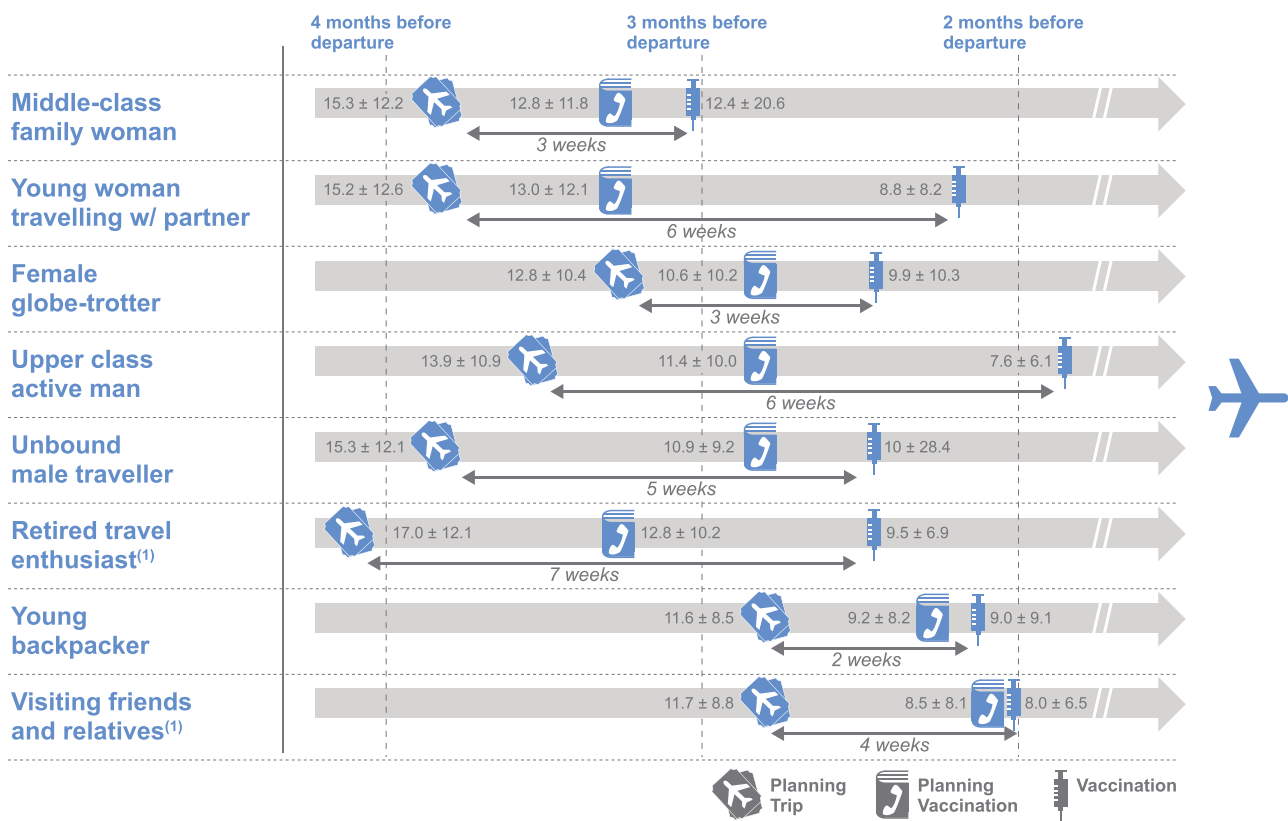


Figure 2. Traveller journey—from the start of planning to vaccination and departure (average durations with standard deviations) (online survey)

The ‘middle-class family woman’ and ‘upper-class active man’ personas were the most active groups in online conversations regarding disease and vaccine-related topics, representing 40% and 19% of all comments, respectively.

The level of anxiety or confidence about endemic diseases was highly dependent on the traveller persona and type of disease (Appendix 1, Supplementary Table S3). Overall, the ‘middle-class family woman’ and ‘young backpacker’ personas were less anxious about diseases than the average global level of anxiety expressed across the personas, whereas the ‘young woman travelling with partner’, ‘upper-class active man’, and ‘visiting friends and relatives’ personas were more anxious. The ‘female globe-trotter’ and ‘single male traveller’ personas’ level of anxiety were the same as the average global level of anxiety. The ‘retired traveller’ persona’s anxiety was highly dependent on the disease. For the specific diseases mentioned, the ‘middle-class family woman’, ‘young woman travelling with partner’ and ‘female globe-trotter’ all had high anxiety for meningitis and Zika, with the latter two also having high anxiety for typhoid and dengue, or dengue, respectively. The ‘upper-class active man’ persona had high anxiety for encephalitis and yellow fever, but the ‘single male traveller’ had the same level of anxiety for all diseases. The ‘retired traveller’ persona had high anxiety for encephalitis and hepatitis; the ‘young backpacker’, for malaria and dengue; and ‘visiting friends and relatives’, for meningitis and yellow fever.

There were no major differences in the level of anxiety or confidence about endemic diseases per travel destination by traveller persona.

Perception of disease risk, source of vaccination-related information and reluctance to receive vaccination (online survey)

The perception of disease risk was highest for travel to Africa (excluding North Africa) [rated 2.4 on a scale from 0 (no risk) to 10 (high risk)] and lowest for North Africa and the Middle East (rated 1.4); the main perceived risks were malaria and hepatitis A, and hepatitis A and flu in the two regions, respectively. The perceived risks in Central and South America, and Asia were rated similarly, at 2.1, with malaria, hepatitis A and tetanus perceived as risks in Central & South America, and hepatitis A and B in Asia.

Travellers checked on average 2.1 information sources to inquire about vaccinations, ranging from 1.5 sources on average for the ‘visiting friends and relatives’ persona to 2.3 for the ‘middle-class family woman’ and ‘young backpacker’ personas. The main source of information for vaccine-related topics was the general practitioner among all personas (Figure 3).

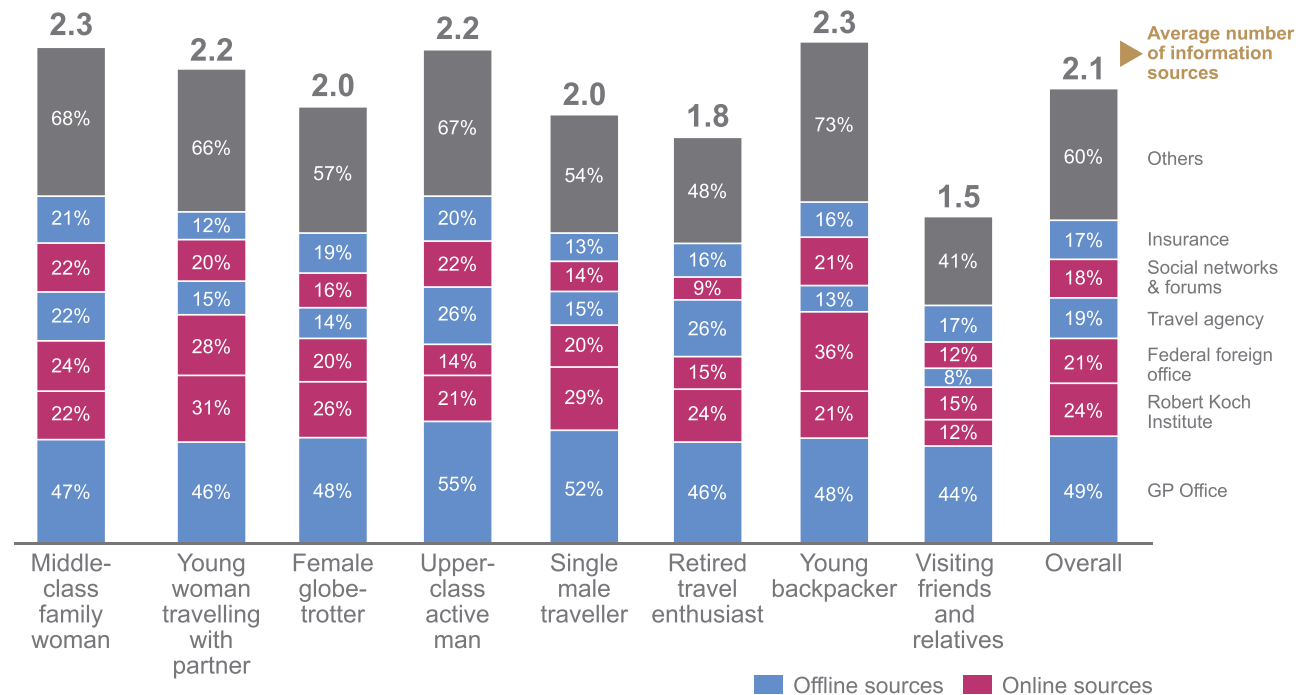
Only 33% of those who travelled to destinations outside the European Union and North America felt they had enough information on vaccination. The majority of respondents (61%) had received at least one vaccination for travel-related vaccine-preventable diseases; most vaccinations were against hepatitis B (43%) and hepatitis A (41%). Among those not vaccinated, 42% and 49% consider that they were at low/medium risk for hepatitis A and B, respectively.

Reasons for reluctance to receive vaccination included the perception that the risk of disease exposure was low (20%), price (14%), fear of side effects (12%) and number of vaccines (11%)

Table 2. Comments mentioning endemic disease in the web scraping dataset (total number of comments in thousand presented)

	Malaria	Rabies	Dengue	Yellow Fever	Hepatitis	Typhoid	Encephalitis	Others
Middle-class family woman	5174	2547	1954	2264	1609	715	227	808
Upper-class active man	2696	1201	1224	507	572	432	196	435
Young backpacker	1277	1207	698	229	544	316	221	280
Female globe-trotter	970	205	185	849	239	62	2	222
Single male traveller	841	402	210	180	243	162	127	152
Visiting friends and relatives	773	587	359	175	213	208	118	138
Young woman travelling with partner	496	250	226	262	202	98	84	138
Retired traveller	504	195	178	107	160	120	44	110

Note: Others include Cholera, Japanese Encephalitis, Zika and Ebola.

**Figure 3.** Vaccine-related information sources by persona and information source (percentages shown for information source) (online survey)

(Appendix 1, Supplementary Table S4). The personas with more vaccine reluctance were the ‘visiting friends and relatives’ and ‘retired traveller’, with the main reason for not vaccinating with these two persona being the perception that disease risks (32% and 29%, respectively) were low.

Discussion

In our study, we defined a number of distinct traveller personas based on common features associated with their online interactions/comments, preferences and behaviours; these were predominantly leisure travellers using online travel forums in Germany, not previously characterized. For example, we characterized three distinct groups of female leisure travellers: ‘middle-class family woman’, ‘young woman travelling with partner’ and ‘female globe-trotter’. We additionally characterized two distinct groups of male leisure travellers, ‘upper-class active man’ and ‘single male traveller’, as well as ‘retired traveller’ and ‘young backpacker’, and those ‘visiting friends and relatives’. Although ‘backpackers’ are a well-recognized group, there are ambiguities

and inconsistencies in their characterization [e.g. by age group (15–24 years or 18–30 years), travel characteristics (budget accommodation/youth hostel stays), duration of travel (several months) and economic criteria].²⁹ Nonetheless, our characterization of ‘young backpacker’ (aged below 35 years, mostly stayed in hostels, and most took long duration trips) was broadly consistent with previous criteria reported across the literature. In addition, those ‘visiting friends and relatives’ are also generally well recognized and described,^{9,11,12,30–32} although there may be differences in the demographic characteristics between first-generation immigrants and other travellers (including second-generation immigrants) visiting friends and relatives.³¹ The advantages of empirically defining the traveller personas in our study include that they would be less biased to perceived stereotypes as these were based on real-life experiences and behaviours.

Vaccination rates against endemic diseases across the traveller personas varied from 54% to 71%, highest among ‘young backpacker’ and lowest among ‘visiting friends and relatives’ personas. Reasons for reluctance to receive vaccination included low perception of exposure to disease (21%), price (14%), fear

of side effects (12%) and number of vaccines (11%). To our knowledge, reports of reasons for refusal to vaccinate among international German travellers are lacking. A study of German travellers' preference for travel vaccines found strong preferences for vaccination than refusal, among four types of travellers: business travellers; travellers visiting friends and relatives; leisure travellers; and backpackers.³² All traveller personas attached the greatest importance to disease risk, health impact and vaccine cost for their travel vaccine decisions, with other disease and vaccine attributes less important.

Studies assessing the characteristics, travel preferences and attitudes and travel-related health information seeking activities of older travellers (typically aged ≥ 65 years) are limited. Older travellers represent a significant proportion of international travellers [though underrepresented in our online survey (10%) and web scraping (3%) profiles], and their numbers are likely to increase with aging populations particularly in developed countries.³³ Moreover, they represent a population that may have poor protection against some infectious diseases due to immunosenescence,³⁴ and typically face high premiums for travel insurance often due to high prevalence of preexisting medical conditions. Older travellers would likely focus more on health risk in the current COVID-19 pandemic, as the virus causes worse outcomes and higher mortality in this age group and those with preexisting medical conditions.³⁵ We did not assess the occurrence of preexisting medical conditions among the traveller personas. Of concern, the 'retired traveller' persona in our study generally had a low perception of disease risk and a low proportion received travel vaccinations. Our observations are consistent with a systematic review of pretravel advice seeking behaviour that found perceived low risk of infection to be a major reason for not seeking or lack of adherence to pretravel health advice.³⁶ However, they tended to take less adventurous trips than other personas. Indeed, older travellers have been shown to have a lower willingness to take recreational risks during their travels than backpackers and young travellers.³⁷

We showed that vaccinations were generally received >2 months before departure. This finding appears inconsistent with a European airport survey of intercontinental travellers ($n = 5465$) undertaken between September 2002 and September 2003, including those from Germany³⁸; the majority of those surveyed (59%) reported starting preparations <2 months before departure. Thus, there is a disparity in how travellers responded about their timelines to vaccination/planning before their last trip depending on the source of traveller assessed.

A study of imported infectious diseases in Munich, Germany, between 1999 and 2014 after travel from the sub-tropics found that the number of infectious diseases was significantly elevated among backpackers (particularly those returning from Asia) than the other travellers assessed (business and all-inclusive travellers).³⁹ It is possible that backpackers are at high risk of infectious disease because they are more often faced with low hygienic conditions and over a longer duration of stay. Backpackers had the longest travel duration (≥ 21 days) among the travellers assessed; in our study also, 'young backpackers' were among two personas with the longest duration of travel. Of concern, backpackers and young travellers were found to have a higher willingness to take recreational risks than other

travellers (luxury travellers and older travellers), with males more so than females.³⁷ Thus, it is imperative that 'young backpackers' are fully immunized and receive all necessary travel vaccines, as well as being informed on best practice to avoid infectious diseases circulating at their destinations.

There is also a need to ensure appropriate disease awareness for a given region so as to minimize inappropriate anxiety. For example, several traveller personas in our study were concerned by tick-borne encephalitis in regions with no such risk and despite the disease being endemic in parts of Germany.⁴⁰ The global level of anxiety about endemic disease at travel destinations in our study differed by persona, and of note, the different personas were not anxious about the same diseases. Our results suggest a need to adapt discussions between healthcare professionals and travellers to focus on the diseases that are associated with anxiety for the different traveller profiles to reassure and vaccinate (or initiate appropriate prophylaxis) as required.

Our study has a number of limitations. The online format limited the population to those with internet access and who were active on social media portals. The proportion of respondents in the online survey in some age groups was limited, and thus our results may lack generalizability. Travellers' actual vaccine decisions in practice may depend on contextual factors that are beyond the scope of our study. In addition, we did not ascertain the ethnic composition of the personas. We did not identify those specifically travelling for religious or health purposes; it is possible that these types of travellers may not use the travel forums assessed. In addition, there was a higher level of missing information in the web-scraped dataset, and we could not confirm if this group undertook any travel. A strength of this study is that the information was obtained in the 'big data' environment.

To the best of our knowledge, this is the first time that social media listening and data mining techniques have been used to assess travellers' knowledge, attitudes and practices. Nonetheless, the mining of social media information has increasingly been used to assess public knowledge, attitudes and practices, as well as for 'infoveillance' for a host of public health issues including but not limited to obesity and dieting, electronic nicotine delivery systems, heart disease, mental health, infectious diseases (Zika; COVID-19; influenza; West Nile virus) and fitness inspiration.⁴¹⁻⁵¹ Such 'big data' analyses may help inform healthcare organizations on the level of public misinformation and reluctance to adopt best practice, and the need to monitor and deliver reliable healthcare guidance. Additional research would be required to determine the utility of social media to influence travel medicine practice. However, social media should be used to augment, rather than replace pretravel consultation, to ensure a high level of knowledge of the travel-related health risks.⁵²

The COVID-19 global health crisis has had an unprecedented impact on international travel, with about 96% of destinations worldwide imposing some form of travel restriction in response to the pandemic by April 2020.⁵³ Health crises lead to anxiety among tourists and influence travel intention,⁵⁴ and travel-related health risk perception has increased as a result of COVID-19.^{55,56} The focus on health risk leads to more travellers seeking or discussing health-related information online to

prepare for possible health risks at their destination. Our survey allowed us to evaluate opinions and perceptions that shaped travellers' decisions before the COVID-19 pandemic. However, COVID-19 has likely greatly influenced travellers' decisions, with greater emphasis on safety and less crowded destinations since the pandemic began.⁵⁷ Emerging mobile phone technology may also help better track human movements and transmission (importation or exportation) of infectious diseases at multiple temporal and spatial scales to allow for more accurate estimates of the health risks during travel or on return to home countries.⁵⁸ Post-COVID-19, it would be interesting to repeat our study to assess how the online interactions/comments, preferences and behaviours among the traveller personas have been affected.

In conclusion, this study provides new and innovative ways to collect useful information on health-related opinions and perceptions for a number of typical travellers, which should help better guide counselling by healthcare professionals, and may help reduce the risk of infection and disease transmission among these traveller groups.

Supplementary data

Supplementary data are available at *JTM* online.

Author statements

Author contributions

CB, VBC, SZ, P-AB and CM contributed to the conceptual design of the study; P-AB was involved with data acquisition. All authors contributed to the interpretation of the data and participated in the drafting and critical revision of this report, approved the final version and are accountable for its accuracy and integrity.

Acknowledgements

Editorial assistance with the preparation of the manuscript was provided by Richard Glover, inScience Communications, Springer Healthcare Ltd, Chester, UK and was funded by Sanofi Pasteur. The authors also thank Michel Ferracci and Nacira Abbassi for their strong support on the design and preparation of this project as well as Roopsha Brahma PhD, for editorial assistance and manuscript coordination on behalf of Sanofi-Pasteur.

Funding

This work was supported by Sanofi Pasteur.

Conflicts of interest: CB, VBC, SZ, CM are employees of Sanofi Pasteur and may hold shares and/or stock options in the company. P-AB was a former employee of Roland Berger, who received funding from Sanofi Pasteur to conduct the study. MF has no relevant conflict of interest to report.

References

- Zamarron Fuertes P, Perez-Ayala A, Perez Molina JA *et al.* Clinical and epidemiological characteristics of imported infectious diseases in Spanish travelers. *J Travel Med* 2010; 17:303–9.
- Odolini S, Parola P, Gkrania-Klotsas E *et al.* Travel-related imported infections in Europe, EuroTravNet 2009. *Clin Microbiol Infect* 2012; 18:468–74.
- Sambri V, Cavrini F, Rossini G *et al.* The 2007 epidemic outbreak of chikungunya virus infection in the Romagna region of Italy: a new perspective for the possible diffusion of tropical diseases in temperate areas? *New Microbiol* 2008; 31:303–4.
- Gautret P, Botelho-Nevers E, Brouqui P, Parola P. The spread of vaccine-preventable diseases by international travellers: a public-health concern. *Clin Microbiol Infect* 2012; 18:77–84.
- Freedman DO, Chen LH. Vaccines for international travel. *Mayo Clin Proc* 2019; 94:2314–39.
- Duffy MR, Reed C, Edelson PJ *et al.* A survey of US travelers to Asia to assess compliance with recommendations for the use of Japanese encephalitis vaccine. *J Travel Med* 2013; 20:165–70.
- Alqahtani AS, Rashid H, Heywood AE. Vaccinations against respiratory tract infections at hajj. *Clin Microbiol Infect* 2015; 21:115–27.
- Wilder-Smith A, Khairullah NS, Song JH *et al.* Travel health knowledge, attitudes and practices among Australasian travelers. *J Travel Med* 2004; 11:9–15.
- Lammert SM, Rao SR, Jentes ES *et al.* Refusal of recommended travel-related vaccines among US international travellers in global TravEpiNet. *J Travel Med* 2016; 24:taw075. <https://doi.org/10.1093/jtm/taw075>.
- Khan NM, Jentes ES, Brown C *et al.* Pre-travel medical preparation of business and occupational travelers: an analysis of the global TravEpiNet consortium, 2009 to 2012. *J Occup Environ Med* 2016; 58:76–82.
- Heywood AE, Watkins RE, Iamsirithaworn S *et al.* A cross-sectional study of pre-travel health-seeking practices among travelers departing Sydney and Bangkok airports. *BMC Public Health* 2012; 12:321. <https://doi.org/10.1186/1471-2458-12-321>.
- Zimmermann R, Hattendorf J, Blum J *et al.* Risk perception of travelers to tropical and subtropical countries visiting a Swiss travel health center. *J Travel Med* 2013; 20:3–10.
- World Tourism Organization (UNWTO). *International recommendations for tourism statistics 2008*. https://unstats.un.org/unsd/publication/seriesm/seriesm_83rev1e.pdf (14 November 2020, date last accessed).
- Sarker A, Nikfarjam A, Gonzalez G. Social media mining shared task workshop. *Pac Symp Biocomput* 2016; 21:581–92.
- Lee J, Kim J, Hong YJ *et al.* Health information technology trends in social media: using twitter data. *Healthc Inform Res* 2019; 25:99–105.
- Correia RB, Wood IB, Bollen J, Rocha LM. Mining social media data for biomedical signals and health-related behavior. *Annu Rev Biomed Data Sci* 2020; 3:433–58.
- Lazard AJ, Scheinfeld E, Bernhardt JM *et al.* Detecting themes of public concern: a text mining analysis of the centers for disease control and prevention's Ebola live twitter chat. *Am J Infect Control* 2015; 43:1109–11.
- Abd-Alrazaq A, Alhuwail D, Househ M *et al.* Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. *J Med Internet Res* 2020; 22:e19016.
- Al-Abri SS, Abdel-Hady DM, Al-Abaidani IS. Knowledge, attitudes, and practices regarding travel health among Muscat international airport travelers in Oman: identifying the gaps and addressing the challenges. *J Epidemiol Glob Health* 2016; 6: 67–75.

20. Van Herck K, Zuckerman J, Castelli F *et al.* Travelers' knowledge, attitudes, and practices on prevention of infectious diseases: results from a pilot study. *J Travel Med* 2003; 10:75–8.
21. Toovey S, Jamieson A, Holloway M. Travelers' knowledge, attitudes and practices on the prevention of infectious diseases: results from a study at Johannesburg international airport. *J Travel Med* 2004; 11:16–22.
22. Subramanian, D. *A simple introduction to k-nearest neighbors algorithm*. <https://towardsdatascience.com/a-simple-introduction-to-k-nearest-neighbors-algorithm-b3519ed98e> (10 November 2021, date last accessed) 2019.
23. Kanungo T, Mount DM, Netanyahu NS *et al.* An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans Pattern Anal Mach Intell* 2002; 24:881–8892.
24. Yadav J, Sharma M. A review of k-mean algorithm. *Int J Eng Trends Technol* 2013; 4:2972–6.
25. Thorndike RL. Who belongs in the family? *Psychometrika* 1953; 18:267–76.
26. IBM. *Tone analyzer: understand emotions and communication style in text*. <https://www.ibm.com/watson/services/tone-analyzer/> (1 September 2020, date last accessed) 2019.
27. Norman KP, Govindjee A, Norman SR *et al.* Natural language processing tools for assessing progress and outcome of two veteran populations: cohort study from a novel online intervention for posttraumatic growth. *JMIR Form Res* 2020; 4:e17424.
28. Ottenhoff JSE, Kortlever JTP, Teunis T, Ring D. Factors associated with quality of online information on trapeziometacarpal arthritis. *J Hand Surg Am* 2018; 43:889–896 e885.
29. Dayour F, Kimbu AN, Park S. Backpackers: the need for reconceptualisation. *Ann Tour Res* 2017; 66:191–93.
30. Paudel P, Raina C, Zwar N *et al.* Risk activities and pre-travel health seeking practices of notified cases of imported infectious diseases in Australia. *J Travel Med* 2017; 24:1–7. <https://doi.org/10.1093/jtm/tax044>.
31. Leder K, Tong S, Weld L *et al.* Illness in travelers visiting friends and relatives: a review of the geosentinel surveillance network. *Clin Infect Dis* 2006; 43:1185–93.
32. Poulos C, Curran D, Anastassopoulou A, De Moerloose L. German travelers' preferences for travel vaccines assessed by a discrete choice experiment. *Vaccine* 2018; 36:969–78.
33. Patterson I, Balderas A. Continuing and emerging trends of senior tourism: a review of the literature. *J Popul Ageing* 2020; 13: 385–99.
34. Jilg W. Vaccines for older travelers. *Interdiscip Top Gerontol Geriatr* 2020; 43:158–81.
35. Shahid Z, Kalayanamitra R, McClafferty B *et al.* COVID-19 and older adults: what we know. *J Am Geriatr Soc* 2020; 68:926–9.
36. Kain D, Findlater A, Lightfoot D *et al.* Factors affecting pre-travel health seeking behaviour and adherence to pre-travel health advice: a systematic review. *J Travel Med* 2019; 26:taz059. <https://doi.org/10.1093/jtm/taz059>.
37. Farnham A, Ziegler S, Blanke U *et al.* Does the dospert scale predict risk-taking behaviour during travel? A study using smartphones. *J Travel Med* 2018; 25:1–7. <https://doi.org/10.1093/jtm/tay064>.
38. Van Herck K, Van Damme P, Castelli F *et al.* Knowledge, attitudes and practices in travel-related infectious diseases: the European airport survey. *J Travel Med* 2004; 11:3–8.
39. Herbinger KH, Alberer M, Berens-Riha N *et al.* Spectrum of imported infectious diseases: a comparative prevalence study of 16,817 German travelers and 977 immigrants from the tropics and subtropics. *Am J Trop Med Hyg* 2016; 94:757–66.
40. Liebig K, Boelke M, Grund D *et al.* Tick populations from endemic and non-endemic areas in Germany show differential susceptibility to TBEV. *Sci Rep* 2020; 10:15478.
41. Miller M, Banerjee T, Muppalla R *et al.* What are people tweeting about Zika? An exploratory study concerning its symptoms, treatment, transmission, and prevention. *JMIR Public Health Surveill* 2017; 3:e38.
42. Bragazzi NL, Bacigaluppi S, Robba C *et al.* Infodemiological data of West-Nile virus disease in Italy in the study period 2004-2015. *Data Brief* 2016; 9:839–45.
43. Chew C, Eysenbach G. Pandemics in the age of twitter: content analysis of tweets during the 2009 H1N1 outbreak. *PLoS One* 2010; 5:e14118.
44. Carrotte ER, Prichard I, Lim MS. "fitspiration" on social media: a content analysis of gendered images. *J Med Internet Res* 2017; 19:e95.
45. McClellan C, Ali MM, Mutter R *et al.* Using social media to monitor mental health discussions - evidence from twitter. *J Am Med Inform Assoc* 2017; 24:496–502.
46. Eichstaedt JC, Schwartz HA, Kern ML *et al.* Psychological language on twitter predicts county-level heart disease mortality. *Psychol Sci* 2015; 26:159–69.
47. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. *Am J Prev Med* 2011; 40:448–53.
48. Coogan S, Sui Z, Raubenheimer D. Gluttony and guilt: monthly trends in internet search query data are comparable with national level energy intake and dieting behavior. *Palgrave Commun* 2018; 4. <https://doi.org/10.1057/s41599-017-0055-7>.
49. Daughton AR, Paul MJ. Identifying protective health behaviors on twitter: observational study of travel advisories and Zika virus. *J Med Internet Res* 2019; 21:e13090.
50. Daughton AR, Shelley CD, Barnard M *et al.* Mining and validating social media data for COVID-19-related human behaviors between January and July 2020: infodemiology study. *J Med Internet Res* 2021; 23:e27059.
51. Li J, Xu Q, Cuomo R *et al.* Data mining and content analysis of the Chinese social media platform weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. *JMIR Public Health Surveill* 2020; 6:e18700.
52. Rozenberg G, Petersiel N, Korytny A *et al.* Standard pre-travel consultation versus shorter consultation combined with smartphone support: a randomized controlled trial. *J Travel Med* 2019; 26:taz025. <https://doi.org/10.1093/jtm/taz025>.
53. UNWTO & data partners. COVID-19 - a global insight on travel and tourism impacts. https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2020-03/21_4_Tourism_COVID19_Data_Coalition_npptx.pdf (13 November 2020, date last accessed) .
54. Senbeto DL, Hon AHY. The impacts of social and economic crises on tourist behaviour and expenditure: an evolutionary approach. *Curr Issue Tour* 2020; 23:740–55. <https://doi.org/10.1080/13683500.2018.1546674>.
55. Neuburger L, Egger R. Travel risk perception and travel behaviour during the COVID-19 pandemic 2020: a case study of the DACH region. *Curr Issue Tour* 2021; 24:1003–16. <https://doi.org/10.1080/13683500.2020.1803807>.
56. Nazneen S, Hong X, Ud Din N. COVID-19 crises and tourist travel risk perceptions. 2020. Available at SSRN. <https://doi.org/10.2139/ssrn.3592321>. SSRN Electronic Journal.
57. Flores-Ruiz D, Elizondo-Salto A, Barroso-González MO. Using social media intourist sentiment analysis: a case study of Andalusia during the COVID-19 pandemic. *Sustainability* 2021; 13:3836. <https://doi.org/10.3390/su13073836>.
58. Lai S, Farnham A, Ruktanonchai NW, Tatem AJ. Measuring mobility, disease connectivity and individual risk: a review of using mobile phone data and mHealth for travel medicine. *J Travel Med* 2019; 26:taz019. <https://doi.org/10.1093/jtm/taz019>.