PLOS Science Wednesday: Hi Reddit, we're Nick and Cori Ruktanonchai, and we published a paper in PLOS Computational Biology on how mobile phone data can target malaria elimination efforts – Ask Us Anything!

PLOSScienceWednesday [1] and r/Science AMAs[1]

[1]Affiliation not available

April 17, 2023

Abstract

Hi Reddit! I'm Nick Warren Ruktanonchai, a postdoctoral research fellow at the University of Southampton. I'm interested in understanding how people move, which helps us predict when, where, and why some people become exposed to areas with infectious diseases. And I am Cori Warren Ruktanonchai, a PhD student in Geography & Environment at the University of Southampton–as you may have noticed by the names, I also happen to be Nick's wife! I'm interested in using spatial statistics to better locate pregnant women, mothers and newborns at risk of adverse health outcomes. We recently published an article titled "Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data" in PLOS Computational Biology, mapping where people got malaria based on their travel patterns. We combined data from 1.19 million mobile phones in Namibia with a map of malaria prevalence to predict areas where the most people get infected. We hope that by targeting these hotspots, elimination efforts can both send help where it's most needed and reduce transmission nationwide. Call and text locations from mobile phones are a great tool for knowing where people have been. At Flowminder, we've used mobile phone data to not only help governments and NGOs predict the spread of disease, we've also used it to understand how people move after catastrophes, including a PLOS Medicine paper on the 2010 Haiti earthquake and a PLOS Currents paper on the 2015 Nepal earthquake. We'll be answering your questions at 1pm ET – Ask Us Anything! Don't forget to follow Nick on Twitter at @nruktanonchai and Cori at @cwruktanonchai. Also, the Flowminder Foundation can be found at @flowminder, and the WorldPop Project at @WorldPopProject!

the WINNOWER

REDDIT

PLOS Science Wednesday: Hi Reddit, we're Nick and Cori Ruktanonchai, and we published a paper in PLOS Computational Biology on how mobile phone data can target malaria elimination efforts -- Ask Us Anything!

PLOSSCIENCEWEDNESDAY R/SCIENCE

Hi Reddit!

I'm Nick Warren Ruktanonchai, a postdoctoral research fellow at the University of Southampton. I'm interested in understanding how people move, which helps us predict when, where, and why some people become exposed to areas with infectious diseases. And I am Cori Warren Ruktanonchai, a PhD student in Geography & Environment at the University of Southampton--as you may have noticed by the names, I also happen to be Nick's wife! I'm interested in using spatial statistics to better locate pregnant women, mothers and newborns at risk of adverse health outcomes.

We recently published an article titled "Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data" in PLOS Computational Biology, mapping where people got malaria based on their travel patterns.

We combined data from 1.19 million mobile phones in Namibia with a map of malaria prevalence to predict areas where the most people get infected. We hope that by targeting these hotspots, elimination efforts can both send help where it's most needed and reduce transmission nationwide.

Call and text locations from mobile phones are a great tool for knowing where people have been. AtFlowminder, we've used mobile phone data to not only help governments and NGOs predict the spread of disease, we've also used it to understand how people move after catastrophes, including a PLOS Medicine paper on the 2010 Haiti earthquake and a PLOS Currents paper on the 2015 Nepal earthquake.

We'll be answering your questions at 1pm ET -- Ask Us Anything!

Don't forget to follow Nick on Twitter at @nruktanonchai and Cori at @cwruktanonchai. Also, the Flowminder Foundation can be found at @flowminder, and the WorldPop Project at @WorldPopProject!

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:

DATE RECEIVED:
May 26, 2016

DOI:
10.15200/winn.146418.80643

ARCHIVED:
May 25, 2016

CITATION:
PLOSScienceWednesday , r/Science , PLOS Science Wednesday: Hi Reddit, we're Nick and Cori Ruktanonchai,

Hi! As someone who researches both Haiti and new media your article about the use of Haitian mobile phone data for modeling how people move after earthquakes was really interesting. I had some questions, though!

- How does metered mobile phone usage impact how you can track population movements?
- You say you only tracked phones with calls made - does that exclude calls received? (It costs to call but not to receive so many Haitians I know only take calls. But they are on their phone with family in the diaspora all the time!)
- How do gendered, age, and class differences in mobile phone ownership impact your results? Is there data on which demographics are more likely to be the one who would be carrying the phone? (extended families living together [a *lakou*] often share a single phone)

and we published a paper in PLOS Computational Biology on how mobile phone data can target malaria elimination efforts -- Ask Us Anything!, *The Winnower* 3:e146418.80643 , 2016 , DOI: 10.15200/winn.146418.80643

- In addition to your suggestions for using this kind of data to better conduct disaster management and infectious disease management, what other issues can this shed light upon? I'm thinking about displacement due to manmade disasters (war), modeling how mobs form and dissipate, tracking nomadic communities, etc. But I'm curious about other possibilities!

firedrops

NWR: Great questions!! I'll address one by one.

- The possibility of people using the phones more/less often at particular times/places is definitely a concern. It's possible that people use their phones more often when they're away from home, which would mean our analyses overestimate overall travel patterns. Or, maybe they call more often when they're in urban areas, partly because of better cell coverage. While we didn't do anything to explicitly account for this in this study (other than only using the last call/text of each day), we have done some thinking about how we might account for this. One possibility: In other datasets, the receiving SIM location is recorded, so we could look at SIMs that call really frequently (perhaps telemarketers or related groups) and use their calls as "pings" that are independent of where the receiver is, or what they're doing.

- In this particular dataset, we only have calls/text info from the originating telephone. So, those sorts of patterns (people who might only receive calls, for example) wouldn't be picked up in the dataset in this study, though there are other datasets where we could explore this effect. Another important factor in these settings that we haven't explicitly considered, I think, is the possibility of multiple SIM ownership--we assume that each SIM is one person.

- There are definite biases in these data, as most certainly mobile phone ownership is biased towards wealthier populations. Some previous studies such as this one by Wesolowski et al suggest that the movement patterns found in these data are reasonably robust to these biases, but it's definitely still a concern. As a bit of a tangent, the demographics of mobile phone ownership and of individuals in these data are a particular interest for me, though this dataset doesn't really inform much of that. We know that certain groups of people are particularly at-risk of vector-borne disease (including, for example, pregnant women when considering Zika), and we can hypothesize that they probably move differently from other subgroups of people. To fully incorporate those differences into predictions of mosquito-borne disease spread, we need to link movement patterns from mobile phone data to individual-level socioeconomic and other characteristics. It's not possible with this dataset but we could start addressing it with, say, top-up information, which could be a proxy for socioeconomic status.

- Yes! I totally agree with the displacement possibility, and indeed we're using similar datasets from Nepal (and other places) to understand how people respond to disaster. On top of that, if we can develop algorithms to identify when disasters are occurring (based on how people are moving in real-time) we could perhaps predict when disasters are happening, before the actual knowledge of it makes it through the traditional channels. I think it could also help with infrastructure development--people move through time, and census-based population estimates don't really capture this. If we build, say, hospitals or roads based on where we think people will be based on a static census estimate, those services could be underused during some seasons and overburdened during others. By using data like these to inform where people are *seasonally*, we could refine these population estimates through time and better design infrastructure. Here's a good example of where my colleagues have used similar data to map population densities seasonally in France and Portugal.

What percentage of the Namibian population has a mobile phone? Was it difficult to get access to the call data?

nate

CWR: Great question! The denominator is always an important consideration. According to the Demographic and Health Surveys (freely available, nationally representative surveys of households across countries which provide a range of indicators for monitoring & evaluation in the areas of population, health, and nutrition), 88.5% of households in Namibia possessed a mobile phone in 2013. Now that comes with its own caveats, most importantly that this number represents households and not individuals, but we know that mobile phone ownership in low- and middle-income countries is only continuing to grow over the years.

Within our specific data set, we worked with the leading mobile provider in Namibia, Mobile Telecommunications Limited (MTC), who reported 1.5 million subscribers in 2011, and a 90% market share. These numbers, too, have caveats including multiple SIM ownership.

Not surprisingly, it was difficult to get initial buy-in from the mobile operators for privacy and security reasons, but we worked with them to set up security protocols and ethical considerations. Now that we've worked with them some, they've really started to engage with us because they've never really thought about using these data in public health applications (as opposed to simply market research, for example).

What are some of the movement trends you noticed from using mobile data?

And how exactly do you guys interpret the data to reach those solutions?

watches_u_poop

NWR: In this analysis, we looked at the locations associated with the calls/texts for each SIM, and assigned proportions of time to each tower. So, for example, if someone made a call from tower A on Sunday, then tower B on Tuesday, and a third call from tower on Wednesday with no other calls in the week, we assumed they moved halfway between the calls--which, for this example means they spent Sunday, part of Monday, and Weds-Sat at A, and part of Monday and Tuesday at B. (So, overall, ~75% of their time was at A, and 25% at B)

We did this for every SIM, and gave each SIM a "home" location based on where they spent the most time. Then, we averaged across everyone who shared a home location to get an idea of what that community's movement patterns were like, and then inferred where people got infected on a community-level basis.

So, given this particular way of analyzing the data we did find some really fascinating patterns! For example, not surprisingly, people were very likely to spend time in Windhoek (the capital and largest city), but there were also regional patterns--northern Namibia, near the Angolan border was pretty well interconnected. The Zambezi region (the panhandle) was also quite interconnected and less connected to Namibia as a whole, which reflects people in those regions traveling potentially more often to the neighboring countries of Zambia, Botswana, and Zimbabwe than to the rest of Namibia.

Analyzing these data in other ways we've found really fascinating patterns as well! For example, just looking at movement of SIMs over time, there was a really big movement north in December (from Windhoek to the populated area north of Etosha national park), and then a corresponding movement south in January... which likely has to do with holiday-related travel.

Hello, not sure if off-topic but I'll ask anyway. This question is somewhat aimed at Nick, but I'll be happy for any response.

Just from looking at your surname, I could tell that you were from Thailand. Because Thailand is rarely heard of in the scientific field, I would like to ask some questions regarding you and your experiences. If you have not lived in Thailand, then I apologize for making the assumption.

How was your experience studying in Thailand? What do you feel about the current state of science education in the country? What do you think can be done to improve the quality of education for prospective students? Where do you think southeast Asia stands in the global scientific field? Do you think that your data can be of use to improving the public health of other countries?

And, finally, what advice would you give to a freshman biology major?

LesserWraith

NWR: Good call! :) I'm Thai-American actually. My parents moved to the US from Thailand before I was born, so unfortunately I can't really speak too much to your questions!

I have worked with folks in the region (and most of my family lives in Bangkok), so I can say it has been really inspiring to see the growth of science in Thai universities. There's of course room for improvement, but in particular, completely anecdotally, there's been a real move towards and enthusiasm for quantitative biology, which I think is going to be a really important direction for biology into the future.

To the freshmen biology majors out there: Keep an open mind! I absolutely *hated* math for pretty much the first twenty or so years of my life--but then in starting my PhD, realized it could actually be applied to public health and infectious disease, and now the interface between biology, health, and mathematics is fascinating for me. One great thing about the increasing interconnectedness of science (with the internet and other tools) is it means more and more interdisciplinary work. Keeping your mind open to potential links between things that might appear really different will help you make the connections that us old fogeys can't see, that will be fundamental for science into the future!

How close are we to eliminating Malaria, and can your research and methods be used to attack the Zika virus that plagues Latin/South America?

(And can you give us a TL;DR?)

Aquilian_Nights

TL;DR: We can tell where people get malaria from by tracking their calls and texts.

How close are we to eliminating Malaria, and can your research and methods be used to attack the Zika virus that plagues Latin/South America?

(And can you give us a TL;DR?)

Aquilian_Nights

CWR:

Regarding malaria elimination: Not close enough.

Regarding Zika: This framework could definitely be applied to Zika virus! (With modifications, of course) In fact, it can be applied not only to any vector-borne disease, but to any disease that has an underlying environmental reservoir or spatial component.

Conceptually, it's a pretty straight forward epidemiological investigation... you can think of it similar to a

food-borne outbreak. In that situation, epidemiologists would ask, "Okay, what foods have people consumed, and are there foods in particular that sick people ate in common?" Here, we basically ask, "Okay, where have you been in the last year, and are there particular spatial locations that sick people have in common?" So you can see how this sort of question could be applied to malaria, to Zika, to cancer, to lots of stuff!

As an aside, though, I've actually been recently involved with some really great Zika modelling work that other teams are leading at Oxford and the University of Notre Dame. I'd encourage you to check them out!

Messina et al. 2016

Perkins et al. 2016

In addition to gathering data and identifying trouble zones for Malaria transmission, are you at all involved in prescribing solutions based on those findings? Ex: Are you suggesting certain areas be avoided, contacting health organizations, etc?

atbrow00

CWR: We're definitely keen to make sure our research has some policy relevance! While we ourselves avoid coming in and telling countries exactly what to do and how to spend their money and how to make their decisions, we work closely with the Clinton Health Access Initiative (CHAI), among other organisations, as an intermediary for policy advocacy and development. Their role is to act as a bridge between the researchers (us) and policy makers in countries like Namibia, and to translate our findings into the evidence that can be used to make policy decisions. In fact, one of the co-authors (C Lourenço) works with CHAI and was employed for several years in Namibia with the Malaria Control Programme there.

When are you going to do the same for influenza?

nesrekcajkcaj

NWR: The challenge with using this sort of method for a disease like the flu is our analysis assumed that we could identify where people were infected--and that these areas are places where transmission is intrinsically higher. That makes sense for a disease like malaria, because some areas have lots of mosquitoes, and some have none, but with a directly transmitted disease like influenza, location plays less of a role.

That said, human movement is still really important for influenza, and knowing how people move with the flu could help us predict where it'll go next. That's not been a particular emphasis for us (for myself, I've mostly worked with mosquito-borne disease so haven't really personally done anything with influenza), but it is something that people are investigating, if not with mobile phone data, then other sources of information on human movement, like air traffic data.

You've got data from 1.19 million mobile phones in Namibia...How do you address privacy concerns from having data from that many people's phones?

Aquilian_Nights

NWR: Great question! The data we get are completely anonymous--the data we have consist of: a randomized unique ID for the SIM, the time of each call/text event, and the cell tower that the call/text

was routed through. One of the first things we did was combine all cell towers within urban centers into one, so we don't discern location within cities. (This wasn't necessarily just for privacy reasons, it was also for analytic reasons--if you're in a city, you might get routed randomly through one of many different towers within a 5km radius).

That said, even this is quite a bit of individual-level information, and understandably mobile operators and regulatory bodies are wary of the information still contained within these data! So, now we also store these data only locally within the mobile operator's building, on a server that is behind the mobile operator's firewalls. We then do any initial data processing on that in-house server, and make sure that we don't take any individual-level data out. Flowminder has since helped codify these standard practices for such highly sensitive data in collaboration with other organizations and the GSM Association. You can see the general guidelines for working with these data here: http://www.gsma.com/mobilefordevelopment/wp-content/uploads/2014/11/GSMA-Guidelines-on-protecting-privacy-in-the-use-of-mobile-phone-data-for-responding-to-the-Ebola-outbreak-_October-2014.pdf.

In terms of the particular dataset used for this study, they predate the establishment of these guidelines--using mobile phone data is very new and exciting and we are actively sorting out how to best use them while protecting privacy and storing them securely! So, we stored the data for these analyses locally at the University of Southampton on secure servers in a way signed off by the Ethics and Research Governance body at the university. As per these new GSMA guidelines, though, that's no longer the case and all the individual-level data are stored on MTC's servers.

How does your research differ from that of Caroline Buckee?

TigerBalmer

NWR: We collaborate with Caroline quite a bit! In fact, Elisabeth (one of the coauthors and our collaborators at Southampton) has been regularly visiting Harvard to work on similar datasets from Namibia. So, there's a good bit of overlap (well, collaboration) there. If I were to say how this line of research differs from the sorts of things Dr Buckee's work, I'd say while they're doing stellar work using models of human mobility to understand the factors that drive movement and applying that in disease systems, this line of research takes the movement patterns *per se* and contextualizes them within a model of malaria transmission.

We don't do anything with the mobile phone data to understand the underlying patterns--we just plug them into an understanding of malaria transmission, and while this ignores a lot of the valuable information present within these movement patterns, I also think it makes minimal assumptions about how movement depends on different factors (as a movement model would necessitate). It also complements Dr Buckee's work quite well (at least, in my opinion!)

Hello. I am a biologist and have focused for the past 5 years on the major malaria vector mosquito in Africa (Anopheles gambiae). As you may know, mosquitoes aquire/transmit malaria primarily while people sleep. Your analysis of the mobility data does not appear to take this time-of-day parameter into account. Can you elaborate on this?--are you only focusing on moments that occur during hours of peak transmission?

RabidMortal

CWR: You're absolutely right about the malaria vector biting predominantly at night. To account for this, our analyses only used the last call/text of the day, with the idea that this is the best proxy we have for an individual's home location, or at least where they spent the night. This probably works

reasonably well for call/texts made at say, 10pm, but it definitely might be less certain for calls/texts made at say, 2pm.

Hi, we were just discussing what must be your work at an OSAC meeting hosting a joint CDC, AFRIMS and Thai Ministry of Public Health event on corporate pandemic risk management. So thanks and keep it up!

Why did you choose Malaria? Is your work applicable to diseases that spread between humans?

upvotersfortruth

CWR: Wow, that's really exciting, and thanks for the question! I sort of addressed it a little above, but the short answer is yes, it can definitely be applicable to other diseases, with modifications to the underlying model and model assumptions.

Conceptually, it's a pretty straight forward epidemiological investigation... you can think of it similar to a food-borne outbreak. In that situation, epidemiologists would ask, "Okay, what foods have people consumed, and are there foods in particular that sick people ate in common?" Here, we basically ask, "Okay, where have you been in the last year, and are there particular spatial locations that sick people have in common?" So you can see how this sort of question could be applied to malaria, to Zika, to cancer, to lots of stuff!

We used malaria in this context because 1) the Ross-Macdonald model is a well-established model for malaria transmission, 2) Namibia is very close to malaria elimination and therefore presented a useful case study, and finally, 3) simply from a pragmatic point of view, Nick is most familiar with malaria as a disease system because it comprised his PhD work.

Does the study account for other factors such as local geography and weather?

aconitine-

NWR: It doesn't, but for me, that made the results particularly fascinating. The areas predicted to have lowest transmission reasonably well correlated (at least on an eye test) with the areas we'd expect to have low transmission--such as the Skeleton Coast, and the drier areas southeast of the Etosha area (those areas being so dry that we wouldn't really expect many mosquitoes at all). So, to me it was encouraging that even though we didn't incorporate any environmental factors like rainfall or temperature into our model, this method still predicted that dry areas probably don't have much transmission!

Was movement data utilized strictly as it occured, or was there any prediction or pattern development for partially active users involved?

For example, if a particular cell phone stopped producing data midway throughout the time range evaluated, was there any filtering or weighting taken into account versus a cell phone which produced consistent or regular results over the entire time range being examined ?

BODHISATTVABDO

NWR: We didn't do anything too complicated, partly by design--by using the data mostly as-is, we were thinking to preserve whatever patterns were in the data without smoothing over them with an associated model. However, if people disappeared from the dataset, we did remove them from the dataset, and we only initiated the analysis with the first call/text each SIM made. If there were big gaps

between calls/texts, we simply assumed they moved between towers halfway between call/text events.

Hi! This seems cool, but it seems like the most impoverished areas (which are most likely to lack phones) will probably have quite a bit of malaria. Do you have a plan for this?

BookAnnelida

CWR: Thanks for the question! Mobile phone ownership is actually exploding around the world, particularly in areas that are most at risk of malaria. For example, in Africa there were over 400 million mobile phone subscribers in 2010 alone, and in Namibia 88.5% of households reported mobile phone ownership in 2013 according to the DHS.

These numbers are only continuing to grow, as mobile phones become cheaper to make and produce, and the good news is that this technology is diffusing into even the most poor and most rural of areas. In fact, there has been a lot of research in the public health community around trying to utilise this explosion of technology, and more increasingly we're beginning to see "mHealth" type of interventions that can bring services directly to historically neglected populations.

Check out the UN Foundation for more!

I'd like to know more about the predictive abilities of your findings. Are you using the mobile users' movement data to determine where the next malaria outbreaks will occur?

cooterwoober

NWR: What could be considered outbreaks in Namibia are mostly caused by infected people coming in from neighboring countries. We don't have any way of predicting how people move across national borders (the phone data is restricted to the mobile operator we have data from, and mobile operators generally only work within a given country), so we haven't really tried to predict where outbreaks will start. That said, what we *can* do is predict what happens *after* they start, and we're working on some simulation models addressing exactly that: If malaria shows up somewhere, where is it going to go next?

Nick, aren't you from Pike County, Kentucky :) - Keep on killing it, brother.

lawdog22

NWR: Word. We need more folks from Appalachia in public health! There's no malaria of course, but public health issues are an important concern for the region, and who knows about them better than the children of coal country?

Does human mobility data qualify as a "target"?

jbsinger

NWR: The targets we talk about are places where we could prioritize for bednet distributions, or drug allocation, or other interventions. So, the mobility data themselves aren't a "target", but they help us identify areas with the most malaria transmission, which are spatial targets for malaria control efforts.

Whats your career goal? Are you hoping to be professors?

bobthemagiccan

CWR: I'm only in the beginning stages of my PhD, so I imagine if you were to ask me this question again in 3 years it might be radically different! But for me, personally, I really enjoy working with policy makers to help inform the decision making process, and I could definitely see myself working with governments toward that end. That being said, I think teaching undergraduates oftentimes gets undervalued by overworked professors within a publish or perish culture. It's an incredibly impactful and rewarding avenue, so there's a lot of appeal for me there, as well.

Whats your career goal? Are you hoping to be professors?

bobthemagiccan

NWR: I'd say so... Academia provides a nice outlet for researchers to work with policymakers, but also have some time for just pursuing interesting questions because they're cool... and those can sometimes lead to the most impact!

Computational biologist here.

Where did you get the data? What kind of cleaning was necessary? Is it something that anyone would have access to?

southernstorm

NWR: We got the data through a data-sharing agreement with the mobile operator in Namibia, MTC. Unfortunately (well, perhaps fortunately if you think from the perspective of the user!) these data are highly proprietary and sensitive, so they're not generally made available for wider use without similar sorts of agreements. We'd like to fit movement models to these data and make the fitted models available for others to use, however, which won't be ideal, but better than nothing! I should note that we've provided aggregated versions of the data as a supplement on the paper--not the individual-level information (or even tower-level), but at least pairwise connectivities at the second administrative unit level.

I (thankfully) wasn't that involved with the data cleaning, but it definitely required a good bit of legwork at first! Even the geographic coordinates of the towers was sometimes misreported, so someone had to go in, make sure those matched, make sure any nonsense results were addressed (such as days with zero calls reported across the entire country), etc.