**Research Strategy**

**INSTRUCTIONS:**

*Instructions are taken directly from the* [*NIH SF424 Application Guide*](https://grants.nih.gov/grants/how-to-apply-application-guide/forms-d/general-forms-d.pdf)*. For internal use only, do not distribute. Please delete prior to submission.*

**Format:** 6 page maximum, 11pt font or larger (suggest fonts - Arial, Garamond, Georgia, Helvetica, Palatino Linotype, Times New Roman, Verdana), at least 0.5” margins, single column formats are highly encouraged. Attach this information as a PDF file.

**Content:** Organize the Research Strategy in the specified order and use the instructions provided below unless otherwise specified in the FOA. Start each section with the appropriate heading – Significance, Innovation, Approach. Cite published experimental details in the Research Strategy attachment and provide the full reference in R.220 - R&R Other Project Information Form, Bibliography and Reference Cited.

**Note for Applications Proposing the Involvement of Human Subjects and/or Clinical Trials:**

* Use the Research Strategy section to discuss the overall strategy, methodology, and analyses of your proposed research, but do not duplicate information collected in the PHS Human Subjects and Clinical Trials Information form.
* The PHS Human Subjects and Clinical Trials Information form will capture detailed study information, including eligibility criteria; inclusion of women, minorities, and children; protection and monitoring plans; and statistical design and power.
* You are encouraged to refer to information in the PHS Human Subjects and Clinical Trials Information form as appropriate in your discussion of the Research Strategy (e.g., see Question 2.4 Inclusion of Women, Minorities, and Children).

**Note for Applicants with Multiple Specific Aims:** You may address the Significance, Innovation, and Approach either for each Specific Aim individually or for all of the Specific Aims collectively.

1. Significance

* Explain the importance of the problem or critical barrier to progress that the proposed project addresses.
* Describe the scientific premise for the proposed project, including consideration of the strengths and weaknesses of published research or preliminary data crucial to the support of your application.
* Explain how the proposed project will improve scientific knowledge, technical capability, and/or clinical practice in one or more broad fields.

2. Innovation

* Explain how the application challenges and seeks to shift current research or clinical practice paradigms.
* Describe any novel theoretical concepts, approaches or methodologies, instrumentation or interventions to be developed or used, and any advantage over existing methodologies, instrumentation, or interventions.
* Explain any refinements, improvements, or new applications of theoretical concepts, approaches or methodologies, instrumentation, or interventions.

3. Approach

* Describe the overall strategy, methodology, and analyses to be used to accomplish the specific aims of the project. Describe the experimental design and methods proposed and how they will achieve robust and unbiased results. Unless addressed separately in the Resource Sharing Plan, include how the data will be collected, analyzed, and interpreted, as well as any resource sharing plans as appropriate.
* Discuss potential problems, alternative strategies, and benchmarks for success anticipated to achieve the aims.
* If the project is in the early stages of development, describe any strategy to establish feasibility, and address the management of any high risk aspects of the proposed work.
* Explain how relevant biological variables, such as sex, are factored into research designs and analyses for studies in vertebrate animals and humans. For example, strong justification from the scientific literature, preliminary data, or other relevant considerations, must be provided for applications proposing to study only one sex.
* Refer to [NIH Guide Notice on Sex as a Biological Variable in NIH-funded Research](https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-102.html) for further consideration of NIH expectations about sex as a biological variable.
* If your study(s) involves human subjects, the sections on Inclusion of Women and Minorities and Inclusion of Children can be used to expand your discussion on inclusion and justify the proposed proportions of individuals (such as males and females) in the sample, but it must also be addressed here in the "Approach" section of the "Research Strategy" attachment.
* Point out any procedures, situations, or materials that may be hazardous to personnel and the precautions to be exercised. A full discussion on the use of select agents should appear in the Select Agent Research attachment.
* If research on Human Embryonic Stem Cells (hESCs) is proposed but an approved cell line from the NIH [hESC Registry](https://grants.nih.gov/stem_cells/registry/current.htm) cannot be chosen, provide a strong justification for why an appropriate cell line cannot be chosen from the registry at this time.

**RESEARCH STRATEGY**

**General background**

**The impact of dengue in Latin America and Honduras.** Dengue hemorrhagic fever (DHF) is one of the most abundant and rapidly-spreading vector-borne diseases globally. Its causative agent, the dengue virus, is transmitted mainly by *Aedes aegypti* mosquitoes, which also transmit chikungunya, yellow fever and Zika. The majority of infections are asymptomatic or result in a brief flu-like illness, but a small proportion of patients, often those with prior exposure to dengue virus, develop a potentially lethal complication called severe dengue (WHO 2018). There is no specific treatment for dengue/ severe dengue, but early detection and proper medical care lowers fatality rates below 1%. Severe dengue is a leading cause of serious illness and death among children in Latin America. From 1952 to 1965, 19 South and Central American countries were free of *A. aegypti*, but dengue re-emerged in the late 60s and spread dramatically throughout the region (Tapia-Conyer 2012). The number of dengue cases in Latin America has risen from 1,033,417 annually in the 1980s to 2,725,405 in the 1990s, and 4,759,007 between 2000 and 2007. Between 2001 and 2009, Venezuela, Brazil, Costa Rica, Colombia, Honduras and Mexico accounted for more than 75% of all cases in the region. Honduras is one of the countries in central America with the highest incidence of this disease, and suffered its worst epidemic of dengue dengue in 2010 (Brathwaite 2012), with 66,814 suspected cases nationally (3,266 severe cases), primarily young individuals between the ages of five and nineteen, and 83 deaths. A second major epidemic occurred in 2013 with a total of 34,803 suspected cases (COHA 2013, PAHO 2018).

**Dengue diagnostic and surveillance capacity:** The research in this proposal will be carried in collaboration with the University Teaching Hospital (HEU) at the Autonomous National University of Honduras (UNAH) in Tegucigalpa (www.hospitalescuela.edu.hn). HEU, inaugurated in 1978, is a national reference hospital and a center for pre and post-graduate medical education. It is the only hospital in Tegucigalpa, the capital of Honduras, with emergency service in all specialties, including infectology (infectious disease), for children and adults with 24 hours coverage and care throughout the year. Currently, dengue PCR diagnostics for suspected cases are conducted at UNAH’s reference laboratory, and complemented with RDTs at private clinics. Test results are reported to a national database, while epidemiological information (e.g. patient’s demographics, signs and symptoms) are recorded in paper forms kept at HEU. To serve the general population in the Central District, HEU offers three Peripheral Clinics (CLIPER) that provide support in the resolution of low complexity emergency pathologies. With regards to incidence, up to week 32 of 2018, a total of 104 hospitalized (171 ambulatory) cases were treated at HEU compared to a total of 76 hospitalized cases reported in 2017, a 27% increase. Of the total cases, 88% (92/104), were reported in children under 18 years of age. Women account for 40% (42/104) of the incidence and men 60% (62/104). 83% (86/104) are from the Francisco Morazán Department where Tegucigalpa is located, with remaining cases reported from the rest of the country.

**Per-aim Strategy**

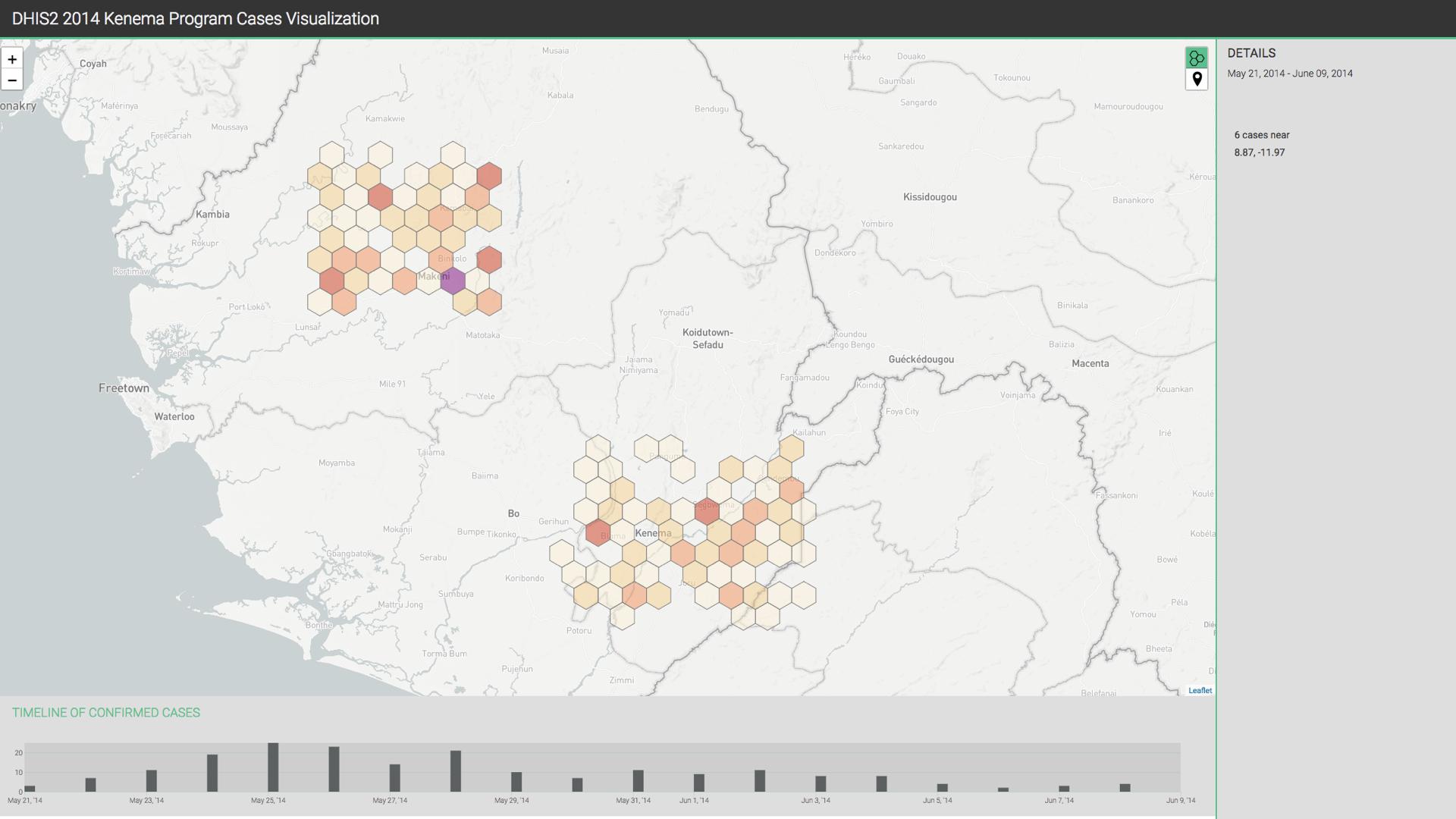
**Aim 1: Develop an open-source platform for collecting and integrating laboratory diagnostics results with epidemiological data from suspected dengue cases reported at the University Teaching Hospital.**

**Significance:** Diagnostic results are not effectively integrated with epidemiological data at HEU. Suspected case counts from the clinic are communicated to the infectology department by fax on a weekly basis. When laboratory testing is unavailable at HEU, which can happen in over 50% of cases, clinical suspicion is not quickly resolved. In that situation, patients need to get tested at private clinics and bring their results back to HEU. Clinics can communicate test results directly to HEU by email, but this is uncommon and poses privacy risks. A digital data-entry platform available from web and mobile will allow secure synchronization of clinical and laboratory data, faster coordination across clinics and laboratories, and will make it easier to conduct analyses for actionable insights. This level of data integration would be highly valuable: *“In a digital age, it is possible to link data from diagnostic laboratories and POC test readers and devices to provide data on testing coverage, disease trends and timely information for early warning of infectious disease outbreaks to inform design or optimisation of disease control and elimination programmes. Data connectivity also allows control programmes to monitor the quality of tests and testing, and optimise supply chain management; thus, increasing the efficiency of healthcare systems and improving patient outcomes.”* (Peeling 2015)

**Innovation:** Our proposed platform, *DengueData*, will complement efforts in accelerating dengue diagnosis and linkage to treatment at HEU by improving the expediency of testing services and strengthening the quality management systems. Specific examples include the use of validated testing algorithms and strategies, retesting individuals diagnosed positive before initiating treatment, and providing clinicians with auxiliary information to identify false negative results, particularly from RDTs (Miller 2015, Johnson 2017). In order to maximize capacity building in Honduras, we will follow the EpiHack model (www.epihack.org) of workshops, which engage local stakeholders in the process (developers, public health officials, healthcare providers, researchers). This has been applied in various settings across the globe, to develop syndromic surveillance tools such as *SickSense* in Thailand and *AfyaData* in Tanzania. These tool aims at faster disease detection to prevent and decrease the impact of outbreaks (Susumpow 2014, Karimuribo 2017). We will organize smaller-scale workshops and engage medical trainees and laboratory staff from HEU, UNAH researchers, and representatives from the Honduran Secretary of Health, to discuss the specific features of the platform and integration with existing national databases. This model of user-centered design for health information systems is fundamental, as engagement, perceived functionality, perceived ease of use, and efficient information management have been identified as cornerstone concepts for the acceptance of these systems (Kim 2012).

We plan to adopt the District Health Information Software version 2 (DHIS2 2018) as the data back-end of *DengueData*. The Malaria Case Surveillance Application (PSI 2016) is a DHIS2-based existing platform that provides a relevant reference for this aim. This mobile application allows the reporting of positive cases of malaria detected throughout Cambodia in near real-time. The application is used by healthcare providers in the field after they have conducted a Malaria RDT test. Test results and basic information –location, type of malaria detected, age/gender- are uploaded into a DHIS2 database. Once this data is pushed to DHIS2, a surveillance team is able to to track malaria cases and immediately identify hot spots, liaising with the regional and national health agencies to determine the appropriate course of action. The app reduces the time to move data through the system and also alleviates data-entry errors. Adoption of DHIS2 for malaria indicator data reporting in other countries has shown sustained improvements in the completeness of data (Githinji 2017). Motivated by such systems, we envision a unified mobile/desktop interface to upload both diagnostics and epidemiological data, at the testing laboratories and the hospital respectively, while reducing error and increasing completeness. We require a more complex data model than that of the Malaria application in order to include all relevant epidemiological data currently recorded in paper forms, while also ensuring that different sources of dengue diagnostics (PCR from the reference laboratory and RDTs from private clinics) can be properly harmonized. In that regard, research on smartphone based RDT readers (Priye 2017, Laktabai 2018) provides another useful reference in the design and implementation of the data-entry functionality for dengue diagnostics.

*DengueData* will provide a visualization and reporting module, *DengueVis*, that will allow evaluating key indicators at HEU. First of all, the introduction of new dengue RDTs in the field, including CRISPR-based assays from our lab (Myhrvold 2018), requires validation against the gold standard represented by PCR and strong safeguards against false positive results. *DengueVis* will allow us to quickly compare PCR and RDT diagnostics in detail, and to control for confounding factors such as testing clinic and patient characteristics. Despite growing use of RDTs in research and clinical settings, there is still limited data to demonstrate usefulness under routine conditions and cost effectiveness (Osorio 2015, Lim 2017). Our platform could provide evidence to that end, especially when combined with the available epidemiological data. Existing visual analytics platforms for infectious diseases surveillance and response in LMICs demonstrate the impact of such platforms (Ali 2016).

**Approach:** Our general approach to implement the HEU platform will consist of involving the local stakeholders from the beginning and throughout the entire process of design, development, and rollout of the platform. We expect this process to occupy months 1 through 12 of the grant. We will conduct an initial quality assessment of the data as it is handled with the current protocols at HEU, in terms of completeness and errors (months 1-3). We will also quantify metrics relevant for medical practice, such as the overhead due to handling of paper forms and delays in receiving test results from the laboratory or clinics. Following this initial quality and performance assessment, we will organize a series of participatory workshops at UNAH where HEU trainees, technical staff, researchers, and representatives from the Secretary of Health will be invited to come together to discuss the specific needs that the new platform should satisfy in terms of technical requirements, data management regulations, and expected reporting functionality (month 4). These workshops will result in the creation of “user journeys” (McCarthy, 2016) and culminate with the synthesis of the information in a service blueprint where the relationships between different service components —in this case, patients, hospital staff, data entry devices, existing clinical protocols— are directly tied to the instances where the users interact with the system in a specific journey. Once the reference performance and service blueprint of the platform are defined, we will use the DHIS2 data model and web Application Program Interface (API) to implement the mobile/desktop data entry interfaces, and a new epi+diagnostics visualizer that integrates clinical case records with results from PCR and RDTs (months 5-10). Extending existing open-source technologies already in use, in particular DHIS2, will ensure long-term maintenance and local governance. We will create a new set of web-based apps, to complement the default visualization tools in DHIS2. We already have in-house experience using DHIS2’s data model and web API (Figure 1), which will allow us to start developing working prototypes quickly and iterating the interface design. Months 11 and 12 will be devoted exclusively to evaluating the platform by conducting usability studies and surveys with the trainees from HEU, who will be the primary users of the platform at this stage. Their feedback will inform further iterations of testing and refinement, which will also be informed by re-evaluating the quality and performance metrics obtained initially before introduction of the platform (completeness, error, overhead).

***Figure 1:*** *Prototype of a DHIS2-based tool web visualization of epi+genomic data from Ebola cases in the 2014-16 outbreak in West Africa, developed in-house at the Sabeti lab. Source code available at:*

*https://github.com/broadinstitute/sabeti-dhis2-clinical*

**Aim 2: Identify the barriers and challenges of mobile-based participatory syndromic surveillance systems in Honduras and pilot a new self-reporting app for dengue symptoms.**

**Significance:** Passive dengue surveillance provides a baseline for disease surveillance that can be strengthened with novel tools and indicators such as electronic event-based syndromic surveillance and shifts in dengue serotypes/genotype (Runge-Ranzinger 2014). The first two decades of the 21st century have seen the emergence of participatory syndromic surveillance systems for infectious diseases. These systems capture voluntarily submitted syndromic data from the general public, which can be aggregated and communicated in near real-time. This capacity is a major advantage over traditional healthcare-based surveillance, which is affected by delays in seeking treatment by sick individuals, manual reporting at various administrative levels, and laboratory testing prior to notification. Existing mobile platforms for influenza-like illnesses (e.g. *Flu Near You* and *Kidenga* in the US, *FluTracking* in Australia) could be adapted for dengue surveillance in Honduras. However, participatory surveillance systems face several challenges. Recruiting and maintaining participants is difficult, and reporting fidelity is variable, with some participants reporting sporadically over time and others very frequently. Moreover, by virtue of being inherently syndromic without any laboratory testing, data can lend itself to be inaccurate unless complementary data sources are incorporated: *“One of the key aspects that needs to be addressed for participatory surveillance systems to gain greater acceptance and credibility in the field of public health is data validation. This can be accomplished by comparison to traditional surveillance systems, or by laboratory testing of biological samples from participants reporting symptoms”* (Wójcik 2014).

**Innovation:** We propose the creation of a custom mobile app-based platform for dengue syndromic surveillance in Honduras, *DengueDoctor*, built upon two key efforts. The first is to fully engage with the potential users of a dengue surveillance app, by conducting user studies, surveys, and local outreach campaigns. Experience from existing participatory surveillance systems (Dalton 2017) will provide important insights on recruitment methods and principles that can enable user engagement and retention. These include “obvious design” and repeated usability testing of new and current participants of the system. Our second key effort will be to validate the data to be generated with the app; we will design a validation and notification protocol involving the three HEU peripheral clinics (CLIPERs), which conduct passive syndromic surveillance among the general population in the Central District in Honduras. Currently, communication of the surveillance data collected at the CLIPERs to the reference hospital is conducted over telephone or fax, which hinders rapid response. However, the personnel at these clinics have the needed medical and situational knowledge to assess the significance of voluntary symptom reporting. Tapping into this prior knowledge, we plan to introduce a novel validation loop into the proposed app platform, dubbed *DengueAlert*, by which clinicians will receive summary information of the symptoms being reported though the app. Based on their expertise and complementary data provided by the platform (e.g. geographical location of the reports, frequency of symptoms, etc.), clinicians will be able to apply their medical judgment to flag those reports that appear more urgent for immediate follow up. We will also consider automated decision systems (Sa-ngamuang 2018) to pre-screen reports and only alert clinicians on those that pass a certain detection threshold.

**Approach:** Developing a new app platform for participatory syndromic surveillance entirely from scratch falls outside the scope of this proposal, and also represents an unnecessary duplication of efforts, as several of these platforms already exist and are open source. Therefore, we plan to collaborate with the team behind the *Kidenga* self-reporting app for mosquito transmitted diseases (www.kidenga.org) to base our app on their backend infrastructure. Since *Kidenga* has been designed to monitor the spread of Zika, dengue, and chikungunya in Texas, Florida, Arizona, and California, with its team planning to deploy the system in Mexico and Costa Rica, our app for syndromic surveillance of dengue in Honduras would provide a complementary effort. Any improvements and additions resulting from this aim will be contributed back to *Kidenga*’s open-source code base. Initial work for this aim will be conducted during months 9-12, as aim 1 winds down, during which we will organize community outreach campaigns to recruit members of the general population in Tegucigalpa and surrounding areas at HEU to conduct user study interviews. We will also conduct regular workshops, surveys, and design sessions with the personnel at HEU and the CLIPPERs to determine the level of syndromic data required to make reasonable assessments on case follow-up and the need for clinic visits. This approach will allow us to identify factors that would influence interest in, and long-term use of, self-reporting mobile apps for dengue symptoms among the local population, and to obtain generalizable knowledge on local stakeholder engagement with participatory syndromic surveillance systems in LMICs. This will be complemented by the insights gained from the clinicians, not only from the perspective of understanding what summary data they need, but also what kind of notifications the users should receive from the app to increase its perceived usefulness, and therefore users’ continued engagement with the self-reporting app.

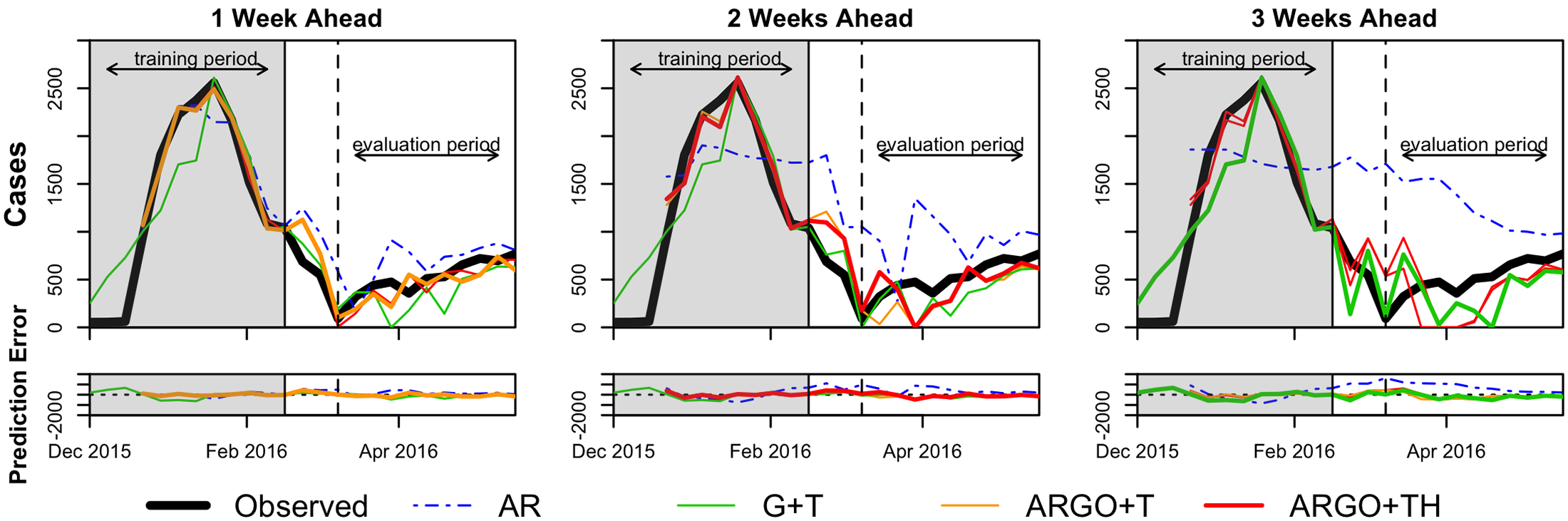
Months 13 - 18 will be occupied with the implementation of the app and accompanying server infrastructure. As mentioned, the app will be designed to satisfy the specific needs of the Honduran population and HEU staff, following the conclusions from the initial user studies. We will also incorporate the proposed notification mechanism. Clinicians at CLIPERs will receive anonymized summaries of reports being uploaded through the app, including location, basic demographics (age, sex), and presence of alert symptoms that may be indicating severe dengue cases. The clinicians will evaluate these reports and will decide whether the users originating them should be notified to come to the clinic for examination and testing, if needed. Implementation of this feature, in addition to enabling data-validation by contrasting self-reported symptoms with clinical assessment and diagnostic results, could result in increased user engagement and long-term adoption, since the app will be providing a tangible benefit to its users by facilitating communication with the healthcare workers at the CLIPERs. During months 19 - 24, we will begin a pilot rollout of the app, restricted to residents of the Central District region, and by the end of that period, we will have some preliminary data that would allow us to quantify its adoption and initial retention and compare with results from similar platforms (Baltrusaitis 2017).

**Aim 3: Integrating novel data sources into forecasting models to strengthen traditional epidemiological surveillance of dengue in Honduras.**

**Significance:** The ability to predict the evolution of a dengue epidemic is crucial to inform effective public health interventions, such as sustainable vector control and education of individual and household protection strategies (WHO 2012). Official government-led case count in Latin America is often plagued by delays of several days or weeks, which makes up-to-date monitoring of the disease difficult. Thus, timely forecasting systems are needed to evaluate interventions designed to reduce disease transmission. Due to the difficulty of parameterizing mechanistic models based on detailed knowledge of the biology of dengue virus transmission, models for large-scale dengue early warning systems have focused mainly on two components, temporal autocorrelation and association with weather and climate (Johansson 2017). Recent research (McGough 2017, Yang 2017) has demonstrated that flu and Zika forecasting models combining traditional surveillance data with novel sources such as internet search trends, social media, and news report data can reduce error in predictions and used as timely and complementary ways to assess the dynamics of the outbreak.

**Innovation:** We propose to develop an integrated forecasting model for dengue in Honduras, combining the data sources implemented in aims 1 and 2, namely hospital records, PCR and rapid diagnostics, and participatory syndromic surveillance, as well as potentially complimenting them with internet search, social media, and weather data (Zambrano 2012). The results from the models for flu and Zika mentioned earlier suggest that an integrated modeling strategy based on high-quality, timely-updated sources could enable vastly improved dengue outbreak prediction in the country (Figure 2). We conducted a systematic literature search but did not find any publications implementing such an integrated strategy for dengue in Honduras or elsewhere. We only identified models that rely on standard autocorrelation and weather time-series regression (Withanage 2018, Ramadona 2016), or internet searches (Guo 20167) and social media streams (Marques-Toledo 2017) in isolation. These types of models show mixed performance (Reich 2016) across different geographical regions and time periods, justifying the development of new integrative approaches.

**Approach:** The work in this aim will be carried out between months 13 and 24. The initial development of the models will take place as the diagnostics and epidemiological platform from HEU becomes operational, and the dengue self-reporting app is designed and developed. By month 19, data from the app should start to become available as well, allowing us to test a full model including all de-identified sources. Even before that time, we will be able to test reduced versions of the model using the data that is available. We will follow the methodology described in McGough 2017, by generating weekly reports of confirmed and suspected cases from the HEU platform, and then incorporating syndromic data from the self-reporting app. We will evaluate the integration of additional sources, such as weather time-series, Google search queries, Twitter microblogs, and HealthMap digital dengue surveillance (which aggregates news reports of the disease). Our ultimate goal is to develop a family of multivariable models to estimate and forecast weekly suspected cases of Dengue across the country, by considering different combinations of input variables represented in the data. The selection of the most predictive variables will be conducted using penalized regression. At each epidemiological week, we will analyze transformations of the input variables that would increase their correlation with the output variable, namely the predicted case count. In addition to the models that will use the aforementioned data streams as input, we will also develop a set of baseline models for comparison and context. We will consider models that only use historical observation of Dengue cases to predict cases on the subsequent weeks and models that incorporate information from the various data streams that we obtain from the work in aims 1 and 2.



***Figure 2.*** *Prediction results for Zika forecasting in Honduras using models combining traditional disease surveillance with search, social media, and news report data [18]. AR: autoregressive-only model, G+T: Google+Twitter data, ARGO+T: autoregressive with Google+Twitter data, ARGO+TH: autoregressive with Google+Twitter+Healthmap data.*

**Pitfalls and Challenges**

We described an integrative data platform for dengue surveillance that strongly depends on institutional and population buy-in. This poses challenges in terms of incorporating new digital systems into existing patient and data management protocols at HEU and local clinics, and engaging individuals into syndromic self-reporting on a long-term basis. In order to address these challenges, we are planning to apply a participatory, user-centered process of the tools and platforms proposed here. The EpiHack model will allow us to work closely with local institutional, technical, and scientific stakeholders to produce a set of specifications for the *DengueData* platform that reflects local needs and constraints. By adopting *Kidenga*’s data backend we will be leveraging on their expertise, however, we anticipate that the self-reporting app *DengueDoctor* will have to be significantly customized in order to be engaging and useful to the the local population. Smartphone adoption in Honduras is a key factor for the success of the app. It appears to be relatively high at least 40% according to reports from the National Telecommunications Commission (El Heraldo 2016), with overall overall mobile telephony penetration over 90% (CONATEL 2018). Capturing the appropriate syndromic data into the app is another challenge. We plan on applying a validation sample approach where we will enroll people in the waiting room of the HEU into the app, they type in their symptoms while they are waiting and then they see the doctor. The doctor could be blinded to who generates what data in the app and then they review those data later in bulk and make their assessments. Finally, aims 1 and 2 are independent, which allows us to still make progress with one even if we face unforeseen difficulties with the other. Aim 3 is dependent on aims 1 and 2, but work towards this aim could still be performed if only one of those first two aims is completed, since the forecasting models could be initially trained with fewer data inputs, and later extended to incorporate additional sources.

**Note:** The names *DengueData*, *DengueDoctor*, *DengueAlert* are preliminary and only meant to help with the review process. The final names of all the components of the proposed platform will be decided with the local stakeholders as part of the participatory design process.