PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Seattle Flu Study: a multi-arm community-based prospective study protocol for assessing influenza prevalence, transmission, and genomic epidemiology
AUTHORS	Chu, HY; Boeckh, Michael; Englund, Janet; Famulare, Michael; Lutz, Barry; Nickerson, Deborah; Rieder, Mark; Starita, Lea; Shendure, Jay; Bedford, Trevor; Adler, Amanda; Brandstetter, Elisabeth; Frazar, Chris; Han, Peter; Gulati, Reena; Hadfield, James; Jackson, Michael; Kiavand, Anahita; Kimball, Louise; Lacombe, Kirsten; Logue, Jennifer; Lyon, Victoria; Newman, Kira; Sibley, Thomas; Zigman Suchsland, Monica; Wolf, Caitlin

VERSION 1 – REVIEW

REVIEWER	Elodie Ghedin New York University USA
REVIEW RETURNED	16-Apr-2020

GENERAL COMMENTS	This is a very timely protocol for the detection and analysis of influenza--and other respiratory pathogens--in community and hospital settings from subjects presenting with ILI. The protocol is clear and well designed. The specimen collections (nasal swabs and environmental sampling at childcare locations) and the analysis methods proposed (molecular epidemiology studies, clinical assessments) will lead to a much clearer picture of influenza transmission in a metropolitan area, and help inform policy for intervention strategies. This protocol also provides a solid framework that can be repurposed for the surveillance of other emerging infections, like COVID-19.

REVIEWER	Stephen Kissler Harvard T.H. Chan School of Public Health, Boston MA, USA
REVIEW RETURNED	22-May-2020

GENERAL COMMENTS	This is a well-written and comprehensive protocol for an ongoing study to map the spread of influenza and other respiratory pathogens in Seattle. The authors rightly indicate that their findings will inform surveillance and response to both seasonal and pandemic outbreaks of respiratory viruses. This work is important, of broad interest, and is being conducted by a team with relevant expertise and a clear understanding of best practices for respiratory specimen collection and data analysis. I have a few minor comments and suggestions:

	In the Introduction, lines 33-43: How would rapid genome sequencing and the subsequent improved understanding of the strain diversity of influenza in a given season assist with targeted interventions? To my knowledge, the influenza strains that circulate in a given season are rarely distinct enough to merit different responses. I think that more clarity about the link between real-time viral genomic analysis and policy would help here. Line 46: "unprecedented intensity" - not quite sure what this means - is this "resolution"? The authors mention asymptomatic transmission in the second and third paragraphs of the introduction, but the study seems to only relate to symptomatic influenza (Line 53). Will there be any attempt to infer rates of asymptomatic infection? If not, this could be added as a limitation of the study. Related to the previous comment: there is no clear discussion of the study's limitations, either in the Abstract (the strengths and limitations appear to only include strengths) or in the Discussion. The protocol would benefit by pointing out these limitations, e.g. pertaining to asymptomatic infections, convenience sampling, restricted geographic scope, etc. Statistical methods (page 10, line 45): How will the authors account for test sensitivity/specificity? What are the sensitivities/specificities of the tests that they will use? General question: How is the protocol being adjusted for the current SARS-CoV-2 pandemic? Are there strategies in place to ensure the safety of those collecting specimens, or will the study be delayed? Will SARS-CoV-2 be added to the list of pathogens they're testing for? Given that the study is motivated in part by the need to improve pandemic response, some mention of the current pandemic would be worthwhile.

REVIEWER	Casey Zipfel Georgetown University, USA
REVIEW RETURNED	27-May-2020

GENERAL COMMENTS	Summary: Influenza surveillance typically relies on healthcare seeking, which results in many influenza cases being unreported, hindering our understanding of influenza dynamics and how to best curtail transmission through pharmaceutical and non-pharmaceutical interventions. This article describes the Seattle Flu Study, which aims to overcome typical limitations in influenza surveillance through a community-based study with multiple different arms that are integrated together to improve our understanding of influenza in near-real-time. They describe each arm and the protocols that each arm will utilize in great detail. They anticipate being able to provide spatial mapping, infection profiles based on demographic and immune characteristics, and gene sequencing of influenza viruses. This study is a huge undertaking that will greatly improve our understanding of influenza dynamics in a US city, and the study is thoughtfully designed, and likely to provide a wealth of data for future analysis. The many components and large quantity of resulting data, however, make their hypotheses and goals somewhat undefined and open-ended. This paper could expand upon the

	explanation of their choices, provide more context and motivation, expand on the integration of the different study arms, and more clearly define their expected results. The paper should also address the limitations of their study. Comments: • The article summary indicates that there will be a list of strengths and limitations. This list does not seem to include any limitations. This is something that is missing from the whole paper; addition of limitations and potential pitfalls in the discussion section will be a useful addition. • In the introduction, the novelty of individual-level influenza data from the community, and the associated demographic information, should not be overlooked. This will overcome the biases of healthcare seeking, which are very significant for influenza, and specific to the United States. Generally, I think overcoming this bias is one of the greatest strengths of this study, and I think this should be more fully considered in the statistical analysis. Can the researches weight the resulting data to account for their own testing biases? Will the results be representative of the population? Will the demographics of Seattle impact these findings? • The aims and hypotheses are not clearly defined and are too open-ended. More specific hypotheses regarding anticipated findings would be helpful to more clearly understand the methods and their motivations. Something that will help the reader to understand the hypotheses is more background on previous community based studies, and what this study provides that moves further than prior studies. The text mentions that integration of community based survey data with inpatient hospital and ambulatory care, which is a limitation of other studies, but it is not clear to me what this integration achieves and why it is important. • A clearer definition of the goals of the different branches would be very helpful. Is each branch characterizing influenza in a different population? What is the significance of this? How will they be integrated, and what will this reveal that other studies may have missed? • This paper is lacking in the discussion. What are the broader impacts of this study? What are the limitations? Is this design scalable or transferable to other areas? Specific Comments: • The expected findings related to asymptomatic individuals are somewhat unclear. The introduction mentions asymptomatic infection as a limitation that community studies can overcome, but then ARI symptoms are a requirement to receive a test in all of the branches. If this is not the case, and some of the branches are able to capture asymptomatic infection, this should be specifically identified. • Page 5, line 38 is missing a word. Possibly should be, “identify new and emerging influenza strains and their transmission dynamics” • The importance of the environmental swabs and their purpose is unclear. • For the data quality, the authors could consider dataset validation, by comparing the influenza dynamics in their data with another dataset, like the CDC influenza testing or influenza-like illness tracking. If both datasets provide the same spatiotemporal signals, that will increase confidence in the representativeness of this dataset. • A suggestion in the statistical methods: the authors could consider using an n-mixture model, which accounts for measurement

	processes based on measurement effort and coverage. • Page 11, line 33: Missing a word. Possibly should be “if they experience any participation-related harm.” • Will the TaqMan RT-PCR Test be able to test for COVID-19? If so, this could be added to table 2, since many of symptoms for COVID-19 would fall into the symptoms. • Will test results be communicated to those tested? This could be important to limit transmission.

VERSION 1 – AUTHOR RESPONSE

Reviewer 2 Comments:

In the Introduction, lines 33-43: How would rapid genome sequencing and the subsequent improved understanding of the strain diversity of influenza in a given season assist with targeted interventions? To my knowledge, the influenza strains that circulate in a given season are rarely distinct enough to merit different responses. I think that more clarity about the link between real-time viral genomic analysis and policy would help here.

> **Response:** We agree that on a virologic level for an individual, the differences are unlikely to change clinical management, but our intent is to utilize deep sequencing and phylogenetics to understand how influenza is spreading through a population to understand, at a population level, which groups should be targeted for public health interventions. One example of this is the work of Dr. Bedford and others from our group on SARS-CoV-2 using the NextStrain platform, which allowed rapid understanding of how SARS-CoV-2 was being introduced into cities and which outbreaks were epidemiologically related, without having to do extensive contact tracing. We have added a reference to NextStrain to help readers better understand the type of real-time platforms we will use for influenza (19).
>
> The benefits of rapid genome sequencing have also been further highlighted in the Discussion section of the manuscript:
>
> *“We utilize broad, multiplex molecular testing for viral and bacterial pathogens to characterize the molecular epidemiology of respiratory pathogens in the community. For samples with influenza detected, we perform whole genome sequencing to evaluate transmission patterns within a community, and identify target populations for public health interventions. This strategy is translatable to the current SARS-CoV-2 pandemic. We utilized this platform to first identify community transmission of SARS-CoV-2 in the United States (19, 21).”*

Line 46: "unprecedented intensity" - not quite sure what this means - is this "resolution"?

> **Response:** Thank you for this comment. We intended to emphasize the magnitude of sample collection and have re-written this line to clarify the intended message so that it now describes the Seattle Flu Study as *“a multi-armed, regional study of influenza at a city-wide scale that integrates community, ambulatory care, and inpatient surveillance at a large magnitude.”*

The authors mention asymptomatic transmission in the second and third paragraphs of the introduction, but the study seems to only relate to symptomatic influenza (Line 53). Will there be any attempt to infer rates of asymptomatic infection? If not, this could be added as a limitation of the study.

> **Response:** Thank you for these comments. In the second year of our study, we began enrollment of asymptomatic participants. However, that collection does not fall within the

study period described in this paper. We have removed the mentions of asymptomatic transmission from the introduction and added this as a limitation of Year 1 of the study:

"Our community-based sampling required participants to have at least two ARI-associated symptoms, and therefore did not allow us to examine the role of asymptomatic respiratory viral transmission. To address this limitation, we broadened the community-based eligibility criteria to include individuals without symptoms (22)."

Related to the previous comment: there is no clear discussion of the study's limitations, either in the Abstract (the strengths and limitations appear to only include strengths) or in the Discussion. The protocol would benefit by pointing out these limitations, e.g. pertaining to asymptomatic infections, convenience sampling, restricted geographic scope, etc.

Response: We have added the following limitations to the Abstract, which are further explained in the Discussion section:

- Convenience sampling of participants from the community
- Data collected from large, metropolitan city and may not be representative of suburban or rural viral transmission patterns
- Under-sampling of populations such as non-English speakers, older adults, and racial and ethnic minorities
- Study inclusion criteria do not allow for examining the role of asymptomatic respiratory viral transmission

Statistical methods (page 10, line 45): How will the authors account for test sensitivity/specificity? What are the sensitivities/specificities of the tests that they will use?

Response: We added the following paragraph to the statistical methods section of the manuscript:

"Respiratory pathogen prevalence is calculated as the number of cases detected out of the total number of episodes with testing. We do not make any statistical adjustments to account for specificity and sensitivity for the following reasons. Previous sensitivity analyses have verified that the TaqMan and Open Array are highly specific and so false positives are rare relative to true positives. For specificity, we have also compared our Open Array results for samples with clinical test results and found them to be highly concordant with Cepheid and BioFire Film Array for detection of influenza (96%) and RSV (92%), with discordance most common for samples with concentrations near the detection thresholds for each assay. Thus, test prevalence without adjustment is comparable to standard reporting from clinical labs."

General question:
How is the protocol being adjusted for the current SARS-CoV-2 pandemic? Are there strategies in place to ensure the safety of those collecting specimens, or will the study be delayed? Will SARS-CoV-2 be added to the list of pathogens they're testing for? Given that the study is motivated in part by the need to improve pandemic response, some mention of the current pandemic would be worthwhile.

Response: With the first detection of community spread of SARS-CoV-2 in the United States, the study was put on hold. After all staff received thorough training in donning and doffing proper personal protective equipment (PPE) and the collection and handling of potential SARS-CoV-2 positive samples, certain portions of the community cross-sectional study resumed. This has been addressed in the final paragraph of the manuscript, with additional references to direct translation of the community-based platform for SARS-CoV-2 surveillance:

"*The first year of the Seattle Flu Study demonstrated the utility of an integrated surveillance system for pandemic preparedness. Our research team did not anticipate that a global pandemic would arise while the study was being conducted. However, with Seattle as an early hotspot of the SARS-CoV-2 pandemic, the Seattle Flu Study was well-positioned to provide the infrastructure for early identification and mapping of transmission. During the pandemic, we rapidly developed an assay for SARS-CoV-2, eliminated in-person community-based sampling due to widespread "Stay-at-Home" orders, and scaled up the online and at-home sampling strategies (22, 24)."*

Reviewer 3:

The article summary indicates that there will be a list of strengths and limitations. This list does not seem to include any limitations. This is something that is missing from the whole paper; addition of limitations and potential pitfalls in the discussion section will be a useful addition.

Response: We have added limitations to the abstract as well as the discussion section, as mentioned above.

In the introduction, the novelty of individual-level influenza data from the community, and the associated demographic information, should not be overlooked. This will overcome the biases of healthcare seeking, which are very significant for influenza, and specific to the United States. Generally, I think overcoming this bias is one of the greatest strengths of this study, and I think this should be more fully considered in the statistical analysis. Can the researchers weight the resulting data to account for their own testing biases? Will the results be representative of the population? Will the demographics of Seattle impact these findings?

Response: Thank you for this comment. We can compare demographic data from study participants to the underlying population being sampled into the study (for example: from the American Community Survey) to assess the representativeness of participants in our study. Depending on the goals of specific analyses, the study sample can be weighted based on representation of the underlying Seattle-area population to obtain incidence estimates that are representative of the region. We expect our results to be representative of respiratory viral transmission in the greater Seattle area. However, Seattle is a large urban area, and the micro-epidemiology of respiratory viruses is notecessarily expected to generalize from this study to areas with different population densities or contact patterns.

We have aimed to emphasize the importance of individual-level, community-based influenza data in the Introduction and Discussion sections:

Introduction:
"The integration of individual-level, community-based sampling is crucial to augment samples collected in clinical settings, particularly in the United where care-seeking behavior is impacted by many factors, including health insurance status."

Discussion:
"Current respiratory virus-based transmission models are largely based on clinical data. The Seattle Flu Study integrates the sociodemographic, clinical, and geospatial characterizations of individuals with respiratory illnesses from diverse sources, including community-based, ambulatory, hospital-based, and environmental sampling. This large-scale, multi-arm sampling strategy allows us to overcome biases introduced by influenza care-seeking trends in the United States and generates a more real-time mapping of community transmission dynamics. Further, the use of community-based ARI surveillance generates novel, individual-level respiratory pathogen data in individuals who may not seek care."

The aims and hypotheses are not clearly defined and are too open-ended. More specific hypotheses regarding anticipated findings would be helpful to more clearly understand the methods and their motivations. Something that will help the reader to understand the hypotheses is more background on previous community based studies, and what this study provides that moves further than prior studies. The text mentions that integration of community based survey data with inpatient hospital and ambulatory care, which is a limitation of other studies, but it is not clear to me what this integration achieves and why it is important.

> **Response:** Current influenza based transmission models are largely based on data from care-seeking individuals. Community sampling allows us to also capture data from individuals who may not seek care for their illness. A major premise of our study design is that detection of transmission of respiratory viruses earlier in the season permits greater opportunity to assess genetic transmission chains and assess the rate and magnitude of change compared to that assessment based solely on samples derived from individuals seeking medical care during time periods when respiratory virus transmission is often occurring at much higher rates. Furthermore, rates of viral transmission may occur differentially in various patient populations such as the elderly, children attending childcare or school, or the immunocompromised. In order to have the most robust understanding of community transmission dynamics, it is important to combine the data from care-seeking and non-care seeking individuals, to give us a much more complete picture of real-time influenza burden within a community.

A clearer definition of the goals of the different branches would be very helpful. Is each branch characterizing influenza in a different population? What is the significance of this? How will they be integrated, and what will this reveal that other studies may have missed?

> **Response:** We added a section in the methods and analysis called "Study Arms", which includes a brief description of the overall aim of each study arm. We have also added discussion as to why the study is sampling across multiple groups. In short, the overall purpose is to represent the diversity of the region in community surveillance while also allowing simultaneous understanding of viral dynamics in high-risk groups like children, hospitalized individuals, and individuals in congregate living facilities.

This paper is lacking in the discussion. What are the broader impacts of this study? What are the limitations? Is this design scalable or transferable to other areas?

> **Response:** We have added a more thorough discussion of the Seattle Flu Study platform to discuss the strengths and limitations of the described study design as well as the transportability of these methods to the current SARS-CoV-2 pandemic.
>
> *"We present the study design and infrastructure for a large-scale assessment of the burden of ARI attributable to influenza and other respiratory pathogens. The overarching goal of this study is to develop and implement strategies for actionable pathogen surveillance in a major metropolitan area. Cumulatively, these strategies will facilitate the early identification of novel pathogens and allow for targeted deployment of public health resources and interventions at the community-level during pandemics.*
>
> *This study design has several strengths. Current respiratory virus-based transmission models are largely based on clinical data. The Seattle Flu Study integrates the sociodemographic, clinical, and geospatial characterizations of individuals with respiratory illnesses from diverse sources, including community-based, ambulatory, hospital-based, and environmental sampling. This large-scale, multi-arm sampling strategy allows us to overcome biases introduced by influenza care-seeking trends in the United States and generates a more real-time mapping of community transmission dynamics. Further, the use of community-based ARI*

surveillance generates novel, individual-level respiratory pathogen data in individuals who may not seek care.

We utilize broad, multiplex molecular testing for viral and bacterial pathogens to characterize the molecular epidemiology of respiratory pathogens in the community. For samples with influenza detected, we perform whole genome sequencing to evaluate transmission patterns within a community, and identify target populations for public health interventions. This strategy is translatable to the current SARS-CoV-2 pandemic. We utilized this platform to first identify community transmission of SARS-CoV-2 in the United States (19, 21).

This study has several limitations. We used convenience sampling for enrollment, and under-sampled populations such as non-English speakers, older adults, and racial and ethnic minorities. In the second year of the Seattle Flu Study, we enhanced our sampling strategy so that it would be better representative of the diversity of the Seattle metropolitan population through targeted in-person and online marketing recruitment. Our community-based sampling required participants to have at least two ARI-associated symptoms, and therefore did not allow us to examine the role of asymptomatic respiratory viral transmission. To address this limitation, we broadened the community-based eligibility criteria to include individuals without symptoms (22). Finally, the in-person community-based recruitment restricted our geographic scope, thus resulting in data that may not be representative of suburban or rural transmission patterns. To account for this, we subsequently added an online-based, at-home enrollment strategy so that individuals could participate in the study from home through home-based delivery of self-collection kitsk (23).

The first year of the Seattle Flu Study demonstrated the utility of an integrated surveillance system for pandemic preparedness. Our research team did not anticipate that a global pandemic would arise while the study was being conducted. However, with Seattle as an early hotspot of the SARS-CoV-2 pandemic, the Seattle Flu Study was well-positioned to provide the infrastructure for early identification and mapping of transmission. During the pandemic, we rapidly developed an assay for SARS-CoV-2, eliminated in-person community-based sampling due to widespread "Stay-at-Home" orders, and scaled up the online and at-home sampling strategies (22, 24).

In conclusion, the study design presented here may provide guidance for establishment of a respiratory pathogen surveillance system for current and future pandemics."

The expected findings related to asymptomatic individuals are somewhat unclear. The introduction mentions asymptomatic infection as a limitation that community studies can overcome, but then ARI symptoms are a requirement to receive a test in all of the branches. If this is not the case, and some of the branches are able to capture asymptomatic infection, this should be specifically identified.

Response: In the second year of our study, we began performing collection in asymptomatic participants. However, that collection does not fall within the timeline for the period of the study covered in this paper. We have removed the mentions of asymptomatic transmission from the Introduction and added it as a limitation of the study in the Discussion:

"Our community-based sampling required participants to have at least two ARI-associated symptoms, and therefore did not allow us to examine the role of asymptomatic respiratory viral transmission. To address this limitation, we broadened the community-based eligibility criteria to include individuals without symptoms (22)."

Page 5, line 38 is missing a word. Possibly should be, "identify new and emerging influenza strains and their transmission dynamics"

Response: Thank you for catching this omission, we have corrected the sentence to read:

"...identify new and emerging influenza strains and associated transmission dynamics."

The importance of the environmental swabs and their purpose is unclear.

Response: We have added a brief description of the environmental swabs and their added benefit in the "Study arms" section:

*"The **environmental sampling arm** involves specimen collection from high-touch surfaces and bioaerosol sampling at community enrollment sites and childcare sites to characterize the extent of environmental pathogen detection and the concordance between detection on environmental samples and participants' respiratory samples at those sites."*

For the data quality, the authors could consider dataset validation, by comparing the influenza dynamics in their data with another dataset, like the CDC influenza testing or influenza-like illness tracking. If both datasets provide the same spatiotemporal signals, that will increase confidence in the representativeness of this dataset.

Response: We are able to cross-validate the timing of influenza circulation in our study population with several data sources, including Washington State influenza surveillance (both influenza-like-illness and laboratory-confirmed influenza) and hospital-based surveillance such as the University of Washington's Clinical Virology laboratory and the Seattle Children's Hospital Virology Report, which report weekly numbers of patients with different respiratory viruses detected.

A suggestion in the statistical methods: the authors could consider using an n-mixture model, which accounts for measurement processes based on measurement effort and coverage.

Response: Thank you for this comment. Our group has not worked with n-mixture models, but in reading more about them and speaking with colleagues, it seems as if they could be a helpful addition to our statistical methods and something we will certainly consider in subsequent analyses.

Page 11, line 33: Missing a word. Possibly should be "if they experience any participation-related harm."

Response: Thank you for catching this error. We have re-written the sentence as follows:

"Participants receive contact information for the study team and are encouraged to report any participation-related harm."

Will the TaqMan RT-PCR Test be able to test for COVID-19? If so, this could be added to table 2, since many of symptoms for COVID-19 would fall into the symptoms.

Response: The TaqMan RT-PCR will be able to test for SARS-CoV-2. The scope of this paper is October 2018 - May 2019, which is prior to the current pandemic. However, we have added a brief mention of this to the Discussion of this paper, including a reference to published work describing the addition of SARS-CoV-2.

"However, with Seattle as an early hotspot of the SARS-CoV-2 pandemic, the Seattle Flu Study was well-positioned to provide the infrastructure for early identification and mapping of transmission. During the pandemic, we rapidly developed an assay for SARS-CoV-2, eliminated in-person community-based sampling due to widespread "Stay-at-Home" orders, and scaled up online and at-home sampling strategies (22, 24)."

Will test results be communicated to those tested? This could be important to limit transmission.

> **Response:** We did not have permission to return results to individuals enrolled in Year 1 of the study. We agree that return of results may be important in helping to limit transmission, and have obtained permission to return participants' influenza test results for Year 2 of the study.

VERSION 2 – REVIEW

REVIEWER	Stephen Kissler Harvard T.H. Chan School of Public Health
REVIEW RETURNED	03-Jul-2020

GENERAL COMMENTS	The manuscript is much improved. My only comment is a suspected typo at the end of Paragraph 4 of the Discussion (kitsk -> kiosk, I think).

REVIEWER	Casey Zipfel Georgetown University, USA
REVIEW RETURNED	09-Jul-2020

GENERAL COMMENTS	This manuscript describes a study that integrates community-based and healthcare-based sampling for influenza in the Seattle area. The integration of the multiple arms will provide novel, important insights for influenza dynamics. Overcoming the limitation of healthcare seeking in influenza sampling will fill a gap that is a drawback to many healthcare-based infectious disease datasets. The study is described thoroughly and clearly. The authors mention that the study played a role in identifying community-based transmission of COVID-19, which is a proof of concept that this study is able to detect viral strains that arise, and help to create and implement public health plans. Minor comment: Remove comma, page 7, line 23: "recruited into the study, with"