Concept Proposal: Assessment of Candidate Genetic Variants in Orofacial Clefts

Abstract

Non-syndromic cleft lip and/or palate (NSCL/P) account for over 70% of all cleft cases globally, and although the contribution of genetic components to the cleft development has been strongly established, very little is understood about the role copy-number variants (CNVs) play in shaping their onset. CNVs detected in NSCL/P patents will be compared with a null distribution of CNVs in the human genome and a distribution of CNVs across the genome in general population to determine whether the CNVs in patients are likely to have occurred by chance. Genes within the CNVs will be prioritized according to whether multiple patients have deleted or duplicated copies of the gene and the level of expression of the genes in developing lip and palate in mice. Biological pathways enriched in the prioritized genes will be considered as pathways that are potentially involved in NSCL/P.

Purpose

This study aims to explore the contribution and potential roles that CNVs might have in non-syndromic orofacial cleft development.

Background

Orofacial clefts remain one of the most common birth defects globally[1] and rank as the second most common birth defect in the United States[2]. Prevalence remains disproportionately high relative to other birth defects, with an estimated occurrence of about 1.5 cases per 1,000 live births[3]. Orofacial clefts result from disruptions in the normal embryogenetic processes that causes the tissues of the face to fuse properly[2], and phenotypically present as a cleft in the lip and/or palate (CL/P), or a cleft in the palate only (CPO)[2, 4]. Non-syndromic cleft lip and/or palate comprises a subgroup of orofacial clefts that is not associated with any other congenital anomaly or clinical morbidity. Among all orofacial clefts cases, non-syndromic cases are the most reported, with incidence as high as 70% in cleft lip cases and more than half of cleft palates alone[1]. The onset of orofacial clefts is influenced by heterogeneous, complex interactions between genes and the environment, making it difficult to establish a unique cause. However, several risk factors that impact cleft development have been identified, including smoking, pregestational and gestational diabetes, alcohol abuse, and the use of some anticonvulsants[5, 6, 7]. There is also a strong genetic component associated with cleft development, with over 200 and 400 genetic syndromes uniquely linked with cleft lip and cleft palate, respectively[8]. Structural aberrations in the genome with significant disease-causing potential include micro-deletions and micro-duplications of segments of DNA sequence; mostly presenting as CNVs.

Relative to other changes, CNVs are the most likely to impact cleft development due to the wide variability in the potential duplications and deletions that could occur in the genome. CNVs arise from the gain and/or loss of genetic material, causing the number of repeat copies of segments of DNA sequence to vary in a population. In contrast to single nucleotide polymorphism (SNPs), which affect only one nucleotide, CNVs are much larger, ranging from 50 base pairs to several megabases in size[10]. Large CNVs may contain genes, resulting in gene duplication or deletion. CNVs are usually inherited but can also arise de novo (although a rare event). It is estimated that the human genome contains ~20,000 CNVs[11] that cover about 12% of the human genome[12].

Recently, there have been more focused studies aiming to reveal the genetic architecture of orofacial cleft development, particularly to elucidate which regions in the genome most likely make one susceptible to orofacial cleft. In this study, we ask how CNVs are implicated and might drive cleft development, particularly non-syndromic ones.

Methods

1.0 Study Group and Study Sample

Non-syndromic cleft cases will be extracted from the publicly available dataset previously published by *Landson et al.,* (2023)[9]. The extracted samples will consist of 1,021 individuals; including patients seen during a surgical screening in the Philippines (792) and patient samples evaluated at the University of Iowa (229) with different European ancestry. All samples comprise 58,535 CNVs; 47,390 from the Philippines cohort and 11,145 from the European cohort. The subgroups of our study samples are presented in table 1 below.

Table 1: A summary of study subjects with oral clefts and copy-number variants to be included in our study.

Ancestry	Orofacial cleft phenotype						Total	
	Cleft lip		Cleft lip with cleft palate		Cleft palate only			
	Subjects (*n*)	CNVs (*n*)	Subjects (*n*)	CNVs (*n*)	Subjects (*n*)	CNVs (*n*)	Subjects (*n*)	CNVs (*n*)
Philippines	207	13,166	531	31,037	54	3,187	792	47,390
Europe	33	1,846	166	7,997	30	1,302	229	11,145
Total	240	15,012	697	39,034	84	4,489	1,021	58,535

1.1 Study Design

This will be a retrospective study using an already-collected patient data from a publicly available dataset on cleft cases from a previous publication.

1.2 Data Analysis

All analysis will be performed with Rstudio[13], BEDTOOLS[14], and on Marshfield Clinic's BIR20-LC Linux Server.

1.2.1 Identify Overlapping CNV Regions in Patients

We aim to first determine the common genomic regions shared by patient CNVs, which include the start-end genomic coordinates of each unique overlap. Thereafter, the total number of overlaps in each CNV common genomic region will be determined. This, in essence, will represent the number of patients that share that CNV common region/overlap. The following steps of the analysis will focus on the CNV common genomic regions. The underlying assumption of the analysis is that CNV common regions shared by more than one patient are more relevant to orofacial clefts than CNV regions observed in only a single patient.

1.2.2 Assess whether CNV overlaps in patients occur by chance, using an empirical p-value.

Random sampling of a reference genome will be performed n times (we propose 1,000) to obtain a list of CNVs of the same number and sizes as the CNV common genomic regions previously determined in the patients. This will produce 1,000 lists of random CNVs across the reference genome. The mean number of overlapping CNVs in each list will then be calculated. With the use of the 1,000 means, the overall mean and standard deviation of the number of overlapping CNVs among the entire set of 1,000 samples will be calculated. A z-score for each randomized list will be calculated as:

$$z_i = (\mu_i - \bar{\mu})/\sigma \quad \text{for } i \in R \qquad (1)$$

where μ_i represents the mean number of overlaps of a specific randomized list (i), $\bar{\mu}$ indicates the overall mean number of overlaps among all lists (average of all 1,000 means), σ indicates the overall standard deviation of the number of overlaps among all lists, and R represents the set of 1,000 randomized lists. The set of 1,000 z-scores will form a null distribution of z-scores. A z-score for patient overlapping CNVs will be calculated using the same formula (1) that was applied to the randomized lists. To establish statistical significance for patient overlapping CNVs, an empirical p-value will be calculated by counting the number of z-scores from the 1,000 lists that matched or exceeded the patient z-score.

Data interpretation: The distribution of z-scores from the 1,000 randomized lists represent the number of CNV overlaps observed by chance. If the z-scores from the 1,000 randomized lists match or exceed the patient z-score, this means that the patient z-score falls within the null distribution of z-scores and the number of CNV overlaps in patients could have occurred by chance. If the patient z-score does not fall within the null distribution of z-scores, then the number of CNV overlaps in patients would be unlikely to have occurred by chance, with a p-value < 0.001. If the number of CNV overlaps in patients are unlikely to be due to chance, then one or more of these CNV overlapping regions in patients are potentially relevant to orofacial clefts.

1.2.3 Prioritize genes encompassed by overlapping CNV regions in patients.

A list of the locations of all genes throughout the human reference genome will be generated using the UCSC Genome Browser[15]. This list will be filtered using three main steps. Step 1 will be to obtain the subset of genes that coincide with the location of patient CNV common genomic regions. Step 2 will be to exclude genes identified in step 1 that are deleted or duplicated in only one patient. Step 3 will be to retain only the genes from step 2 that are highly expressed in embryonic mouse facial and palate structures. The retained genes from step 3 will then be our list of prioritized genes. Next, we will use a hypergeometric test to determine whether the prioritized genes are enriched in genes known to be associated with orofacial clefts in scientific literature.

Data interpretation: If the list of prioritized genes shows statistically significant enrichment of genes known to be associated with orofacial clefts, this will suggest that patient CNV overlapping regions coincide with genes of relevance to orofacial clefts.

1.2.4 Assess the likelihood of Patient CNVs to occur in regions of the genome with high variability.

This analysis will determine the distribution of the number of CNVs across the genome in the general population, and the upper extreme of this distribution will be considered to represent genomic regions with high variability. The count of patient overlapping CNVs will be compared with this distribution to determine whether the patient data falls within or beyond the upper extreme of the distribution.

The genomic coordinates of CNVs occurring in the general population will be obtained from the NCBI's dbVar[16] database. Our reference human genome will be split into regions (windows) of one mega base in size, and the number of CNVs present in the general population in each window will be counted. Windows with a higher count will be considered as regions of the genome with higher variability. The mean and standard deviation of the counts across all windows will also be calculated. Next, a z-score for each window will be calculated according to:

$$z_i = (c_i - \bar{c})/\sigma \quad \text{for } i \in W \tag{2}$$

where c_i represents the number of CNVs in a specific window (i), $\bar{c}$ indicates the mean number of CNV across all windows, σ indicates the standard deviation of the number of CNVs across all windows, and W represents the set of all windows across the genome. The obtained z-scores will then be used to generate a frequency distribution of the count of CNVs per mega base in the general population. Now, using the known location of genes in the genome, the z-scores for the windows that coincide with prioritized genes and known genes associated with orofacial clefts also will be determined. Finally, the z-scores of the genes will be compared with the distribution of z scores from the general population to determine whether the z-scores of the genes fall within or beyond the upper extreme of the distribution of z-scores from the general population.

Data interpretation: If the genes are located in regions with high variability, then patient CNVs that coincide with these genes will be considered as having arisen in regions of the genome where CNVs are observed with very high frequency in the general population. This would suggest that the patient CNV overlapping regions could be due to chance alone. If the genes are not located in regions with high variability, then this would suggest that patient CNV overlapping regions are unlikely to have occurred due to chance and are potentially relevant to orofacial clefts.

1.2.5. Perform Gene Ontology Analysis to Determine Pathways influenced by High Confidence CNVs

Lastly, Gene ontology analysis will be performed to identify the biological pathways enriched in our prioritized list of genes. The analysis will be performed directly from the Gene ontology website[16], with the use of the PANTHER over-representation test[17] from the PANTHER gene classification resource[18].

Data interpretation: Any statistically significant biological pathways identified in Gene ontology analysis will be considered as pathways that are potentially impacted by patient overlapping CNVs and are involved in orofacial clefts.

Project Timeline

Time period	Activities
May 29-Jun 9	Plan project with mentor, literature review, finalize exposures, prepare and submit concept proposal.
Jun 10-30	Complete data request, conduct practice statistical training and begin preliminary statistical analysis.
Jul 1-21	Analyze final dataset, create figures and tables to summarize results.
Jul 24-Aug 4	Compile findings into a presentation and paper.
Aug 6-7	Finalize and practice presentation.
Aug 8	Present findings at the MCRI SRIP Annual Research Symposium

References

1. Babai, A., & Irving, M. (2023). Orofacial Clefts: Genetics of Cleft Lip and Palate. *Genes*, *14*(8), 1603. https://doi.org/10.3390/genes14081603
2. Nasreddine, G., El Hajj, J., & Ghassibe-Sabbagh, M. (2021). Orofacial clefts embryology, classification, epidemiology, and genetics. *Mutation research. Reviews in mutation research*, *787*, 108373. https://doi.org/10.1016/j.mrrev.2021.108373
3. Watkins, S. E., Meyer, R. E., Strauss, R. P., & Aylsworth, A. S. (2014). Classification, epidemiology, and genetics of orofacial clefts. *Clinics in plastic surgery*, *41*(2), 149–163. https://doi.org/10.1016/j.cps.2013.12.003
4. Dixon, M. J., Marazita, M. L., Beaty, T. H., & Murray, J. C. (2011). Cleft lip and palate: understanding genetic and environmental influences. *Nature reviews. Genetics*, *12*(3), 167–178. https://doi.org/10.1038/nrg2933
5. Margulis, A. V., Mitchell, A. A., Gilboa, S. M., Werler, M. M., Mittleman, M. A., Glynn, R. J., ... & Study, N. B. D. P. (2012). Use of topiramate in pregnancy and risk of oral clefts. *American journal of obstetrics and gynecology*, *207*(5), 405-e1.
6. Honein, M. A., Rasmussen, S. A., Reefhuis, J., Romitti, P. A., Lammer, E. J., Sun, L., & Correa, A. (2007). Maternal smoking and environmental tobacco smoke exposure and the risk of orofacial clefts. *Epidemiology*, *18*(2), 226-233
7. Correa, A., Gilboa, S. M., Besser, L. M., Botto, L. D., Moore, C. A., Hobbs, C. A., ... & Study, N. B. D. P. (2008). Diabetes mellitus and birth defects. *American journal of obstetrics and gynecology*, *199*(3), 237-e1.
8. Wong, F. K., & Hägg, U. (2004). An update on the aetiology of orofacial clefts. *Hong Kong Medical Journal*.
9. Lansdon, L. A., Dickinson, A., Arlis, S., Liu, H., Hlas, A., Hahn, A., Bonde, G., Long, A., Standley, J., Tyryshkina, A., Wehby, G., Lee, N. R., Daack-Hirsch, S., Mohlke, K., Girirajan, S., Darbro, B. W., Cornell, R. A., Houston, D. W., Murray, J. C., & Manak, J. R. (2023). Genome-wide analysis of copy-number variation in humans with cleft lip and/or cleft palate identifies COBLL1, RIC1, and ARHGEF38 as clefting genes. *American journal of human genetics*, *110*(1), 71–91. https://doi.org/10.1016/j.ajhg.2022.11.012
10. Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., ... & Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, *464*(7289), 704-712.
11. Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., ... & 1000 Genomes Project. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*, *470*(7332), 59-65.
12. Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., ... & Hurles, M. E. (2006). Global variation in copy number in the human genome. *nature*, *444*(7118), 444-454.
13. RStudio: https://posit.co/download/rstudio-desktop/
14. BEDTOOLS: https://bedtools.readthedocs.io/en/latest/
15. UCSC Genome Browser: https://genome.ucsc.edu/
16. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2021 Jan 8;49(D1):D325-D334. doi: 10.1093/nar/gkaa1113. PMID: 33290552; PMCID: PMC7779012.
17. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 2019 Jan 8;47(D1):D419-D426. doi: 10.1093/nar/gky1038. PMID: 30407594; PMCID: PMC6323939.
18. Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, Thomas PD. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat Protoc. 2019 Mar;14(3):703-721. doi: 10.1038/s41596-019-0128-8. Epub 2019 Feb 25. PMID: 30804569; PMCID: PMC6519457.

Concept Proposal

Adverse Childhood Experiences in Rural and Farm Families in Wisconsin

Objectives:

The objectives of this study are to:

1. Estimate current trends in Adverse Childhood Experiences (ACEs) in north-central Wisconsin children and adolescents – The prevalence of ACEs, including emotional, physical, and sexual abuse, between 2017-2023, will be reported.
2. Examine associations between ACEs and farm residence – The odds of ACEs will be compared between children/adolescents in farm vs non-farm households. Secondary analyses will also compare ACEs between children/adolescents in rural vs. non-rural households and by specific ACEs components. Analyses will be conducted without a priori hypothesis.

Background:

According to CDC 2024, Adverse childhood experiences, or ACEs, are potentially traumatic events that occur in childhood[1]. Numerous studies of ACEs include the following types of adversities: physical, sexual and emotional abuse, as well as exposure to domestic violence, parental divorce or separation, or having resided with someone who abused drugs or alcohol, was incarcerated, or had a mental illness[2,3]. Early life exposures to ACEs has been linked to an increased risk of cancer, heart disease, liver disease, depression, diabetes, or other chronic illnesses later in life. Moreover, exposure to ACEs can often lead to intergenerational implications, with parental exposure to ACEs having an impact on their children[4].

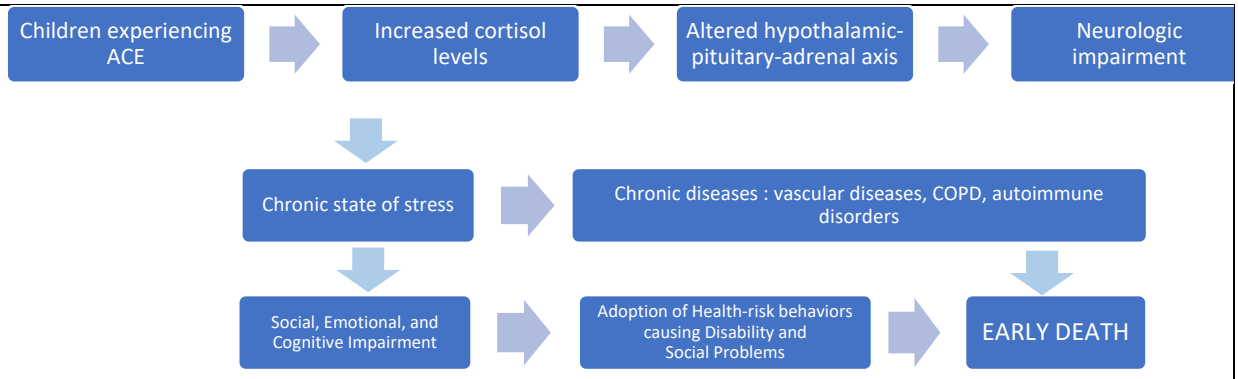

Figure 1. Developmental changes in children experiencing ACEs

ACE showed a linear association with child development: higher number of ACEs, lower the development scores[5]. Findings indicate that addressing the effects of childhood abuse on adult socioeconomic status will not significantly reduce health risks[3], highlighting the need for early interventions.

Burden of ACEs among children in rural areas

Over 13.4 million children under the age of 18 lived in the rural areas[6] and about 1.5 million children under the age of 20 live or work on farms, in the United States[7]. Data from the National Survey on Children's Health (NSCH) 2011–2012 found that 28.9 % of children living in small rural areas experienced ≥ 2 ACEs

compared with 21.3 % of urban children[8]. Children in large rural areas were more likely to have had at least 1 ACEs, than their peers in urban or small rural areas; and children in urban areas were less likely than those in rural areas to have had ≥ 2. However, another study showed that rural areas had a significantly higher percentage of individuals without ACEs (44.7%) compared to the ones in urban areas (40.5%)[15], showing increased ACEs in rural areas than urban areas[9,10].

Children living in rural areas experience higher rates of child poverty, under-utilization of preventive health services, and increased risk of death compared to urban children. The recent opioid epidemic in rural America has increased the risk for maltreatment among rural children[11].

Gaps in literature

Unfortunately, ACEs are too common in U.S. children and adolescents. The evidence on how rural residency impacts the risk of ACEs, however, is unclear. Some studies suggest rural children are at greater risk of ACE, while others suggest they are at lower risk than their urban or suburban counterparts. Furthermore, no studies have examined the burden of ACEs in youth who live on farms. According to the CDC, agriculture is one of the most hazardous labor sectors in the U.S[12], and farms are often a major component of rural regions, thus there remains a clear need to expand our knowledge on the burden of ACEs in this area. Much of the prior literature is also based on retrospective self-reporting of ACEs by adults, which may be subject to recall and self presentation or detection biases[3,4,13]. The general purpose of this study is to examine ACEs in medical records to determine how the burden of medically attended ACEs differs in farm vs. non-farm children and adolescents.

Methods:

Design and Setting

We will conduct a cross-sectional analysis using clinical and sociodemographic data from the electronic health records (EHR) of the Marshfield Health Clinic Health System (MCHS), as well as linked data on farm residence from the Wisconsin NCCRAHS Agricultural Injury Surveillance (WINS) cohort.[14] The source population will include children and adolescents who reside in a 20-county region of north-central Wisconsin, and who have capture of their medical care within MCHS data systems.

Figure 4. 20-county target population of north-central Wisconsin farm and non-farm households.[14]

Participants

Eligibility criteria for participants will include all children and adolescents who were under surveillance in 2017-2023 as part of the WINS study. The WINS cohort is described in greater detail elsewhere[14], but briefly, includes those who, between 1/01/2017 and 12/31/2023:

(1) were age 0-17 years for ≥ 90 continuous days,
(2) had ≥ 1 medical encounter, and
(3) had reasonably complete capture of medical care within MCHS data systems as evidenced by:

(a) "medically homed" to an MCHS medical center (i.e., ≥ 2 preventive or well-child visits over the previous 3 years or an assigned MCHS primary care provider,
(b) member of the MCHS-affiliated Security Health Plan (SHP) of Wisconsin or
(c) resident of the Marshfield Epidemiologic Study Area (MESA).[16]

All procedures were approved in advance by the MCHS Institutional Review Board, including an approval to wave documentation of informed consent by HIPAA authorization.

Sample Size

There is no known prior research on ACEs in farm populations to guide assumptions needed for precise sample size calculations. Statistical power, however, is expected to be robust for the planned analysis. In any given year, there are approximately 215,000 unique individuals in the study sample. Of these, about 8,500 (4%) are in the farm group and the remaining 206,500 (96%) are in the non-farm group. The estimated prevalence of ACEs is unknown, particularly in medical records data. Based on self-report data of adults from multiple countries, approximately 60% report having had at least one ACE in their lifetime[17]. The prevalence of ACEs in the EHR among children and adolescents is expected to be far lower, perhaps as low as 1% or 2% (depending on the case definition used). Using this available sample, we calculated the effect size needed to detect significant differences between the farm vs. non-farm groups. We would need to observe a very modest Cohen's w effect size of w=0.01 to achieve 80% power in a two-tailed one-degree of freedom test. This would, for example, roughly translate into an ACE prevalence of 1.5 % in farm vs 1.0 % in non-farm individuals (odds ratio = 1.5).

Adverse Childhood Experiences

The primary outcome will be the presence of ACEs. ACEs will be ascertained using similar methods outlined by Elia et.al[18], including extraction of Systematized Nomenclature of Medicine Clinical Terms (SNOMED) for emotional abuse, physical abuse, and sexual abuse. SNOMED codes essentially map and 'role-up' International Classification of Disease (ICD)[19] codes indicative of ACEs[23]. SNOMED is a systematically organized collection of medical terms linked to codes, synonyms and definitions used in clinical documentation, and it provides a standard approach to indexing, storing, and retrieving medical data across specialties and sites of care[21]. Any individual with ACEs during the 2017-2023 study timeframe will be defined as a case. In addition, the number of ACEs during the study timeframe will also be collected in secondary analyses to examine differences in case definitions whereby ACEs are observed in multiple clinical encounters.

Farm and Rural Residence:

The primary exposure will be farm and rural residence, per WINS address data. Briefly, that study has identified active farms within the target population, as evidenced by an agricultural

production address listed on the register of licensed dairy producers from Wisconsin's Department of Agriculture, Trade, and Consumer Protection, and/or a commercially available purchased listing of farm producers (www.dtn.com/agriculture/). Limitations of these farm identifiers are described elsewhere[14]. Participants who do not have evidence of farm residence will be categorized in the non-farm comparison group. Secondary analyses will also split the non-farm comparison group into those who live in rural and non-rural areas, per Rural Urban Commuting Area (RUCA) scores[22]. RUCA defines rural and urban-grounded residencies based on the Census Bureau's definitions of such areas by using a combination of criteria including population density and commuter patterns.

Covariates:

There are numerous sociodemographic and clinical covariates that will be extracted from the EHR. The model will consider several a priori specified covariates based on their potential to confound ACE-farm/rural associations, including age (quartiles), sex, race/ethnicity, health insurance, residential ZIP code, time resided in the source population, number of medical encounters in the past three years, and various medical comorbidities (e.g., cardiovascular disease, diabetes, etc.).

Analyses:

Descriptive statistics will be reported for ACEs and all study covariates. Multivariable logistic regression will be used to examine associations between farm/rural residence and ACEs, including adjustment for potential confounders, while accounting for the clustering of participants within households. Secondary analyses will also examine rural vs. non-rural (non-farm) comparison groups, disaggregated ACE outcome measures, as well as sensitivity analyses restricted to only new ACE cases in the 2017-2023 timeframe. All analytical procedures will be conducted using SAS.

Project Timeline:

To be Completed	Date
Project overview, literature review, and concept proposal	May 28 – June 7
Access dataset of interest and begin initial analysis	June 10 – June 28
Analyze final dataset, record results, and prepare abstract.	July 1 – July 12
Finalize abstract and prepare presentation of results	July 15- July 22
Finalize presentation	July 23 – August 07
Present findings at MCRI SRIP Research Symposium	August 08

References:

1. About adverse childhood experiences. Adverse Childhood Experiences (ACEs). Published April 9, 2024. https://www.cdc.gov/aces/about/index.html
2. Merrick MT, Ford DC, Ports KA, et al. Vital signs: Estimated proportion of adult health problems attributable to adverse childhood experiences and implications for prevention — 25 states, 2015–2017. *Morbidity and Mortality Weekly Report*. 2019;68(44):999-1005. doi:10.15585/mmwr.mm6844e1
3. ClinicalKey. https://www.clinicalkey.com/#!/content/playContent/1-s2.0-S0145213415001933?returnurl=https:%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0145213415001933%3Fshowall%3Dtrue&referrer=https:%2F%2Fpubmed.ncbi.nlm.nih.gov%2F
4. Crouch E, Radcliff E, Probst JC, Bennett KJ, McKinney SH. Rural-Urban differences in adverse childhood experiences across a national sample of children. *the Journal of Rural Health/the Journal of Rural Health*. 2019;36(1):55-64. doi:10.1111/jrh.12366
5. Rocha HAL, Sudfeld CR, Leite ÁJM, et al. Adverse childhood experiences and child development outcomes in Ceará, Brazil: a population-based study. *American Journal of Preventive Medicine*. 2021;60(4):579-586. doi:10.1016/j.amepre.2020.08.012
6. US Census Bureau. Our rural communities. Census.gov. Published August 31, 2022. https://www.census.gov/library/spotlights/2020/rural.html
7. Health and Safety for Kids on the Farm | NIOSH | CDC. https://archive.cdc.gov/www_cdc_gov/niosh/topics/childag/kidsag.html#:~:text=Approximately%201.5%20million%20children%20under,and%20migrant%20and%20seasonal%20workers
8. U.S. Department of Health and Human Services, Health Resources and Services Administration: The health and wellbeing of children in rural areas: A portrait of the Nation. 2011-2012. Available at: The Health and Well-Being of Children in Rural Areas: A Portrait of the Nation 2011-2012 (hrsa.gov)
9. Maine Rural Health Research Center : Adverse childhood experiences in rural and urban contexts: Research and policy brief. Available at: https://muskie.usm.maine.edu/Publications/rural/Adverse-Childhood-Experiences-Rural.pdf
10. Chen Y. Adverse childhood experiences and behavioral problems in early adolescence: An empirical study of Chinese children. *Frontiers in Psychiatry*. 2022;13. doi:10.3389/fpsyt.2022.896379
11. Keyes KM, Cerda M, Brady JE, Havens JR, Galea S. Understanding the rural-urban differences in nonmedical prescription opioid use and abuse in the United States. Am J Public Health. 2014;104(2):e52-e59. doi:10.2105/AJPH.2013.301709
12. Agriculture worker safety and health. Agricultural Safety. Published May 16, 2024. https://www.cdc.gov/niosh/agriculture/about/?CDC_AAref_Val=https://www.cdc.gov/niosh/topics/aginjury/default.html
13. Rosado JI, Ramirez A, Montgomery J, Reyes E, Wang Y. Adverse childhood experiences and its association with emotional and behavioral problems in U.S. children of Latino immigrants. *Child Abuse & Neglect*. 2021;112:104887. doi:10.1016/j.chiabu.2020.104887

14. VanWormer JJ, Berg RL, Burke RR, Barnes KL, Weichelt BP. Regional surveillance of medically-attended farm-related injuries in children and adolescents. *Frontiers in Public Health*. 2022;10. doi:10.3389/fpubh.2022.1031618
15. Chanlongbutra A, Singh GK, Mueller CD. Adverse childhood experiences, Health-Related quality of life, and chronic disease risks in rural areas of the United States. *Journal of Environmental and Public Health*. 2018;2018:1-15. doi:10.1155/2018/7151297
16. Destefano F, Eaker ED, Broste SK, et al. Epidemiologic research in an integrated regional medical care system: The Marshfield Epidemiologic Study Area. *Journal of Clinical Epidemiology*. 1996;49(6):643-652. doi:10.1016/0895-4356(96)00008-x
17. Madigan S, Deneault AA, Racine N, Park J, Thiemann R, Zhu J, Dimitropoulos G, Williamson T, Fearon P, Cénat JM, McDonald S, Devereux C, Neville RD. Adverse childhood experiences: a meta-analysis of prevalence and moderators among half a million adults in 206 studies. World Psychiatry. 2023 Oct;22(3):463-471
18. Elia J, Pajer K, Prasad R, et al. Electronic health records identify timely trends in childhood mental health conditions. *Child and Adolescent Psychiatry and Mental Health*. 2023;17(1). doi:10.1186/s13034-023-00650-7
19. ICD- 10 - CM International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM). https://www.cdc.gov/nchs/icd/icd-10-cm.htm#:~:text=The%20ICD%2D10%20is%20used,%2C%20effective%20October%201%2C%202015
20. Vikström A, Nyström M, Ahlfeldt H, Strender LE, Nilsson GH. Views of diagnosis distribution in primary care in 2.5 million encounters in Stockholm: a comparison between ICD-10 and SNOMED CT. *Journal of Innovation in Health Informatics*. 2010;18(1):17-29. doi:10.14236/jhi.v18i1.750
21. Kostick K. SNOMED CT integral part of quality EHR documentation. *ResearchGate*. Published online October 1, 2012.
22. Hart LG, Larson EH, Lishner DM. Rural Definitions for health policy and research. *American Journal of Public Health*. 2005;95(7):1149-1155. doi:10.2105/ajph.2004.04243