Evaluating Proposed Phenotypic Predictors of Recurrence Risk in Unaffected Individuals with a Family History of Isolated Cleft Lip with or without Cleft Palate

by

Charlotte Ernst Skinner

BS, Biology, Christopher Newport University, 2011

Submitted to the Graduate Faculty of

the Department of Human Genetics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science
/
Master of Public Health

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Charlotte Ernst Skinner

It was defended on

June 10, 2019

and approved by

Andrea Durst, MS, DrPH, LCGC, Assistant Professor, Department of Human Genetics, Licensed Genetic Counselor, Associate Director, Genetic Counseling Program, Co-Director, Public Health Genetics Program, Graduate School of Public Health, University of Pittsburgh

John R. Shaffer, PhD, Assistant Professor, Department of Human Genetics, Graduate School of Public Health, Assistant Professor, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh

Seth M. Weinberg, BA, MA, PhD, Associate Profession, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh

Thesis Advisor: Mary L. Marazita, BS, PhD, FACMG, Director, Center for Craniofacial and Dental Genetics, Professor, Department of Oral Biology in the School of Dental Medicine; Professor, Department of Human Genetics, Graduate School of Public Health; Professor, Clinical and Translational Science and Department of Psychiatry, School of Medicine, University of Pittsburgh

Evaluating Proposed Phenotypic Predictors of Recurrence Risk in Unaffected Individuals with a Family History of Isolated Cleft Lip with or without Cleft Palate

Charlotte Ernst Skinner, MS

University of Pittsburgh, 2019

Abstract

Cleft lip ± cleft palate (CL/P) is a common birth defect and public health concern, with significant, long-term health impacts, which require many healthcare resources and increase costs of care. The majority (>70%) of cases are isolated and non-syndromic (NSCL/P). Single-gene syndromes, chromosome anomalies, or teratogens account for the remainder. There is a strong genetic component in NSCL/P; individuals with a positive family history are more likely to have an affected child. The complex, multifactorial nature of NSCL/P complicates recurrence risk assessment. Currently, genetic counseling for NSCL/P is based on empiric recurrence risk figures from large cohort studies. Thus, risk assessment cannot be personalized. Developing predictive models based on established risk factors could refine estimates. Several subtle phenotypic differences have been identified among unaffected relatives of an individual with NSCL/P, compared to individuals with no family history. Phenotypes consistently associated with NSCL/P may be useful as markers of recurrence risk. The present study investigated the effect of seven such traits, on the odds of participants' case versus control family status. Three models were built using logistic regression. Model 1 included orbicularis oris muscle defects and velopharyngeal dysfunction, and achieved an area under the receiver operating curve (AUC) of 0.63. Model 2 included maximum facial and intercanthal width, upper and lower facial height, and midface depth measurements, and achieved an AUC of 0.64. Model 3 included all variables from Models 1 and 2, and achieved an AUC of 0.72. AUC values below 0.7 indicate poor model discrimination.

Models 1 and 2 had poor predictive performance, while Model 3 had fair performance. The public health significance of this work is that it adds to the body of research on risk factors and recurrence risk assessment in NSCL/P. It also works towards creating risk models. Reliable risk models allow for the identification of high-risk individuals, therefore making it possible to design and/or implement primary and secondary prevention interventions.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Completing a thesis is a long process, filled with high points and low points, from which you emerge with new confidence as a professional in your field. The high points are the moments you decide on a great idea, start devouring literature with excitement, have a great writing day, or an awesome critical thinking session, during which things just "click". The low points happen when your data feels like a mess, you spend hours trying to figure out "simple" code with your novice statistical software knowledge, or when you just can't write a sentence for anything! Through the high points, you have individuals who celebrate you; through the low points, those same people rally, with the wisdom of a seasoned researcher who has been there, the confidence of a trusted colleague or advisor who knows you can do it, or the love of friend or partner who recognizes your true grit and won't let you lose sight of it.

Thank you to those who have been there for my own high and low points. My committee: Mary Marazita, Andrea Durst, John Shaffer, and Seth Weinberg. Thank you for your investment in my academic development. Joel Anderton, special thank you for the database help! Classmates: Kelsey Bohnert, Natasha Berman, Caitlin Russell (and her partner, Luke), and especially, Alyssa Azevedo, my co-worker at the CCDG. Thank you for the coffee shop working sessions, celebrating my successes, and hearing my frustrations. Family and friends, near and far: Cara and EJ, Shaunbay, Angela, Kat, Nickole, Sean, Alex, Justin, and my parents, Craig and Kathleen. You all are strong sources of love and support, and you also get me out once in a while. I am truly fortunate to have you. Lastly, my steadfast husband, Ben, who spent a lot of time making me tea and smiling knowingly at me when I had days of doubt or discouragement – knowingly, because he never had any doubt. This process would not have been as successful or pleasant without your support.

1.0 Introduction

Cleft lip with or without cleft palate (CL/P), with a prevalence of approximately 1 in 700 births worldwide, is among the most commonly observed congenital anomalies in humans, making it a significant public health issue (Mai et al., 2014). Non-syndromic forms of CL/P are separated from syndromic forms of CL/P based on the presence of additional congenital anomalies. Non-syndromic CL/P (NSCL/P) accounts for greater than 75 percent of all observed cases of CL/P, with the remaining cases caused by a known Mendelian genetic variant, chromosome anomaly, or teratogen (IPDTOC, 2011; Mai et al., 2014). Individuals born with NSCL/P have complex medical management needs from birth, into early adulthood.

Starting in the neonatal period, newborns struggle with feeding, are less likely to be breast fed, and are more likely to experience failure to thrive (Kaye et al., 2017). Additionally, sleep disordered breathing (SDB) is common among individuals with CL/P (MacLean, Tan, Fitzgerald, & Waters, 2013; Muntz, Wilson, Park, Smith, & Grimmer, 2008). Multiple corrective surgeries to restore function and improve aesthetic are required as a child grows, in addition to surgeries and/or therapies for speech impediments and delays, hearing loss, and dental/orthodontic issues (Leslie & Marazita, 2013). Depending on the cleft type and severity, revision surgeries may be needed well into adulthood (Ranganathan et al., 2016). Certain cancers are also more common among individuals with NSCL/P, and all-cause mortality is greater across all age groups (Leslie & Marazita, 2013).

Treatment and supportive therapies continue to improve the lives of individuals born with orofacial clefts (OFCs), but studies have suggested there can be long-term psychological effects which result in an overall reduction in quality of life and an increase in mental health problems in

the NSCL/P population (Strauss & Cassell, 2009; Wehby & Cassell, 2010). Financially, there is also a burden on affected individuals and their families. Individuals with OFCs experience higher costs of care in the newborn period, increased hospital utilization compared to unaffected siblings, and need expensive surgeries and treatments (Wehby & Cassell, 2010). Despite relatively high prevalence, significant impact on quality of life, and a large body of literature, an understanding of how genetic and environmental factors, and interactions between them, affect occurrence and recurrence of NSCL/P remains incomplete.

NSCL/P is currently thought to be a heterogeneous, multifactorial trait, with several environmental and genetic risk factors, which makes it complicated to provide accurate familial risk assessment and genetic counseling (Collins et al., 2013; Grosen, Chevrier, et al., 2010). Recurrence is increased over the general population risk in the first-degree relatives of an individual with a NSCL/P, with lower recurrence risks for second and third-degree relatives (Grosen, Chevrier, et al., 2010). Furthermore, reported concordance rates for NSCL/P in monozygotic twins range from 40 to 60 percent, versus three to five percent in dizygotic twins (Grosen et al., 2011). Increased risk for recurrence with increased degree of relatedness and high concordance in monozygotic twins, suggests a strong genetic component in NSCL/P.

While currently considered a multifactorial trait, studies on the genetics of NSCL/P in families with multiple affected members (multiplex families) continue to identify the transmission of highly penetrant, deleterious variants in genes and gene regions associated with NSCL/P (Collins et al., 2013; Pengelly et al., 2016). It is currently estimated that, of the 30 percent heritability of NSCL/P, about 25 percent is accounted for by 24 known common variants (single nucleotide polymorphisms or SNPs; see Table 1) (Ludwig et al., 2017). The identification of highly penetrant variants in multiplex families, and the known effects of 24 common variants associated

with NSCL/P suggest that at least some subset of NSCL/P cases have a more predictable pattern of inheritance that is yet to be discovered.

Over several decades, the spectrum of physical expression in NSCL/P has been extended from apparent orofacial clefts (OFCs), to include a wide range of subphenotypes, including "microform" clefts. Microform clefts are visible, but mild, manifestations of NSCL/P which can still indicate increased recurrence risk. In addition, there has been consideration of including certain occult traits, not visible with the naked eye, in the OFC spectrum. These "subclinical phenotypes" are observed more frequently in individuals who have a family history of NSCL/P versus those who do not, and tend to track with the cleft in the family (Marazita, 2007; Roosenboom et al., 2015). Studies focused on unaffected relatives of an individual with NSCL/P, who are expected to have some genetic susceptibility to NSCL/P, have evaluated several subclinical phenotypes for their utility as markers for increased recurrence risk. Subclinical phenotypes evaluated for association with NSCL/P include: elevated levels of facial asymmetry, slight deviations in the size and shape of certain facial features, increased non-right handedness, altered dermatoglyphic patterns, discontinuity of the orbicularis oris muscle (OOM), anomalies of the teeth, olfactory deficits, and mild velopharyngeal dysfunction/incompetence/insufficiency (VPD) (Chollet, DeLeon, Conrad, & Nopoulos, 2014; Howe et al., 2015; Klotz et al., 2010; Marazita, 2007; Scott, Weinberg, Neiswanger, Brandon, & Marazita, 2005; Weinberg et al., 2009; Weinberg et al., 2006).

The aim of the present study is to expand upon prior studies and investigate the combined predictive value of multiple phenotypes previously associated with NSCL/P with predictive modeling. The development of this model is supported by successful past investigations of facial morphology and NSCL/P risk, in combination with recent attempts to develop polygenic risk

scores (PRS) based on genetic loci associated with facial variation and NSCL/P (L. J. Howe et al., 2018; Wilson-Nagrani, Richmond, & Paternoster, 2018). While several phenotypic markers have been associated with increased risk for NSCL/P, the individual strength of each marker to provide meaningful estimates of NSCL/P risk is limited (Wilson-Nagrani et al., 2018). The present study assessed the individual and combined predictive value of the following risk phenotypes: lower face height, upper face height, intercanthal width, maximum facial width, midface depth, VPD, and OOM discontinuity, on the odds of an unaffected individual being from a case family versus a control family.

This study will add potentially useful information for improving risk assessment and genetic counseling for individuals with a family history of NSCL/P. Better individualized risk assessment is indispensable in counseling patients about their reproductive risks, as it allows for informed reproductive decision making (Nusbaum et al., 2008). Additionally, determining the strongest phenotypic predictors of underlying susceptibility to NSCL/P may guide/support studies on the genetics of NSCL/P. Predictive modeling for birth defects also has the potential to identify significant risk factors for which public health interventions and/or improved therapies for individuals with OFCs can be designed.

The specific aims of the present study are: (1) evaluate the effect of each previously proposed phenotype of interest on the odds of an unaffected individual being from a case family and (2) evaluate the effect of combinations of phenotypes from Aim 1, on the odds of case family status.

2.0 Literature Review

2.1 Epidemiology

NSCL/P is a relatively common birth defect, with an average incidence of 1 in 700 live births. Incidence varies from 1 in 500 to 1 in 2500 live births depending on race. Worldwide, NSCL/P is most prevalent in the Native American (3.6 in 1000) and Asian (2 in 1000) populations; this is two to three times as high as in the non-Hispanic white population (1 in 1000) (Wyszynski, 2002; Dixon, Marazita, Beaty, & Murray, 2011; IPDTOC, 2011). NSCL/P is least prevalent in Africans, with 0.4 in 1000 individuals affected at birth (Conway et al., 2015). Prevalence varies between the sexes as well. Males are roughly twice as likely to have CL/P, and two-thirds less likely to have isolated cleft palate (Watkins, Meyer, Strauss, & Aylsworth, 2014).

OFCs present across a wide phenotypic spectrum, and are typically categorized into three main subphenotypes depending upon involvement of the lip and hard/soft palate: clefts involving only the lip (CL), clefts involving both the lip and palate (CL/P), and clefts involving only the palate (CPO) (See Figure 1) (Jiang, Bush, & Lidral, 2006; Yoon, Chung, Seol, Park, & Park, 2000). There is also variability within each of these cleft subtypes, including bilateral or unilateral (left or right) presentation; of note, left-sided unilateral clefts are twice as common as right-sided (Dixon, Marazita, Beaty, & Murray, 2011a). The subtypes of OFCs are not only heterogeneous in phenotypic expression, but also in terms of risk factors. Family studies of recurrence show predominance of the same cleft type, evidence for different developmental processes, and the discovery of unique genetic risk factors for clefts of the lip and/or palate and clefts of only the palate have provided evidence for distinct etiologies behind the two cleft subtypes (Grosen et al.,

2011; Ludwig et al., 2017; Sivertsen et al., 2008; Wen & Lu, 2015). It is generally accepted that many of the genetic variants underlying CL/P differ from those underlying CPO. Thus, many studies of recurrence risk, including the present one, differentiate between CL/P and CPO. There are also subtler subphenotypes of OFCs, including microform cleft lip and submucous cleft palate.

Microform clefts are milder, but still visible, clefts of the lip or palate; these can appear as a notch, raised seam of tissue, or a scar-like mark on the upper lip (microform cleft lip), or a bifid (split) uvula (microform cleft palate) (M. Marazita, 2007). Microform clefts of the lip have been shown to affect the organization of the tissue below the surface of such a seam or scar, possibly due to healing or correction of a cleft late in the development of the lip (M. Marazita, 2007). Bifid uvula often, but not always, indicates the presence of a submucous cleft palate (Shprintzen, Schwartz, Daniller, & Hoch, 1985). Submucous cleft palates are characterized by abnormal division of the soft palate and a notched or U-shaped posterior hard palate, covered and concealed by the oral mucosa (Shprintzen et al., 1985). In addition to these microform cleft types, there are subclinical features, not readily visible, that may represent the mildest end of the OFC spectrum (Klotz et al., 2010; M. Marazita, 2007; Martin et al., 2000; Watkins et al., 2014).

There are many known genetic and environmental risk factors in NSCL/P. The heterogeneous presentation of OFCs and the complexity of the cellular processes involved in facial development make it difficult to identify precisely which steps are disrupted in any particular case of NSCL/P (Jiang et al., 2006). Research continues to look for answers to the question of which factors cause the disruption.

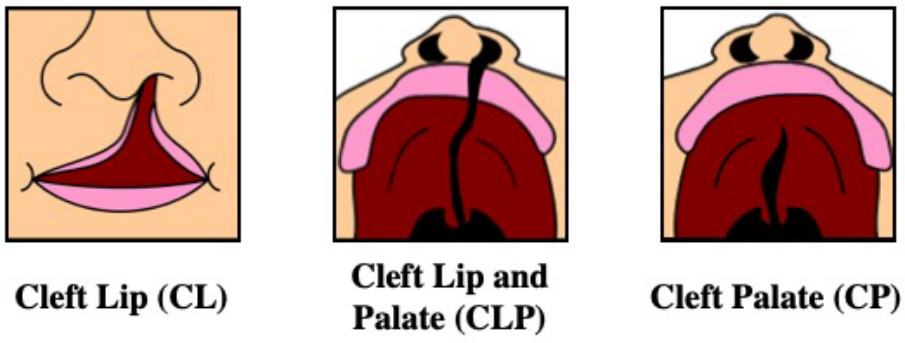

Figure 1 Main Types of Cleft Lip and/or Cleft Palate
(Wikimedia Commons, open access)

2.1.1 Environmental Risk Factors

There are clear environmental influences on the development of NSCL/P, but the strength of effect for each risk factor is somewhat unclear, and consistency across studies is poor. Population or regional differences may account for some differences, in addition to study design inconsistencies. Of note, most of the well-studied factors' effect sizes in relation to NSCL/P risk are small to moderate (OR<1.5). Many of these factors are also difficult to separate from confounding interactions, therefore limiting the ability to discern the truly causal factors driving increased risk. exposure to cigarette smoke, direct or second-hand, is consistently identified as a risk factor for NSCL/P, with odds ratios ranging from weak to moderate effects. A recent study suggests second-hand smoke may result in an odds ratio of 1.3 to 2.5, depending on frequency of exposure (Pi et al., 2018). Another recent meta-analysis of 29 case-control and cohort studies conducted between 1966 and 2015, suggests a moderate risk for OFCs based on maternal active smoking (OR=1.368 for CL/P and 1.241 for CPO) (Xuan et al., 2016). Fetal exposure to alcohol is less consistently associated with OFCs across studies. Maternal consumption of large quantities of alcohol in the first trimester have been associated with increased risk (Dixon et al., 2011a).

However, in a more recent meta-analysis of 33 case-control and cohort studies on maternal alcohol consumption in the first trimester and OFCs, no relationship was found between odds of CL/P or CPO in mothers who drank little to no alcohol (reference group), versus those with moderate to heavy consumption, and binge drinking (Bell et al., 2014). Bell et al. acknowledge that there was significant heterogeneity in study design, including how alcohol consumption levels were measured and defined, time of exposure, and whether OFCs were separated as isolated or syndromic; there was also variability in how well studies controlled for confounding. This heterogeneity may have impacted the results.

Maternal nutrition likely plays a role in NSCL/P. The effects of various vitamins and minerals on OFCs has been heavily studied, including folate, vitamin E, vitamin A, zinc, and combinations found in multivitamins (Hozyasz, 2010). Study issues of small sample size, dosage, nutritional intake from food, and failure to separate syndromic from non-syndromic forms of OFC have led to inconsistency across these studies (Dixon et al., 2011a; Hozyasz, 2010).

Retinoic acid (Vitamin A) is protective against CL/P at appropriate doses, but is also known to be teratogenic, and has been shown to cause CL/P at doses >10,000 IU (Hozyasz, 2010; Watkins et al., 2014). Vitamin E may have a protective effect in NSCL/P and zinc deficiency may lead to an increased cleft risk (Hozyasz, 2010). Multivitamin use has had mixed support across studies, but may be a protective factor in clefting (Dixon et al., 2011a; Hozyasz, 2010). Studies of supplementation may also be confounded by associated factors, such as lower socioeconomic status, access to healthcare, or health-related behaviors, leading to difficulty assessing the true impact of supplementation in and of itself (Mossey, Little, Munger, Dixon, & Shaw, 2009).

The effect of folate on occurrence and recurrence of OFCs has received much attention, partly due to its protective effect in open neural tube defects (e.g. spina bifida). Studies, however,

are inconsistent and often conflicting, with researchers reporting no effect at all, to a 74% decrease in CL/P as a result of folate supplementation (Wehby & Murray, 2010). A recent meta-analysis of 15 studies on folic acid fortification and OFC prevalence found no significant reductions in the prevalence of OFCs as a whole (RR=0.97; 95% CI 0.92, 1.02). However, when they looked at the few studies which separated syndromic from non-syndromic clefts, they found a significant reduction for NSCL/P, but not for OFCs as a whole or for CPO (RR=0.88; 95% CI 0.81, 0.96) (Millacura, Pardo, Cifuentes, & Suazo, 2017). Folate antagonist drugs also increase the risk for fetal development of a cleft. Folate is important in gene methylation, which plays an important role in regulating expression of genes involved in early development (epigenetics) (Millacura et al., 2017). This may explain observed increases in risk with folate deficiency or folate antagonist drugs.

Non-nutrition related risk factors for NSCL/P include illness and fever, maternal stress, and illicit drug use during pregnancy, parental education level, and socioeconomic status (Dixon et al., 2011a; Xu et al., 2018). Maternal obesity and diabetes have been fairly consistently associated with a small increased risk (OR=~1.3) for clefting in offspring (Cedergren & Källén, 2005; Stott-Miller, Heike, Kratz, & Starr, 2010). Maternal obesity and diabetes may have similar risks and related causal pathways; it is difficult to separate the effects, as there is significant overlap in the two conditions.

In a large population-based study in Washington state of nearly 300,000 births between 1987 and 1990, researchers found an increased risk for isolated CL/P in mothers below the age of 20 years versus mothers aged 25-29 years (RR=2.0, CI 1.3, 2.9) (DeRoo, Gaudino, & Edmonds, 2003). However, in a more recent meta-analysis of parental age as a risk factor for NSCL/P, maternal age greater than 40 was significantly associated with 1.56-fold increase in risk, while

there was no associated risk with maternal age younger than 20 years (de Q. Herkrath, Herkrath, Rebelo, & Vettore, 2012). Paternal age as a risk factor has been debated across studies, with some associating increased risk with paternal ages greater than 40 years and other reporting no association; significant heterogeneity existed across and was therefore excluded from meta-analysis.

In summary, environmental factors are an important part of public health efforts to prevent NSCL/P, as many are modifiable. It is important to continue studying large populations to strengthen the understanding of environmental risk factors; doing so may lead to the ability to develop robust risk prediction models which can account for these. The use of environmental risk factors is beyond the scope of the current study.

2.1.2 Genetic Risk Factors

Initially, studies using linkage analysis attempted to uncover single genes for NSCL/P, by studying segregation in large families with multiple affected individuals (Dixon et al., 2011b). As the idea of a single causal locus for NSCL/P shifted to favor a model of more complex inheritance, genome-wide association study (GWAS) analyses became more common. GWASs, having the ability to evaluate common variants (>5% minor allele frequency) with small effect sizes, are valuable in the study of common, complex diseases and traits (Basha et al., 2018). With the availability of GWASs, additional susceptibility loci have been associated with NSCL/P.

NSCL/P heritability, that is, the amount of liability based on an individual's genetic profile, is estimated at 30% (Ludwig et al., 2017), with the remaining liability being due to other factors such as environment, epigenetic modifications, and gene by environment interaction. A small

proportion of the heritability in NSCL/P is accounted for by the 20 to 30 known risk loci, however, there is a significant amount of "missing heritability" (Basha et al., 2018).

GWASs have been able to narrow in on several regions of the genome strongly associated with clefting, but have often fallen short of identifying the true causal variants within those regions (Leslie et al., 2015). Combining GWASs with additional methods such as linkage analysis of candidate genes, has provided additional information on specific variants. Currently, about 20 loci have been associated with NSCL/P (Leslie et al., 2015; Ludwig et al., 2017; Wilson-Nagrani et al., 2018) (see Table 1, below, for a list of proposed candidate genes within these loci).

A few of these discovered regions overlap with areas of the genome known to contain causal variants for syndromes that include CL/P as a key feature; this is of interest to researchers, as it hints at a causal variant lying within the region. A recent whole exome sequencing (WES) investigation of 46 multiplex NSCL/P families identified five families (10%) with mutations in genes associated with syndromic OFCs (*TBX1, LRP6, TP63, GRHL3*); of note, two of the five families (unrelated) had identical mutations in *GRHL3* (Basha et al., 2018). The mutations segregated with the clefts and followed an autosomal dominant pattern of inheritance. The individuals were carefully screened, then four of five families were re-screened, for additional features of the syndromes associated with the genes in which mutations were found. All were confirmed non-syndromic. Two of 15 mutation carriers identified were clinically unaffected, indicating a penetrance of 87%. The researchers concluded there was support for their hypothesis; missing heritability in NSCL/P may be explained by families with variants that segregate in a Mendelian pattern, but exhibit a phenotype consistent with incomplete penetrance and the same variability in cleft expression as mutations causative of known syndromic clefts. Tangentially, one finding is of particular interest to the matter of subclinical phenotypes in NSCL/P: within two of

the families studied, three carriers of the familial mutation for clefts showed only a pronounced VPI, with no overt clefting.

The embryological disruption that leads to clefting, combined with the observed differences in facial shape among families with OFCs, implicate the genes involved in normal facial development. Therefore, a second method of dissecting the genome for causal variants, is to investigate loci involved in the normal development of the face (Howe et al., 2018; Wilson-Nagrani et al., 2018). Howe et al. (2018) analyzed the genetic overlap between unaffected individuals' normal facial morphology variation and previously proposed risk variants for NSCL/P; decreased philtrum width was seen as a result of risk variants for NSCL/P. Wilson-Nagrani et al. (2018) assessed the effects of 19 variants associated with NSCL/P on normal lip morphology; Cupid's bow lip shape and decreased philtrum width were associated with the risk variants. These two studies support the hypothesis that risk variants for NSCL/P are associated with normal facial variation in the general population. One significant limitation of the most robust investigations of facial morphology and NSCL/P is the lack of racial diversity among the exclusively European Caucasian study populations typically used.

Interpreting the results of GWASs and determining the causal variants is a difficult task. Several of the significantly associated loci identified in NSCL/P lie within non-coding regions, suggesting they regulate the function of another unknown gene (Leslie et al., 2015; Ludwig et al., 2017). Additionally, there may be a combination of rare and common variants behind the complex trait of NSCL/P, in addition to potential roles of gene and environment interactions and epigenetic changes. Despite complexities in the genetics of NSCL/P and partially unexplained heritability, polygenic risk scores (PRSs) show potential in predicting NSCL/P (L. J. Howe et al., 2018; Ludwig et al., 2017). It is expected that further development of these risk scores will improve with the

discovery of additional loci and may one day be clinically useful. In the meantime, more readily measured traits that may represent the underlying genetic mechanisms could be more clinically useful.

Table 1 Summary of Genes Conferring Risk for NSCL/P

Associated Locus	Candidate Gene in Region	Associated Phenotype	Analysis Method	References for genome-wide significance
1p36.13	*PAX7*	CL/P	GWAS meta-analysis, GWAS replication	(2-4)
1p36	*GRHL3*	CP	GWAS, GWAS replication	(5, 6)
1p22	*ARHGAP29*	CL/P	GWAS	(4, 7)
1q32.2	*IRF6*	CL/P	GWAS, Linkage	(4, 7-10)
2p13	*TGFA*	CL/P	Linkage	(8)
2p21	*THADA*	CL/P	GWAS meta-analysis	(2)
2p24	*FAM49A*	CL/P	GWAS	(4)
2p25.1	*TAF1B*	CL/P	GWAS	(11)
3p11	*EPHA3*	CL/P	GWAS meta-analysis	(2)
3q12	*COL8A1/FILIPIL*	CL/P	GWAS replication	(3)
3q27-28	*TP63*	CL/P	Linkage	(8)
4p16.2	*MSX1*	CL/P	GWAS	(11)
4q28.1	*SPRY1/LOC285419*	CL/P	GWAS	(11)
6p24.3	*OFCC1*	CL/P	Linkage & GWAS	(11)
8p11.23	*FGFR1*	CL/P	GWAS	(11)
8q21.3	*DCAF4L2*	CL/P	GWAS meta-analysis, GWAS replication	(2-4)
8q22.3	*BAALC*	CP & multivitamins	GWAS x E	(12)
8q24	Gene Desert	CL/P	GWAS	(4, 7, 9, 10, 13)
9q22.32	*PTCH1*	CL/P	GWAS	(11)
9q22.33	*FOXE1*	CL/P and CP	Linkage	(8, 14)
9q31.1	*SMC2*	CP & maternal alcohol	GWAS x E	(12)
10q25.3	*VAX1*	CL/P	GWAS	(4, 7, 10)
12q14	*TBK1*	CP & maternal smoking	GWAS x E	(12)
12q21.1	*THAP2*	CL/P	GWAS	(11)
13q31.2	*SPRY2*	CLP	GWAS meta-analysis	(2)
14q21-24	*PAX9, TGFB3, BMP4*	CL/P	Linkage	(8)
15q22	*TPM1*	CL/P	GWAS meta-analysis	(2)
15q24	*ARID3B*	CL/P	GWAS	(4)
16p13	*CREBBP*	CL/P	GWAS	(15)
16q24	*CRIPSLD2*	CL/P	Linkage	(8)
17p13.1	*NTN1*	CL/P	GWAS meta-analysis, GWAS replication	(2-4, 16)
17q21.32	*WNT3/WNT9B*	CL/P	GWAS	(11)
17q22	*NOG*	CL/P	GWAS	(4, 10, 16)
17q23.2	*TANC2*	CL/P	GWAS	(4)
18q22	*ZNF236*	CP & maternal smoking	GWAS x E	(12)
19q13.11	*RHPN2*	CL/P	GWAS	(4)
20q12	*MAFB*	CL/P	GWAS	(4, 7)

Adapted from Table 2 in Leslie and Marazita, 2015. See **Appendix A** for list of above references.

2.2 Embryology

The development of CL/P in humans occurs early in pregnancy, after the development of the face begins at four weeks' gestation, until approximately twelve weeks' gestation (See Figure 2) (Jiang et al., 2006; Yoon et al., 2000). During embryogenesis, the formation of the lips and primary palate occurs through fusion between five regions of tissue, derived from the neural crest ectoderm: the frontonasal prominence, a pair of lateral maxillary prominences, and a pair of mandibular prominences (Jiang et al., 2006). A complex, time-bound sequence of cell growth and death, migration, and merging of the developing tissues must happen to accomplish full fusion of the prominences (Jiang et al., 2006). Development and differentiation of the secondary palate starts at six to nine weeks' gestation, with vertical growth of two palatal shelves, which then turn, horizontally, and fuse to form the roof of the mouth (Yoon et al., 2000). Fusion of these shelves leads to full differentiation of the hard and soft palate, which compartmentalize the nose and mouth; if complete fusion is not accomplished between the prominences or the shelves, an OFC can result (Yoon et al., 2000).

Several genes are involved in the pathways driving these processes. The complexity of the cellular processes in facial development leads to difficulty in identifying exactly which steps are disrupted and which factors are the most disruptive: genetic or environmental. OFCs present across a wide spectrum, from mild to severe, and are typically categorized into three types depending upon involvement of the lip and hard/soft palate: clefts involving only the lip (CL), clefts involving both the lip and palate (CL/P), and clefts involving only the palate (CPO) (Jiang et al., 2006; Yoon et al., 2000). Each of these cleft types also have variability, including bilateral or unilateral presentation.

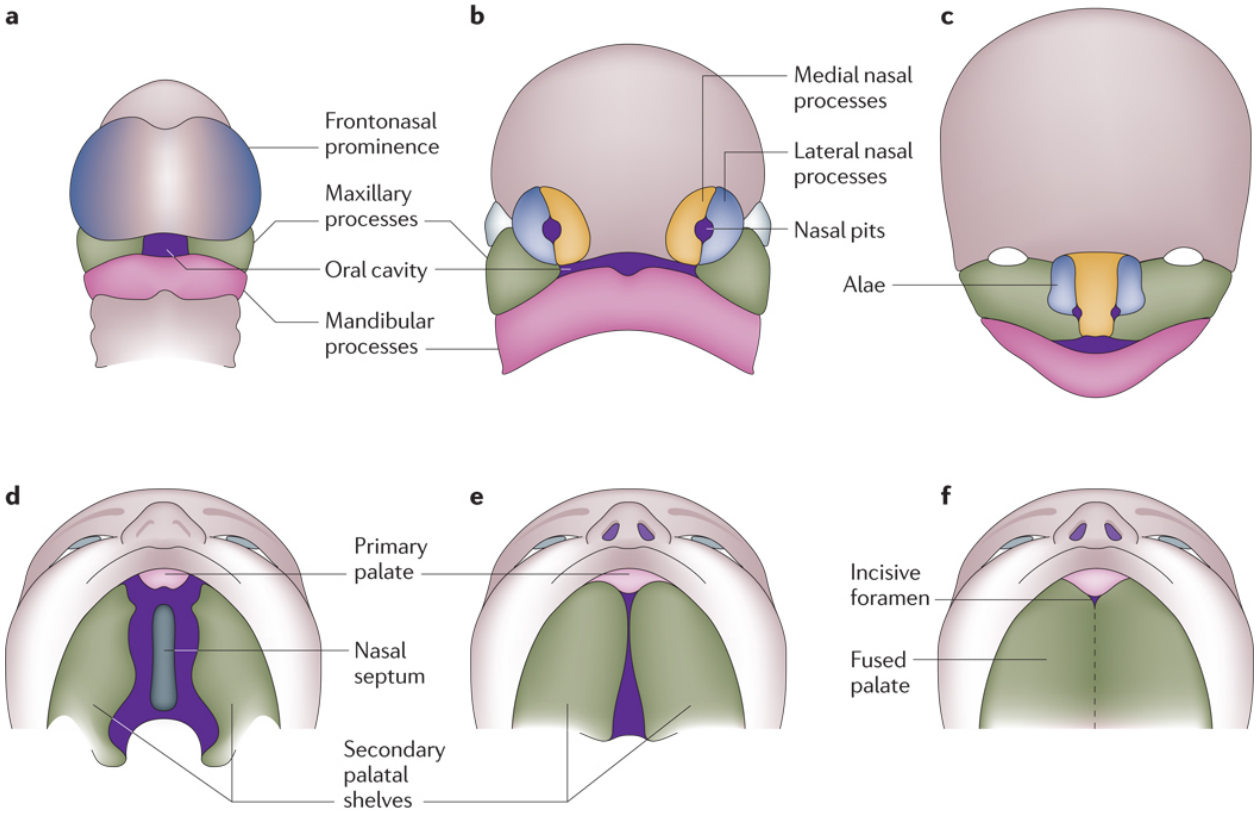

Figure 2 Development of the Lip and Palate

(A) The developing frontonasal prominence, paired maxillary processes and paired mandibular processes surround the primitive oral cavity by the fourth week of embryonic development. (B) By the fifth week, the nasal pits have formed, which leads to formation of the paired medial and lateral nasal processes. (C) The medial nasal processes have merged with the maxillary processes to form the upper lip and primary palate by the end of the sixth week. The lateral nasal processes form the nasal alae. Similarly, the mandibular processes fuse to form the lower jaw. (D) During the sixth week of embryogenesis, the secondary palate develops as bilateral outgrowths from the maxillary processes which grow vertically down the side of the tongue. (E) Subsequently, the palatal shelves elevate to a horizontal position above the tongue, contact one another and commence fusion. (F) Fusion of the palatal shelves ultimately divides the oronasal space into separate oral and nasal cavities. Reused from Dixon et al., 2011, with permission.

2.3 Inheritance

Although there has been much debate about the inheritance pattern in NSCL/P across studies, classification under a multifactorial threshold model, with both genetic and environmental causes, remains most consistent (Grosen, Chevrier, et al., 2010; L. J. Howe et al., 2018). The key features of a threshold model of inheritance are: 1. Most affected children have unaffected parents; 2. Recurrence risk increases as number of affected children in the family increases; 3. Recurrence risk increases with severity of the defect; 4. Consanguinity slightly increases the risk for an affected child; 5. Risk of affected individuals falls quickly as the degree of relationship extends; 6. When two sexes have a different probability of being affected, the least likely sex, if affected, is the most likely sex to produce an affected offspring (Grosen, Chevrier, et al., 2010).

Several features of NSCL/P are most consistent with the threshold model (Beaty, Marazita, & Leslie, 2016). Males and females have different likelihoods of being affected (Watkins et al., 2014). There is a high relative risk of recurrence among first-degree relatives of an affected family member; studies have reported a range of 24 to 82-fold increases in recurrence risk for first degree relatives of an individuals with NSCL/P, with recurrence returning almost to the population risk in third-degree relatives (Klotz et al., 2010; Sivertsen et al., 2008). Increased severity of clefting is also associated with higher risk for recurrence. However, some observations, such as that of most affected children having unaffected parents, could be explained by other inheritance phenomena, such as reduced penetrance or variable expressivity of the phenotype.

In a large cohort study out of Denmark, of approximately 54,000 relatives of individuals with NSCL/P, Grosen et al. demonstrated a risk for recurrence of about 4% (RR=17; 95% CI 3.5-

5.8%) for individuals with a first degree relative affected by NSCL/P. Second and third-degree relatives were reported to have four times and three times the recurrence risk, respectively (RR=4; 95% CI 3-5 and RR= 3; 95% CI 2-4). Increased risk has been proposed in the case of multiple affected siblings (See Table 2) (Basha et al., 2018). Grosen et al. did observe some increase with additional affected siblings, but findings were not statistically significant. Important to note, here, is that these relative risks are based on the Danish population, with an estimated NSCL/P prevalence of 2.1 per 1,000 births; as such, using these numbers to counsel patients of different racial and ethnic background should be done with some caution (Lees, 2008). The degree of heritability can also differ between populations based on risk allele frequencies, which affects recurrence risk estimates (Spence, Westlake, Lange, & Gold, 1976).

The researchers from Denmark also found that recurrence of the same cleft type (CL, CL/P, or CPO) was most common and that severity was positively correlated with likelihood of recurrence. Unlike Grosen et al., an earlier Norwegian cohort study by Sivertsen et al. did not observe predominant recurrence of the same cleft type in affected families, causing the researchers to challenge the multifactorial threshold model of inheritance in NSCL/P (Sivertsen et al., 2008). Grosen et al. repeated their analysis method on the cohort from the Sivertsen et al. study and concluded that recurrence of the same cleft type did indeed predominate within families. The threshold model of inheritance remains the most supported mechanism in NSCL/P, despite some disputes in the literature, which lack sufficient evidence to fully reject it.

Table 2 Recurrence Risk to Siblings Based on Affected First-degree Relatives

Affected Individual(s)	Recurrence Risk to Future Sibling
One child	4%
Two Children	9%
One Child + One Parent	17%

2.4 NSCL/P in the Newborn Period

Most often, babies born in developed countries with isolated NSCL/P do not have any life-threatening complications at birth, and do not account for a significant portion of infant mortality in developed countries (Lewis, Jacob, Lehmann, & Health, 2017). The most significant problems include trouble creating proper intraoral suction, leading to an inability to latch for feeding (Kaye et al., 2017) and, less commonly, concerns for airway obstruction and sleep disordered breathing (Muntz et al., 2008). Outcomes in the newborn period invariably improve for babies when their parents receive counseling and feeding instruction, prenatally or at birth (Hubbard, Baker, & Muzaffar, 2012; Smedegaard, Marxen, Moes, Glassou, & Scientsan, 2008).

Feeding problems are the most common concern among both parents and healthcare providers in the neonatal care of a patient with CL/P (Lewis et al., 2017). Without the ability to create the necessary pressure with their lips (due to cleft lip) and suction with their mouths (due to cleft palate), babies with CL/P can have unproductive, prolonged feedings that can result in a host of problems, from failure to thrive, to frustration and consequent poor maternal child bonding (Kaye et al., 2017). These prolonged feedings are strenuous for mother and baby, with the baby often expending more energy than it consumes in a feeding, leading to poor weight gain (Kaye et al., 2017). Newborns with CL/P usually need specially designed feeding accessories, such as bottle nipples which release formula or breastmilk in response to lower pressure and suction than standard nipples (Lewis et al., 2017). Additionally, parents need training from knowledgeable healthcare providers, such as lactation consultants, specialized occupational therapists, or a primary care provider familiar with CL/P, to support effective bottle or breast feeding (Hubbard et al., 2012; Kaye et al., 2017).

Recently, Kaye et al. looked at a common measure of infant nutritional health, the rate of return to birthweight, in an effort to more precisely evaluate success of feeding in 100 newborns referred to a cleft craniofacial center for CL/P (2017). The researchers evaluated by cleft type, finding that babies with CL had a return to birth weight in an average of 13.58 days, which is in line with the recommended time of 14 days, while babies with CLP or cleft palate only (CPO) had delayed rates (15.88 and 21.93 days, respectively). In several cases, babies were using inappropriate feeding devices (19.4% of babies with CL and 14.3% babies with CP). There was a significant relationship between older age at first assessment by the cleft team and slower return to birthweight. These last two findings highlight the value of early assessment and intervention by a specialized cleft team in reducing failure to thrive.

Longer hospitalization after birth and higher rates of neonatal intensive care unit (NICU) admission have been reported in CL/P patients, in some cases due only to poor feeding and hospital inexperience with OFC care (Hubbard et al., 2012). NICU admission is costly and unnecessary in most newborns with isolated, NSCL/P (Hubbard et al., 2012). A small, retrospective study of 43 children born with cleft lip and palate (CLP) aimed to evaluate the effect of prenatal counseling, including feeding instruction, on rates of NICU admission solely for the issue of feeding. Twenty of the 43 children's parents received counseling and instructions on care and feeding of newborns with CLP. Two out of 20 newborns (10%) whose parents had received prenatal counseling versus five out of 23 newborns (21%) whose parents received no prenatal counseling were admitted to the NICU; however, this result was not statistically significant, possibly due to small sample size (Hubbard et al., 2012).

Further support for early intervention can be drawn from Denmark, where all parents of a child born with an OFC receive counseling by trained providers, with the goal of increasing the

parents' confidence and ability to bond with the child (Smedegaard et al., 2008). In a prospective study of 115 infants with OFCs, Smedegaard et al. showed that there was no difference between babies with or without OFCs in the length of hospitalization or growth at 5 and 12 months. The researchers attributed the improved outcomes to the counseling parents received.

In summary, while NSCL/P does not typically present with additional birth defects, there is still an impact on an individual's overall health in infancy. Outside physical health concerns, without adequate support, there can be psychological concerns for parents, creating barriers to parent-child bonding. Best ensuring the health and well-being of individuals born with OFCs requires an extensively interdisciplinary healthcare approach, focused on the whole person rather than only their physical birth defect. Medical management ideally includes the necessary specialists who educate and support parents and family members, openly communicate with the family and other providers about treatment plans and follow-up, and see that patients receive emotional support from the beginning (Cleft Palate-Craniofacial Association, 2009). This type of planning is best done before there is a newborn child with a cleft born in the hospital, when there are many competing concerns (Cleft Palate-Craniofacial Association, 2009).

Cleft-craniofacial centers bring together craniofacial, oral, and plastic surgeons, audiologists, speech language pathologists, breast feeding specialists, social workers and psychologists, child development and education specialists, orthodontists, geneticists, and others. These centers stay up to date with treatment guidelines, best practices, and are recognized for their expertise in coordinating the complex care of patients with OFCs, effectively. Patients and families with access to cleft-craniofacial centers have the necessary support to thrive, physically, mentally, and emotionally. It is highly recommended for individuals to be seen for evaluation in the first few days or weeks of life (Cleft Palate-Craniofacial Association, 2009).

2.5 Impacts of OFCs in Infancy, Childhood and Adulthood

After the newborn period, there are other important considerations for the ongoing care of an individual with an OFC. In early childhood, the development of language should be carefully monitored, with awareness of any delays requiring intervention (Berryhill, 2016; Cleft Palate-Craniofacial Association, 2009). There can be frequent ear infections with the potential to damage hearing; hearing loss can also lead to or exacerbate issues with language development (Berryhill, 2016). Individuals with OFCs are known to be at a slightly increased risk for learning disability, which can be partially mitigated with proper attention to speech and hearing problems (Berryhill, 2016).

Some children, teens, and adults experience poor self-image and difficulty with socialization as a result of real or perceived speech difficulties and/or facial asymmetry that remains after surgery (Glener et al., 2017; Kaye & Lybrand, 2016a). In addition, research suggests there are other health problems associated with OFCs, such as an increased cancer risk (Bille et al., 2005), which can further affect quality of life. The right care in every life stage can go a long way in ensuring ideal quality of life for individuals with OFCs.

2.5.1 Surgical Repair of Clefts

Surgery for NSCL/P is complex and goes far beyond improving physical appearance. The goal is to restore function of disrupted tissue and improve aesthetic. OFCs may disrupt the arrangement and function of the facial and oropharyngeal muscles, the structure and growth patterns of the facial bones, velopharyngeal function, and the separation of facial cavities (mouth,

nose) and the sinuses (Nahai, Williams, Burstein, Martin, & Thomas, 2005). Surgeons must consider the muscle arrangement and future growth in the primary repair.

Surgery is not typically performed immediately after birth. In the time before surgery, important presurgical manipulation of particularly wide clefts or protruding nasal or alveolar tissue can be performed using taping, lip adhesion, or devices fitted to the palate. This step brings the tissue closer together and improves alignment prior to surgery, which can reduce tension on the sutures post-surgery (Nahai et al., 2005). Clefts of the lip are typically repaired once the baby gains a bit of weight, at around two to three months of age, with palate and alveolus repairs later, around seven months of age (Gatti, Freda, Giacomina, Montemagni, & Sisti, 2017; Nahai et al., 2005). At the time of palate repair, additional surgeries may be necessary to address eustachian tube dysfunction/ear infections (Berryhill, 2016). Clefts are seldom repaired with a single surgery, but instead multiple are required at different stages of growth. In a recent retrospective study of 71 patients at a single U.S. children's hospital who underwent repair of an OFC between 1988 and 2014, patients had an average of five to six (for CL) and nine to ten procedures (for CLP or CPO) (McIntyre et al., 2016).

About 15% of patients need a secondary surgery to correct VPD (Bickham et al., 2017). There may also be minor surgical revisions around the time a child enters school, to improve lip contour or address a widened scar, for example; this is often requested by parents who are concerned about their child's social integration (Nahai et al., 2005). At age seven to nine, children with clefts involving the palate need an alveolar bone graft to provide bone for teeth to erupt and to stabilize the midface during subsequent growth. Jaw surgery may be needed in the middle teen years to correct the position of the maxilla. Lastly, the nose is often affected by a cleft, leading to asymmetry or shape irregularities like flattening or widening (Nahai et al., 2005). Rhinoplasty is

performed in the middle to late teenage years and is often the final revision surgery. Along the way, therapies are also introduced to support the function of tissues affected by clefts, including speech therapy (beginning, ideally, at around 14 to 16 months of age) and orthodontics (Cleft Palate-Craniofacial Association, 2009).

Management of an OFC is essentially a life-long process. The process requires significant coordination, which can be difficult for families, who may have to travel long distances or relocate to access a cleft craniofacial center. There are 174 approved cleft craniofacial teams in the U.S. ("ACPA Approved Teams - ACPA Family Services," n.d.). Centers are often aware of these challenges and may have some resources available to help with travel and accommodations (Kaye & Lybrand, 2016b). Timing of surgical interventions, experience of the surgeon, and long-term follow-up are critical to achieving optimum long-term outcomes (Nahai et al., 2005).

2.5.2 Associated Physical Health Problems

Over time, associations with health problems not directly related to OFCs have been observed in affected individuals. In a large Danish population-based study, certain types of cancer were observed at higher rates than expected in the OFC population. Specifically, female breast in individuals with CL/P, male lung in individuals with CL/P, and primary brain in females with CPO, were seen at an increased incidence (Bille et al., 2005). There was no evidence for a generally increased cancer risk, as no other cancer types were seen more frequently than expected.

Developmental delay is slightly more common in individuals with OFCs (Chollet et al., 2014). This relationship can be difficult to isolate, as children with complex childhood medical needs may have some related developmental and learning delays due to time in the hospital or

under anesthesia. However, there is also evidence of altered brain patterning, which may increase the odds of learning delays or disabilities in the OFC population (Chollet et al., 2014).

Dental health is critically important at an early age in the OFC population. It is an area of early intervention by the cleft team, as dental caries and other dental problems are more common in children with OFCs, and the consequences of poor dental health on alveolar integrity can have long-term effects on the permanent dentition and orthodontic outcomes (Nahai et al., 2005). Additionally, dental caries were found to be a determinant of health-related and oral-health related quality of life (QOL) in a recent study (Corrêa de Queiroz Herkrath, Herkrath, Bessa Rebelo, & Vettore, 2018).

2.5.3 Psychosocial Impact

QOL is a subjective measure of outcomes in healthcare and includes aspects of the physical, mental and social functioning perceived by an individual ("WHO | WHOQOL," n.d.). One key to this definition is the word "perceived". Historically, treatment outcomes for OFCs have been measured objectively by clinicians who compare certain post-treatment measurements, functional assessments, or photos to pre-treatment conditions. This objectivity serves an important function in measuring the clinical success of procedures and looking for ways to improve them, but patients often have more complex feelings about the outcomes (Herkrath, Herkrath, Rebelo, & Vettore, 2015). Even with optimal surgical and therapeutic outcomes, patients with OFCs have lingering social and emotional concerns (A. P. C. Herkrath et al., 2015). QOL assessments may be useful to supplement objective assessments and find areas in which patients can be better supported.

Previous studies of QOL in the OFC population suggest that psychosocial functioning is the most impacted area of health, with the majority of affected individuals reporting less of a perceived effect on their physical health (A. P. C. Herkrath et al., 2015). While individuals with OFCs do not, on average, appear to experience profound mental health problems, there is evidence to suggest some level of negative impact on areas such as socialization, satisfaction with physical appearance, and self-esteem (Mani, Carlsson, & Marcusson, 2010).

In the post-partum period, parents often experience challenges coping with and processing the diagnosis, prognosis, and necessary treatment of an OFC; many express desire for education and support (Kaye & Lybrand, 2016b). It is especially important to provide emotional support to mothers of children with an OFC in the post-partum period, particularly prior to surgical repair. This is a vulnerable period for mother-infant bonding, interactions, and infant attachment (Murray et al., 2007).

In early childhood, speech difficulties can cause issues with verbal communication between a child with an OFC and their family members. When a child feels they are not being well-understood, it can lead to frustration and acting out; behavioral interventions can support the child and the family functioning (Kaye & Lybrand, 2016b). Later, in the pre-teen years, children with facial differences report the most psychological and social stress during middle school (A. P. C. Herkrath et al., 2015). Middle school-aged children in general often report disruptions in social life and increased bullying, so it is difficult to completely isolate the factors for children with facial differences; regardless, middle school-aged children with OFCs report lower QOL than children of the same age without facial differences as well as elementary school-aged children and adults with OFCs (Glener et al., 2017). Knowing this is a particularly difficult period can guide clinical mental health interventions to better facilitate peer-integration and relationship-building.

The mental health of adults with OFCs may be impacted, even after formal surgery and treatments have been completed. A cross-sectional study of 86 individuals with unilateral CL/P, with a 35-year average long-term follow-up, suggested that adults do not have poorer overall health-related QOL, but do report poorer mental health (Hamlet & Harcourt, 2015). Hamlet and Harcourt found that younger adults (20 to 32 years of age) and women report more significant differences in mental health than those aged 33 to 47 years and men. Other studies also report that females with OFCs have more dissatisfaction with appearance and may be at higher risk for poorer QOL (Herkrath et al., 2018; Herkrath et al., 2012; Paganini, Hörfelt, & Mark, 2018).

It is important to note that because of the somewhat subjective nature of measuring QOL, study design can vary significantly between analyses; this has led to inconsistency and a subsequent lack of consensus in the literature as to the degree of impact and determinants related to QOL in individuals with OFCs. Underscoring this point is a recent cross-sectional study of 134 NSCL/P patients between the ages of four and 14 years in a single German cleft treatment center, which actually showed better QOL among the treatment group compared to the German population-based comparison group (Naros, Brocks, Kluba, Reinert, & Krimmel, 2018). The authors concluded with a suggestion that the coordinated care of the patients within a single cleft center, including support from speech therapists and other medical professionals, may partly explain the better QOL observed. They suggest that this centralized care may have facilitated more positive family interaction, better communication skills, and higher self-esteem among patients. If this is true, it provides another strong incentive for ensuring access to centralized care for patients with OFCs. Regardless of the heterogeneity, studies on QOL in the OFC population largely conclude that it is important for providers to place the same emphasis on mental health as they do on physical health in the treatment plan.

2.6 Economic Impact of NSCL/P

There is a shortage of reliable information available on the economic impact resulting from the care of OFCs (Yazdy, Honein, Rasmussen, & Frias, 2007). This is partly due to the various possible ways to assess cost. Studies report costs relative to different parties, such as insurance companies' claims, hospitals/healthcare facilities' billing, and patients' out-of-pocket expenses. Differences in the type of cost reported results in an inability to compare results from smaller studies to assess cost on a larger, population scale. Additionally, current studies vary in terms of how comprehensive the assessment of cost is. There is an incomplete understanding of surgical costs, costs of additional therapies (speech, occupational), mental health services, health-related quality of life, and loss of productivity for individuals and caregivers (Yazdy et al., 2007). Further complicating estimates, many studies do not separate syndromic versus non-syndromic OFCs; this is problematic because syndromic forms often occur with comorbidities that inflate overall healthcare costs (Boulet, Grosse, Honein, & Correa-Villaseñor, 2009). Assessment of economic burden is an essential piece of public health planning and allocation of resources for prevention, research, and treatment of OFCs (Yazdy et al., 2007).

A recent study looked at the difference in costs between children with and without an OFC, using data collected by the MarketScan® Commercial Claims and Encounters database from 2000-2004, on claims reported by about fifty organizations with private health insurance plan enrollees. Using information on both insurance reimbursements and out-of-pocket payments by families for health services, medical care, inpatient and outpatient visits, and prescription drugs, the researchers reported costs for the care of children (age 10 years or younger) with an OFC that were eight times higher than those of children without an OFC (Boulet et al., 2009) . The researchers employed a medical geneticist to review cases and separate non-syndromic and syndromic forms of CL/P, a

major strength of the study. This analysis was also repeated for data from the same database for years 2006-2010 with similar results.

Aiming to characterize expenditures for children with OFCs in the Medicaid population, a retrospective case-control study in North Carolina measured total medical expenditures, including inpatient, outpatient, mental health, home health, dental, and well-child care for the period of 1995 to 2002 (Cassell, Meyer, & Daniels, 2008). During the first five years of life, expenditures were five to six times higher for children with an OFC than without. Children with CL/P and CPO incurred higher healthcare costs than those with isolated cleft lip. This study did separate children born with isolated OFCs and those born with additional congenital anomalies, finding higher costs with the latter group. The mean first-year expenditures for a child with an isolated OFC amounted to $10,099, compared to $3,900 for a control.

2.7 Modeling Risk

Predictive modeling for genetic conditions has two main applications: clinical (healthcare provision and management) and public health (population screening and directing interventions). In a genetic counseling setting, predictive models are used to refine the odds that a patient, who is already known to be at an increased risk based on personal or family history, will develop or have a child with a given condition. A risk estimate based on family history alone may increase or decrease after accounting for the presence or absence of additional risk and protective factors. The Gail model is one such model; it is used to estimate breast cancer risk for women based on history of breast cancer in her first-degree relatives (genetic factor), plus medical and reproductive history information, such as the age at which she began menstruating (environmental factor) (Gail et al.,

1989). The Gail model was developed using a large population database on breast cancer incidence and personal characteristics of the women who developed it. Gail started with a large list of characteristics, eventually paring it down to the most reliable predictors. This process is typical of model building.

Models are only as good as the data used to develop them, so there are always limitations based on the quality and type of available information. The Gail model, for example, was initially best at predicting breast cancer risk in Caucasian women who were receiving routine breast screening as recommended because the women in the data set used to derive RR figures shared those characteristics (Euhus, 2001). Development of any model starts with rigorous definition of relative risk (RR) for each factor of interest, as RR figures will form the basis of the mathematical adjustments made by the model (Lee, Bang, & Kim, 2016). Important to note, here, is that RR must be derived from a prospective study design, in which individuals without an outcome of interest are followed to observe incidence (Trevethan, 2017). Retrospective study designs provide ORs based on prevalence; ORs can only be approximated to RR in the case of rare outcomes, not in common ones. As will become evident later in this manuscript, this forms a significant limitation in the case of NSCL/P.

2.7.1 The Value of Predictive Modeling for NSCL/P

A predictive model for NSCL/P would be most pertinent in the pre-conception or prenatal genetic counseling setting. Technically, prenatal screening already occurs at the population level for OFCs in the United States. OFCs can be detected prenatally by routine, second trimester transabdominal ultrasound. Unfortunately, detection rates vary widely across medical centers, leading to a significant number of missed diagnoses (Maarse et al., 2010). Milder clefts of the lip

are less likely to be detected, and CPO is rarely diagnosed prenatally (Maarse et al., 2010). Therefore, using a predictive model for OFCs in the prenatal setting may improve chances of prenatal diagnosis by prompting more careful ultrasound examination and/or appropriate referral to an experienced tertiary center for follow-up. The most significant benefit of predictive modeling for individuals at increased risk for having a child with NSCL/P, however, would be the ability to perform it prior to conception.

Predictive modeling for NSCL/P would have a substantial impact on genetic counseling for recurrence. Identifying individuals at increased genetic risk prior to conception, whether by genotype or representative phenotypes, will allow for informed parental reproductive decision-making, education and resource identification, and preparation. In addition, awareness of the increased risk allows for the opportunity to reduce that risk to the extent possible, by addressing some of the modifiable risk factors for NSCL/P, such as smoking status and folic acid supplementation. Though the effects of currently known modifiable factors on risk may be small, parents may still benefit from taking steps to reduce risk. Genetic counselors focus on not only the physical health outcomes of genetic counseling, but the psychological ones, such as informed decision-making and adjustment. Providing some control in a situation in which there is little can provide patients with some self-efficacy. This empowerment, regardless of the physical health benefits, can have positive effects on mental health, such as coping and adapting to life with a chronic condition (McAllister, Dunn, Payne, Davies, & Todd, 2012).

Aside from the potential benefits in patient care, modeling could also help in NSCL/P research. Evaluating the effect of various factors on risk could help to clarify the degree of genetic, teratogenic, and environmental influences on the development of OFCs. Identifying environmental factors with strong effects on risk can also prompt studies to identify the genes which interact with

that particular factor during development; this has already been seen with studies implicating genetic variants in *MTHFR* and *RBP4* in NSCL/P risk, because they are involved in the biological processing of folate and vitamin A, respectively (Zhang et al., 2018). If robust modeling reveals new risk factors, especially modifiable ones, it could inform primary prevention efforts.

In the present study, the individuals from case families are considered "high-risk", that is, they have an increased baseline risk for recurrence of NSCL/P based on having a first, second, or third degree relative who is affected. Models developed for a high-risk population do not always perform satisfactorily in the general population (public health setting), where the baseline risk is assumed to be much lower. In the population setting, there are additional considerations for screening. The present study focused on a future clinical application in a high-risk group (i.e. those with family history); population screening is explored in the final chapter of this manuscript (see Public Health Essay).

2.7.2 Potential Benefits of Better Risk Estimates in NSCL/P

While there is no known literature on the benefit of pre-conception counseling for NSCL/P, specifically, a qualitative study by Nusbaum et al. (2008) highlights several advantages and disadvantages of being able to diagnose CL/P prenatally. An even number of parents interviewed either knew about the CL/P prenatally or found out when the baby was born. The advantages may include: allowing caregivers to prepare themselves psychologically, seek out information, plan for care needs, seek additional genetic testing/screening for associated anomalies, and allow room for alternative reproductive decisions, including termination. The disadvantages may include: emotional distress experienced during the pregnancy, lack of prenatal surgical intervention, inability to predict type and severity with high certainty, and incurred financial costs (Nusbaum et

al., 2008). Families may also misunderstand the diagnosis and its implications (Nusbaum et al., 2008).

Analysis of parents' experiences with the unexpected birth of a child with CL/P revealed several adverse outcomes. Parents reported that reactions of the medical team can add to the alarm they already feel; other times providers minimize the cleft, giving parents unrealistic expectations about the simplicity and timing of surgical repair and follow-up (Nusbaum et al., 2008). Many times, those in the delivery room are not well-versed in caring for babies with CL/P, practically or in terms of sensitivity. A hospital may not have necessary items available for optimal care, such as specialized bottle nipples and devices to help the newborn feed (Hubbard et al., 2012; Kaye et al., 2017).

Regardless of whether parents are anticipating the birth of a baby with CL/P or are surprised at birth, it is important that psychosocial considerations are made to help them cope with the range of positive and negative emotions they may feel (Lewis et al., 2017). Better risk estimates could affect the ability to predict the chances of NSCL/P before conception and/or diagnose a cleft prenatally. The timing of CL/P diagnosis (prenatally or postnatally) has the potential to affect parents' psychological adjustment and preparation, as well as the immediacy and quality of medical care the child born with CL/P receives (Hubbard et al., 2012; Kaye et al., 2017; Lewis et al., 2017; Nusbaum et al., 2008).

2.7.3 Previous Attempts at Modeling Risk in NSCL/P

Due to inconsistent findings across studies for NSCL/P risk factors, a lack of robust population level information on known risk factors, and small effect sizes, few studies have attempted to leverage the combined effect of genetic, phenotypic, and environmental risk factors

for NSCL/P in predicting occurrence. Prospective studies at the population level to provide relative risk figures have not been undertaken. However, many studies of facial morphology have been performed, in attempts to design models able to correctly classify parents by case versus control family status, using select facial features only.

Several important lessons have been learned from these studies, and methods have evolved (reviewed in Wyszynski, 2002). Early studies did not consider sex-related differences between parents; most early studies averaged all measurements between parents (Kurisu et al., 1974). Then, Ward et al. changed this, with a study in which they performed analysis differently, finding that the grouping of individuals pointed towards different cleft etiologies. In some pairs, both parents had atypical measurements (15%), in others, only one parent had atypical facial measurements (54%); both observations indicate some level of genetic predisposition. In a substantial portion of pairs, however, neither parent had atypical measurements (31%), possibly indicating truly sporadic cases with environmental causes (Ward, Bixler, & Raywood, 1989).

In 1987, Crawford and Sofaer developed an asymmetry score for CL/P, based on previous studies suggesting bilateral asymmetry as a marker for predisposition to clefting (Crawford & Sofaer, 1987). Variables included in the score were determined by stepwise logistic regression; significant variables included were measures of molar size, *atd* angle (a palm of the hand measurement between triradii points a, d, and t), fifth finger length discrepancy, and ridge count (fingerprints). Using unaffected participants from familial (i.e. multiplex) and sporadic (i.e. simplex) families, the scoring system indicated that 85% of individuals from the familial group showed high genetic predisposition, compared with only 26% in the sporadic group. However, it is not clear if there was a cross-validation step to assess the model.

A 1998 study by Mossey et al. evaluated the cephalometric measurements of 83 case parents— parents of a child with an OFC— (35 parents of a child with CP and 48 parents of a child with CL/P) and 99 control parents — parents of an unaffected child. The goal was to test the hypothesis that genetic and morphometric factors can indicate a predisposition to OFCs. Using radiographs of the participants' skulls, the researchers compared case and control parents across a set of craniofacial measurements thought to be wholly representative of the skeletal craniofacial morphology. They found that fathers of a child with a cleft had smaller mandibular, symphyseal, and maxillary areas and shorter palatal length. Mothers had greater anterior facial height, total facial length, clivus length, and anterior cranial base (Mossey, Arngrimsson, Mccoll, Vintiner, & Connor, 1998). Seventy-six parents in the same study were also genotyped for allelic heterogeneity at the *TGF-alpha* locus, revealing a significantly higher prevalence of the C2 allele of the *TGF-alpha*/*Taq*I polymorphism in parents of a child with a CPO (RR= 4.50, P<0.01) or CL/P (RR= 3.79, P<0.05) in comparison to controls. *TGF-alpha* polymorphisms are no longer considered strong genetic factors in clefting, which may explain why it did not add much predictive value to this model (see Table 1).

The second goal of Mossey et al.'s (1998) study was to show differences between parents of CPO versus CL/P. Using three of the measurements that differentiated best between CPO and CL/P with genotype, the model they created accurately classified parents of a child with CPO 76% of the time and parents of a child with CL/P 94% of the time. Without genotype data, the classification accuracy went down to 78% in CL/P parents and 72% in the CPO parents. They hypothesized that these results, with the predictive accuracy of CL/P being further reduced than CPO in the absence of genotype, indicated a greater genetic influence in the development of a CL/P and a more substantial environmental influence in the development of CPO. Mossey at al.

(1998) demonstrated that there are craniofacial differences between both parents of a child with any cleft type, but also between the parents of children with CPO versus CL/P (1998). In light of increased evidence of different etiologies since Mossey's study, models likely need to consider the phenotypic spectrum.

2.7.4 Exemplar Predictive Model for a Congenital Condition

Using population-based data on non-syndromic birth defects from the National Birth Defects Prevention Study, a group of researchers attempted to create a predictive model for the occurrence of open neural tube defects (ONTDs) (Agopian et al., 2012). Although ONTDs are a common birth defect, with significant health impact and healthcare costs, and multiple risk factors have been consistently linked with their development, this was the first attempt to combine multiple risk factors into a predictive model and assess its performance. Agopian et al. (2012) combined several known risk factors, such as family history, diet, obesity, and ethnicity, into a multivariate logistic regression model (Agopian et al., 2012). The model's ability to correctly distinguish between mothers of children born with an ONTD and mothers of children born without an ONTD was poor (area under the receiver operating curve was 0.55-0.59). Despite the many studies that have been completed to assess for risk factors for ONTDs, the poor predictive ability of this model highlighted the need for additional studies to identify significant risk factors for ONTDs.

The Agopian et al. (2012) study is relevant to the development and exploration of a similar model for OFCs. ONTDs have a similar incidence to OFCs (CDC, 2018b), and both birth defects are considered multifactorial, having multiple environmental and genetic factors associated with their occurrence. Furthermore, OFCs and ONTDs each have at least two, somewhat distinct

manifestations, ranging from mild to severe. ONTDs occur as varying degrees of spina bifida and anencephaly, and OFCs occur as CL/P and CPO. Both spina bifida and anencephaly, and CL/P and CPO, are thought to have some unique risk factors, while also having overlapping ones within the overarching category of ONTDs and OFCs, respectively. The Agopian et al. study investigated the predictive ability of the model in a composite group (all ONTDs) and in component groups (spina bifida versus anencephaly). It may be reasonable to assess future models for OFCs in a similar way, using all OFCs in a composite group and CL/P and CPO as component groups. Indeed, Mossey et al. (1998) also made this distinction in their models.

Predictive modeling has been used extensively in epidemiological studies surrounding conditions such as cancer and heart disease, but not as extensively in the study of birth defects (Agopian et al., 2012). There is substantial value in developing such models. Accurate predictive modeling can aid in identifying individuals at the highest risk, which provides couples the ability to make family planning decisions and prepare themselves for the long-term care of a child with an OFC. Conversely, when predictive models perform poorly, it highlights the ongoing need for research on risk factors, despite studies already undertaken. Discovering how much of a given disease or birth defect can be explained by known risk factors, both genetic and environmental, is valuable in itself. If a model with many proposed risk factors included performs poorly, it may prompt a shift in direction with new and creative study approaches. Finally, if a predictive model does show some promise, but is underpowered or has other significant limitations, understanding those issues can allow for planning of repeat studies to attempt model improvement.

2.8 Subclinical Phenotypes and Markers for Recurrence Risk

It is not a new idea to use the features of the face to infer genetic information or to evaluate human disease. Dysmorphologists have long used the structure of the face and body to detemine the underlying genetic conditions behind them. In recent years, there have been even more robust attempts to use phenotypes to guide patient care and diagnosis. Technology capable of evaluating the face three-dimensionally may have clinical utility (Gurovich et al., 2019). The utility of such technology depends on robust research linking phenotype to human disease. In regards to NSCL/P, specifically, several decades of research have centered on cataloging features that appear to co-segregate with OFCs.

In phenotype studies of families with NSCL/P, researchers have been looking for the specific features associated with NSCL/P that may serve as risk factors for recurrence. Across many studies, several significant differences in phenotype have emerged among close relatives of an individual with NSCL/P, versus individuals from unaffected families. If the threshold model of inheritance holds in NSCL/P, genetic susceptibility is based on inheriting several genetic variants, each conferring variable levels of risk. Based on previous findings and the concept of a threshold model of inheritance, it is reasonably likely that recurrence risk varies among individuals with a family history of NSCL/P, even within the same family. Therefore, if there are phenotypic markers for genetic risk, the individual recurrence risk of a given relative is expected to range from higher to lower, depending on the presence or absence of certain phenotypes.

Successful subclinical phenotyping of these relatives may one day allow healthcare providers to calculate personalized recurrence risk. Currently, NSCL/P reproductive genetic counseling is based on empiric recurrence risk figures, calculated from multiple large cohort studies of mostly white populations (Grosen, Chevrier, et al., 2010). Thus, it is not possible to

individualize risk assessment. However, it may be possible to develop a prediction model, using established phenotypic and environmental risk factors. Additionally, studying and better understanding transmission of these subclinical traits may help to separate out the truly sporadic cases of NSCL/P (i.e. those with low risk for recurrence) from those with significant potential for recurrence. The underlying assumption of this concept is that subclinical phenotypes can serve as a proxy for the number of risk alleles carried by a given family member.

2.8.1 Body Patterning and Asymmetry

Observation of certain bilateral asymmetries among families affected by NSCL/P may indicate some character of fetal development that underlies cleft susceptibility. This is based on the concept of developmental instability. Developmental instability describes an organism's suboptimal adaptive response to developmental stresses (Livshits, Kobyliansky, Opitz, & Reynolds, 1987). Fluctuating asymmetry, or non-size related variance between bilateral body structures, can be a sign of developmental instability (Dongen, 2006).

Dermatoglyphics is the study of fingerprints. There is a long history of studying fingerprints in relation to genetic conditions because they can provide information about early fetal development (Weinberg, Neiswanger, et al., 2006). They are of interest in NSCL/P partly due to similar timing in embryological development. Fingerprints found in individuals with a family member affected by NSCL/P show higher levels of asymmetry (Neiswanger et al., 2002; Singh & Nathani, 2017) and possibly different patterns of loops, arches, and whorls, than individuals with no family history of OFC (Singh & Nathani, 2017). Of note, the results found by Singh and Nathani may be affected by the presence of consanguinity in a large proportion of parents from affected families.

Cheiloscopy (study of lip prints) reveals patterning differences in individuals with NSCL/P and their family members, compared to unaffected individuals and families (Neiswanger et al., 2009). Lip prints are typically seen as a pattern of mostly vertical lines. Crosshatching (reticular), semi-circular whorls, branching lines, and subtle pits have been observed at a two to five-fold increase in frequency in families with a history of NSCL/P, depending on how narrow the definition of atypical lip prints is (Neiswanger et al., 2009; Singh & Nathani, 2017).

Atypical dentition (tooth formation and arrangement) is another heavily studied phenotype in NSCL/P. Differences in tooth shape, size, eruption, and positioning are well-described in individuals affected by NSCL/P. The question, however, has been whether or not physical forces and surgical intervention resulting from the cleft lead to observed differences, or if they are caused by the same primary – perhaps, genetic – cause of the cleft (B. J. Howe et al., 2015). Analysis of dental anomalies in the unaffected relatives of an individual with a cleft has produced mixed results. A few studies report no differences from the expected rate in the frequency of dental differences among unaffected family members versus controls (Anderson and Moss, 1996; Sanders, 2000). However, these studies had small sample sizes. The largest study to date included 1,922 unaffected relatives of 660 individuals with NSCL/P and reported a trend in higher rates of dental anomalies, but no statistically significant differences after correction for multiple tests (Howe et al., 2015). Thus, taken together with previous investigations, Howe et al. (2015) concluded that families with clefts do not likely have a genetic predisposition to atypical dentition.

2.8.2 Olfactory Deficits

Olfactory (sense of smell) deficits, a well-known symptom of Kallman syndrome (which often includes OFCs), were first identified in individuals with isolated CL/P in the 1980s (May,

Sanchez, Deleyiannis, Marazita, & Weinberg, 2015). This observation in affected individuals had two probable explanations: surgical repair of the cleft or a separate primary impairment concurrent/coincident with the cleft. Later, impaired sense of smell was shown to also occur more frequently in the unaffected family members of individuals with CL/P than in control populations. May et al. used a validated test, the UPSIT (University of Pennsylvania Smell Identification Kit), to evaluate olfactory capacity in 60 parents of a child with NSCL/P. Forty-one percent of the unaffected parents studied showed some deficit (microsmia) versus 12.6% of reference controls ($P<0.001$) (May et al., 2015). The microsmia (reduced sense of smell) was mild in all but one parent, who had moderate microsmia. This finding in unaffected individuals supported olfactory deficit as a separate primary defect, rather than a result of surgical repair. May et al. (2015) concluded that impaired sense of smell could be a result of structural changes to the nasal passages or olfactory processing areas of the brain.

Impaired sense of smell may indeed be at least partly structural. Olfactory deficit has been linked to the face shape of individuals with a family history of NSCL/P. Roosenboom et al. demonstrated a significant association between olfactory deficits and midface retrusion, in which the midpoint of the face, surrounding the nose, between the brows and mouth, is deepened (Roosenboom et al., 2015). In 30 unaffected first-degree relatives of an individual with a cleft, 33.3% showed impaired sense of smell versus 9.1% of controls ($P=0.04$). Furthermore, the upper nasal area of the relatives with olfactory impairment was smaller, with an especially narrowed bridge of the nose.

Based on the findings of these studies, it is possible that either facial measures or measures of olfactory function could prove useful in modeling recurrence risk. Intercanthal distance (related to nasal bridge width) and midface retrusion are factors considered in the present study. Thus, if

these factors have a significant effect on risk, future studies could evaluate whether the addition of olfactory function adds to the effect size or if this is accounted for by the facial measures. At that point, the more easily obtained measurement could be focused on for clinical use.

2.8.3 Discontinuity of the Orbicularis Oris Muscle

The orbicularis oris muscle surrounds the lips and plays a role in normal lip movements such as those necessary for speech. In 1993, Martin et al. first proposed that the spectrum of cleft lip be expanded to include the mildest cleft type reported, a "subepithelial cleft," which presents with a mild disruption of the of the OOM that does not include the surface of the lip (Martin, Jones, & Benirschke, 1993). Martin et al. (1993) proposed this after studying fetal lip morphology, and demonstrating histological differences in the pattern of the lip muscle fibers among two of 30 otherwise normal fetuses. Normal lip morphology includes a smooth, continuous, horizontal pattern of OOM and connective tissue (Martin et al., 1993). These two fetuses had apparently normal lips, but upon dissection, interruptions to the normal smooth, horizontal pattern of the tissues were noted. Drawing from earlier studies of minimal clefts, the mechanism was suspected to be an issue with maxillary process development.

A second study by Martin et al. (2000) used ultrasound to evaluate the first-degree relatives of 61 individuals with NSCL/P for OOM discontinuity; they compared the frequency observed to that observed in 52 controls, selected for absence of family history or any other minimal cleft characteristics (Martin et al., 2000). There was a significant increase in the frequency of OOM discontinuity among first-degree relatives versus controls (40% versus 13%, respectively, $P<0.002$). The researchers concluded OOM discontinuity may provide better power in genotyping

studies of NSCL/P, by including those at the mildest end of the cleft spectrum. This was a preliminary study with a small sample size, but it provided a basis for larger studies.

Since then, additional studies have indeed supported OOM discontinuity as a subclinical (i.e. not readily visible) form of a cleft lip, in which there was potential for a full cleft (M. Marazita, 2007). OOM discontinuity is more prevalent in families with than without a family history of NSCL/P and may therefore be indicative of cleft susceptibility (Klotz et al., 2010; Neiswanger et al., 2007). In a study of OOM discontinuities in case versus control families, case families had significantly more individuals with OOM discontinuities; 54 of 525 (10.3%) non-cleft relatives versus 15 of 257 (5.8%) controls had OOM discontinuities (Neiswanger et al., 2007). Male relatives had significantly higher rates of OOM defects than male controls; in females, the rate was still higher, but not significantly so. Neiswanger et al. (2007) concluded, OOM discontinuity is the mildest manifestation of the lip portion of CL/P. They also suggested including OOM discontinuity in the phenotypic spectrum in CL/P may provide more accurate recurrence risk estimation for families as well as improve the power of genetic studies by including individuals who may carry risk alleles, despite not having an overt cleft. The feasibility of OOM discontinuity screening makes it an attractive option for use in risk assessment, as many offices have access to ultrasound, which is used to visualize OOM discontinuity. However, the system of rating used in this study was based on the evaluation of images by several raters, which would be a limitation in clinical practice; a more straightforward approach to evaluating the images would need to be developed (Neiswanger et al., 2007).

Klotz et al. (2010) strengthened the evidence for OOM discontinuity as a part of the cleft phenotype spectrum. Assessing the relationship between OOM discontinuity and the odds of NSCL/P occurrence, they found that the chance for NSCL/P to occur in at least one first-degree

relative of a proband with an OOM discontinuity was 7.3% for any first-degree relative, and 3.3% for siblings. The recurrence risks observed in the study are not significantly different from those reported in the literature for NSCL/P, supporting the notion that OOM discontinuity signifies risk for recurrence. This was also the first study to evaluate recurrence of OOM discontinuity among first degree relatives: 16.4% for all first-degree relatives and 17.2% for siblings (Klotz et al., 2010). The study conclusion was that OOM discontinuity may be useful in modifying recurrence risk.

Leslie et al. (2017) performed analysis of OOM discontinuity in 63 mono- and dizygotic twin pairs, in which only one co-twin was affected by NSCL/P. The aim of the study was to evaluate whether the discordance rate could be reduced by confirming the presence of OOM discontinuity in apparently unaffected co-twins. A trend towards higher rates of OOM discontinuity was observed in unaffected co-twins versus controls (12.5% versus 6-7%), but this was not significant, possibly due to small sample size (Leslie et al., 2017). Discordant twins may carry the same risk alleles, but with variable expression of the cleft phenotype, with one displaying an overt cleft and the other an attenuated phenotype expressed as a subclinical feature such as OOM discontinuity. Larger and more diverse studies are needed to further evaluate this hypothesis.

2.8.4 Velopharyngeal Dysfunction

VPD refers to a dysfunction in which the tissues of the soft palate (velum and lateral pharyngeal walls) at the back of the throat do not completely separate the nasal passages (nasopharynx) from the mouth (oropharynx) during speech and swallowing (See Figure 3) (Chernus et al., 2018). Anatomical defects of the velum in NSCL/P can lead to VPD, which can result in a nasal quality to the sound of the speech (hypernasal), nasal air emission, and nasal regurgitation of liquids when drinking. VPD can occur in the setting of CL/P and CPO (repaired

or unrepaired) and has some genetic etiology in common with CL/P (Sweeney, Lanier, Purnell, & Gosain, 2015). VPD has been proposed as a subclinical phenotype of NSCL/P, but little systematic evaluation has taken place.

Two recent studies reached conflicting conclusions. Chernus et al. (2018) performed a genome wide association study (GWAS) of velopharyngeal dysfunction, finding five significant and 15 additional suggestive loci, a few of which overlap with regions implicated in overt clefts (Chernus et al., 2018). This led them to conclude VPD is supported as a hypothesized subclinical phenotype of NSCL/P. Of note, participants with VPD, but also velopharyngeal incompetency, and velopharyngeal mislearning, were included in analysis; these dysfunctions do not all involve anatomical defects of the velum/palate. Studying VPD in 189 unaffected first degree relatives of an individual with NSCL/P (from Pittsburgh OFC Study) and 207 matched Australian controls, Boyce and colleagues found an increase in VPD between relatives of an individual with a cleft versus controls (3% in unaffected relatives versus 0.5% in controls), but this increase was not statistically significant (P= 0.172) (Boyce et al., 2019). Boyce et al. conclude that VPD is not supported as a subclinical phenotype in NSCL/P; however, the Boyce finding may have been affected by lack of power. To date, there are not any additional cohorts available for analysis of VPD as a marker of risk in non-syndromic OFCs.

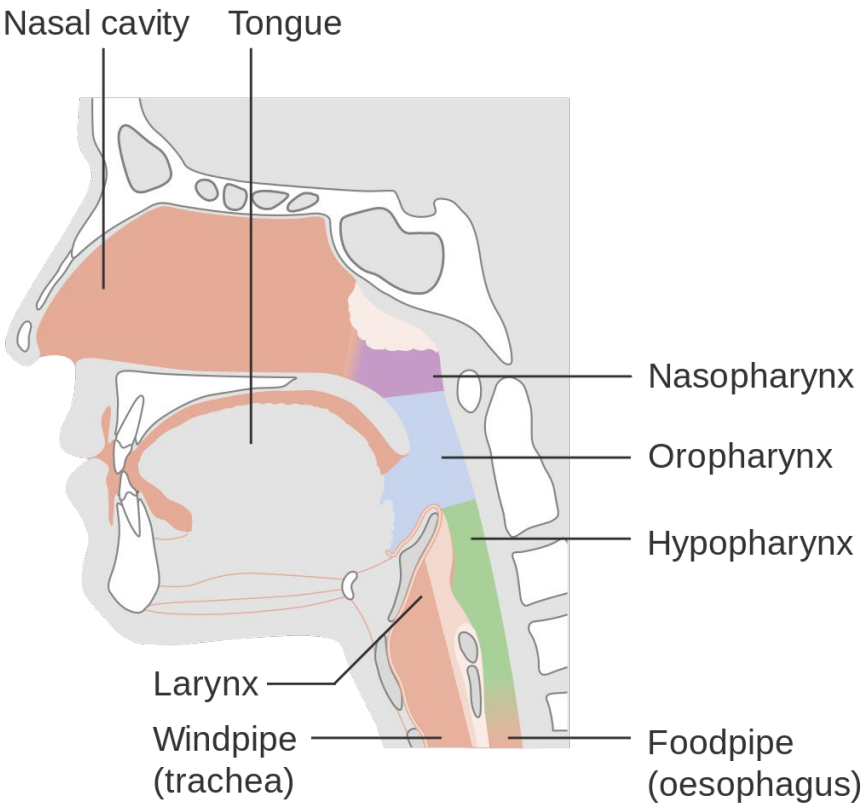

Figure 3 Cross-Sectional View of Oral and Nasal Cavities
(Cancer Research UK, Wikimedia Commons)

2.8.5 Craniofacial Morphology

There is an extensive body of literature concerning the craniofacial morphology of individuals with OFCs and their unaffected family members, spanning a history greater than 50 years. This investigation in humans stemmed from multiple animal model studies, revealing differences in the spatial alignment of facial prominences during embryonic facial development in mouse strains susceptible to orofacial clefting versus mouse strains not susceptible to clefting (Weinberg et al., 2009). There is general consensus within the literature, that unaffected parents of children with a CL/P have craniofacial measurements and an overall face shape that varies from the those in the general population (Weinberg, Neiswanger, et al., 2006). The specifics of exactly which measurements, however, are not unanimous across studies. The uniting idea behind these studies, is that face shape can serve as a proxy for measuring the underlying genetic tendency towards OFC development.

Phenotype studies of NSCL/P rely on the morphometric, or face-shape, hypothesis, which theorizes that certain craniofacial shapes confer risk for OFCs in offspring (Weinberg, Maher, & Marazita, 2006). A caveat of this assumption is that an observed morphometric trait may or may not be a direct or reliable representation of an underlying causal morphogenic factor. Effects of multiple developmental genes and modifier genes, as well as environmental and epigenetic influences on gene expression may make interpreting craniofacial shape directly, problematic (Weinberg et al., 2006). However, it is also possible that phenotype encompasses the complexities of multifactorial interactions in a readily accessible way, compared to dissecting the effects of multiple genes, environmental factors, and epigenetic changes, individually. The latter possibility makes the using phenotype to estimate recurrence attractive.

Over the last several decades, studies using morphometrics (measurements of radiographs), anthropometrics (direct measures including the soft tissue) and, more recently, three-dimensional imaging, have provided information on several craniofacial traits which differ more often among individuals with a family history of clefting, than those without (Mossey, Arngrimsson, Mccoll, Vintiner, & Connor, 1998; Jasmien Roosenboom et al., 2017; Weinberg et al., 2008; Weinberg et al., 2009). Three-dimensional surface imaging has been shown superior to caliper measurements, as it can consider the face along multiple planes, making it more sensitive to the subtle variations of facial morphology than measurements across a single plane (Weinberg, Naidoo, et al., 2006). However, caliper measures of simple linear distances are comparable to those derived from three-dimensional surface scans (Weinberg, Naidoo, et al., 2006). Increased intercanthal distance, facial width, and lower facial height; and decreased upper facial height and midface convexity (i.e. midface retrusion), are supported across multiple studies (Wyszynski, 2002). Notably, Weinberg et al. (2006) point out that, while specific findings have varied, all known studies of unaffected parents of a child with a cleft have reported some differences in craniofacial morphology, compared to control parents. Inherent racial and ethnic facial variation across studies performed in different populations, in addition to substantial differences in study methods, likely explain some of the dissimilarity in findings.

Many reviews of the literature have pointed out the significant differences and discrepancies in craniofacial studies of NSCL/P, and a 2006 meta-analysis of nine case-control studies quantifies them (Weinberg et al., 2006). Key findings included observations of several common, subtly atypical traits. Eighty-three percent of significant facial measures had small (<0.49) or very small (<0.20) effect sizes; effects size also varied considerably across studies. While some findings appeared to be sex-specific between mothers and fathers (upper facial height

only in fathers and more profound mandibular protrusion in mothers), average parental measures consistently included the following traits, with small or very small effect size: wider faces, narrowing of the cranial vault, shorter upper faces, longer lower faces, and more mandible protrusion, compared to controls. Two measures in the study produced moderate effect sizes (>0.49): nasal cavity width and relative nasal cavity width (see Figure 4).

Mandible protrusion leads to a loss of convexity in the midface, which has been observed in subsequent studies; midface retrusion has repeatedly been seen as a significant feature in unaffected parents of a child with NSCL/P and in both sexes of unaffected twins from cleft discordant twin pairs (Roosenboom et al., 2017; Weinberg et al., 2009). Males at risk for cleft recurrence appear to have more significant facial atypia; more variability exists in females and fewer phenotypes provide strong effects or predictive value (Mossey et al., 1998; Roosenboom et al., 2017; Weinberg et al., 2006).

Weinberg et al. (2009) were the first to combine three-dimensional imaging with geometric morphometrics, or statistical tests specifically for evaluating biological shape using landmark coordinates. Uniquely, unaffected parent participants were also divided into subgroups depending on whether or not they had a prior family history of NSCL/P. Significant differences were observed in the unaffected parents with a family history and included: flattened facial profile (loss of convexity), shorter middle and upper face heights, longer lower face heights, and wider interorbital distances (more pronounced in males). Sex-specific findings included a narrowing of the area around the nose and mouth (nose, mouth, philtrum) in males versus minimal variation from normal in females. Substantially increased width of the outer part of the nostril (nasal ala) was unique to females.

Recently, a study of twin pairs in which only one co-twin was affected by NSCL/P further confirmed the fairly consistent pattern of facial variation detected in the previously discussed studies of unaffected relatives. Roosenboom et al. (2017) evaluated 45 Caucasian twins (11 monozygotic and 34 dizygotic) and compared them to 241 sex and age matched controls. The pattern of midface retrusion, increased facial width, and mandibular protrusion was observed for both the combined mono- and dizygotic twins and in separate analysis of the monozygotic twins only. Of additional note, when compared to CPO twin pairs, these characteristic findings were only found in the NSCL/P unaffected co-twins.

In summary, most recent studies agree on the following facial phenotypic variations associated with NSCL/P: midface retrusion or flattening of the facial profile, an increased interorbital distance, widened faces, and changes to the height of the face, with taller lower faces and shorter upper faces. This consensus, in addition to the investigations of OOM discontinuity and VPD as risk phenotypes, is used to guide the current effort to investigate the effects of several risk phenotypes on the odds of an unaffected individual being from a case family.

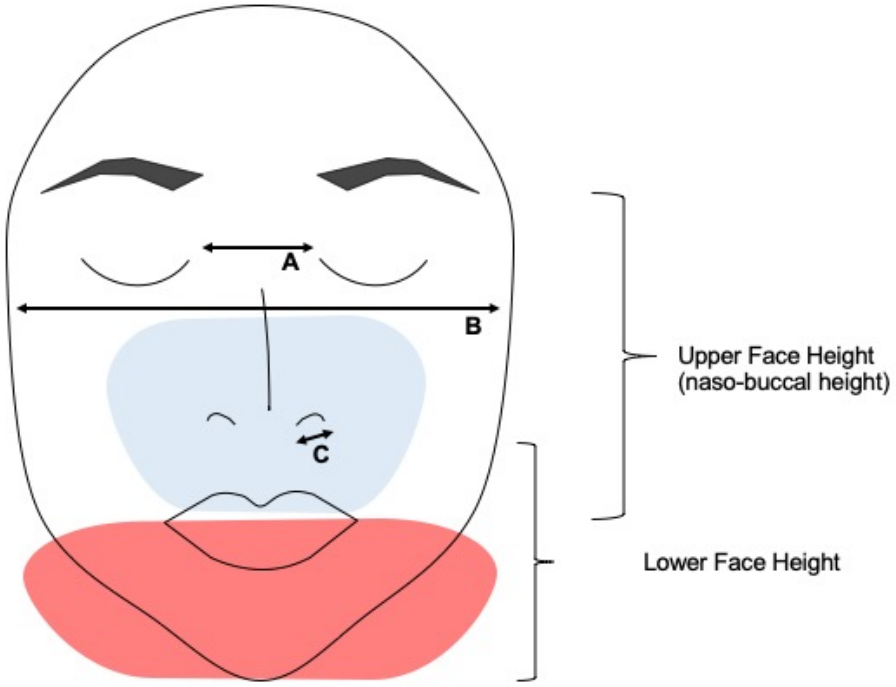

Figure 4 Facial Phenotypes of Interest in Previous Studies of NSCL/P Unaffected Relatives

A: Intercanthal Width; B: Facial Width; C: Nasal Cavity Width; Blue-shaded: Region of the face that is convex/deepened as a result of changes to the area in red. Red: Mandibular protrusion, in which the jaw juts out.

3.0 Manuscript

3.1 Background

Cleft lip with or without cleft palate (CL/P) is the most common congenital malformation of the face and the second most common birth defect overall, affecting 8.1 to 10.63 per 10,000 (roughly 1 in 1,000) live births in the United States (Mai et al., 2014; U.S. Department of Health and Human Services, 2014). Rates vary by race and ethnicity, with Asians and American Indians having the highest prevalence (~1/500), followed by European Caucasians (~1/1000), and Africans, with the lowest reported prevalence (~1/2500); worldwide, CL/P affects about 1 in 700 live births (Dixon et al., 2011b). CL/P can present as an isolated finding or can be part of a larger, multisystem genetic syndrome. Isolated CL/P can occasionally be attributed to a specific teratogenic exposure; however, the majority of CL/P is multifactorial in nature, having several genetic and environmental risk factors. About 70% of all CL/P cases are non-syndromic (NSCL/P). CL/P is etiologically distinct from cleft palate only (CPO) (Dixon et al., 2011b); it is also rarer and more likely to be part of a syndrome.

Individuals born with NSCL/P have complex medical management needs from birth to early adulthood (Ranganathan et al., 2016). Treating NSCL/P is intensive and long-term, and requires a multidisciplinary team including surgeons, speech language pathologists, orthodontists, social workers, and others (Cleft Palate-Craniofacial Association, 2009). Newborns struggle with feeding, are less likely to be breast fed, and are more likely to experience failure to thrive (Kaye et al., 2017). Multiple surgeries are required as a child grows, to restore the structure and function of disrupted tissues and improve the facial appearance; the average patient may undergo five to

six repairs for isolated cleft lip and nine to ten for CL/P or CPO (McIntyre et al., 2016). Primary surgery is typically performed at around three months of age for the lip repair and around seven months for the palate. Secondary revisions occur throughout later childhood, usually ending with rhinoplasty to address nasal asymmetry in later teenage years (Nahai et al., 2005). In addition to surgeries, speech therapy for speech impediments and delays, and dental/orthodontic interventions for disrupted tooth eruption and jaw positioning are commonly necessary (Nahai et al., 2005).

In addition to health problems directly related to the cleft, there are other physical and mental health issues associated with NSCL/P. Children with NSCL/P are at slightly increased risk for developmental delay and learning difficulties; historically speech delays and social difficulties in childhood were proposed explanations for this, but it is also possible that brain patterning differences result in a biological predisposition (Chollet et al., 2014). According to a large, population-based cohort study out of Denmark, the NSCL/P population may also be more likely to develop certain types of cancer, including female breast and male lung in individuals with CL/P, and primary brain in females with CPO (Bille et al., 2005). Regarding mental health, while individuals with orofacial clefts (OFCs) do not, on average, appear to experience profound mental health problems, there is substantial evidence to suggest some level of negative impact on quality of life (QOL) in areas such as socialization, satisfaction with physical appearance, and self-esteem (Mani et al., 2010).

Awareness of the potential reduction in QOL in the OFC population at all ages, highlights the need for a similar emphasis on mental health as physical health in the treatment plan. Studies on the determinants of QOL can help guide interventions. Early and consistent comprehensive care through a cleft craniofacial center can improve long-term outcomes in NSCL/P patient satisfaction and self-esteem (Naros et al., 2018).

Management of an OFC is essentially a life-long process. It requires significant coordination, which can be difficult for families, who may have to travel long distances or relocate to access a cleft craniofacial center. Timing of surgical interventions, quality of surgical repair, and long-term follow-up are critical to long-term outcomes (Nahai et al., 2005).

Costs associated with OFCs affect individuals and their families, as well as the broader healthcare system. Individuals with OFCs experience higher costs of care in the newborn period, increased hospital utilization compared to unaffected siblings, and need expensive surgeries and treatments from birth through early adulthood (Wehby & Cassell, 2010). There is not a comprehensive understanding of the lifetime costs associated with NSCL/P, including surgical costs; costs of additional therapies (speech, occupational), mental health services, health-related QOL; and loss of productivity for individuals and caregivers.

Two robust studies, one in the privately insured population and one in the Medicaid population, give estimates based on insurance reimbursement and out-of-pocket payments, and expenditures, respectively (Boulet et al., 2009; Cassell et al., 2008). In the privately insured group, children with a non-syndromic OFC had healthcare costs eight times higher than children without an OFC in the period measured from the years 2000 to 2004 (Boulet et al., 2009). A North Carolina Medicaid-insured group of children with isolated OFCs monitored between 1995 and 2002 had expenditures five to six times higher than unaffected children, with CL/P and CPO associated with higher healthcare costs than CL (Cassell et al., 2008). The mean first-year expenditures for a child with an isolated OFC amounted to $10,099, compared to $3,900 for a control. The significant impacts on health, QOL, and finances, combined with the relatively common occurrence of NSCL/P, make it a significant public health issue.

Genetic counseling for recurrence risk in NSCL/P is based on empiric estimates of risk, derived from large cohort studies of mostly Caucasian individuals (Basha et al., 2018; Lees, 2008) (See Table 2). Empirical estimates are likely not accurate in every family, as not all families with isolated cases of NSCL/P experience the same recurrence rates; some cases may be truly sporadic or mostly caused by environmental factors, while others have increased risk due to an underlying genetic predisposition (Crawford & Sofaer, 1987). The ability to refine estimates and fit them to these situations appropriately could have a significant impact on genetic counseling and the provision of public health.

Better risk assessment for NSCL/P would be indispensable in the pre-conception genetic counseling setting, where parents could benefit from informed reproductive decision-making, education and resource identification, and practical/psychological preparation. Additionally, though the effects of known modifiable risk factors may be small, parents could benefit from taking steps to reduce the risk to the extent possible (e.g. via vitamin supplementation and smoking cessation). Genetic counselors focus on not only physical health outcomes, but also on psychological outcomes. Providing some control in a situation in which there is little can provide patients with some self-efficacy. This empowerment, regardless of the physical health benefits, can have positive effects on mental health, such as coping and adapting to life with a chronic condition (McAllister et al., 2012).

In terms of public health, modeling could also help in NSCL/P research. Evaluating the effect of various factors on risk can help to clarify the degree of genetic, teratogenic, and environmental influences on the development of OFCs. Identifying environmental factors with strong effects on risk can also prompt studies to identify the genes which interact with that particular factor during development; this has already been seen with studies implicating genetic

variants in *MTHFR* and *RBP4* in NSCL/P risk, because they are involved in the biological processing of folate and vitamin A, respectively (Zhang et al., 2018). If robust modeling reveals new risk factors, especially modifiable ones, it could inform primary prevention efforts.

Several decades of research on the specific genetic and environmental factors associated with NSCL/P have revealed several which may affect the chances of recurrence for unaffected family members of an individual with NSCL/P. Substantial literature supports the observation of subtle phenotype differences, or "subclinical phenotypes" of OFCs, among close relatives of an individual with NSCL/P compared to individuals from unaffected families (Crawford and Sofaer, 1987; Martin et al., 1993; Mossey et al., 1998; Neiswanger et al., 2007; Roosenboom et al., 2017; Weinberg et al., 2006; Weinberg et al., 2009; Weinberg, Neiswanger, et al., 2006). If the proposed threshold model of inheritance holds in NSCL/P (Beaty et al., 2016; Grosen, Chevrier, et al., 2010; Howe et al., 2018), genetic susceptibility is based on inheriting several genetic variants, each conferring variable levels of risk. Therefore, it is reasonably likely that recurrence risk varies among individuals with a family history of NSCL/P, even within the same family. If phenotypic markers prove useful for quantifying this variable genetic risk, the empirical recurrence risk of a given relative could be modified and refined through mathematical modeling, based on the effect of presence or absence of certain physical traits on risk. The underlying assumption of this concept is that subclinical phenotypes can serve as a proxy for the number of risk alleles carried by a given family member.

Most studies agree on the following facial phenotypic variations associated with NSCL/P: midface retrusion or flattening of the facial profile, an increased interorbital distance, widened faces, and changes to the height of the face, with taller lower faces and shorter upper faces (Weinberg et al., 2006). This relative consensus was used to guide the current phenotype

investigation and subsequent model development. There is also substantial support for the inclusion of defects/discontinuity of the orbicularis oris muscle (OOM) of the upper lip as a subclinical manifestation of NSCL/P, based on increased occurrence in the unaffected relatives of individuals with NSCL/P (Klotz et al., 2010; Martin et al., 2000; Neiswanger et al., 2007). Anatomical defects of the velum leading to VPD can occur in the setting of CL/P and CPO (repaired or unrepaired); due to some shared genetic etiology with CL/P (Sweeney et al., 2015), VPD has been proposed as a subclinical phenotype of NSCL/P (Chernus et al., 2018; Sweeney et al., 2015). Little systematic evaluation of VPD as a marker for recurrence has taken place, however.

There may be potential for developing a prediction model for NSCL/P, using established phenotypic and environmental risk factors. The present study was an attempt to develop prediction models for NSCL/P based on the seven subclinical phenotypes (listed above) previously proposed as risk factors for NSCL/P recurrence. This study also aims to evaluate the models' ability to correctly classify the unaffected relatives in the data set as relatives of an individual with NSCL/P versus control.

3.2 Methods

Data used in the present study was collected as part of The University of Pittsburgh Oral-Facial Cleft (POFC) study. Data collection for the POFC study began in 1993, with an original focus on recruiting families with multiple individuals affected by NSCL/P (multiplex families) for genotyping studies, in order to discover genes involved in NSCL/P across diverse populations (Weinberg, Neiswanger, et al., 2006). The project expanded in 1998, to collect information on

NSCL/P-associated phenotypes, with a secondary goal of expanding and better defining the phenotype. At that time, families with only one known individual with NSCL/P (simplex families) were recruited along with multiplex families. Due to the original focus on multiplex families, there is a bias in this data towards those families.

Physical assessments and questionnaire information was collected from participants at various centers around the world, including China, Denmark, Nigeria, the Philippines, Puerto Rico, Hungary, Guatemala, and several in the United States. Study protocols were approved by the University of Pittsburgh Institutional Review Board (IRB) (See IRB approval letters in Appendix D), the coordinating center, as well as the appropriate local IRBs in other U.S. and international sites. Written informed consent was obtained from all study participants; questionnaires and forms were translated to the appropriate languages for each site, when necessary. This work was funded by multiple grants from the National Institutes of Health, primarily the National Institute for Dental and Craniofacial Research (NIDCR).

3.2.1 Participants

Case participants were recruited in a variety of ways depending on site, including through registries, cleft craniofacial centers, and advertisement/word of mouth. Case participants are screened for additional birth defects and features, to identify possible syndromic cleft families to the greatest extent possible. Control participants are recruited from the general population of each site, mainly from advertisement/word of mouth. Control participants are screened for family history of CL/P in their immediate family (and in family extended beyond the third degree of relation, if known), major birth defects, or genetic syndromes. It is important for the present study, to note that cleft status is the only selection criteria in recruitment of families; additional subclinical

phenotypes of interest (e.g. OOM and VPI) were not used to ascertain families. Control participants have no known family history of craniofacial anomalies; family history of craniofacial anomalies is the only exclusion criteria for controls.

The total data set included 11,658 individuals from 5,159 families. The focus of the present study was a subset of unaffected relatives from case families. The following criteria were used to select individuals of interest. First, affected individuals were excluded. Next, families with cleft palate only or unknown cleft types were excluded, leaving only families with CL/P, due to the genetically distinct etiologies proposed between CPO and CL/P. Case families were also excluded based on the presence of additional congenital anomalies or features in affected individuals, which suggested a syndromic form of CL/P when reviewed by researchers at the coordinating center. Next, unaffected relatives from case families were excluded if they had non-cleft syndromes (e.g. Downs syndrome or Noonan syndrome), due to the atypical facial morphology caused by the syndrome. Control families were excluded, if, upon review by researchers at the coordinating center, they were found to have family history of clefting, or features suspicious of a cleft syndrome.

The final data set included 8,487 unaffected individuals from 4,685 families (see Figure 5, below). The age range of participants was 0 to 89 years. The large majority of participants were first-degree relatives (93%; 67% parents, 33% siblings), followed by second-degree relatives (6%), and third-degree relatives (1%). Sex was approximately even between males and females. See Figure 7 in Results section, below, for a summary of participants by race.

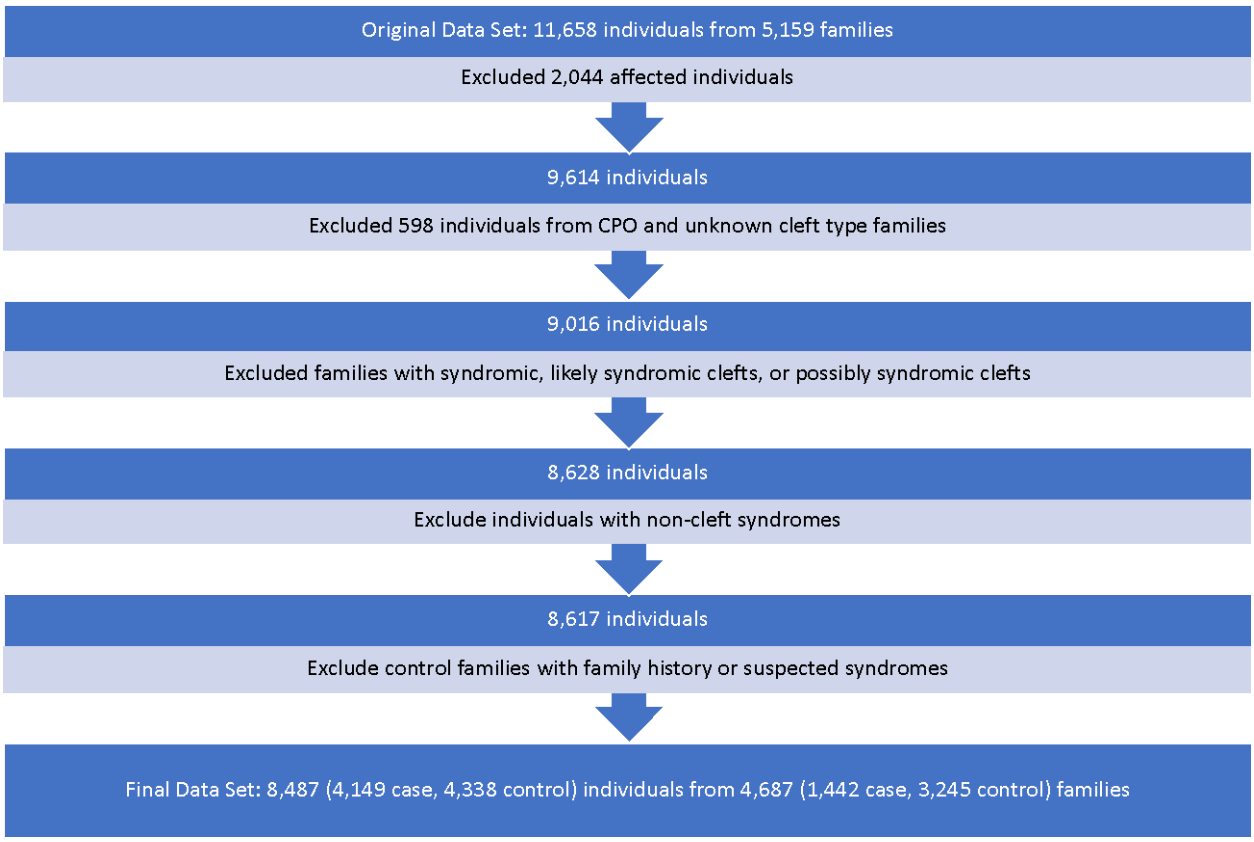

Figure 5 Inclusion and Exclusion Criteria Applied

3.2.2 Phenotype Evaluation

OOM defects were evaluated for via high resolution ultrasound scans of the upper lip by a trained assessor as fully described previously (Neiswanger et al., 2007). Scores were given for the presence or absence of OOM fiber disorganization (0= No; 1= Yes). If raters could not come to a consensus on the rating, or if a video was poor quality, the participant was not given a final score (-4= Unable to Rate).

The presence or absence of VPD was evaluated by an experienced speech pathologist, either in person, or by using a video-recorded passage read by the participants as detailed previously (Chernus et al., 2018). The Pittsburgh Weighted Speech Scale was used to evaluate

speech for audible nasal emission and nasal turbulence, nasality, phonation and articulation patterns indicative of VPD. Past medical history, including speech pathology or surgery was also collected on each patient.

Soft-tissue facial measurements included intercanthal distance, maximum facial width, upper and lower facial height, and midface depth. These standard facial measurements were obtained through either direct caliper measurement (anthropometry) or computer-derived linear distances, derived from 3D facial surface images. Direct anthropometry was necessary to capture maximum facial width because this measurement cannot be obtained reliably from 3D surface images. Detailed description of the imaging systems, full capture protocol, and landmarking procedures were previously described (Weinberg et al., 2008). In the present study, caliper and computer-derived measurements were included together. Figure 6 shows the facial measurements used in the present study. Of note, right and left midface depth measurements were taken; the average of the two was used in the models.

In addition to the specific phenotypes collected, demographic information including sex, race, height, weight, and age were available for most participants. This information was important in making corrections for facial measurements based on participants' overall body size. Demographic information was also used to evaluate and control for inherent differences in facial morphology between males and females and individuals of different racial groups.

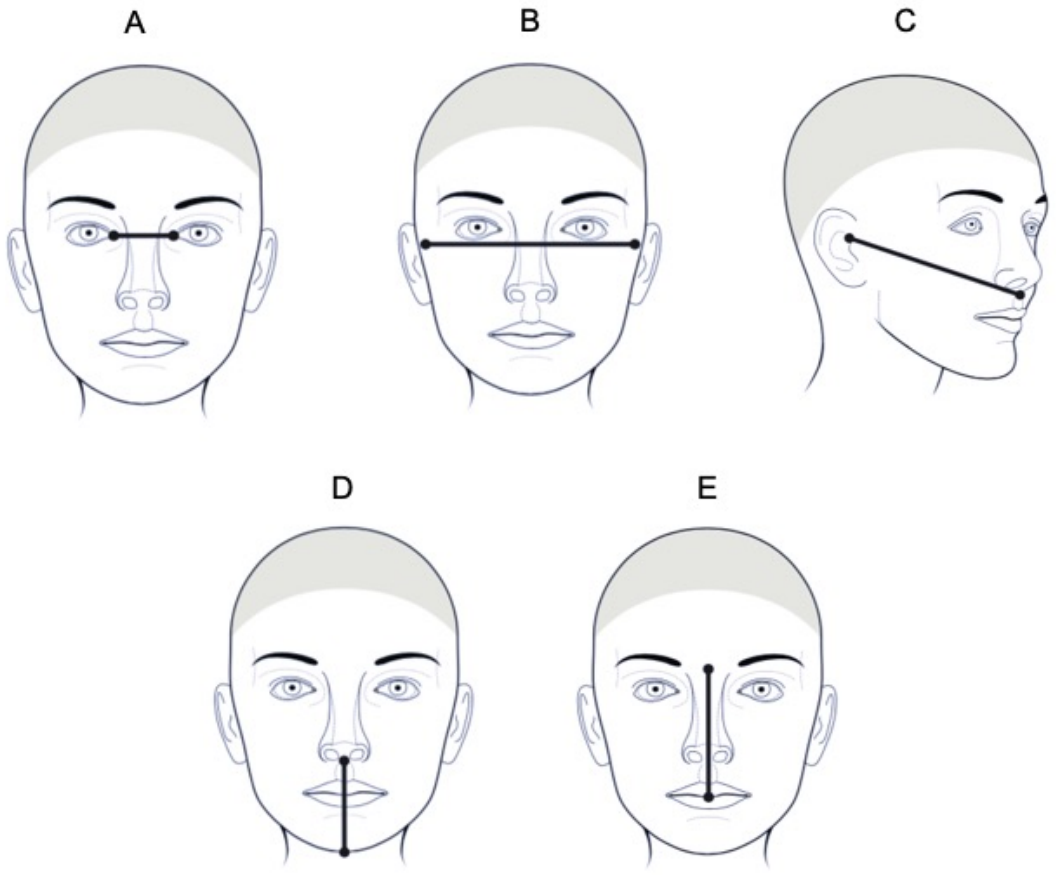

Figure 6 Facial Phenotypes Used in the Present Study

A: Intercanthal Width; B: Maximum Facial Width; C=:Midface Depth; D: Lower Face Height; E: Upper Face Height. Images from the 3D Facial Norms study, reproduced here with permission (Weinberg et al., 2016).

3.2.3 Statistical Methods

The present study's overarching aim was to evaluate the effect of several factors previously associated with NSCL/P, on the likelihood of being from a case family versus a control family. Logistic regression is useful when the research goal is to evaluate possible predictors (independent variables) of a *binary* outcome (dichotomous dependent variable). It is a versatile technique that allows the user to simultaneously control for known confounders, evaluate associations between independent variable(s) and an outcome of interest, and assess predictive ability (Stoltzfus, 2011). Therefore, the aims of the present study were accomplished through constructing predictive statistical models, using multivariate logistic regression. Logistic regression results in an output that falls between 0 and 1, that is, a probability that can be interpreted in reference to the binary outcome (in the present study: odds of being a relative of an individual with an OFC versus a control family member). Only those traits previously found at higher rates in the unaffected family members of individuals with NSCL/P versus control participants were considered for the present study.

In the first step of the present analysis, logistic regression was performed individually for OOM discontinuity, VPD, maximum face width, upper face height, lower face height, midface depth, and intercanthal width. Regressions for each individual phenotype were corrected for age, sex, and race (as an indicator variable including Caucasian, Asian, African, and Mixed/Other). In addition, regressions for each of the five facial measurements were also corrected for body size confounding (height and weight). In univariate logistic regression for a complex trait, there can be significant unexplained variance within the model due to missing important covariates, thus it is common to use a cutoff of p=0.25 for including a covariate in the multivariate model (Hosmer &

Lemeshow, 2000; Stoltzfus, 2011). Backwards stepwise regression was also performed, using a p-value cut off of 0.25. All covariates were ultimately included in the multivariate analysis, despite some p-values >0.25, due to previous studies implicating them as risk factors. Additionally, because no previous studies have analyzed this precise combination of phenotypes, the possibility of interaction between them in a multivariate model could not be anticipated prior to analysis.

In the second step of analysis, three multivariate models were constructed. Model 1 included OOM and VPD, plus the potential confounding factors of age, sex, and race. Model 2 included the five facial measurements, plus potential confounding factors of age, sex, race, weight, and height. Model 3 included OOM discontinuity, VPD, and the five facial measures, plus the potential confounding factors from Model 2.

Univariate analyses from the first step and the models from the second step were assessed for fit via The Hosmer-Lemeshow goodness of fit (GOF) test, Wald chi-squared test, and the likelihood ratio each evaluate model fit (calibration). Each of these tests measure GOF in slightly different ways, but the results from the three tests should be concordant (see Table 6 for a summary) (Peng, Lee, & Ingersoll, 2002).

A model will always perform somewhat well on the data set with which it was developed, therefore, the ideal method of predictive model-building is to build the model with one data set, then apply it to a set of data (test set) which was not used in the building of the model. In the present study, this was not feasible due to sample size constrictions. The next best validation of model prediction is internal cross-validation. Internal cross-validation simulates a test data set by holding out a subset of the data, performing logistic regression on the rest of the data, then applying it to the withheld test subset. Five-fold, internal cross-validation was performed in the present study. After cross-validation, classification (using a cut off of 0.5), area under the receiver

operating curve (AUC) achieved, sensitivity, specificity, positive (PPV), and negative predictive (NPV) values were assessed. Brier score decomposition was also performed for each model to evaluate forecast performance and over-confidence.

Classification involves comparison of predicted and observed outcomes; from the resulting confusion matrix (2x2 table of predicted and observed outcomes), the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) can be calculated. Classification has a default cut off of 0.5, meaning if the predicted probability is greater than 0.5, the subject is classified as positive. There are ways to maximize sensitivity and specificity by choosing a cut off based on the point at which the lines cross on a plot of sensitivity versus specificity; this alters classification performance, as there is a trade-off between sensitivity and specificity, and the true and false positive rates achieved by the model (Trevethan, 2017).

Receiver Operating Characteristic (ROC) curves are plots of sensitivity (i.e. true positive) versus "1 – specificity" (i.e. the. false positive rate) across an entire range of cut points (not just one, as in classification). The area under the ROC curve (AUC) shows a model's ability to discriminate between two outcomes. A value of less than 0.7 indicates poor discrimination; values between 0.7 and 0.8, acceptable discrimination; values between 0.8 and 0.9, excellent discrimination; and values above 0.9, outstanding discrimination (Hosmer & Lemeshow, 2000).

A Brier score is a decomposition metric that accounts for model forecast performance and over or under-confidence, in the context of the level of uncertainty within the data set (Brier score = (Reliability – Resolution) + Uncertainty) (StataCorp, 2017a). Uncertainty affects classification; for example, if there is a higher proportion of positive outcomes than negative outcomes, there are better base chances of predicting the positive outcome. Higher levels of uncertainty occur when outcomes are close to 50/50 in a data set. High uncertainty leads to higher Brier scores, because

there is less confidence in correctly predicting the negative outcome. The closer to zero a Brier score is, the better predictive ability the model has. Stata was used to perform all statistical analyses (StataCorp, 2017b).

3.3 Results

3.3.1 Univariate Models

In the first step of the analysis, an exploratory univariate logistic regression was performed separately for each of the seven phenotypes in the study: OOM defects, VPD, midface depth, upper face height, lower face height, maximum face width, and intercanthal width. Table 5 summarizes the number of participants from case and control family status for each phenotype (univariate analyses) and for each of the three models (multivariate analyses). VPD and increased intercanthal width were significantly associated with increased odds of being from a case family (at a significance level of p=0.05). Maximum face width, upper, and lower face height had p-values <0.25. OOM discontinuity and midface depth had p-values greater than 0.25 (see Table 4).

3.3.2 Multivariate Models

Table 3 contains a summary of unaffected participants from Model 1 by case versus control family status and positive versus negative status for OOM discontinuity and VPD. Results are summarized in Table 6 and validation statistics are summarized in Table 7. Hosmer-Lemeshow GOF tests were not significant for any of the three models, indicating good calibration and fit.

Models 1 and 2 achieved AUC values of 0.6344 and 0.6406 on the cross-validated data-set, indicating poor ability to discriminate between participants from case versus control families. AUC values of 0.5 indicate chance prediction and AUC values below 0.7 indicate poor model discrimination. Model 3 achieved an AUC value of 0.7192 on the cross-validated data set. AUC values of 0.7 to 0.8 indicate acceptable discrimination; therefore, Model 3 falls within the lower range of acceptable.

Intercanthal width was the most consistently associated phenotype with case family status; association was statistically significant in univariate analysis and in multivariate Models 2 and 3. Odds ratios (ORs) for intercanthal width ranged from 1.12 to 1.17 across the models, suggesting that for a one-millimeter increase in intercanthal width, a subject was 12 to 17% more likely to be from a case family. VPD was significantly associated with an increased likelihood of case family status in the univariate analysis and in Model 1, with an OR of about 1.8. However, when combined with additional phenotypes in Model 3, it became insignificant. No other phenotypes were consistently associated with increased likelihood of case family status.

OOM discontinuity was associated with a significant decrease in the likelihood of being from a case family (OR=0.65), however, this association became insignificant in Model 3. Maximum facial width was significantly associated with case family status in multivariate Model 2 (OR=1.03), suggesting a small increase in risk for every millimeter of increased facial width. Upper face height was significantly associated with case family status in Model 3; in the univariate analysis, upper face height was nearly significant at an alpha level of 0.05 (p=0.054). Looking at only adults (age>18), or only first-degree relatives, did not result in any changes to the models.

3.3.3 Differences in Phenotype by Sex

Previous studies report some differences in facial phenotype between mothers and fathers of an individual with NSCL/P (Weinberg et al., 2006). Separating Model 2 by mothers versus fathers did not allow for enough participants per covariate to provide meaningful results (model approached saturation/was overfit). Separating by mothers and fathers in Model 1 was also not feasible due to the small number of individuals with positive OOM and VPD status. However, investigating this phenomenon by sex only was possible in Models 1 and 2; separating by sex in Model 3 led to overfitting. After separating Model 1 by sex, only VPD was associated with case family status, and the likelihood was higher when males had VPD (Males: OR=2.10; 95% CI 1.14, 3.89; p=0.018; Females: OR=1.66; 95% CI 1.06, 2.62; p=0.027). In males, a negative association with case family status for OOM discontinuity was nearly significant at an alpha level of 0.05 (OR=0.6; 95% CI 0.35, 1.03; p=0.062).

Separating by sex in Model 2 also revealed phenotypic differences between males and females. In the original Model 2, wider maximum face and intercanthal widths were associated with increased likelihood of case family status. After separating by sex, the association of intercanthal width remained only in males, with a slightly larger OR (OR=1.2; 95% CI 1.11, 1.32; p=<0.001). Maximum face width, conversely, remained associated only in females (OR=1.04; 95% CI 1.02, 1.06; p=0.001). In males, larger midface depth measurements (i.e. reduced midface retrusion) were associated with a slightly decreased likelihood of case family status (OR=0.94; 95% CI 0.89, 0.99; p=0.03). In females, increased lower face height was associated with increased likelihood of case family status (OR=1.06; 95% CI 1.01, 1.11; p=0.02).

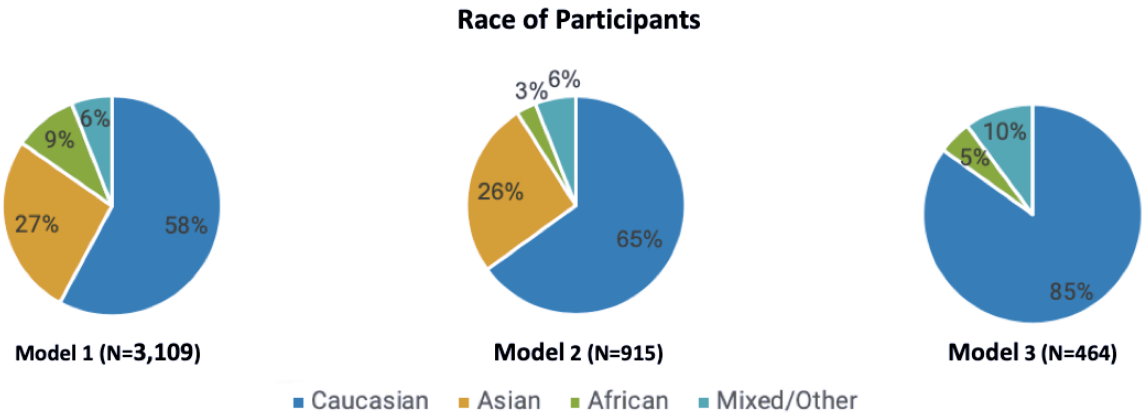

Figure 7 Summary of Participants by Race for Models 1, 2, and 3

Table 3 Distribution of Unaffected Individuals Included in Model 1 by Case Versus Control Family Status and Positive Versus Negative Status for the Presence of OOM Discontinuity and VPD

Total N=3,109	**OOM Discontinuity**		**VPD**	
Individual Status	+	–	+	–
Case Family	50 (3.4%)	1438 (96.6%)	73 (4.9%)	1415 (95.1%)
Control Family	89 (5.5%)	1532 (94.5%)	58 (3.6%)	1563 (96.4%)
Total	4.7%	95.3%	4.4%	95.6%

Table 4 Summary Statistics for Additional Model Variables

Phenotype	**Mean**	**Standard Deviation**
Age (years)	28.9	16.4
Maximum Face Width (mm)	127.2	14.1
Intercanthal Width (mm)	32.3	3.5
Midface Depth (mm)	122.5	9.2
Upper Face Height (mm)	73.0	7.1
Lower Face Height (mm)	68.0	7.0

Table 5 Number of Unaffected Individuals Included in Univariate and Multivariate Logistic Regression Anlayses, by Case versus Control Family Status

Phenotype	Case Family Status	Control Family Status
OOM	3,410	3,222
VPD	1,792	2,310
Maximum Face Width	1,443	970
Intercanthal Width	892	311
Midface Depth	737	275
Upper Face Height	900	310
Lower Face Height	855	299
Model 1	1,617	1,492
Model 2	651	264
Model 3	310	154

Table 6 Logistic Regression Modeling

Univariate Analysis						
Predictors	*N*	*OR*	*SE*	*95% CI*		*p-value*
VPD	4,102	1.8	0.3	1.29	2.51	0.001
OOM	6,632	0.89	0.1	0.71	1.1	0.324
Maximum Face Width	2,413	0.99	0.005	0.984	1.004	0.217
Intercanthal Width	1,203	1.13	0.026	1.08	1.18	<0.001
Midface Depth	1,012	1.01	0.016	0.985	1.048	0.322
Upper Face Height	1,210	1.03	0.016	0.999	1.063	0.054
Lower Face Height	1,154	1.03	0.014	1.006	1.063	0.016
Multivariate Analysis						
Predictors	*N*	*OR*	*SE*	*95% CI*		*p-value*
Model 1: OOM + VPD	3,109					
Constant		0.71	0.07	0.59	0.86	0.86
Age		1.003	0.002	0.999	1.008	0.10
Sex		0.91	0.07	0.79	1.06	0.22
Race (Caucasian Reference)	1,800 (58%)					
Asian	832 (27%)	2.38	0.21	2.00	2.83	<0.001
African	291 (9%)	0.34	0.05	0.27	0.48	<0.001
Mixed/Other/Unknown	186 (6%)	1.72	0.27	1.27	2.33	0.001
OOM		0.65	0.123	0.45	0.943	0.023
VPD		1.78	0.329	1.24	2.558	0.002
Model 2: Facial Measurements	915					
Constant		0.27	0.47	0.008	8.47	0.454
Age		0.99	0.006	0.98	1.006	0.372
Sex		1.06	0.19	0.75	1.5	0.744
Race (Caucasian Reference)	595 (65%)					
Asian	238 (26%)	0.71	0.16	0.46	1.12	0.14
African	27 (3%)	0.98	0.49	0.37	2.58	0.97
Mixed/Other/Unknown	55 (6%)	1.14	0.38	0.59	2.21	0.69
Height		0.97	0.008	0.96	0.99	0.002
Weight		0.99	0.005	0.99	1.01	0.943
Maximum Face Width		1.03	0.009	1.01	1.05	<0.001
Intercanthal Width		1.12	0.03	1.05	1.18	<0.001
Midface Depth		0.97	0.018	0.94	1.008	0.131
Upper Face Height		1.01	0.02	0.97	1.05	0.618
Lower Face Height		1.02	0.017	0.98	1.05	0.289

Table 6 Continued

Predictors	*N*	*OR*	*SE*	*95% CI*		*p-value*
Model 3: Model 1 + Model 2	464					
Constant		1.17	3.05	0.007	194.2	0.953
Age		0.97	0.01	0.96	0.99	0.005
Sex		1.00	0.26	0.60	1.66	0.994
Race (Caucasian Reference)	394 (85%)					
Asian	N/A					
African	23 (5%)	1.34	0.74	0.45	3.95	0.600
Mixed/Other/Unknown	47 (10%)	2.39	1.14	0.94	6.10	0.067
Height		0.96	0.01	0.93	0.98	0.001
Weight		1.01	0.01	0.999	1.03	0.071
OOM		1.02	0.77	0.23	4.49	0.983
VPD		0.71	0.31	0.30	1.68	0.433
Maximum Face Width		1.01	0.02	0.97	1.06	0.514
Intercanthal Width		1.17	0.04	1.09	1.26	<0.001
Midface Depth		0.96	0.03	0.90	1.01	0.109
Upper Face Height		1.06	0.03	1.00	1.13	0.033
Lower Face Height		1.02	0.02	0.97	1.07	0.474

Table 7 Model Validation Summary

Model Assessment						
	Wald χ^2	*Hosmer-Lemeshow GOF (8 df)*	*Likelihood Ratio Test*	*Area Under ROC*	*Classification*	*Brier Score*
Model 1: OOM + VPD	14.34 (p<0.001)	5.92 (p=0.6560)	14.61 (2df; p=<0.001)	0.6405, 0.6344	61.13%, 60.68%	0.2332
OOM	5.16 (p=0.02)					
VPD	9.79 (p=0.002)					
Model 2: Face Measurements	38.89 (p<0.001)	10.05 (p=0.2613)	42.59 (5 df; p=<0.001)	0.6703, 0.6404	70.49%, 69.84%	0.1920
Maximum Face Width	15.48 (p<0.001)					
Intercanthal Width	17.02 (p<0.001)					
Midface Depth	2.28 (p=0.1314)					
Upper Face Height	0.25 (p=0.618)					
Lower Face Height	1.12 (p=0.289)					
Model 3: Model 1 + Model 2	29.58 (p<0.001)	6.36 (p=0.6065)	36.03 (7 df; p<0.001)	0.7544, 0.7192	75.60%, 74.23%	0.1854
Maximum Face Width	0.43 (p=0.5135)					
Intercanthal Width	17.52 (p=<0.001)					
Midface Depth	2.57 (p=0.1092)					
Upper Face Height	4.55 (p=0.0329)					
Lower Face Height	0.51 (p=0.4738)					
OOM	0.00 (p=0.9831)					
VPD	0.61 (p=0.4330)					

Wald chi-square tests null hypothesis that removing a given covariate does not harm model fit; significant p-value indicates that a covariate/group of covariates affects model fit. Hosmer-Lemeshow (chi-square) GOF evaluates ratio of predicted and observed outcomes over ten-centile groups; insignificant p-values indicate acceptable fit. Likelihood ratio test compares base model to full model performance; significant p-value indicates fit difference. AUC shows level of discrimination achieved by model on full data set, test data set (after 5-fold cross-validation). Classification indicates percentage of individuals classified correctly (0.5 cutoff) by specified model in full, test data set. Brier score indicates probability forecast error across participants. Brier scores range from 0 (no gap between predicted probability and outcome – a perfect forecast) to 0.25 (same performance as chance – a forecast having 50/50 odds of specifying correct outcome).

3.4 Discussion

Over the past several decades, numerous research studies have identified potential genetic and environmental risk factors associated with NSCL/P. However, there have been few attempts to build predictive models using these risk factors. There are several possible reasons for this, which are explored in the context of the present study.

Non-syndromic OFCs are highly heterogeneous in terms of their phenotypic expression, and there is general consensus that the genetic risk factors between the subtypes differ (Grosen, Bille, et al., 2010; Ludwig et al., 2017; Wen & Lu, 2015). In predictive modeling, this heterogeneity makes it difficult to obtain sample sizes large enough to make meaningful comparisons between all of the cleft subtypes. For this reason, subtypes are often lumped into two groups: CL/P versus CPO. The present study involved subtypes of NSCL/P only; families with isolated cleft palate were not included in the analysis. Wen and Lu recently demonstrated that allowing for the heterogeneity in modeling, rather than placing individuals in pre-determined groups based on empirical data, may result in better discrimination between OFC subtypes, at least for genetic variants (Wen & Lu, 2015). Future studies in predictive modeling could explore how the same methods affect prediction across OFCs using subclinical phenotypes.

Finding risk factors with good predictive ability across diverse populations is also proving to be complicated. Genetic risk variants for NSCL/P appear to vary by population not only in prevalence, but also possibly in direction of effect on risk. Zhang et al. performed genetic risk assessment in the Han and Uyghur populations in China, finding that the 43 candidate SNPs from previous GWAS studies were better at predicting NSCL/P status in the Han population (AUC 0.90 in Han and 0.64 in Uyghur). Furthermore, they found that 13 SNPs appeared to be risk factors in

one group while being protective in the other (Zhang et al., 2018). This is just one example of how even within one racial group, there can be substantial variation in risk factors.

It is less clear whether the same phenomenon occurs with non-genetic risk factors, as studies are less consistent (see Environmental Risk Factors section). If subclinical phenotypes are assumed to be a proxy for risk alleles, it is logical to suppose that it does. Indeed, facial morphology is known to differ based on characteristics such as race and sex (Weinberg et al., 2006). On the other hand, it is also possible that despite differences in specific risk alleles in a given population, the effect of those variants on phenotype are the same – a resultant OFC or a subclinical phenotype – and therefore measuring phenotypes, as in the present study, rather than genotypes could capture risk across those different variants. If the latter is true, using phenotypic risk factors of unaffected parents could prove extremely valuable in recurrence risk assessment.

Another way to address genetic variation across geographically diverse populations, is to analyze data with sufficient racial diversity. If data is representative, findings should be generalizable. The participants in the present study were recruited as part of the Pittsburgh Orofacial Cleft (POFC) study, which attempted to assemble a robust, diverse data set for studying risk factors for NSCL/P. Thus, the present study is more racially diverse than previous studies of phenotypic risk factors performed exclusively in Caucasians (See Table 5 for the effect of race as a covariate; see Figure 7 for a summary of participants by race). This diversity allows for some degree of representation; however, the proportion of individuals from underrepresented in biomedical research groups (e.g. Africans, Native Americans) with complete measurements available for the phenotypes of interest was still relatively small, making it difficult to extend findings to these groups with high confidence.

The most significant barrier in developing a useful model, as is often the case with research, is the availability of a large cohort with complete information for a set of uniform NSCL/P risk factors. There are some large-scale efforts to collect data on birth defects in the U.S., such as the National Birth Defects and Prevention Network (NBDPN) surveillance program, which works with state-based programs to collect pregnancy, birth, and demographic information for the purposes of tracking the prevalence of birth defects and collecting information on potential risk factors. The NBDPN collects information on OFCs, and began separating by CL, CL/P, and CPO in 2014 (Mai et al., 2014). There is an effort to separate OFCs by isolated versus in addition to other congenital anomalies, however, this is not always interpreted correctly to rule out all syndromic OFCs. There is not another robust data set outside that collected as part of the POFC study which includes information on phenotype (i.e. measurements and assessments used in the present study), environmental risk factors, and genetic risk factors. Thus, the POFC data is a valuable resource for developing predictive models for NSCL/P. Future modeling using this data set could incorporate additional facial measurements, environmental risk factors (e.g. prenatal environment), and possibly a polygenic risk score component.

3.4.1 Model Performance

To the best of this researcher's knowledge, there has only been one attempt to create a similar discriminant analysis type prediction model for NSCL/P reported in the literature. Li et al. conducted a case-control study (113 cases, 226 controls) of mothers of children born with NSCL/P in the Hunan Province of China (Li et al., 2016). The risk factors explored included family income, maternal and paternal occupational hazards and exposures, dietary factors (strong tea drinking,

milk/soymilk intake in first trimester of pregnancy), and family history. There are several methodological concerns surrounding the validity of the findings in this study.

Li et al. (2016) also employed logistic regression, but their sample size to variable ratio is problematic for producing a model that is overfit to the data and therefore expected to have poor prediction stability if used on another population. Furthermore, there was no out-sample validation (e.g. cross-validation) due to sample size, so the AUC reported (0.846) relates only to the sample on which the model was developed. The initial step used to choose variables for the multivariate model was univariate analysis; variables were excluded based on an alpha level of 0.05, potentially leaving out important variables and leading to excess unexplained variance, further reducing the validity of model. Finally, well-studied risk factors including tobacco, alcohol, and medication use during pregnancy, as well as folate supplementation and maternal stress were not assessed for their predictive values, while less generalizable variables were assessed, such as parental strong tea drinking and whether or not there was a premarital medical exam performed. Nevertheless, it represents an attempt to use risk factors to predict NSCL/P and suggests the possibility of creating a model which incorporates family history in addition to environmental exposures.

Logistic regression requires a fairly large number of individuals per variable to ensure reliable performance because of the binary outcome (more room for misclassification error when there are only two outcomes). The events per variable rate (EPV) is a measure commonly used to assess logistic models for the expected level of instability (how differently it may perform across different, but similar data sets) and optimism (overconfidence in prediction, inflation of AUC). EPV is the quotient of the number of events in the smaller group between two binary outcomes, divided by the number of covariates used in the model. Historically, the suggestion for EPV in logistic regression was 10 to 15; however, newer research into modeling suggests EPV should be

between 20 and 50 to ensure high stability and low optimism (Austin & Steyerberg, 2017; van der Ploeg, Austin, & Steyerberg, 2014).

All logistic analyses in the present study meet or exceed an EPV rate of 20, with the exception of Model 3 (EPV 12.8), due to the smaller number of control family members (N=154) and the use of 12 covariates. The Brier score decomposition also supports a moderate degree of overconfidence in prediction (Table 7). Models 1 and 2 of the present study are expected to have good stability and less optimism, compared to Model 3. However, Models 1 and 2 had poorer predictive performance by AUC than Model 3, so further investigation of the variables in Model 3 may be warranted to determine if the predictive ability holds true in another data set.

3.4.2 Predictors

Increased intercanthal width was consistently associated with higher odds of participant case family status, with the measurements for males appearing to drive the effect. Males with deeper midfaces also had higher odds of being from a case family. These observations are both consistent with previous findings in principal component analysis (PCA) of 3D facial scans from unaffected relatives of an individual with NSCL/P (Weinberg et al., 2009). Wider faces have been consistently associated with NSCL/P, and were associated with increased odds of case family status in Model 2, with females appearing to drive the effect. Increased upper face height, especially in females, was associated with higher likelihood of case family status; this is not consistent with previous findings indicating that shorter upper face height is associated with NSCL/P (Weinberg et al., 2009).

Somewhat unexpectedly, OOM discontinuity did not significantly affect the odds of being from a case family in univariate analysis; even more surprising, it was associated with decreased

odds of case family status in Model 1 and nearly significantly associated with decreased odds for males in Model 3. There is substantial support in the literature for OOM as a subclinical phenotype of OFCs (Klotz et al., 2010; Martin et al., 2000; Neiswanger et al., 2007). Statistically, it is possible the small number of individuals positive for OOM discontinuity (about 4%) resulted in reduced model effect; in addition, there was no significant difference in the number of positive individuals between the case and control families, but there were slightly more positive individuals in the control family group. Biologically, perhaps individuals with OOM defects have protective factors that prevent full manifestation of the OFC, leading to decreased risk of recurrence in the family. Indeed, Klotz et al. (2010) found a higher recurrence risk for an OOM defect than for NSCL/P within families positive for an OOM defect. Additionally, in a study of monozygotic twins discordant for NSCL/P, a trend towards higher OOM discontinuity was seen in the unaffected co-twins, but this was not significant, possibly due to sample size (Leslie et al., 2017). Finally, the nearly significant difference in OR for males with OOM defects is in line with the observation of a higher proportion of unaffected male relatives with OOM defects in a previous study, however, the direction of effect differs (Neiswanger et al., 2007).

In exploring this unexpected outcome, race resulted in statistically significant differences in the proportion of individuals with OOM discontinuity; Africans had the lowest proportion of individuals affected by NSCL/P, but made up the largest proportion of the individuals with OOM discontinuity, potentially lending support to the idea of protective factors associated with OOM discontinuity (8.9% versus 3.2-3.7% for the other three groups). Secondly, since intercanthal width was a consistent risk factor, a comparison between those measurements and OOM discontinuity was made. A trend in increased intercanthal width was seen in individuals negative for OOM defects, with decreased measurements in those who were positive, but this was not statistically

significant after accounting for confounding factors. It is also possible that the effect of the OOM defect differs based on the added effects of additional factors in the model.

The very wide confidence interval (CI) for the constant term in Model 3 hints at an issue of sparse data and separation, in which the very small number of individuals who are positive for OOM tend to be from the control group by chance (Greenland, Mansournia, & Altman, 2016). Further supporting this notion, is the wide OR CI for OOM status (0.23, 4.49), which hints at the fact that the odds are erratic, from a large reduction in odds of case family status to a large increase, depending on each individuals' combination of factors in addition to OOM status. There are only seven individuals with OOM defects present among the individuals included in Model 3.

VPD resulted in increased odds of being from a case family in univariate analysis and in Model 1 (OR=1.8). As discussed above, there are few studies on VPD and NSCL/P recurrence risk. As with OOM discontinuity, the effect was not seen when combined with additional phenotypes in model 3. Possibly for similar reasons of sparse data and separation. Regarding OOM and VPD in Model 1, it is possible that the presence of one phenotype affects the likelihood of the other being present; however, few (~3% overall) of the participants had both VPD and an OOM defect, so it was not possible to evaluate interaction (5 in control families, 6 in case families).

3.4.3 Developing a Clinical Useful Model

Predictive models for a future health outcome can be extremely beneficial to patient care, but developing them is complex. Currently, there are significant barriers to developing a rigorously tested model for NSCL/P. First and foremost, is the issue with the type of data available. The mostly cross-sectional cohort studies available for the measurements used in this analysis are not the best-suited for building a model to predict an incident event (i.e. the future birth of a child with

NSCL/P); the best data for this would be derived from a longitudinal or prospective cohort study (Lee et al., 2016). Ideally, we would collect all measurements of interest from individuals with a family history prior to conception, then follow-up with them several years later, to record the affected or unaffected status of their children. This data could then be used to create a predictive model for the incidence of NSCL/P. No such data currently exists for the phenotypes used in the present study.

The other difficult aspect of predictive modeling, is the absence of any hard and fast rules surrounding acceptable model performance. Screening in a clinical, high-risk group generally leaves more room for test error than in a low-risk general population, simply because the baseline suspicion of a positive result (i.e. the prevalence) is already higher. The main considerations for determining whether a model has acceptable performance depend upon the situation in which it is designed to be employed (Trevethan, 2017). The consequences of a false negative versus false positive result, for example, may differ greatly in predicting a birth defect, such as NSCL/P, versus predicting the chance of an adult developing diabetes. Most screening is designed to be highly sensitive, that is, to err on the side of false positive results and avoid false negative results (Harris, Sawaya, Moyer, & Calonge, 2011).

False positives, however, have a set of negative consequences, such as parental stress and anxiety and possibly, in the worst case scenario, pregnancy termination (Risks, Andrews, Fullarton, Holtzman, & Motulsky, 1994) (although, even in prenatally diagnosed cases of NSCL/P, termination is rare (Nusbaum et al., 2008)). Perhaps first trimester maternal serum screening (for chromosomal aneuploidy) can provide useful guidance in determination of acceptable predictive performance for a NSCL/P predictive model; there is a false positive rate of about 5%, but a much lower false negative rate. However, in the high-risk population, motivation for some answers could

motivate individuals to desire models even at the expense of some accuracy. Many considerations need to be made in the case of employing a predictive model for NSCL/P, especially surrounding test performance. In-depth consideration of these issues is beyond the scope of the current manuscript.

3.4.4 Limitations and Future Directions

The present study was performed on data containing a higher proportion of individuals with a positive family history of NSCL/P, therefore, it is not representative of the population prevalence of NSCL/P. Additionally, the data was imbalanced; more individuals with complete measurements available for the phenotypes of interest fell into the case family group. While this allows for increased precision in terms of the coefficients/ORs reported by logistic analysis, it does not allow for the model to have very robust discrimination within the control families, which hurts the overall specificity and negative predictive value of the model.

The data also included some families with multiple blood-relative participants (e.g. father, mother, and sibling). Logistic analysis assumes independent sampling, so the related individuals in the data violate this assumption. Additive effects of relatives who share a common risk phenotype may have impacted the performance of the model. Mixed effect logistic regression, and other statistical methods for non-independent data, can control for the "clustering" of family members with similar traits, however, this is computationally burdensome and was not performed for the purposes of the present study. An analysis including only parents would also help to address this issue, however, sample size did not permit such analysis in this case.

Finally, there were many individuals with missing data points for one of the phenotypes of interest. This led to a sharp reduction in sample size with the combination of multiple variables.

Unfortunately, there is not yet another data set available for which to test this model on. In future studies, alternative measurements or possible factor analyses will be considered to minimize the number of covariates and maximize sample size. Many additional facial measurements are available, as well as information on environmental risk factors. Nasal cavity width, for example, was shown to have a moderate effect on risk for NSCL/P and could be an important predictor (Weinberg et al., 2006). A future model could include such variables and evaluate their predictive performance, possibly in addition to genotype information for proposed genetic risk variants.

3.5 Conclusions

The present study investigated the use of several proposed risk phenotypes to predict case versus control family status of individuals with a positive family history of NSCL/P. Despite limitations in sample size and power, modeling revealed several traits which appeared to significantly affect the odds of being from a case family and fairly acceptable AUC estimates.

Intercanthal width was consistently associated with higher odds of case family status, with an especially pronounced effect for males with wider intercanthal distances. Therefore, intercanthal width may prove to be a useful marker for modeling recurrence risk for NSCL/P. VPD is worth further consideration, as well, particularly based on the results observed in males in the present study. Maximum facial width, midface depth, and upper/lower facial heights have unclear significance per the results of modeling, however, there were significant effects observed in a sex-dependent manner, thus modeling for parents separately may be most helpful in future model development attempts. Lastly, the effect of OOM in this study was not significant, but was associated with a trend towards lower odds of case family status. This warrants further

investigation, as modeling not only attempts to identify those at increase odds, but also those who may be at decreased odds.

Future studies including additional phenotypes could reveal relative associations between the facial phenotypes that could not be determined in the present study. Based on the ease with which most of these measurements can be obtained, if future studies confirm their predictive value, there could be significant benefit to patients seeking risk assessment.

4.0 Research Significance to Genetic Counseling and Public Health

Providing accurate risk assessment and anticipatory guidance are key functions of comprehensive genetic counseling. Providing accurate risk assessment for recurrence in the setting of a complex genetic trait, such as NSCL/P, is currently difficult. Providing preconception risk counseling to individuals with a personal or family history of NSCL/P is less than precise; the only risk figures available to provide are empirical in nature, and generally based on large prospective studies of mostly Caucasian cohorts (Grosen, Chevrier, et al., 2010). It is known that not all families with isolated cases of NSCL/P have the same recurrence – some cases may be truly sporadic or environmental, while others have increased risk due to an underlying genetic predisposition (Crawford & Sofaer, 1987). The ability to separate these two situations would allow for individualized risk assessment, not solely reliant on positive family history. If a predictive model could be built using risk factors in addition to family history, genetic counseling has the potential to improve for recurrence risk assessment in NSCL/P. The present study takes an important first step in the development of such a model.

A model is only clinically useful if it has an acceptable trade-off between sensitivity and specificity; highly sensitive tests allow practitioners to more confidently rule out disease if a highly sensitive test is negative (Trevethan, 2017). Conversely, a test which is highly specific will rarely result in a false negative, so a positive result provides reasonable confidence that a condition is truly present (Trevethan, 2017). Even the model with the best performance in the present study, Model 3, would result in a large number of false positives. The sensitivity was 87.1% and the specificity was 33.8%. This test performance means that nearly 30% of positives would be false

positives and nearly 50% of negatives would be false negatives, making it of little use in clinical practice.

Clinically useful methods of assessing risk can only be developed in the setting of robust research on the genetic and environmental risk factors. The present study evaluated only a handful of the many proposed risk phenotypes, finding cause for further consideration. Many more risk factors have been proposed in the literature that may be worth focusing on for developing a risk assessment tool with greater predictive ability, possibly even with the addition of a PRS component. Adding other risk factors may have the potential to improve model performance.

Providing accurate risk figures to couples in the preconception stage could allow for family planning, education, and preparation (both practical and emotional) for parenting a child with NSCL/P. Prenatally, risk assessment could provide cause for more careful ultrasound evaluation, plus the same ability to prepare, allowing parents to seek out resources and meet with cleft craniofacial centers to create a plan.

Predictive models can also serve an important public health service at the population-level, if they are able to aid in identifying subgroups of individuals at increased risk for recurrence (Agopian et al., 2012). Identifying those at increased risk to have a child with NSCL/P prior to conception could provide the opportunity to facilitate risk reduction steps (e.g. smoking cessation, folate and vitamin A supplementation) in an effort at primary prevention. Furthermore, the process of model development has the potential to reveal novel risk factors with significant predictive ability, providing new avenues for designing public health interventions. Research, such as that undertaken in the present study, is an essential component of public health services, as it allows for evidence-based policy development for population screening.

Additionally, one of the most critical points in laying the foundation for effective long-term care of a child with NSCL/P is in the first few days to weeks of the child's life (Cleft Palate-Craniofacial Association, 2009). There are effects on physical, mental, and emotional health for individuals with NSCL/P and their caregivers; some of these effects can remain long-term issues without proper treatment (Sischo, Wilson-Genderson, & Broder, 2017). In the newborn period, infants with NSCL/P are at risk for failure to thrive without the proper feeding interventions; babies seen earlier in life by a cleft craniofacial center are more likely to have the right feeding support from the beginning (Kaye et al., 2017), they are also less likely to be admitted to the NICU.

It is important that caregivers have the support of providers who are adept at facilitating successful navigation of the surgical and therapeutic interventions for the OFC, specifically, and also of the mental and emotional difficulties they may feel along the way. Additionally, children with OFCs need access to emotional support through their families and the medical team to adapt to potential differences in their facial appearance and speech and to cope with the many surgical procedures and therapies they made need over the course of their lives. Cleft craniofacial centers are an invaluable resource in every life stage and step in treatment of an OFC. However, there may not be a center available immediately in certain areas of the country, as there are only about 175 centers in the U.S. Knowledge of risk status prior to delivery could help in coordinating with the nearest center, with the goal of an early visit with the team to plan for immediate needs (e.g. feeding), pre-surgical interventions, and initial surgery. In this case, predictive modeling could better the ability to provide another essential public health service: linking to or providing care.

5.0 Public Health Essay

5.1 Introduction

Congenital malformations (i.e. birth defects) are a significant public health concern; they affect approximately 1 in 33 babies born in the United States (US) (CDC, 2018a). They are costly to the healthcare system and account for a substantial portion of neonatal mortality (20%) and morbidity (CDC, 2018a). Orofacial clefts (OFCs) are one of the most common birth defects in the U.S., occurring in 8.1 to 10.63 per 10,000, or roughly 1 in in 1,000, births (Mai et al., 2014; U.S. Department of Health and Human Services, 2014). OFCs include isolated clefts of the palate (openings in the roof of the mouth) and clefts of the lip (unilateral or bilateral gaps in the upper lip which may extend to the nose). Non-syndromic or isolated forms of cleft lip, with or without cleft palate (NSCL/P), account for the majority of OFCs: 70% of all cases of isolated cleft lip with or without cleft palate, and 50% of isolated cleft palate cases are non-syndromic (IPDTOC, 2011; Mai et al., 2014). While NSCL/P is not often life-threatening to newborns in the U.S. and other developed countries, it has a significant, long-term effect on health (Lewis et al., 2017).

Recognizing the public health significance of OFCs, due to relatively high prevalence and the resulting increase in healthcare use and costs, the Centers for Disease Control (CDC) assembled a group of experts in 2006 to examine the state of public health-related research surrounding OFCs (Yazdy et al., 2007). The working group sought to identify gaps in knowledge, prioritize the most critical issues, and create a public health research agenda. Key areas the panel identified included (1) the relationship between maternal health and OFC etiology, (2) the psychosocial impact of OFCs in childhood, (3) the impact on quality of life for children with OFCs

and their families, and (4) the healthcare costs associated with OFCs (Yazdy et al., 2007). Major goals of creating this agenda surround improving the ability to prevent OFCs and improve long-term treatment outcomes and quality of life (Yazdy et al., 2007).

Another area of OFC research that appears to have gaps is recurrence risk assessment. Currently, genetic counseling and risk assessment for NSCL/P are based only on empiric risk estimates (Basha et al., 2018; Grosen, Chevrier, et al., 2010). While many predictive models exist for predicting occurrence of conditions such as breast cancer (e.g. the Gail model), and a few common, complex diseases even have polygenic risk scores (PRSs) (e.g. heart disease and diabetes), birth defects, in general, have received less attention in this regard (Agopian et al., 2012). In the past decade, a couple of research groups have noted this and explored predictive modeling methods for two congenital conditions: neural tube defects and congenital heart defects (Agopian et al., 2012; Luo et al., 2017).

Models built for open neural tube defects by Agopian et al. (2012), using known maternal risk factors in a large, population-based cohort, had poor predictive performance, leading the researchers to conclude that more research into the risk factors is needed to create a useful model. However, Luo et al. (2017) had reasonably good results predicting risk for congenital heart defects using the models they built. Their models included nine, broad risk factors: maternal age at delivery, annual per capita income, family history, maternal history of pre-conception illness, inadequate nutrition/folic acid intake, maternal illness during pregnancy, medication use during pregnancy, environmental risk factor exposures during pregnancy (e.g. radiation, pesticides), and unhealthy maternal lifestyle during pregnancy (e.g. smoking and alcohol use). The sample was large and population-based; it included 78 cases of congenital heart defects among 33,831 births between 2006 and 2008 in ShangXi, China. The true positive rate for the models was 65% and the

true negative rate was 95%, which the researchers concluded was sufficient for consideration of using them to identify women at high risk in the general population (Luo et al., 2017). OFCs have received even less attention in terms of predictive modeling using known risk factors; only one study is known, to date, and it was performed on a small sample, using risk factors that are difficult to generalize, and the statistical methodology was unsound (see section 3.4.1) (Li et al., 2016).

There have been attempts to develop PRSs for NSCL/P, with some success (Ludwig et al., 2017). However, it is expected that there are still several unknown genetic variants, gene by environment interactions, and epigenetic phenomenon that have yet to be explained in the heritability of NSCL/P (Basha et al., 2018; Ludwig et al., 2017). Regardless, it is likely that genetic and environmental risk factors have predictive value in estimating the recurrence risk of NSCL/P (Dixon et al., 2011b; Ludwig et al., 2017; S. M. Weinberg et al., 2006) and could therefore be used to create a predictive model or PRS for reproductive screening purposes. In considering the development and implementation of such a model, it is important to analyze the public health implications associated with preconception screening for NSCL/P.

The overarching goals of this essay are to (1) consider the public health implications, including public health utility, selected ethical, legal, and social issues (ELSI), and economics, of preconception population screening for NSCL/P, (2) evaluate NSCL/P, using the principles set forth by established frameworks/criteria for evaluating public health programs, and (3) conclude whether such screening warrants consideration, and, if so, in which setting it would be most appropriate to implement.

Evaluation of a population screen includes thorough assessment of test performance, which is a critical issue when a screen is used in the general population, in which prevalence of a health outcome of interest is expected to be low. Many screening tests are designed to be highly sensitive,

so they often have a significant false positive rate (Harris et al., 2011). Accomplishing the above goals within the scope of this manuscript required assuming that a method of screening with acceptable population test performance had already been developed. In order to include a brief discussion of how test performance impacts evaluation of harms and benefits, the hypothetical screen for NSCL/P will be assumed to have similar performance to other screening protocols already in use. Therefore, the starting point for discussion is a hypothetical situation, in which a reliable pre-conception screen for NSCL/P is available. This assumption allows for a thoughtful consideration of the implications of successfully developing a model, such as that described in section 3.

The following is a limited discussion of the history and current state of perinatal population screening, meant to provide background and context for evaluating preconception population screening for NSCL/P. Where there are gaps in specific literature on OFCs, other relevant examples of well-established screening programs are used to facilitate discussion of a hypothetical screening tool for NSCL/P. Various criteria, developed for evaluating public health screening programs, are also examined for their relevance and limitations in regards to screening for a congenital malformation.

5.2 Brief History and Overview of Perinatal Population Screening in the U.S.

Assessing the likelihood of disease presence or development in apparently healthy individuals started with infectious diseases (Wilson & Jungner, 1968). In 1936, syphilis became the first prenatal screen to be mandated by federal law, with several states enacting state laws in the 1940's (Institute of Medicine (US) & National Research Council (US) and Institute of

Medicine (US) Board on Children, Youth, and Families, 1999). Around the time this legislature was developed, syphilis accounted for the largest proportion of pregnancy losses and stillbirths in the U.S., making it a significant public health problem (Institute of Medicine (US) & National Research Council (US) and Institute of Medicine (US) Board on Children, Youth, and Families, 1999). Since the mid-twentieth century, perinatal population screening in the US has grown beyond syphilis to encompass non-communicable, chronic conditions (Wilson & Jungner, 1968), which are detectable in individuals who are pre-symptomatic. This arguably began with newborn screening, specifically mandatory testing for phenylketonuria in newborns in the 1960's (Institute of Medicine (US) & National Research Council (US) and Institute of Medicine (US) Board on Children, Youth, and Families, 1999). The primary benefit in both situations lies in prevention, either primary or secondary, to reduce morbidity and/or mortality, and to minimize the economic impact of disease (Wilson & Jungner, 1968).

The association of maternal health with infant health, both prior to and during pregnancy, has been understood to some degree for hundreds of years (Freda, Moos, & Curtis, 2006). In a paper analyzing the history of preconception care, Freda, Moos, and Curtis (2006) make this point by quoting Plutarch, who described that in ancient Sparta, women were encouraged to be physically fit and strong to ensure the same in their offspring. Despite the ancient roots of an intuitive understanding of the link between maternal health and the health of babies, in the U.S. it was not until fairly recently that there was a purposeful shift in the focus of prevention of neonatal morbidity and mortality (Freda et al., 2006).

In the 1980's, public health interventions began to move away from a main focus on interventions during pregnancy and infancy, to those that could be more effective if employed prior to a mother becoming pregnant (Freda et al., 2006). One classic example of this type of

primary prevention via a public health intervention involves an effort to reduce the occurrence of a birth defect. The mandatory fortification of cereal grain products with folic acid became an important public health intervention to drastically reduce women's risks of having a child with an open neural tube defect; folic acid fortification was first implemented between 1996-1998 in the U.S. (Crider, Bailey, & Berry, 2011) and continues to date.

Beyond fortification, a second public health intervention intended for women with a higher than population risk for open neural tube defects based on family history is folic acid supplementation. All women are encouraged to take in 400 micrograms per day of folic acid from diet and supplementation prior to and during pregnancy, and it is recommended that women who have a previous pregnancy or close family member affected by an open neural tube defect, or have certain health conditions, take an additional supplement to achieve an intake of 4 milligrams per day (Copp et al., 2015; Farahi & Zolotor, 2013). This simple example illustrates the basic idea of any public health intervention: (1) identify a population at increased risk for a congenital condition and (2) employ an intervention.

Preconception screening, counseling, and perinatal healthcare (collectively referred to as family planning), is well-recognized for its valuable role in improving population health; it is one of the top ten public health achievements recognized by the CDC (1999). Family planning results in improved maternal and infant outcomes and reduces healthcare costs (Institute of Medicine (US) & National Research Council (US) and Institute of Medicine (US) Board on Children, Youth, and Families, 1999; Johnson et al., 2006; Lu, 2007). One major benefit of effective preconception care, including family planning, is the ability of women to avoid unintended, including untimed and undesired pregnancies, (Institute of Medicine (US) Committee on a Comprehensive Review of the HHS Office of Family Planning Title X Program, 2009). On average, compared to women with

planned pregnancies, women who have unintended pregnancies come later to prenatal care and are more likely to smoke and consume alcohol; exposure to smoking, alcohol, and drugs often occurs at a vulnerable point in embryonic development (Coles, 1994), that is, during the first two months of pregnancy (embryonic period and organogenesis) before a woman knows she is pregnant (Institute of Medicine (US) Committee on a Comprehensive Review of the HHS Office of Family Planning Title X Program, 2009). This becomes an important point in the discussion of NSCL/P, as maternal smoking and exposure to second-hand smoke is a well-established risk factor and alcohol has also been associated with increased risk (Bell et al., 2014; Pi et al., 2018; Xuan et al., 2016).

Evaluating the setting in which a screen will take place is an important consideration for any screening test (Becker et al., 2011), therefore, understanding what pre-conception counseling already includes is an important place to start. Nearly 50% of all pregnancies in the U.S. are unintended; therefore primary care for all women of reproductive age should routinely include preconception care and family planning considerations (Farahi & Zolotor, 2013). National organizations have put forward guidelines for preconception care including general issues that should be included. Recommendations include assessment of environmental exposures (e.g. workplace exposure to toxic chemicals), medication use, psychiatric illness, psychosocial health (including intimate partner violence screening), and substance use. Nutrition, infectious disease screening (e.g. Syphilis, Tuberculosis), and immunization status should be assessed. Management of chronic diseases such as diabetes, hypertension, and thyroid disease, and maintaining a healthy body weight should be discussed. Finally, family genetic history, including personal or family history of congenital anomalies or heritable genetic conditions, should be obtained; patients with positive family history should receive genetic counseling on modifiable risk factors and

testing/screening options (American College of Obstetricians and Gynecologists, 2019; Farahi & Zolotor, 2013). Of note, these current guidelines for preconception care in the primary care setting, if followed effectively, already address the currently known modifiable risk factors for NSCL/P.

5.3 Frameworks for the Evaluation of Population Screening

While the goals of any public health screen are similar, as screening has moved into new areas of application and screening methods have evolved, it has become evident that there are important distinctions depending on each screen's goal. For example, the goal of traditional population screening for a condition which is already present, but is in a pre-symptomatic or very early stage (case-finding/early detection), differs from the goal in predictive screening for conditions which *may* develop in the individual being tested (primary prevention). Even more so, predictive testing in an adult, who will be directly impacted by a given condition, differs from screening for a condition which may occur in a future child. An understanding of these differences has led to the development and adaptation of criteria for evaluating the utility (i.e. benefits and harms) of a proposed screening program.

The Wilson and Jungner criteria, which sets forth ten "principles" for evaluating a condition for population screening has historically been considered the gold standard (Harris et al., 2011). Principles is in quotations, as Wilson and Jungner made it clear they did not intend for them to be taken as dogmatic or set in stone, but rather intended for them to be used to help standardize and stimulate thinking about important aspects of the evaluation process (Wilson & Jungner, 1968). Since the 1960's when these criteria were developed, however, public health needs and the many interventions available have changed. Additional guidelines have been proposed for

evaluating genetic screens in a public health context (Becker et al., 2011). There is an extensive body of literature on evaluating population screening with many authors critiquing Wilson and Jungner principles for being limited in regards to new screening methods and the current setting of these methods (Andermann, Blancquaert, Beauchamp, & Déry, 2008; Botkin, 2009; Harris et al., 2011). A limited summary of four main criteria and critiques is presented below.

The original Wilson and Jungner criteria include ten "principles" of good public health screening; these are summarized in Table 8 below (Wilson & Jungner, 1968). While these principles still serve as an important basis for evaluating conditions for inclusion in population screening, there are several critiques of their limitations in regards to emerging public health applications, including that they are too broad (e.g. what is an "important" health problem?) and that they leave out several possible benefits of screening (Botkin, 2009). Of interest in the case of screening for reproductive risks, for example, Botkin (2009) points out that these criteria do not consider the benefits of reproductive information for parents.

Andermann and colleagues (2008) performed a literature review of screening criteria that have been proposed over the last 40 years. They presented a summary of the overarching themes of these proposed criteria (summarized in Table 8), which they point out often overlap with those originally proposed by Wilson and Jungner (Andermann et al., 2008). Key differences or clarifications of the Wilson and Junger principles that are included in the Andermann et al. (2008) summary center around concepts of ensuring equity and access, elements of declaring intent of screening and planning for rigorous evaluation of the program, and a movement towards a consideration of the balance between benefits and harms, rather than a "checklist". In a more recent analysis, similar in intent to that in Andermann et al. (2008), four current and former members of the U.S. Preventive Services Task Force (USPSTF) presented their idea of moving away from a

checklist for screening evaluations, instead performing a careful and complete evaluation of benefits and harms (Harris et al., 2011), see Table 8.

Finally, the ACCE (Analytic validity, Clinical validity, Clinical utility, and Ethical, legal, and social issues (ELSI)) framework (see Table 8) was funded by the Office of Public Health Genomics at the CDC. It was developed to provide a framework for formal evaluation of DNA and related testing for genetic conditions; it includes the processes of collecting, evaluating, interpreting and reporting (Haddow & Palomaki, 2011). The ACCE framework stresses that the clinical disorder, test, and test setting must be the starting point of the program (Table 8). There are 44 standardized questions included in the ACCE framework, to facilitate thinking about each of the four components of genetic testing (the list of questions is publicly available at https://www.cdc.gov/genomics/gtesting/acce/acce_proj.htm). The questions guide the evaluator through considerations of test performance, such as the test development, reliability within and between laboratories performing the testing, and quality control measures (analytic validity). Questions for clinical validity include sensitivity and specificity, considerations of population prevalence, and positive and negative predictive value.

Clinical utility differs from clinical validity in that it describes a test's resulting risks and benefits (positive and negative health outcomes); clinical utility is evaluated via a series of questions on the natural history of the disorder, impact on patient care (e.g. is there a clinical intervention available to treat, prevent, or slow progression of the disease?), availability and type of follow-up diagnostic tests, and the availability of facilities/personnel to effectively perform the screening. Finally, ELSI considerations can include stigma associated with the condition being screened for, potential discrimination, privacy considerations, plus legal issues such as consent, duty to disclose/report, and proprietary/patented testing methods (Haddow & Palomaki, 2011).

While the ACCE framework is meant to provide important considerations for genomic testing, it also contains many of the same themes as previous frameworks, plus additional aspects (e.g. ELSI), which can be helpful in thinking through screens that are not strictly a genetic test, but which may still have ELSI or familial implications.

Harris et al. (2011) present a framework that is less semi-quantitative than with a checklist approach, but centers around analysis of the magnitude of benefits versus the magnitude of harms, followed by a critical assessment of whether the net benefit is worth the resources required to implement a screen (Table 8). They also include that studying the perception of benefits and harms of an informed public is helpful, but infrequently used in screening evaluation. Harris et al. (2011) criticize what they call a "benefits only" approach to evaluating screening, citing this type of thinking as a problematic factor in the expansion of highly sensitive screens to ever broader populations without consideration of the harms (overdiagnosis, high rate of false positives, resource use at the expense of other public health programs). The basic conclusion appears to be that screening is not an ideal strategy for controlling the population burden of all diseases; the authors believe it is important to rein in the application of screening when there is not clear *net* benefit to justify resources (patient/practitioner time, money), according to rigorous research. The basic process is outlined in Table 8; see Appendix B, Table 9, for full explanation of the minimal evidence sufficient to estimate the benefits and harms of a screen according to Harris et al. (2011).

The frameworks available for evaluating a screening program are generally meant for a biochemical or genetic test; however, Harris et al. (2011) point out that risk stratification tools (i.e. models) can be considered a form of testing, it just requires considering the outcomes, or as they refer to them "predictors of poor health (PPH)" (page 23), slightly differently. So, in the case of NSCL/P, using the framework provided by Harris et al. (2011), the target of the screen, or the

PPH, is the risk category assigned by the model. So, in evaluating a model for predicting the risk category for NSCL/P, it is important to remember that the evaluation of the predictive model is based on how well the identification of an individual as "high risk" translates to actually having a child born with NSCL/P. This framing also allows for thinking beyond a disease or birth defect, to thinking through the adverse health outcomes that come of a PPH. In terms of NSCL/P, it takes the focus from simply considering the birth defect to be the sole cause of adverse health outcomes, to also thinking of it as a risk factor for poorer long-term overall health outcomes depending on the quality of care received.

In the limited literature available on developing predictive models for birth defects, the authors have not included a thorough discussion of potential benefits, harms, and resources that would be required to implement a proposed screening tool (Agopian et al., 2012; Li et al., 2016; Luo et al., 2017). The focus in these studies on predictive modeling surrounds developing a reliable method of screening, which is an important first step. However, it is potentially valuable to analyze the hypothetical implementation of the tool in the early stages of model/tool/test development, in order to decide whether additional valuable resources should be dedicated to its development, or if resources would be better allocated to other efforts to address the actual health outcomes (Haddow & Palomaki, 2011; Harris et al., 2011). This type of analysis can reveal important gaps in research (Haddow & Palomaki, 2011), identify barriers to successful implementation, and provide information on the potential utility versus cost of the screen proposed (Harris et al., 2011).

Just as there was with the start of predictive testing in the prenatal realm, there are unique benefits and harms to be considered in a potential preconception population screen for birth defects that, to the author's knowledge, are not as well-researched or clearly defined as those for existing screenings in the prenatal setting, or in newborns and adults (Solomon, Jack, & Feero, 2008).

While isolated and correctable birth defects differ in a number of ways from systemic conditions/syndromes, research on other conditions for which we screen, such as cystic fibrosis, can be looked to for preliminary guidance on screening for birth defects. Prenatal screening is also well-researched and can provide information which can be potentially extrapolated to birth defects screening tools. Additionally, modeling for cancer (e.g. with the Gail model) provides helpful considerations for risk modeling, in general. In the future, to fully evaluate, it will be important for preconception modeling for birth defects to be studied to elucidate any unique impacts and considerations.

Table 8 Summary and Comparison of Evaluation Criteria

Wilson and Jungner, 1968	Andermann et al., 2008	CDC ACCE Framework Haddow and Palomaki, 2011	Harris et al., 2011 Balance Approach
"The condition sought should be an important health problem. There should be an accepted treatment for patients with recognized disease. Facilities for diagnosis and treatment should be available. There should be a recognizable latent or early symptomatic stage. There should be a suitable test or examination. The test should be acceptable to the population. The natural history of the condition, including development from latent to declared disease, should be adequately understood. There should be an agreed policy on whom to treat as patients. The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole. Case-finding should be a continuing process and not a "once and for all" project" (p. 27-28)	"The screening program should respond to a recognized need. The objectives of screening should be defined at the outset. There should be a defined target population. There should be scientific evidence of screening program effectiveness. The program should integrate education, testing, clinical services, and program management. There should be quality assurance, with mechanisms to minimize potential risks of screening. The program should ensure informed choice, confidentiality, and respect for autonomy. The program should promote equity and access to screening for the entire target population. Program evaluation should be planned from the outset. The overall benefits of screening should outweigh the harms." (p.83; Box 2)	*Disorder and setting should be established first.* Analytic validity (A): • Analytic sensitivity • Analytic specificity • Quality control • Assay Clinical validity (C): • Clinical sensitivity • Clinical specificity • Prevalence • PPV/NPV • Penetrance Clinical utility (C): • Natural history • Effective (benefit) • Quality assurance • Pilot trials • Health risks • Evaluation • Facilities • Education • Monitoring and evaluation Ethical, Legal, and Social (E): • Safeguards • Impediments	Define clearly the adverse health outcome the program is intended to reduce and the population that the program intends to screen. Define the important potential health benefits and harms of the screening program. Perform a systematic review of the evidence to estimate the absolute magnitude of the potential benefits and potential harms of the screening program. Include a disclaimer about uncertain extrapolations of benefits and harms (i.e. those not studied/supported). Answer key question: is the net benefit of screening 1) zero/negative, 2) small, or 3) moderate/substantial? Estimate resources required to successfully implement. Consider implementation in terms of nonevidence/practical considerations outside the evaluation (i.e. what resources are available, are there other priorities or population preferences?).

5.4 Evaluating Preconception Population Screening for NSCL/P

A comprehensive review of benefits and harms of preconception screening for NSCL/P is outside the scope of this manuscript, and is likely not possible at this time due to gaps identified in the public health literature on OFCs (see section 5.1). Using the body of literature reviewed in section 5.2 of this manuscript as a guide, in addition to the frameworks for evaluating population screens (summarized in Table 8), several important public health elements of preconception screening for NSCL/P were considered: (1) screening type, clinical setting, and goals of the screening (including availability of trained personnel and facilities to carry out screening and follow-up and a discussion of diagnostic ultrasound performance); (2) clinical utility (the impact of a positive screen on parents' level of psychological stress and preparation, primary prevention, and the timeliness and quality of patient care – both for parents prior to the birth of a child, and for the affected child); (3) selected ESLI, including pregnancy termination, views of the disability community, and equity/access; and (4) economic impact.

5.4.1 Screening Type, Clinical Setting, and Goals of Screening

As emphasized in the ACCE and Harris et al. (2011) evaluation approaches, determining the target population, screen, and goal of the screen is the most important first step. Screening could potentially take place in the specialty clinic setting (e.g. medical genetics) or in the primary care setting. Specialty clinic, here, is meant to describe a provider sought out specifically for evaluation of NSCL/P risk and who would not routinely provide a patient with care; this care could

be sought through an individual's awareness of family history or a primary care provider's referral. This is in contrast to the use of primary care, which, here, is meant to describe a provider who routinely cares for a patient; this could be a woman's regular obstetrician/gynecologist or family practitioner, for example. It is important to consider that not all individuals have a primary care provider. The screening proposed would only be available in a clinic setting, which would leave out individuals who do not receive preconception care.

Framing this discussion with an example of a current screening protocol may be helpful. Currently, in the U.S., there are two types of recommended carrier screening: ethnicity or ancestry-based and population-wide (i.e. pan-ethnic) screening. Screening for two autosomal recessive conditions, cystic fibrosis and spinal muscular atrophy, should be offered to all women who are pregnant or plan to become pregnant (American College of Obstetricians and Gynecologists, 2017); these are examples of pan-ethnic screening. Additional carrier screening should be offered to women of specific ancestry, due to higher prevalence of carriers; Tay-Sachs Disease carrier screening is offered to individuals of Ashkenazi Jewish, French–Canadian, or Cajun descent and hemoglobinopathy screening via hemoglobin electrophoresis is offered to women of African, Mediterranean, Middle Eastern, Southeast Asian, or West Indian descent. Other carrier screening is offered only to individuals with a family history that indicates risk, e.g. women with a family history of intellectual disability or Fragile-X Syndrome, or women with premature ovarian failure (before age 40), should be offered Fragile-X carrier screening (American College of Obstetricians and Gynecologists, 2017).

There are various reasons for the differences in recommendations. Carrier screening for extremely rare disorders is not feasible in most cases; rare disorders are tested for based on specific case by case situations. Screening, however, is meant for conditions that are common enough in a

target population to justify the costs. Sometimes, a combination of condition severity and carrier prevalence drive screening; spinal muscular atrophy, for example, is both somewhat common and also one of the leading causes of infant death (American College of Obstetricians and Gynecologists, 2017). However, there is also debate about the social justice of carrier screening recommendations based on ethnicity, especially in light of the increasingly admixed racial composition of the population (Langlois, Benn, & Wilkins-Haug, 2015). These examples demonstrate some of the reasons behind choosing a screening and target population, and also the complexity of the considerations.

The setting chosen would determine the target population, to a large extent. Screening could be offered to either individuals already known to be at some increased risk based on family history (likely in the specialty clinic setting), or to all individuals of reproductive age as part of routine preconception care (more likely in the primary care setting). Offering NSCL/P screening as part of routine preconception care for all may have the potential to reach more individuals, but the prevalence of NSCL/P susceptibility must be considered in screening performance. Prevalence is expected to be lower in the population at large than in the high risk group, potentially leading to a greater number of false positive screen results (Trevethan, 2017) and, in turn, unnecessary use of valuable resources and greater potential for parental distress. However, individuals with a family history, especially a previous affected child, may already be more familiar with resources for cleft care and less likely to receive any additional benefit from screening.

The goals of screening in either setting would be to refine the reproductive risk for NSCL/P and provide information about modifiable risk factors. A second goal would include an attempt to improve parental preparedness for the care of a child with a cleft by providing high risk individuals with genetic counseling, and facilitating referral to a cleft craniofacial center. This would provide

parents the opportunity to gather information on treatment and learn about support resources available. Information on prenatal diagnosis via ultrasound could also be provided at this time. There are elements of both primary (reducing risk from modifiable factors) and secondary prevention (preparing for quality care of a child with NSCL/P).

Although family history is a strong risk factor for NSCL/P, many OFCs (possibly 70 to 80% in the Caucasian population) occur in the absence of family history (Grosen, Bille, et al., 2010; M. L. Marazita, 2012). Furthermore, while family history allows only for empiric risk estimation, of the currently known risk factors, it appears to have the strongest effect on risk. Individuals with no family history are still at some level of risk to have a child with NSCL/P; we currently have no means of estimating that risk within the general population, which may make up the largest proportion of individuals who will go on to have a child with a CL/P. The ability to estimate risk for all individuals, regardless of family history of NSCL/P, may have greater public health impact. Thus, for the purposes of the present discussion, we will assume the proposed screening will take place in the primary care setting, with the target population being all individuals of reproductive age, regardless of family history. The goals of the proposed screening are primary prevention of NSCL/P in a future child and secondary prevention surrounding improvements in parental preparedness, coping, and improvements in timeliness and quality of care.

The type of screening to be used is also an important consideration. Implementing a model using risk factors that can be measured non-invasively, such as that described in section 3 (Manuscript) of this document, requires education and training of personnel to take precise physical measurements of the face, assess for VPD, and possibly take an ultrasound of the upper lip to visualize the OOM. This training could be difficult to standardize and would require valuable provider time. A genetic testing component, such as would be required for calculating a PRS,

requires working with laboratories and analysts who can perform the testing, and it is more invasive and expensive. A genetic test can also indirectly reveal genetic risk of other family members who may not wish to know that information (Haddow & Palomaki, 2011). Additionally, there are privacy and discrimination concerns around genetic testing; these concerns arose from the use of single-gene testing with high predictive value (in some cases, near certainty) for a condition such as cystic fibrosis (Haddow & Palomaki, 2011; Janssens & Khoury, 2006).

Janssen and Khoury (2006) point out several issues with extending the same implications of genetic testing for single gene disorders to genetic testing for multifactorial conditions such as NSCL/P. They reason that, since testing for multifactorial conditions (e.g. with PRSs) involves evaluating variants in many genes, each with a small effect on overall risk, there is less implication for family members of an individual who screens positive, partly because the likelihood of two relatives sharing the exact same level of risk (i.e. the same genetic profile of variants for a group of multiple genes) is far less likely than in a monogenic testing scenario. Further, they discuss that there is pleiotropy to consider, in which some variants have effects on risk for multiple conditions, sometimes being a risk factor for one and a protective factor for another; it is possible this could reduce the problem of insurance discrimination, as all individuals are likely to have both risk and protective factors for some conditions (Janssens & Khoury, 2006). Thus, it is possible that PRSs should be considered similar to non-genetic screens that evaluate risk markers. For the purposes of the present discussion, we will discuss screening via a non-genetic testing method, using phenotype and environmental risk factors only.

5.4.2 Clinical Utility

The ability for parents to prepare psychologically and seek information and resources is repeatedly touted as a potential benefit of preconception screening for OFCs. There is no known literature supporting these benefits; however, there is limited literature available on the impact of *prenatal* diagnosis of CL/P which may provide some insight. There is some support for improved parental preparedness with prenatal diagnosis, especially in regards to successfully feeding the newborn with an OFC (Hubbard et al., 2012; Robbins et al., 2010; Smedegaard et al., 2008). There is also an association with prenatal counseling on cleft care and lower rates of NICU admission and poor weight gain in newborns affected by CL/P (see section 2.4). Earlier evaluation with a cleft team has been associated with faster return to birthweight (Kaye et al., 2017).

However, in a population-based study of maternal satisfaction, team care, and treatment outcomes, timing of diagnosis did not have a significant impact (Robbins et al., 2010). Robbins et al. (2010) interviewed 253 mothers from Arkansas, Iowa, and New York about their level of satisfaction with the information and support provided about OFC at diagnosis, perception of the help they received with feeding at birth, perceived severity of the OFC, and satisfaction with their child's post-surgical appearance and speech. The survey also included questions about where the child received care (cleft team or not). About 28% of the infants were diagnosed prenatally (excluding CPO); family incomes of greater than $60,000 were associated with a significantly higher likelihood of receiving a prenatal diagnosis. Mothers who had prenatal diagnosis were significantly less satisfied with their child's appearance, but more likely to report a positive experience with support around feeding. There were no differences in treatment by a cleft team or the perceived quality of information they received. The study population was nearly all white, married mothers; most were privately insured (Robbins et al., 2010). These population factors are

important to consider along with the results, as individuals with lower income, of different races, and those with Medicaid or who are uninsured may have different experiences. No other studies of this kind were identified.

In terms of psychological impact, there is support for psychological distress in parents following diagnosis, including feelings of guilt, anxiety, fear, and sadness (Sreejith, Arun, Devarajan, Gopinath, & Sunil, 2018). However, the majority of parents across studies appear to desire and express satisfaction with prenatal diagnosis after the initial period of difficulty; most parents cite having time to adapt to the diagnosis and the ability to prepare other family members as benefits (Nusbaum et al., 2008; Sreejith et al., 2018). It is important to consider that the utility of prenatal diagnosis or preconception risk assessment may be affected by the quality of social support and counseling received, therefore it is critical to consider the availability of resources after diagnosis. Experience with the recent availability of expanded carrier screening has shown that there can be major issues with provider knowledge, provider and patient understanding and/or interest in benefits of preconception carrier screening, and the degree of appropriate post-test follow-up and counseling (Kraft, Duenas, Wilfond, & Goddard, 2019).

Since research on modifiable risk factors for OFCs is somewhat inconsistent, and those known appear to have a small effect, there is no current way to assure primary prevention of NSCL/P, even with knowledge of risk prior to pregnancy. With monogenic conditions, preimplantation genetic testing and selective embryo transfer based on genetic testing results may be available, but this is not yet an option for complex traits including the majority of OFCs. However, preconception counseling still allows couples the opportunity to avoid unwanted pregnancies, to adopt, and to engage in risk reduction steps with known small effects on risk, such

as multivitamin and folate supplementation, avoidance of certain medications, and other lifestyle changes.

Follow-up diagnostic testing in the event of a positive screen for NSCL/P would be accomplished through second-trimester prenatal ultrasound. While, technically speaking, OFCs can be detected in most cases with current technology (with the exception of most cases of CPO), it is important to consider that detection has been shown to vary widely by center, from less than 10%, up to 100% (Maarse et al., 2010). There also may be socioeconomic factors associated with the likelihood of prenatal diagnosis by ultrasound, with women in higher income levels being more likely to receive a prenatal diagnosis (Robbins et al., 2010). Note that a false negative on ultrasound after a positive preconception screen for NSCL/P could result in more harm than good, as parents would be initially distressed by the screen, then falsely reassured by the ultrasound, and then surprised by the birth of a child with a cleft.

Finally, the availability of treatment and/or benefit to earlier intervention has a strong bearing on the value of preconception screening for NSCL/P. Currently, there is no fetal intervention available during pregnancy, such as fetal surgery to correct the lip/palate. Surgical procedures to correct OFCs have been performed experimentally, and have shown promise in preventing scarring, reducing the number of procedures necessary to correct the cleft, and less disruption to the normal growth of the face, but are not currently considered worth the dangers posed to the mother and fetus (Papadopulos et al., 2005). The availability of in utero surgery could have significant bearing on the potential benefit of preconception screening, both in practice and for the possible impact on public interest in screening.

5.4.3 Ethical, Legal, and Social Implications

In considering the appropriateness of a screening program, understanding the social landscape is important. Among the many possible ELSI associated with screening for birth defects are the issues of legal and financial access to birth control and pregnancy termination services (Institute of Medicine (US) Committee on a Comprehensive Review of the HHS Office of Family Planning Title X Program, 2009; Pergament & Ilijic, 2014), concern surrounding harm to the disability community and stigma for parents who choose to have children with "preventable" conditions (Boardman & Hale, 2018), and issues of access and equity in cleft care (Nidey & Wehby, 2019).

Pregnancy termination due to an isolated OFC is a highly controversial topic surrounding prenatal diagnosis or potential screens for OFCs (Boardman & Hale, 2018). Concern over selective pregnancy termination has been seen in regards to other conditions, including Down syndrome, which has sparked outrage and conversations about eugenics; this has even resulted in policy changes in several states to attempt to legally restrict pregnancy terminations sought for the sole purpose of a Down syndrome diagnosis (Stirone, 2019). In the 1980s and early 1990s, research suggested that over 90% of pregnancies with Down syndrome were terminated; newer research in the past decade suggests that rates of termination have fallen below 70% (de Graaf, Buckley, & Skotko, 2015; Natoli, Ackerman, McDermott, & Edwards, 2012). Pregnancy terminations also appear to occur as a result of isolated birth defects (Svensson et al., 2014). Svensson et al. (2014) were interested in the rate of birth defects among pregnancies terminated, as those numbers are often left out of epidemiological studies, which may result in an underestimation of birth defect incidence. They found that 14% of pregnancies terminated in Denmark between 2007 and 2011

had birth defects compared with 4% of live born infants; 12% of fetuses with ear, face, and neck deformities were terminated.

Knowledge of risk prior to pregnancy, such as is suggested through preconception screening, can allow for avoidance of unwanted pregnancies, which has been shown to reduce elective termination rates (Institute of Medicine (US) Committee on a Comprehensive Review of the HHS Office of Family Planning Title X Program, 2009). However, preconception care and family planning is not available nor accessed by many women who could benefit, including a substantial proportion of women who are low-income, young (18 to 24 years), unmarried, or part of a racial or ethnic minority (Institute of Medicine (US) Committee on a Comprehensive Review of the HHS Office of Family Planning Title X Program, 2009). Of further note, studies have also suggested that Hispanic women account for fewer terminations of fetuses with Down syndrome than Caucasian, Asian, of African American women; the degree to which this may be due to lack of access to services or other healthcare disparities is unknown (Natoli et al., 2012).

Among all women, it is apparent that routine preconception care is not yet occurring at optimal levels in the U.S., despite the recognized importance and value. A study using information from several population-based surveys in the U.S. between 2011 and 2014 revealed that only about 33% of women who had given birth in the preceding two to nine months had talked about their health prior to their pregnancy; additionally, less than half of women of reproductive age and less than 5% of men were counseled on contraception (Pazol, 2017). It is evident that preconception care in the U.S. does not meet even basic needs in many cases; this is an important public health consideration in terms of introducing additional screening recommendations, as it is likely to be less than optimally implemented based on the observation of gaps in current practices.

Parents of children with disabilities often experience reduced self-esteem and worry over the judgement of others (Nidey, Moreno Uribe, Marazita, & Wehby, 2016). Regarding mental health specifically in individuals with NSCL/P, while those with orofacial clefts (OFCs) do not, on average, appear to experience profound mental health problems, there is substantial evidence to suggest some level of negative impact on quality of life (QOL) in areas such as socialization (especially in the middle school years), satisfaction with physical appearance (especially in females), and self-esteem (Herkrath et al., 2018; Herkrath et al., 2015; Mani et al., 2010). Adults with disabilities have varying views on selective reproduction (preconception testing and efforts to avoid having an affected child), often based on their lived experiences. Adults who have experienced severe illnesses, poor quality of life, or low social support may be more likely to support selective reproduction (Boardman & Hale, 2018). There is also potential for long-term impacts on social support, funding for treatments, and peer support in the disability community as a result of greater ability to prevent the birth of a child with a birth defect (Boardman & Hale, 2018).

Finally, there are increasingly obvious issues with access to quality and timely cleft care. Long travel times, issues with insurance coverage, and racial disparities have been recently noted, although studies on specific barriers are still needed to deepen understanding (Nidey & Wehby, 2019). A few key issues in access to care are likely to have a significant impact on the potential benefits of earlier or better care for children with NSCL/P as a result of preconception screening. First, based on the observation made by Nidey and Wehby (2019) that non-white children with OFCs experience longer delays prior to surgery, combined with the issue of racial and ethnic minorities being less likely to receive preconception care, screening has the potential to widen existing health disparities (Institute of Medicine (US) Committee on a Comprehensive Review of

the HHS Office of Family Planning Title X Program, 2009). That is, those most in need of improved access to services and information regarding cleft treatment may also be the least likely to benefit from screening. Secondly, there are issues of disparity depending on insurance coverage; those who are Medicaid-insured, for example, appear to experience issues with finding in-network coverage (e.g. dental care for children with an OFC) and face caps on reimbursement (Nidey & Wehby, 2019). This may mean that despite the availability of providers in a given geographical area (see Figures 9-11), patients may still experience barriers to care based on insurance coverage. It is clear that the current system available for treatment and follow-up for individuals with an OFC has room for significant public health improvement.

5.4.4 Economic Impact

The costs of care for individuals with NSCL/P can be up to eight times higher in the first ten years of life compared to those without NSCL/P due to required surgeries and therapies (Boulet et al., 2009). It is possible that with earlier access to quality cleft team care, NICU admission and feeding issues could be reduced, and the number of surgeries could be minimized to the extent possible; these changes may lead to reduced costs. However, if one of the goals is to ensure quality outcomes by improving care, which includes the use of a multidisciplinary team and all supportive therapies from infancy through adolescence (Cleft Palate-Craniofacial Association, 2009), it is possible that utilization of services, and therefore costs, could increase.

It is also important to consider the education and training of personnel to screen, as well as provider time required to successfully implement a screening program for NSCL/P. Finally, insurance coverage of testing, even for well-established carrier screening programs, can be problematic (Kraft et al., 2019), which may prove to be a financial barrier for patients. There is

insufficient literature to make more specific estimates of the economic impact of screening for NSCL/P, as one of the gaps in the current literature, is an understanding of healthcare use and costs associated with NSCL/P (Yazdy et al., 2007).

5.5 Geographical Distance and Access in Cleft Care

Several studies in the past decade have looked at barriers to effective patient care and follow-up in the OFC population (Cassell et al., 2013; Nidey & Wehby, 2019; Sharif-Askary et al., 2018). One practical concern that may create a barrier for patients is distance from a cleft center. Many studies have reported an association between patients who are lost to follow-up and distance to care; this phenomenon of greater distance and higher incidence of lost to follow-up is well-known in healthcare as distance decay (Sharif-Askary et al., 2018). One group of researchers found that the majority of parents report travel times of over one hour to get to facilities that provide cleft care (Cassell et al., 2013). The majority of respondents in that study (67%) reported that travel to care was not an issue, but there were significant associations, in addition to travel time and distance, with maternal education, the child's age, and cleft type (Cassell et al., 2013). The respondents in the study were 83% non-Hispanic white, the majority had college education, and there were no uninsured individuals. Of note, women with incomes below $50,000 were more than twice as likely to travel more that 60 miles to care; this could highlight an important disparity in OFC care for those less able to travel long distances due to financial constraints (Cassell et al., 2013).

Geographic distance has been proposed as a major barrier to access; however, most studies have been done in a relatively small geographic area (e.g. within a single state). Additional studies

are needed to fully understand the role of distance to care as a barrier to healthcare utilization in the OFC population across the U.S., as existing studies have also found that individuals within a short distance to a cleft center may still be at higher risk for loss to follow up based on lower socioeconomic status and shorter duration of the relationship with a cleft team (Sharif-Askary et al., 2018). The conflicting findings could be hinting at the importance of coordinated care, parental understanding of the importance of long-term follow-up after the initial surgery to repair an OFC is performed, and the need to build trusting relationships with parents, in addition to geographical factors (Nidey & Wehby, 2019; Sharif-Askary et al., 2018).

In order to visualize the coverage of cleft centers in the U.S., the maps in Figures 9-11 were created. These maps give an idea of coverage of ACPA approved cleft teams, in addition to other services and teams that provide cleft care. Coverage of cleft providers has been proposed as a factor in geographical barriers to optimal care (Nidey & Wehby, 2019). The maps are intended to allow for graphic visualization of the large gaps between centers in some areas of the U.S., however, it is not meant to be comprehensive of every provider of cleft care in the U.S.

5.5.1 Map Development Methods

Figures 9 and 10 (below) were created based on the cities in which there are cleft-craniofacial centers recognized by the American Cleft Palate-Craniofacial Association (ACPA) as approved teams (map created using free software available at www.mapchart.net). The ACPA has a voluntary application process for cleft-craniofacial teams who wish to be approved and listed on the ACPA website team list ("ACPA Approved Teams - ACPA Family Services," n.d.). The approved teams demonstrate ACPA's requirements for excellence in patient care by including providers of speech language pathology, surgery, and orthodontics, and access to recommended

supportive therapies (e.g. psychology, genetics, social work. There are also requirements of having a designated patient care coordinator, practices which are culturally competent, and additional qualities recommended for optimal cleft-craniofacial care based on peer-reviewed research (full requirements for team approval available at www.acpa-cpf.org). Figure 10 was created using Google Sheets.

Of note, no ACPA approved centers are located in Kentucky, North Dakota, Wyoming, or Maine (Figures 9-10). Further investigation via a Google search for "cleft lip and palate care" or "cleft lip and palate clinic" for each state revealed that each of these states has a unique state-run program for providing care to individuals with OFCs. These programs are mainly run by the states' departments of health ("Cleft Lip & Palate - Clinics - CSHN; Division of Disease Prevention - Maine CDC: DHHS Maine," n.d.; "Multidisciplinary Clinic Program: Special Health Services - North Dakota Health Dept. - Healthy & Safe Communities," n.d.; "Office for Children with Special Health Care Needs - Cabinet for Health and Family Services," n.d.), with the exception of in Wyoming, where the Wyoming Cleft Lip and Palate Foundation is made up of a team of multi-disciplinary providers based out of Cheyenne Oral and Maxillofacial Surgery ("Wyoming Cleft Lip and Palate Foundation," n.d.).

Figure 11 (below) includes centers recognized by the ACPA, in addition to the state-based programs and those found from additional web searches. A parent-run website called Cleftopedia ("Cleft Teams," n.d.), allows individuals to post about their experiences with providers of cleft care. The providers listed on this site were cross-checked using a Google search, and included as one of the following in figure 11: specialist (any provider with specialty training in craniofacial surgery), cleft-craniofacial center (not ACPA approved), oral and maxillofacial surgeon with some supportive services available on site for patients with OFCs, oral and maxillofacial surgeon with

no supportive services available on site, general plastic surgeon (no specialized training in cleft-craniofacial care), pediatric plastic surgeon, and the state of Wyoming's unique state-run program. Figure 11 was created using Google Maps.

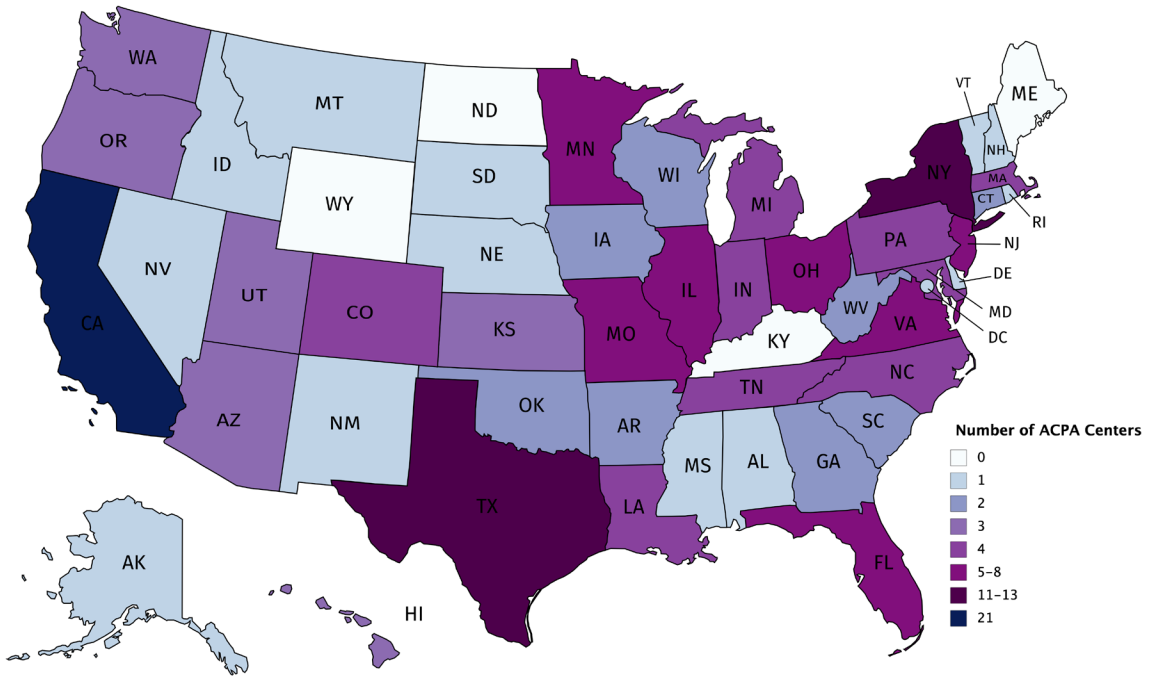

Figure 8 Density of ACPA Approved Cleft-Craniofacial Centers by State

This figure provides a visual depiction of the 174 cleft-craniofacial teams recognized by the ACPA. The colors range from white to purple, to deep blue; as the number of cleft teams per state increases, the color darkens

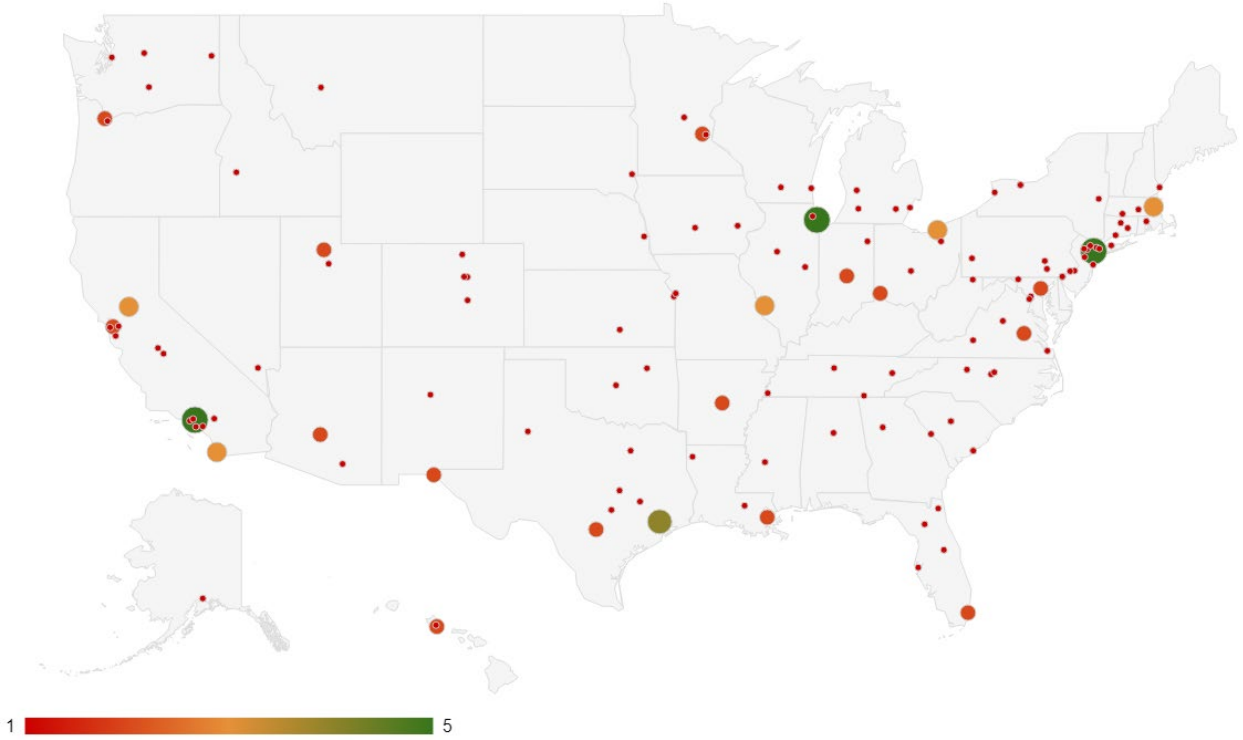

Figure 9 Number of ACPA Approved Cleft-Craniofacial Centers by City

This figure provides a visual depiction of the 174 cleft-craniofacial teams recognized by the ACPA. Larger dots indicate multiple teams within the same city. The scale describes the number of centers by both color and size. Smaller yellow to red dots represent 3 or fewer centers, while larger yellow to green dots represent 3 to 5 centers. Some smaller dots overlap larger ones due to centers which are close to, but not within, the same city

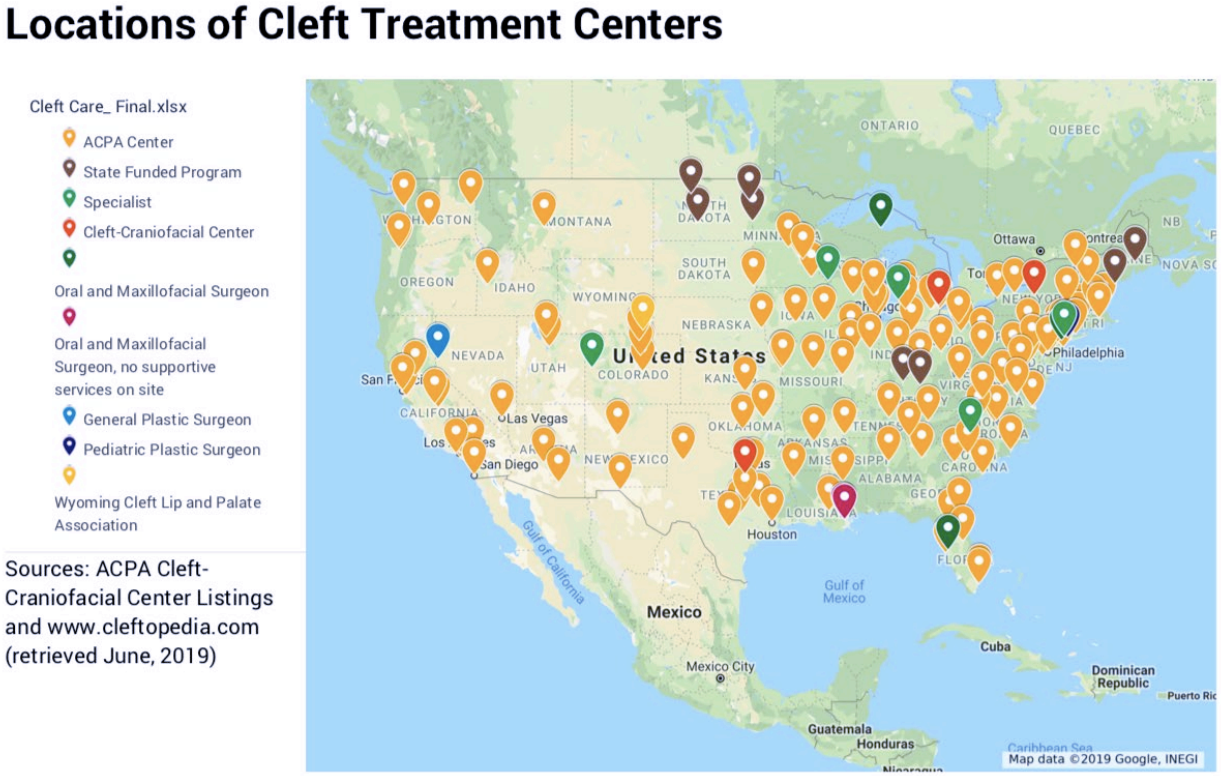

Figure 10 Locations of ACPA and Additional non-ACPA Cleft-Craniofacial Providers by City

An interactive version of this map is available online at: https://drive.google.com/open?id=1NSlYPghetW-0Te942RdKgGrO-_RHFJVO&usp=sharing. This map is meant only to provide an illustrative example of the coverage of major cleft-craniofacial centers, as well as several additional cleft-craniofacial healthcare providers across the U.S.; it is by no means an exhaustive list of all providers available. The ACPA website and a patient-run website called Cleftopedia were used to assemble this list of providers, in addition to other Google searches. The purpose is to provide context to the discussion of access and provide a visual depiction of the sometimes vast distances between major multidisciplinary cleft teams.

5.6 Conclusion and Discussion of Preconception NSCL/P Screening

There is insufficient evidence to support the implementation of a preconception screening test for NSCL/P. Review of the literature available does not support such screening for three main reasons. First, there are significant gaps in knowledge surrounding the benefits and harms of implementing screening for NSCL/P, a significant consideration according to Harris et al. surrounding the evidence needed to adequately evaluate a screening program (2011) (See Table 8). The original Wilson and Jungner criteria also encourage a full understanding of the condition's natural history, which is arguably incomplete for NSCL/P. Chiefly, known modifiable risk factors have a small to moderate effect on overall risk; additionally, studies of the effect of risk factors such as multivitamin use, folate supplementation, maternal/second-hand smoking, maternal alcohol use, diabetes, illness/fever, and medication use on NSCL/P are inconsistent (Bell et al., 2014; DeRoo et al., 2003; Herkrath et al., 2012; Hozyasz, 2010; Pi et al., 2018; Wehby & Murray, 2010). An understanding of these risk factors is essential for finding ways to reduce the risk for NSCL/P.

Aside from risk factors which are important for primary prevention goals, secondary prevention goals of this screening would be affected by the incomplete understanding of determinants of the physical and mental health of individuals with NSCL/P and their caregivers (Herkrath et al., 2018). Some known factors in quality of life include satisfaction with appearance, speech and behavioral therapy early interventions, sex, and possibly, coordinated, multidisciplinary care (Herkrath et al., 2012, 2012; Naros et al., 2018; Paganini et al., 2018). Understanding the determinants of health are important for designing effective public health

interventions in cleft care. There is no current evidence that earlier diagnosis – a goal of the proposed screening – facilitates earlier or higher quality care for parents and children with NSCL/P (Robbins et al., 2010). This last point is a major consideration surrounding the value of a screening program, according to Andermann et al. (2008), that is, that there should be scientific evidence of screening program effectiveness (See Table 8).

Second, there is potential to widen disparities in cleft care if screening is undertaken without a clear understanding of the current issues of equity in access to cleft care. A good screening program should be promote equity across the target population (Andermann et al., 2008). There are important differences in cleft care across individuals depending on their geographic location, insurance-coverage, socioeconomic status, and race/ethnicity which need further study to accurately plan for addressing disparities in access and outcomes (e.g. surgical outcomes) (Nidey & Wehby, 2019). For example, based on a recent population-based study of 456 infants born with birth defects between 1999 and 2007, American Indians and Alaskan Natives (AI/AN) have an incidence of CL/P that is about 70% higher than for non-Hispanic white individuals; AI/AN mothers of the infants born with CL/P were also significantly more likely to have diabetes (a proposed risk factor for NSCL/P (Dixon et al., 2011b)), while being less likely to have attained a high-school education (about 30% of mothers). Based on the characteristics of individuals who are more likely to receive preconception counseling and prenatal diagnosis, it appears there is a risk for further excluding those who are young, low-income, and who belong to a racial minority from screening and any benefits of such screening (Institute of Medicine (US) Committee on a Comprehensive Review of the HHS Office of Family Planning Title X Program, 2009; Robbins et al., 2010). Additionally, because studies of phenotypic risk factors have taken place largely in Caucasian populations (Seth M. Weinberg et al., 2016), and we currently have little knowledge of

whether phenotypic risks are the same across non-white populations, it is possible that models developed with current data sets would be less applicable to non-white individuals. Indeed, this has been seen in previous models, including the first version of the Gail model for breast cancer risk, which was applicable to white women (Gail et al., 1989); only recently has it been improved for use in other populations, such as the Asian population, but there are still concerns about the accuracy in non-white individuals compared to white individuals (Wang et al., 2018). It is important that efforts be made to include underrepresented in biomedical research groups, such as NA/AI, Hispanics, and African Americans, in model development.

Third, and perhaps most critically, in considering the setting and facilities in which the proposed screening for NSCL/P would take place (per Harris et al. (2011); see Table 8), a few major issues become evident. Preconception care does not appear to accomplish the guidelines set forth by national organizations in practice; many women are not having conversations with providers about maternal health and prevention of maternal and fetal complications before pregnancy (Pazol, 2017). An important part of evaluating a screening program, according to Harris et al. (2011), is to consider the practical conditions around implementation. If NSCL/P screening is implemented in the primary care setting as part of routine preconception care, it is unlikely to be more successful than other current preconception practices, which appear in need of public health resources for improvement. Of import, all of the currently known modifiable risk factors for NSCL/P are included as targets of preconception healthcare; thus, if preconception care was highly successful, NSCL/P occurrence/risk would likely be addressed without need for a particular screen. There are several additional benefits and harms which could be further discussed, but given these three main issues, there is sufficient concern that the initial step in screening evaluation, of determining the setting, target population, and goals of screening, do not set a strong enough

foundation for further consideration of implementing a preconception population screening program for NSCL/P.

In conclusion, while there could be future potential for NSCL/P preconception screening to positively impact the care of individuals with NSCL/P and their families, research to date is not robust enough to justify the most important benefits (primary and secondary prevention) of screening. Currently, there appear to be other important priorities in general preconception care, which arguably have the potential to have broader impact on maternal and infant health, while also addressing many of the modifiable factors in NSCL/P. Secondly, advocating for funds to educate community-based providers, develop better policies for health insurance reimbursement of cleft-related care, and determine additional factors behind cleft care disparities in access and equity could likely have a greater impact on the health and quality of life in the NSCL/P population.

Appendix A Table 1 NSCL/P GWAS References

1. Leslie EJ, Marazita ML. Genetics of Orofacial Cleft Birth Defects. In: Stoler JM, editor. Clinical Genetics. 3rd ed. New York: Springer Science; 2015. p. 118-26.

2. Ludwig KU, Mangold E, Herms S, Nowak S, Reutter H, Paul A, et al. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat Genet. 2012 Sep;44(9):968-71.

3. Beaty TH, Taub MA, Scott AF, Murray JC, Marazita ML, Schwender H, et al. Confirming genes influencing risk to cleft lip with/without cleft palate in a case-parent trio study. Hum Genet. 2013 Jul;132(7):771-81.

4. Leslie EJ, Carlson JC, Shaffer JR, Feingold E, Wehby G, Laurie CA, et al. A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. Hum Mol Genet. 2016 Mar 30.

5. Leslie EJ, Liu H, Carlson JC, Shaffer JR, Feingold E, Wehby G, et al. A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. Am J Hum Genet. 2016 Apr 7;98(4):744-54.

6. Mangold E, Bohmer AC, Ishorst N, Hoebel AK, Gultepe P, Schuenke H, et al. Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate. Am J Hum Genet. 2016 Apr 7;98(4):755-62.

7. Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, et al. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat Genet. 2010 Jun;42(6):525-9.

8. Marazita ML, Lidral AC, Murray JC, Field LL, Maher BS, Goldstein McHenry T, et al. Genome scan, fine-mapping, and candidate gene analysis of non-syndromic cleft lip with or without cleft palate reveals phenotype-specific differences in linkage and association results. Hum Hered. 2009;68(3):151-70.

9. Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet. 2009 Apr;41(4):473-7.

10. Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, Herms S, et al. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat Genet. 2010 Jan;42(1):24-6.

11. Yu Y, Zuo X, He M, Gao J, Fu Y, Qin C, et al. Genome-wide analyses of non-syndromic cleft lip with palate identify fourteen novel loci and genetic heterogeneity. Nat Commun. 2017;in press.

12. Beaty TH, Ruczinski I, Murray JC, Marazita ML, Munger RG, Hetmanski JB, et al. Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. Genet Epidemiol. 2011 Sep;35(6):469-78.

13. Grant SF, Wang K, Zhang H, Glaberson W, Annaiah K, Kim CE, et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J Pediatr. 2009 Dec;155(6):909-13.

14. Ludwig KU, Bohmer AC, Rubini M, Mossey PA, Herms S, Nowak S, et al. Strong association of variants around FOXE1 and orofacial clefting. J Dent Res. 2014 Apr;93(4):376-81.

15. Sun Y, Huang Y, Yin A, Pan Y, Wang Y, Wang C, et al. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun. 2015 Mar 16;6:6414.

16. Leslie EJ, Taub MA, Liu H, Steinberg KM, Koboldt DC, Zhang Q, et al. Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. Am J Hum Genet. 2015 Mar 5;96(3):397-411.

Appendix B Supplementary Evaluation Criteria Material

Table 9 Estimating Outcomes of a Screening Program

Consideration	Comments
Magnitude of Potential Benefits	
Probability of an adverse health outcome without screening	It is important to define and focus on the adverse health outcome that the screening program is attempting to reduce. One must also define the specific population that the program intends to screen.
Degree to which screening identifies all people who would suffer the adverse health outcome	The proper target of screening is to detect those people who will suffer the adverse health outcome. Detecting (labeling) people who will not suffer the adverse health outcome is not a benefit.
Magnitude of incremental health benefit of earlier versus later treatment resulting from screening	For a screening program to reduce morbidity or mortality from an adverse health outcome, earlier treatment (after screening detection) must provide more health benefit than later treatment (after clinical detection). It is important to estimate this incremental benefit in absolute terms.
Magnitude of Potential Harms	
Frequency of false-positive screening tests	The best estimate of the frequency of false-positive screening tests is the cumulative percentage of people screened who have at least one false-positive screening test over a period of time, such as 10 years.
Experience of people with false-positive results	A negative experience of people with a false-positive screening test may come from either physical (e.g., risk of complications from a colonoscopy after a positive fecal occult blood test) or psychological (e.g., short-term or long-term anxiety after a false-positive test) causes. Small or infrequent negative experiences, when experienced by many people, may add up to large harms for a population.
Frequency of overdiagnosis	The critical issue in defining overdiagnosis is whether earlier diagnosis (due to screening) compared with later diagnosis (due to clinical detection) leads to increased labeling, diagnostic evaluation, or treatment that has potential adverse effects on health.
Experience of people who are overdiagnosed	It is important to estimate, for the screened population, the absolute frequency and severity of the adverse effect on health due to increased labeling, diagnostic evaluation, or treatment resulting from earlier diagnosis.
Frequency and severity of harms of workup and treatment	Helps determine the harms of overdiagnosis (2b)a.
	For people found to have conditions that would lead to the adverse health outcome, harms from earlier workup or treatment means they will suffer these harms for a longer time, thus reducing net benefits.

(reuse with permission from Harris et al., 2011)

Appendix C Institutional Review Board Approval

University of Pittsburgh
Institutional Review Board

3500 Fifth Avenue
Pittsburgh, PA 15213
(412) 383-1480
(412) 383-1508 (fax)
http://www.irb.pitt.edu

Memorandum

To: Mary Marazita, PhD
From: Sue Beers, PhD
Date: 9/16/2018
IRB#: REN18080277 / IRB0405013
Subject: University of Pittsburgh: Coordinating Center for Oral-Facial Cleft Families: Phenotype and Genetics

Your renewal for the above referenced research study has received expedited review and approval from the Institutional Review Board under:

45 CFR 46.110.(7)

Please note the following information:

Approval Date: 9/14/2018
Expiration Date: 9/27/2019

Please note that it is the investigator's responsibility to report to the IRB any unanticipated problems involving risks to subjects or others [see 45 CFR 46.103(b)(5) and 21 CFR 56.108(b)]. Refer to the IRB Policy and Procedure Manual regarding the reporting requirements for unanticipated problems which include, but are not limited to, adverse events. If you have any questions about this process, please contact the Adverse Events Coordinator at 412-383-1480.

The protocol and consent forms, along with a brief progress report must be resubmitted at least **one month** prior to the renewal date noted above as required by FWA00006790 (University of Pittsburgh), FWA00006735 (University of Pittsburgh Medical Center), FWA00000600 (Children's Hospital of Pittsburgh), FWA00003567 (Magee-Womens Health Corporation), FWA00003338 (University of Pittsburgh Medical Center Cancer Institute).

Please be advised that your research study may be audited periodically by the University of Pittsburgh Research Conduct and Compliance Office.

University of Pittsburgh
Institutional Review Board

3500 Fifth Avenue
Pittsburgh, PA 15213
(412) 383-1480
(412) 383-1508 (fax)
http://www.irb.pitt.edu

Memorandum

To: Mary Marazita, PhD
From: IRB Office
Date: 4/30/2018
IRB#: REN18040174 / IRB0607057
Subject: Oral-Facial Cleft Families: Phenotype and Genetics: (Pittsburgh and Guatemala Sites)

Your renewal for the above referenced research study has received expedited review and approval from the Institutional Review Board under:

45 CFR 46.110.(9)

Please note the following information:

Approval Date: 4/30/2018
Expiration Date: 4/29/2019

Please note that it is the investigator's responsibility to report to the IRB any unanticipated problems involving risks to subjects or others [see 45 CFR 46.103(b)(5) and 21 CFR 56.108(b)]. Refer to the IRB Policy and Procedure Manual regarding the reporting requirements for unanticipated problems which include, but are not limited to, adverse events. If you have any questions about this process, please contact the Adverse Events Coordinator at 412-383-1480.

The protocol and consent forms, along with a brief progress report must be resubmitted at least **one month** prior to the renewal date noted above as required by FWA00006790 (University of Pittsburgh), FWA00006735 (University of Pittsburgh Medical Center), FWA00000600 (Children's Hospital of Pittsburgh), FWA00003567 (Magee-Womens Health Corporation), FWA00003338 (University of Pittsburgh Medical Center Cancer Institute).

Please be advised that your research study may be audited periodically by the University of Pittsburgh Research Conduct and Compliance Office.

Bibliography

ACPA Approved Teams - ACPA Family Services. (n.d.). Retrieved June 17, 2019, from https://cleftline.org/find-a-team/acpa-approved-teams-in-the-us-and-canada/

Agopian, A. J., Lupo, P. J., Tinker, S. C., Canfield, M. A., Mitchell, L. E., & National Birth Defects Prevention Study, the N. B. D. P. (2012). Working towards a risk prediction model for neural tube defects. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, *94*(3), 141–6. https://doi.org/10.1002/bdra.22883

American College of Obstetricians and Gynecologists. (2017). Carrier screening for genetic conditions. Committee Opinion No. 691. *Obstetrics & Gynecology*, *129*, e41–55.

American College of Obstetricians and Gynecologists. (2019). Prepregnancy counseling. ACOG Committee Opinion No. 762. *Obstetrics & Gynecology*, *133*, e78-79.

Andermann, A., Blancquaert, I., Beauchamp, S., & Déry, V. (2008). Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years Public health classics. *Bulletin of the World Health Organization*, *86*(4). https://doi.org/10.2471/BLT.07.050112

Austin, P. C., & Steyerberg, E. W. (2017). Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Statistical Methods in Medical Research*, *26*(2), 796–808. https://doi.org/10.1177/0962280214558972

Basha, M., Demeer, B., Revencu, N., Helaers, R., Theys, S., Bou Saba, S., … Vikkula, M. (2018). Whole exome sequencing identifies mutations in 10% of patients with familial non-syndromic cleft lip and/or palate in genes mutated in well-known syndromes. *Journal of Medical Genetics*, *55*(7), 449–458. https://doi.org/10.1136/jmedgenet-2017-105110

Beaty, T. H., Marazita, M. L., & Leslie, E. J. (2016). Genetic factors influencing risk to orofacial clefts: Today's challenges and tomorrow's opportunities. *F1000Research*, *5*, 2800. https://doi.org/10.12688/f1000research.9503.1

Becker, F., van El, C. G., Ibarreta, D., Zika, E., Hogarth, S., Borry, P., … Cornel, M. C. (2011). Genetic testing and common disorders in a public health framework: How to assess relevance and possibilities. Background Document to the ESHG recommendations on genetic testing and common disorders. *European Journal of Human Genetics: EJHG*, *19 Suppl 1*, S6-44. https://doi.org/10.1038/ejhg.2010.249

Bell, J. C., Raynes-Greenow, C., Turner, R. M., Bower, C., Nassar, N., & O'Leary, C. M. (2014). Maternal Alcohol Consumption during Pregnancy and the Risk of Orofacial Clefts in

Infants: A Systematic Review and Meta-Analysis. *Paediatric and Perinatal Epidemiology*, *28*(4), 322–332. https://doi.org/10.1111/ppe.12131

Berryhill, W. (2016). Otologic Concerns for Cleft Lip and Palate Patient. *Oral and Maxillofacial Surgery Clinics*, *28*(2), 177–179.

Bickham, R. S., Ranganathan, K., Wombacher, N. R., Shapiro, D. N., Carlozzi, N. E., Baker, M. K., … Waljee, J. F. (2017). Speech Perceptions and Health-Related Quality of Life Among Children With Cleft Lip and Palate. *Journal of Craniofacial Surgery*, *28*(5), 1264–1268. https://doi.org/10.1097/SCS.0000000000003718

Bille, C., Winther, J. F., Bautz, A., Murray, J. C., Olsen, J., & Christensen, K. (2005). Cancer Risk in Persons with Oral Cleft—A Population-based Study of 8,093 Cases. *American Journal of Epidemiology*, *161*(11), 1047–1055. https://doi.org/10.1093/aje/kwi132

Boardman, F. K., & Hale, R. (2018). How do genetically disabled adults view selective reproduction? Impairment, identity, and genetic screening. *Molecular Genetics & Genomic Medicine*, *6*(6), 941–956. https://doi.org/10.1002/mgg3.463

Botkin, J. R. (2009). Assessing the New Criteria for Newborn Screening. *Health Matrix*, *19*(163). Retrieved from https://scholarlycommons.law.case.edu/healthmatrix/vol19/iss1/9

Boulet, S. L., Grosse, S. D., Honein, M. A., & Correa-Villaseñor, A. (2009). *Children with Orofacial Clefts: Health-Care Use and Costs Among a Privately Insured Population* (p. 447). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2663882/pdf/phr124000447.pdf

Boyce, J. O., Raj, S., Sanchez, K., Marazita, M. L., Morgan, A. T., & Kilpatrick, N. (2019). Speech Phenotyping in Unaffected Family Members of Individuals With Nonsyndromic Cleft Lip With or Without Palate. *The Cleft Palate-Craniofacial Journal*, 1055665618823936. https://doi.org/10.1177/1055665618823936

Cassell, C. H., Krohmer, A., Mendez, D. D., Lee, K. A., Strauss, R. P., & Meyer, R. E. (2013). Factors Associated with Distance and Time Traveled to Cleft and Craniofacial Care. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, *97*(10), 685–695. https://doi.org/10.1002/bdra.23173

Cassell, C. H., Meyer, R., & Daniels, J. (2008). Health care expenditures among Medicaid enrolled children with and without orofacial clefts in North Carolina, 1995–2002. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *82*(11), 785–794. https://doi.org/10.1002/bdra.20522

CDC. (2018a, April 30). Data & Statistics on Birth Defects | CDC. Retrieved June 5, 2019, from Centers for Disease Control and Prevention website: https://www.cdc.gov/ncbddd/birthdefects/data.html

CDC. (2018b, September 13). Data and Statistics | Spina Bifida | NCBDDD | CDC. Retrieved June 3, 2019, from Centers for Disease Control and Prevention website: https://www.cdc.gov/ncbddd/spinabifida/data.html

Cedergren, M., & Källén, B. (2005). Maternal Obesity and the Risk for Orofacial Clefts in the Offspring. *The Cleft Palate - Craniofacial Journal; Lawrence*, *42*(4), 367–371.

Chernus, J., Roosenboom, J., Ford, M., Lee, M. K., Emanuele, B., Anderton, J., … Marazita, M. L. (2018). GWAS reveals loci associated with velopharyngeal dysfunction. *Scientific Reports*, *8*(1), 8470. https://doi.org/10.1038/s41598-018-26880-w

Chollet, M. B., DeLeon, V. B., Conrad, A. L., & Nopoulos, P. (2014). Morphometric analysis of brain shape in children with nonsyndromic cleft lip and/or palate. *Journal of Child Neurology*, *29*(12), 1616–1625. https://doi.org/10.1177/0883073813510603

Cleft Lip & Palate - Clinics - CSHN; Division of Disease Prevention - Maine CDC: DHHS Maine. (n.d.). Retrieved June 18, 2019, from https://www.maine.gov/dhhs/mecdc/population-health/mch/cshn/cleftlip-palate/clp-clinic.html

Cleft Palate-Craniofacial Association, A. (2009). *Parameters for Evaluation and Treatment of Patients with Cleft Lip/Palate or Other Craniofacial Anomalies*. Retrieved from www.acpa-cpf.org

Cleft Teams. (n.d.). Retrieved June 18, 2019, from Cleftopedia website: http://www.cleftopedia.com/cleft-teams/

Coles, C. (1994). Critical Periods for Prenatal Alcohol Exposure. *RESEARCH WORLD*, *18*(1), 8.

Collins, A., Arias, L., Pengelly, R., Mart﷽nez, J., Brice﷽o, I., & Ennis, S. (2013). The potential for next generation sequencing to characterise the genetic variation underlying nonsyndromic cleft lip and palate phenotypes. *OA Genetics*, *1*(1). https://doi.org/10.13172/2054-197X-1-1-987

Conway, J. C., Taub, P. J., Kling, R., Oberoi, K., Doucette, J., & Jabs, E. W. (2015). Ten-year experience of more than 35,000 orofacial clefts in Africa. *BMC Pediatrics*, *15*. https://doi.org/10.1186/s12887-015-0328-5

Copp, A. J., Adzick, N. S., Chitty, L. S., Fletcher, J. M., Holmbeck, G. N., & Shaw, G. M. (2015). Spina Bifida. *Nature Reviews. Disease Primers*, *1*, 15007. https://doi.org/10.1038/nrdp.2015.7

Corrêa de Queiroz Herkrath, A. P., Herkrath, F. J., Bessa Rebelo, M. A., & Vettore, M. V. (2018). Determinants of Health-Related and Oral Health–Related Quality of Life in Adults With Orofacial Clefts: A Cross-Sectional Study. *The Cleft Palate-Craniofacial Journal*, *55*(9), 1244–1257. https://doi.org/10.1177/1055665618763377

Crawford, F. C., & Sofaer, J. A. (1987). Cleft lip with or without cleft palate: Identification of sporadic cases with a high level of genetic predisposition. *Journal of Medical Genetics*, *24*(3), 163–169. https://doi.org/10.1136/jmg.24.3.163

Crider, K. S., Bailey, L. B., & Berry, R. J. (2011). Folic Acid Food Fortification—Its History, Effect, Concerns, and Future Directions. *Nutrients*, *3*(3), 370–384. https://doi.org/10.3390/nu3030370

de Graaf, G., Buckley, F., & Skotko, B. G. (2015). Estimates of the live births, natural losses, and elective terminations with Down syndrome in the United States. *American Journal of Medical Genetics Part A*, *167*(4), 756–767. https://doi.org/10.1002/ajmg.a.37001

DeRoo, L. A., Gaudino, J. A., & Edmonds, L. D. (2003). Orofacial cleft malformations: Associations with maternal and infant characteristics in Washington State. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *67*(9), 637–642. https://doi.org/10.1002/bdra.10114

Dixon, M. J., Marazita, M. L., Beaty, T. H., & Murray, J. C. (2011a). Cleft lip and palate: Synthesizing genetic and environmental influences. *Nature Reviews. Genetics*, *12*(3), 167–178. https://doi.org/10.1038/nrg2933

Dixon, M. J., Marazita, M. L., Beaty, T. H., & Murray, J. C. (2011b). Cleft lip and palate: understanding genetic and environmental influences. *Nature Reviews. Genetics*, *12*(3), 167–78. https://doi.org/10.1038/nrg2933

Dongen, S. V. (2006). Fluctuating asymmetry and developmental instability in evolutionary biology: Past, present and future. *Journal of Evolutionary Biology*, *19*(6), 1727–1743. https://doi.org/10.1111/j.1420-9101.2006.01175.x

Euhus, D. M. (2001). Understanding Mathematical Models for Breast Cancer Risk Assessment and Counseling. *The Breast Journal*, *7*(4), 224–232. https://doi.org/10.1046/j.1524-4741.2001.20012.x

Farahi, N., & Zolotor, A. (2013). Recommendations for Preconception Counseling and Care. *American Family Physician*, *88*(8), 499–506.

Freda, M. C., Moos, M.-K., & Curtis, M. (2006). The History of Preconception Care: Evolving Guidelines and Standards. *Maternal and Child Health Journal*, *10*(Suppl 1), 43–52. https://doi.org/10.1007/s10995-006-0087-x

Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C., & Mulvihill, J. J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*, *81*(24), 1879–1886. https://doi.org/10.1093/jnci/81.24.1879

Gatti, G. L., Freda, N., Giacomina, A., Montemagni, M., & Sisti, A. (2017). Cleft Lip and Palate Repair. *The Journal of Craniofacial Surgery*, *28*(8), 1918–1924. https://doi.org/10.1097/SCS.0000000000003820

Glener, A. D., Allori, A. C., Shammas, R. L., Carlson, A. R., Pien, I. J., Aylsworth, A. S., … Marcus, J. R. (2017). A Population-Based Exploration of the Social Implications Associated with Cleft Lip and/or Palate. *Plastic and Reconstructive Surgery. Global Open*, *5*(6), e1373. https://doi.org/10.1097/GOX.0000000000001373

Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: A problem hiding in plain sight. *BMJ (Clinical Research Ed.)*, *352*, i1981. https://doi.org/10.1136/bmj.i1981

Grosen, D., Bille, C., Pedersen, J. K., Skytthe, A., Murray, J. C., & Christensen, K. (2010). Recurrence risk for offspring of twins discordant for oral cleft: a population-based cohort study of the Danish 1936-2004 cleft twin cohort. *American Journal of Medical Genetics. Part A*, *152A*(10), 2468–74. https://doi.org/10.1002/ajmg.a.33608

Grosen, D., Bille, C., Petersen, I., Skytthe, A., Hjelmborg, J. von B., Pedersen, J. K., … Christensen, K. (2011). Risk of oral clefts in twins. *Epidemiology (Cambridge, Mass.)*, *22*(3), 313–9. https://doi.org/10.1097/EDE.0b013e3182125f9c

Grosen, D., Chevrier, C., Skytthe, A., Bille, C., Mølsted, K., Sivertsen, A., … Christensen, K. (2010). A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. *Journal of Medical Genetics*, *47*(3), 162–8. https://doi.org/10.1136/jmg.2009.069385

Gurovich, Y., Hanani, Y., Bar, O., Nadav, G., Fleischer, N., Gelbman, D., … Gripp, K. W. (2019). Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine*, *25*(1), 60. https://doi.org/10.1038/s41591-018-0279-0

Haddow, J. E., & Palomaki, G. E. (2011). *An Introduction to Assessing Genomic Screening and Diagnostic Tests*. *46*(4), 7.

Hamlet, C., & Harcourt, D. (2015). Older Adults' Experiences of Living With Cleft Lip and Palate: A Qualitative Study Exploring Aging and Appearance. *The Cleft Palate - Craniofacial Journal; Lawrence*, *52*(2), E32–E40.

Harris, R., Sawaya, G. F., Moyer, V. A., & Calonge, N. (2011). Reconsidering the Criteria for Evaluating Prpopsed Screening Programs: Reflections from 4 Current and Former Members of the U.S. Preventative Services Task Force. *Epidemiologic Reviews*, *33*, 20–35.

Herkrath, A. P. C. de Q., Herkrath, F. J., Rebelo, M. A. B., & Vettore, M. V. (2012). Parental age as a risk factor for non-syndromic oral clefts: A meta-analysis. *Journal of Dentistry*, *40*(1), 3–14. https://doi.org/10.1016/j.jdent.2011.10.002

Herkrath, A. P. C., Herkrath, F. J., Rebelo, M. A. B., & Vettore, M. V. (2015). Measurement of Health-Related and Oral Health–Related Quality of Life among Individuals with Nonsyndromic Orofacial Clefts: A Systematic Review and Meta-Analysis. *The Cleft Palate-Craniofacial Journal*, *52*(2), 157–172. https://doi.org/10.1597/13-104

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (Second). John Wiley & Sons, Inc.

Howe, B. J., Cooper, M. E., Vieira, A. R., Weinberg, S. M., Resick, J. M., Nidey, N. L., … Moreno Uribe, L. M. (2015). Spectrum of Dental Phenotypes in Nonsyndromic Orofacial Clefting. *Journal of Dental Research*, *94*(7), 905–12. https://doi.org/10.1177/0022034515588281

Howe, L. J., Lee, M. K., Sharp, G. C., Smith, G. D., Pourcain, B. S., Shaffer, J. R., … Lewis, S. J. (2018). Investigating the shared genetics of non-syndromic cleft lip/palate and facial morphology. *PLOS Genetics*, *14*(8), e1007501. https://doi.org/10.1371/journal.pgen.1007501

Hozyasz, K. K. (2010). The search for risk factors that contribute to the etiology of non-syndromic cleft lip with or without cleft palate (CL/P) in the Polish population. *Pediatria Polska*, *85*(6), 609–623. https://doi.org/10.1016/S0031-3939(10)70562-X

Hubbard, B. A., Baker, C. L., & Muzaffar, A. R. (2012). Prenatal Counseling's Effect on Rates of Neonatal Intensive Care Admission for Feeding Problems Cleft Lip/Palate Infants. *Missouri Medicine*, *109*(2), 153–156.

Institute of Medicine (US) Committee on a Comprehensive Review of the HHS Office of Family Planning Title X Program. (2009). *A Review of the HHS Family Planning Program: Mission, Management, and Measurement of Results.* (A. S. Butler & E. W. Clayton, Eds.). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK215219/

Institute of Medicine (US), & National Research Council (US) and Institute of Medicine (US) Board on Children, Youth, and Families. (1999). 2, Public Health Screening Programs. In M. A. Stoto, D. A. Almario, & M. C. McCormick (Eds.), *Reducing the Odds: Preventing Perinatal Transmission of HIV In The United States*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK230552/

IPDTOC. (2011). Prevalence at Birth of Cleft Lip With or Without Cleft Palate: Data From the International Perinatal Database of Typical Oral Clefts (IPDTOC). *The Cleft Palate - Craniofacial Journal; Lawrence*, *48*(1), 66–81.

Janssens, A. C. J. W., & Khoury, M. J. (2006). Predictive value of testing for multiple genetic variants in multifactorial diseases: Implications for the discourse on ethical, legal and social issues. *Italian Journal of Public Health*, *3*(3–4). https://doi.org/10.2427/5915

Jiang, R., Bush, J. O., & Lidral, A. C. (2006). Development of the upper lip: morphogenetic and molecular mechanisms. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, *235*(5), 1152–66. https://doi.org/10.1002/dvdy.20646

Johnson, K., Posner, S., Biermann, J., Cordero, J. F., Atrash, H. K., Parker, C. S., … Curtis, M. G. (2006). *Recommendations to Improve Preconception Health and Health Care --- United States: A Report of the CDC/ATSDR Preconception Care Work Group and the Select Panel on Preconception Care* (Morbidity and Mortality Weekly Report No. 55(RR06); pp. 1–

23). Retrieved from CDC website: https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5506a1.htm

Kaye, A., & Lybrand, S. (2016a). *Social Work in Health Care The cleft team social worker*. https://doi.org/10.1080/00981389.2016.1145168org/10.1080/00981389.2016.1145168

Kaye, A., & Lybrand, S. (2016b). The cleft team social worker. *Social Work in Health Care*, *55*(4), 280–295. https://doi.org/10.1080/00981389.2016.1145168

Kaye, A., Thaete, K., Snell, A., Chesser, C., Goldak, C., & Huff, H. (2017). Initial Nutritional Assessment of Infants With Cleft Lip and/or Palate: Interventions and Return to Birth Weight. *The Cleft Palate - Craniofacial Journal; Lawrence*, *54*(2), 127–136. http://dx.doi.org/10.1597/15-163

Klotz, C. M., Wang, X., Desensi, R. S., Grubs, R. E., Costello, B. J., & Marazita, M. L. (2010). Revisiting the recurrence risk of nonsyndromic cleft lip with or without cleft palate. *American Journal of Medical Genetics. Part A*, *152A*(11), 2697–702. https://doi.org/10.1002/ajmg.a.33695

Kraft, S. A., Duenas, D., Wilfond, B. S., & Goddard, K. A. B. (2019). The evolving landscape of expanded carrier screening: Challenges and opportunities. *Genetics in Medicine*, *21*(4), 790. https://doi.org/10.1038/s41436-018-0273-4

Kurisu, K., Mazaheri, M., Niswander, J, & Johnston, M. (1974). Facial Morphology as an Indicator of Genetic Predisposition to Cleft Lip and Palate. *American Journal of Medical Genetics*, *26*, 702–714.

Langlois, S., Benn, P., & Wilkins-Haug, L. (2015). Current controversies in prenatal diagnosis 4: Pre-conception expanded carrier screening should replace all current prenatal screening for specific single gene disorders. *Prenatal Diagnosis*, *35*(1), 23–28. https://doi.org/10.1002/pd.4532

Lee, Y., Bang, H., & Kim, D. J. (2016). How to Establish Clinical Prediction Models. *Endocrinology and Metabolism*. https://doi.org/10.3803/EnM.2016.31.1.38

Lees, M. (2008). Familial risks of oral clefts. *BMJ : British Medical Journal*, *336*(7641), 399. https://doi.org/10.1136/bmj.39470.657685.80

Leslie, E. J., Carlson, J. C., Cooper, M. E., Christensen, K., Weinberg, S. M., & Marazita, M. L. (2017). Exploring Subclinical Phenotypic Features in Twin Pairs Discordant for Cleft Lip and Palate. *The Cleft Palate-Craniofacial Journal : Official Publication of the American Cleft Palate-Craniofacial Association*, *54*(1), 90–93. https://doi.org/10.1597/15-190

Leslie, E. J., & Marazita, M. L. (2013). Genetics of cleft lip and cleft palate. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, *163C*(4), 246–58. https://doi.org/10.1002/ajmg.c.31381

Leslie, E. J., Taub, M. A., Liu, H., Steinberg, K. M., Koboldt, D. C., Zhang, Q., … Murray, J. C. (2015). Identification of Functional Variants for Cleft Lip with or without Cleft Palate in or near PAX7, FGFR2, and NOG by Targeted Sequencing of GWAS Loci. *American Journal of Human Genetics*, *96*(3), 397–411. https://doi.org/10.1016/j.ajhg.2015.01.004

Lewis, C. W., Jacob, L. S., Lehmann, C. U., & Health, S. on O. (2017). The Primary Care Pediatrician and the Care of Children With Cleft Lip and/or Cleft Palate. *Pediatrics*, *139*(5), e20170628. https://doi.org/10.1542/peds.2017-0628

Li, H., Luo, M., Luo, J., Zheng, J., Zeng, R., Du, Q., … Ouyang, N. (2016). A discriminant analysis prediction model of non-syndromic cleft lip with or without cleft palate based on risk factors. *BMC Pregnancy and Childbirth*, *16*(1), 368. https://doi.org/10.1186/s12884-016-1116-4

Livshits, G., Kobyliansky, E., Opitz, J. M., & Reynolds, J. F. (1987). Dermatoglyphic traits as possible markers of developmental processes in humans. *American Journal of Medical Genetics*, *26*(1), 111–122. https://doi.org/10.1002/ajmg.1320260118

Lu, M. C. (2007). Recommendations for Preconception Care - American Family Physician. *Am Fam Physician*, *76*(3), 397–400.

Ludwig, K. U., Böhmer, A. C., Bowes, J., Nikolić, M., Ishorst, N., Wyatt, N., … Mangold, E. (2017). Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip ± cleft palate and cleft palate only. *Human Molecular Genetics*, *26*(4), 829–842. https://doi.org/10.1093/hmg/ddx012

Luo, Y., Li, Z., Guo, H., Cao, H., Song, C., Guo, X., & Zhang, Y. (2017). Predicting congenital heart defects: A comparison of three data mining methods. *PLoS ONE*, *12*(5). https://doi.org/10.1371/journal.pone.0177811

Maarse, W., Bergé, S. J., Pistorius, L., van Barneveld, T., Kon, M., Breugem, C., & Mink van der Molen, A. B. (2010). Diagnostic accuracy of transabdominal ultrasound in detecting prenatal cleft lip and palate: A systematic review. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, *35*(4), 495–502. https://doi.org/10.1002/uog.7472

MacLean, J. E., Tan, S., Fitzgerald, D. A., & Waters, K. A. (2013). Assessing ventilatory control in infants at high risk of sleep disordered breathing: A study of infants with cleft lip and/or palate. *Pediatric Pulmonology*, *48*(3), 265–273. https://doi.org/10.1002/ppul.22568

Mai, C. T., Cassell, C. H., Meyer, R. E., Isenburg, J., Canfield, M. A., Rickard, R., … Kirby, R. S. (2014). Birth defects data from population-based birth defects surveillance programs in the United States, 2007 to 2011: Highlighting orofacial clefts. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *100*(11), 895–904. https://doi.org/10.1002/bdra.23329

Mani, M., Carlsson, M., & Marcusson, A. (2010). Quality of Life Varies With Gender and Age Among Adults Treated for Unilateral Cleft Lip and Palate. *The Cleft Palate - Craniofacial Journal; Lawrence*, *47*(5), 491–498.

Marazita, M. (2007). Subclinical features in non-syndromic cleft lip with or without cleft palate (CL/P): review of the evidence that subepithelial orbicularis oris muscle defects are part of an expanded phenotype for CL/P. *Orthodontics & Craniofacial Research*, *10*(2), 82–87. https://doi.org/10.1111/j.1601-6343.2007.00386.x

Marazita, M. L. (2012). The Evolution of Human Genetic Studies of Cleft Lip and Cleft Palate. *Annual Review of Genomics and Human Genetics*, *13*, 263–283. https://doi.org/10.1146/annurev-genom-090711-163729

Martin, R. A., Hunter, V., Neufeld-Kaiser, W., Flodman, P., Spence, M. A., Furnas, D., & Martin, K. A. (2000). Ultrasonographic detection of orbicularis oris defects in first degree relatives of isolated cleft lip patients. *American Journal of Medical Genetics*, *90*(2), 155–161. https://doi.org/10.1002/(SICI)1096-8628(20000117)90:2<155::AID-AJMG13>3.0.CO;2-V

Martin, R. A., Jones, K. L., & Benirschke, K. (1993). Extension of the cleft lip phenotype: The subepithelial cleft. *American Journal of Medical Genetics*, *47*(5), 744–747. https://doi.org/10.1002/ajmg.1320470529

May, M. A., Sanchez, C. A., Deleyiannis, F. W. B., Marazita, M. L., & Weinberg, S. M. (2015). Evidence of Olfactory Deficits as Part of the Phenotypic Spectrum of Nonsyndromic Orofacial Clefting: *The Journal of Craniofacial Surgery*, *26*(1), 84–86. https://doi.org/10.1097/SCS.0000000000001242

McAllister, M., Dunn, G., Payne, K., Davies, L., & Todd, C. (2012). Patient empowerment: The need to consider it as a measurable patient-reported outcome for chronic conditions. *BMC Health Services Research*, *12*, 157. https://doi.org/10.1186/1472-6963-12-157

McIntyre, J. K., Sethi, H., Schönbrunner, A., Proudfoot, J., Jones, M., & Gosman, A. (2016). Number of Surgical Procedures for Patients With Cleft lip and Palate From Birth to 21 Years Old at a Single Children's Hospital. *Annals of Plastic Surgery*, *76 Suppl 3*, S205-208. https://doi.org/10.1097/SAP.0000000000000765

Millacura, N., Pardo, R., Cifuentes, L., & Suazo, J. (2017). Effects of folic acid fortification on orofacial clefts prevalence: A meta-analysis. *Public Health Nutrition*, *20*(12), 2260–2268. https://doi.org/10.1017/S1368980017000878

Mossey, P. A., Arngrimsson, R., Mccoll, J., Vintiner, G. M., & Connor, J. M. (1998). *Prediction of liability to orofacial clefting using genetic and craniofacial data from parents* (pp. 371–378). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1051310/pdf/jmedgene00234-0019.pdf

Mossey, P. A., Little, J., Munger, R. G., Dixon, M. J., & Shaw, W. C. (2009). *Cleft lip and palate*. *374*, 13.

Multidisciplinary Clinic Program: Special Health Services - North Dakota Health Dept. - Healthy & Safe Communities. (n.d.). Retrieved June 18, 2019, from https://www.ndhealth.gov/cshs/clinic.htm

Muntz, H., Wilson, M., Park, A., Smith, M., & Grimmer, J. F. (2008). Sleep Disordered Breathing and Obstructive Sleep Apnea in the Cleft Population. *The Laryngoscope*, *118*(2), 348–353. https://doi.org/10.1097/MLG.0b013e318158195e

Murray, L., Hentges, F., Hill, J., Karpf, J., Mistry, B., Kreutz, M., … Goodacre, T. (2007). The effect of cleft lip and palate, and the timing of lip repair on mother–infant interactions and infant development. *Journal of Child Psychology and Psychiatry*, *0*(0), 071103060106002–??? https://doi.org/10.1111/j.1469-7610.2007.01833.x

Nahai, F. R., Williams, J. K., Burstein, F. D., Martin, J., & Thomas, J. (2005). The Management of Cleft Lip and Palate: Pathways for Treatment and Longitudinal Assessment. *Seminars in Plastic Surgery*, *19*(4), 275–285. https://doi.org/10.1055/s-2005-925900

Naros, A., Brocks, A., Kluba, S., Reinert, S., & Krimmel, M. (2018). Health-related quality of life in cleft lip and/or palate patients – A cross-sectional study from preschool age until adolescence. *Journal of Cranio-Maxillofacial Surgery*, *46*(10), 1758–1763. https://doi.org/10.1016/j.jcms.2018.07.004

Natoli, J. L., Ackerman, D. L., McDermott, S., & Edwards, J. G. (2012). Prenatal diagnosis of Down syndrome: A systematic review of termination rates (1995–2011). *Prenatal Diagnosis*, *32*(2), 142–153. https://doi.org/10.1002/pd.2910

Neiswanger, K, Cooper, M., Weinberg, S., Flodman, P., Keglovits, A. B., Liu, Y., … Marazita, M. (2002). Cleft lip with or without cleft palate and dermatoglyphic asymmetry: Evaluation of a Chinese population. *Orthodontics and Craniofacial Research*, *5*(3), 140–146. https://doi.org/10.1034/j.1600-0544.2002.02196.x

Neiswanger, Katherine, Chirigos, K. W., Klotz, C. M., Cooper, M. E., Bardi, K. M., Brandon, C. A., … Marazita, M. L. (2009). Whorl patterns on the lower lip are associated with nonsyndromic cleft lip with or without cleft palate. *American Journal of Medical Genetics Part A*, *149A*(12), 2673–2679. https://doi.org/10.1002/ajmg.a.33089

Neiswanger, Katherine, Weinberg, S. M., Rogers, C. R., Brandon, C. A., Cooper, M. E., Bardi, K. M., … Marazita, M. L. (2007). Orbicularis oris muscle defects as an expanded phenotypic feature in nonsyndromic cleft lip with or without cleft palate. *American Journal of Medical Genetics Part A*, *143A*(11), 1143–1149. https://doi.org/10.1002/ajmg.a.31760

Nidey, N., Moreno Uribe, L. M., Marazita, M. M., & Wehby, G. L. (2016). Psychosocial Wellbeing of Parents of Children with Oral Clefts. *Child: Care, Health and Development*, *42*(1), 42–50. https://doi.org/10.1111/cch.12276

Nidey, N., & Wehby, G. L. (2019). Barriers to Health Care for Children with Orofacial Clefts: A Systematic Literature Review and Recommendations for Research Priorities || Oral Health and Dental Studies || Rivera Publications. *Oral Health and Dental Studies*, *2*(1). Retrieved

from http://riverapublications.com/article/barriers-to-health-care-for-children-with-orofacial-clefts-a-systematic-literature-review-and-recommendations-for-research-priorities

Nusbaum, R., Grubs, R. E., Losee, J. E., Weidman, C., Ford, M. D., & Marazita, M. L. (2008). A Qualitative Description of Receiving a Diagnosis of Clefting in the Prenatal or Postnatal Period. *Journal of Genetic Counseling*, *17*(4), 336–350. https://doi.org/10.1007/s10897-008-9152-5

Office for Children with Special Health Care Needs - Cabinet for Health and Family Services. (n.d.). Retrieved June 18, 2019, from https://chfs.ky.gov/agencies/ccshcn/Pages/default.aspx

Paganini, A., Hörfelt, C., & Mark, H. (2018). Gender differences in surgical treatment of patients with cleft lip and palate. *Journal of Plastic Surgery and Hand Surgery*, *52*(2), 106–110. https://doi.org/10.1080/2000656X.2017.1348951

Papadopulos, N. A., Papadopoulos, M. A., Kovacs, L., Zeilhofer, H. F., Henke, J., Boettcher, P., & Biemer, E. (2005). Foetal surgery and cleft lip and palate: Current status and new perspectives. *British Journal of Plastic Surgery*, *58*(5), 593–607. https://doi.org/10.1016/j.bjps.2005.01.014

Pazol, K. (2017). Receipt of Selected Preventive Health Services for Women and Men of Reproductive Age — United States, 2011–2013. *MMWR. Surveillance Summaries*, *66*. https://doi.org/10.15585/mmwr.ss6620a1

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, *96*(1), 3–14. https://doi.org/10.1080/00220670209598786

Pengelly, R. J., Arias, L., Martínez, J., Upstill-Goddard, R., Seaby, E. G., Gibson, J., … Briceño, I. (2016). Deleterious coding variants in multi-case families with non-syndromic cleft lip and/or palate phenotypes. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep30457

Pergament, D., & Ilijic, K. (2014). The Legal Past, Present and Future of Prenatal Genetic Testing: Professional Liability and Other Legal Challenges Affecting Patient Access to Services. *Journal of Clinical Medicine*, *3*(4), 1437–1465. https://doi.org/10.3390/jcm3041437

Peter J. Anderson, & Anthony L.H. Moss. (1996). Dental Findings in Parents of Children with Cleft Lip and Palate. *Cleft Palate-Craniofacial Journal*, *33*(5).

Pi, X., Li, Z., Jin, L., Liu, J., Zhang, Y., Zhang, L., … Ren, A. (2018). Secondhand smoke during the periconceptional period increases the risk for orofacial clefts in offspring. *Paediatric and Perinatal Epidemiology*, *32*(5), 423–427. https://doi.org/10.1111/ppe.12497

Ranganathan, K., Shapiro, D., Aliu, O., Vercler, C. J., Baker, M., Kasten, S. J., … Waljee, J. F. (2016). Health-Related Quality of Life and the Desire for Revision Surgery Among

Children With Cleft Lip and Palate. *Journal of Craniofacial Surgery*, *27*(7), 1689–1693. https://doi.org/10.1097/SCS.0000000000002924

Risks, I. of M. (US) C. on A. G., Andrews, L. B., Fullarton, J. E., Holtzman, N. A., & Motulsky, A. G. (1994). *Social, Legal, and Ethical Implications of Genetic Testing*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK236044/

Robbins, J. M., Damiano, P., Druschel, C. M., Hobbs, C. A., Romitti, P. A., Austin, A. A., … Burnett, W. (2010). Prenatal Diagnosis of Orofacial Clefts: Association With Maternal Satisfaction, Team Care, and Treatment Outcomes. *The Cleft Palate-Craniofacial Journal : Official Publication of the American Cleft Palate-Craniofacial Association*, *47*(5), 476–481. https://doi.org/10.1597/08-177

Roosenboom, J., Saey, I., Peeters, H., Devriendt, K., Claes, P., & Hens, G. (2015). Facial Characteristics and Olfactory Dysfunction: Two Endophenotypes Related to Nonsyndromic Cleft Lip and/or Palate [Research article]. https://doi.org/10.1155/2015/863429

Roosenboom, Jasmien, Indencleef, K., Hens, G., Peeters, H., Christensen, K., Marazita, M. L., … Weinberg, S. M. (2017). Testing the face shape hypothesis in twins discordant for nonsyndromic orofacial clefting. *American Journal of Medical Genetics. Part A*, *173*(11), 2886–2892. https://doi.org/10.1002/ajmg.a.38471

Sanders, S. H., JH Noah, R. (2000). An investigation into the dentition of parents of children with cleft lip and palate. *Cleft Palate-Craniofacial Journal*, *37*(4), 395–405.

Scott, N. M., Weinberg, S. M., Neiswanger, K., Brandon, C. A., & Marazita, M. L. (2005). Hair Whorls and Handedness: Informative Phenotypic Markers in Nonsyndromic Cleft Lip With or Without Cleft Palate (NS CL/P) Cases and Their Unaffected Relatives. *American Journal of Medical Genetics*, *136*, 158–161. https://doi.org/10.1002/ajmg.a.30806

Sharif-Askary, B., Bittar, P. G., Farjat, A. E., Liu, B., Vissoci, J. R. N., & Allori, A. C. (2018). Geospatial Analysis of Risk Factors Contributing to Loss to Follow-up in Cleft Lip/Palate Care. *Plastic and Reconstructive Surgery – Global Open*, *6*(9), e1910. https://doi.org/10.1097/GOX.0000000000001910

Shprintzen, R. J., Schwartz, R. H., Daniller, A., & Hoch, L. (1985). Morphologic Significance of Bifid Uvula. *Pediatrics*, *75*(3), 553–561.

Singh, P., & Nathani, D. B. (2017). Dermatoglyphics and Cheiloscopy as Key Tools in Resolving the Genetic Correlation of Inheritance Patterns in Cleft Lip and Palate Patients: An Assessment of 160 Patients. *The Cleft Palate - Craniofacial Journal; Lawrence*, *54*(5), 588–594. http://dx.doi.org.pitt.idm.oclc.org/10.1597/15-168

Sischo, L., Wilson-Genderson, M., & Broder, H. L. (2017). Quality-of-Life in Children with Orofacial Clefts and Caregiver Well-being. *Journal of Dental Research*, *96*(13), 1474–1481. https://doi.org/10.1177/0022034517725707

Sivertsen, A., Wilcox, A. J., Skjaerven, R., Vindenes, H. A., Abyholm, F., Harville, E., & Lie, R. T. (2008). Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. *BMJ (Clinical Research Ed.)*, *336*(7641), 432–4. https://doi.org/10.1136/bmj.39458.563611.AE

Smedegaard, L. H. V., Marxen, D. R. N., Moes, J. H. V., Glassou, E. N., & Scientsan, C. (2008). Hospitalization, Breast-Milk Feeding, and Growth in Infants with Cleft Palate and Cleft Lip and Palate Born in Denmark. *The Cleft Palate-Craniofacial Journal*, *45*(6), 628–632. https://doi.org/10.1597/07-007.1

Solomon, B. D., Jack, B., & Feero, W. G. (2008). The Clinical Content of Preconception Care: Genetics and Genomics. *American Journal of Obstetrics and Gynecology*, *199*(6 Suppl 2), S340–S344. https://doi.org/10.1016/j.ajog.2008.09.870

Spence, A., Westlake, J., Lange, K., & Gold, D. P. (1976). Estimation of Polygenic Recurrence Risk for Cleft Lip and Palate. *Human Heredity*, *26*(5), 327–336. https://doi.org/10.1159/000152824

Sreejith, V. P., Arun, V., Devarajan, A. P., Gopinath, A., & Sunil, M. (2018). Psychological Effect of Prenatal Diagnosis of Cleft Lip and Palate: A Systematic Review. *Contemporary Clinical Dentistry*, *9*(2), 304–308.

StataCorp. (2017a). Brier Score Decomposition. In *Stata 15 Base Reference Manual*. College Station, TX: Stata Press.

StataCorp. (2017b). Stata Statistical Software: Release 15 (Version 15.1). College Station, TX: StataCorp LLC.

Stirone, A. (2019). *Overview of Legislation and Litigation Involving Protections Against Down Syndrome Discrimination Abortion | Charlotte Lozier Institute*. Retrieved from https://lozierinstitute.org/overview-legislation-litigation-involving-protections-against-down-syndrome-discrimination-abortion/

Stoltzfus, J. C. (2011). Logistic Regression: A Brief Primer. *Academic Emergency Medicine*, *18*(10), 1099–1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x

Stott-Miller, M., Heike, C. L., Kratz, M., & Starr, J. R. (2010). Increased risk of orofacial clefts associated with maternal obesity: Case–control study and Monte Carlo-based bias analysis. *Paediatric and Perinatal Epidemiology*, *24*(5), 502–512. https://doi.org/10.1111/j.1365-3016.2010.01142.x

Strauss, R. P., & Cassell, C. H. (2009). Critical Issues in Craniofacial Care: Quality of Life, Costs of Care, and Implications of Prenatal Diagnosis. *Academic Pediatrics*, *10*(9), 427–432. https://doi.org/10.1016/j.acap.2009.09.019

Svensson, E., Ehrenstein, V., Nørgaard, M., Bakketeig, L. S., Rothman, K. J., Sørensen, H. T., & Pedersen, L. (2014). Estimating the proportion of all observed birth defects occurring in

pregnancies terminated by a second-trimester abortion. *Epidemiology (Cambridge, Mass.)*, *25*(6), 866–871. https://doi.org/10.1097/EDE.0000000000000163

Sweeney, W. M., Lanier, S. T., Purnell, C. A., & Gosain, A. K. (2015). Genetics of Cleft Palate and Velopharyngeal Insufficiency. *Journal of Pediatric Genetics*, *4*(1), 9–16. https://doi.org/10.1055/s-0035-1554978

Trevethan, R. (2017). Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Frontiers in Public Health*, *5*. https://doi.org/10.3389/fpubh.2017.00307

U.S. Department of Health and Human Services. (2014). Birth Defects, Child Health USA 2014. Retrieved May 17, 2019, from https://mchb.hrsa.gov/chusa14/health-status-behaviors/infants/birth-defects.html#source4

van der Ploeg, T., Austin, P. C., & Steyerberg, E. W. (2014). Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. *BMC Medical Research Methodology*, *14*. https://doi.org/10.1186/1471-2288-14-137

Wang, X., Huang, Y., Li, L., Dai, H., Song, F., & Chen, K. (2018). Assessment of performance of the Gail model for predicting breast cancer risk: A systematic review and meta-analysis with trial sequential analysis. *Breast Cancer Research : BCR*, *20*. https://doi.org/10.1186/s13058-018-0947-5

Ward, R. E., Bixler, D., & Raywood, E. R. (1989). A study of cephalometric features in cleft lip-cleft palate families. I: Phenotypic heterogeneity and genetic predisposition in parents of sporadic cases. *The Cleft Palate Journal*, *26*(4), 318–325; discussion 325-326.

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*, 149–163. https://doi.org/10.1016/j.cps.2013.12.003

Wehby, G., & Cassell, C. H. (2010). The Impact of Orofacial Clefts on Quality of Life and Health Care Use and Costs. *Oral Diseases*, *16*(1), 3–10. https://doi.org/10.1111/j.1601-0825.2009.01588.x

Wehby, G., & Murray, J. C. (2010). Folic Acid and Orofacial Clefts: A Review of the Evidence. *Oral Diseases*, *16*(1), 11–19. https://doi.org/10.1111/j.1601-0825.2009.01587.x

Weinberg, S. M., Maher, B. S., & Marazita, M. L. (2006). Parental craniofacial morphology in cleft lip with or without cleft palate as determined by cephalometry: A meta-analysis. *Orthodontics & Craniofacial Research*, *9*(1), 18–30. https://doi.org/10.1111/j.1601-6343.2006.00339.x

Weinberg, S., Naidoo, S., Bardi, K., Brandon, C., Neiswanger, K., Resick, J., … Marazita, M. (2009). Face shape of unaffected parents with cleft affected offspring: Combining three-dimensional surface imaging and geometric morphometrics. *Orthodontics & Craniofacial Research*, *12*(4), 271–281. https://doi.org/10.1111/j.1601-6343.2009.01462.x

Weinberg, Seth M., Naidoo, S., Govier, D. P., Martin, R. A., Kane, A. A., & Marazita, M. L. (2006). Anthropometric precision and accuracy of digital three-dimensional photogrammetry: Comparing the genex and 3dMD imaging systems with one another and with direct anthropometry. *Journal of Craniofacial Surgery*, *17*(3), 477–483. https://doi.org/10.1097/00001665-200605000-00015

Weinberg, Seth M., Neiswanger, K., Martin, R. A., Mooney, M. P., Kane, A. A., Wenger, S. L., … Marazita, M. L. (2006). The Pittsburgh Oral-Facial Cleft study: Expanding the cleft phenotype. Background and justification. *Cleft Palate-Craniofacial Journal*, *43*(1), 7–20. https://doi.org/10.1597/04-122R1.1

Weinberg, Seth M., Neiswanger, K., Richtsmeier, J. T., Maher, B. S., Mooney, M. P., Siegel, M. I., & Marazita, M. L. (2008). Three-dimensional morphometric analysis of craniofacial shape in the unaffected relatives of individuals with nonsyndromic orofacial clefts: A possible marker for genetic susceptibility. *American Journal of Medical Genetics, Part A*, *146A*, 409–20. https://doi.org/10.1002/ajmg.a.32177

Weinberg, Seth M., Raffensperger, Z. D., Kesterke, M. J., Heike, C. L., Cunningham, M. L., Hecht, J. T., … Marazita, M. L. (2016). The 3D Facial Norms Database: Part 1. A Web-Based Craniofacial Anthropometric and Image Repository for the Clinical and Research Community. *The Cleft Palate-Craniofacial Journal*, *53*(6), 185–197. https://doi.org/10.1597/15-199

Wen, Y., & Lu, Q. (2015). Risk prediction models for oral clefts allowing for phenotypic heterogeneity. *Frontiers in Genetics*, *6*, 264. https://doi.org/10.3389/fgene.2015.00264

WHO | WHOQOL: Measuring Quality of Life. (n.d.). Retrieved June 17, 2019, from WHO website: https://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/

Wilson, J. M. G., & Jungner, G. (1968). PRINCIPLES AND PRACTICE OF SCREENING FOR DISEASE. *Public Health Papers*, 168.

Wilson-Nagrani, C., Richmond, S., & Paternoster, L. (2018). Non-syndromic Cleft Lip and Palate Polymorphisms Affect Normal Lip Morphology. *Frontiers in Genetics*, *9*. https://doi.org/10.3389/fgene.2018.00413

Wyoming Cleft Lip and Palate Foundation. (n.d.). Retrieved June 18, 2019, from Wyoming Cleft Lip and Palate Foundation website: https://wyomingcleftfoundation.org/

Wyszynski, D. F. (2002). Cleft Lip and Palate: From Origin to Treatment. In *Cleft Lip and Palate: From Origin to Treatment* (pp. 66–86). Oxford University Press, Incorporated.

Xu, D.-P., Qu, W.-D., Sun, C., Cao, R.-Y., Liu, D.-W., & Du, P.-G. (2018). A Study on Environmental Factors for Nonsyndromic Cleft Lip and/or Palate. *The Journal of Craniofacial Surgery*, *29*(2), 364–367. https://doi.org/10.1097/SCS.0000000000004214

Xuan, Z., Zhongpeng, Y., Yanjun, G., Jiaqi, D., Yuchi, Z., Bing, S., & Chenghao, L. (2016). Maternal active smoking and risk of oral clefts: A meta-analysis. *Oral Surgery, Oral*

Medicine, Oral Pathology and Oral Radiology, *122*(6), 680–690. https://doi.org/10.1016/j.oooo.2016.08.007

Yazdy, M. M., Honein, M. A., Rasmussen, S. A., & Frias, J. L. (2007). Priorities for Future Public Health Research in Orofacial Clefts. *The Cleft Palate - Craniofacial Journal; Lawrence*, *44*(4), 351–357.

Yoon, H., Chung, I. S., Seol, E. Y., Park, B. Y., & Park, H. W. (2000). Development of the Lip and Palate in Staged Human Embryos and Early Fetuses. *Yonsei Medical Journal*, *41*(4), 477–484.

Zhang, S.-J., Meng, P., Zhang, J., Jia, P., Lin, J., Wang, X., … Wei, X. (2018). Machine Learning Models for Genetic Risk Assessment of Infants with Non-syndromic Orofacial Cleft. *Genomics, Proteomics & Bioinformatics*. https://doi.org/10.1016/j.gpb.2018.07.005