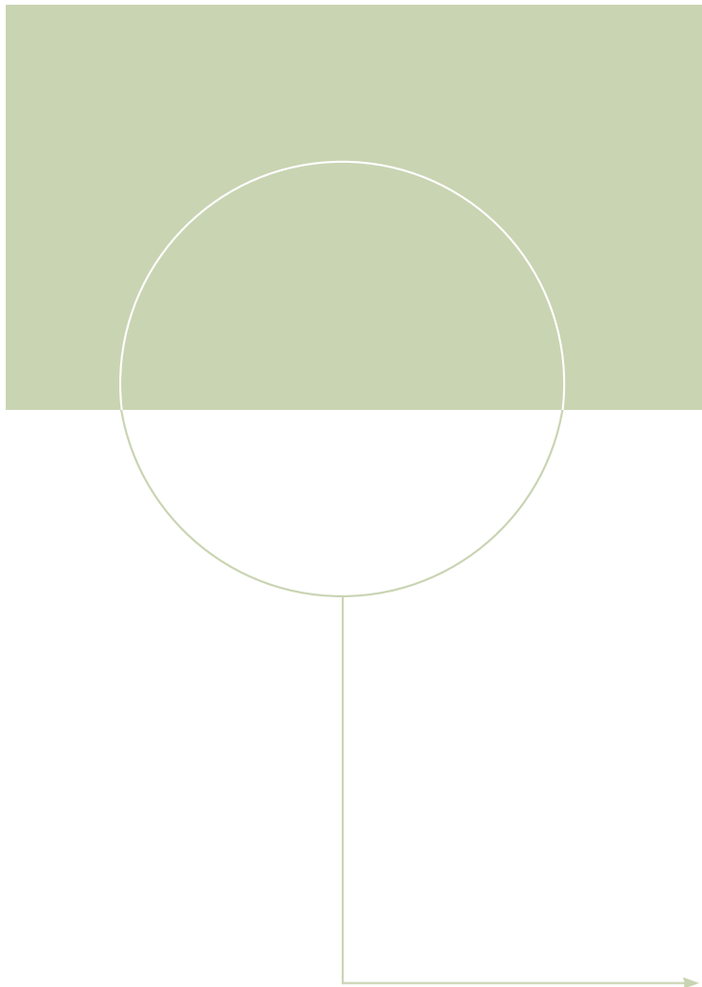


Master's thesis

2021

Natalie Steen

NTNU
Norwegian University of
Science and Technology
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

Natalie Steen

DENTAL CARIES IN CHILDREN

an investigation of levels of two saliva components

June 2021

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU

Norwegian University of
Science and Technology

DENTAL CARIES IN CHILDREN

an investigation of levels of two saliva components

Natalie Steen

Masters in Molecular Medicine, NTNU
Submission date: June 2021
Supervisor: Torunn Børsting (NTNU and Kompetansesenteret Tannhelse Midt)
Co-supervisor: Astrid J Feuerherm (Kompetansesenteret Tannhelse Midt)

Norwegian University of Science and Technology
Department of Clinical and Molecular Medicine

SUMMARY

Despite advances in preventative care, dental caries continues to be a source of poor health in children around the world. The causes of this are complex, however differences in the composition of saliva are suspected to be a factor and is a focus of current research. This thesis describes a literature review of the field, a preliminary process for selection of two saliva components with known or suspected roles in caries in children, pilot studies assessing saliva pre-analytical and ELISA methods, ELISA evalution of the two analytes of interest and a cross-sectional investigation into associations between dental caries experience and the expression levels of the two analytes in the saliva of children involved in a broader public health research program in Norway.

This thesis finds that cathelicidin (LL37) and statherin are plausible saliva components with a role in caries aetiology, the levels of LL37 and statherin can be readily measured in saliva samples using commercial ELISA kits and that this returns similar results to the published literature. The caries experience in this population is in line with national trends and shows a positive association with salivary statherin concentration and no significant association with salivary LL37 concentration by the use of logistic regression. The strengths of association were not altered by controlling for confounders identified by Directed Acyclic Graph analysis. This study also finds that the relationships between saliva volume and protein concentrations are complicated, including the observation that LL37, but not statherin, is diluted when large volumes of saliva are spat out after a chewing stimulus. The importance of preanalytical sample handling and analyte level normalising methods selected are discussed, as are the advantages and challenges of saliva as a sample material, and prospects for future research.

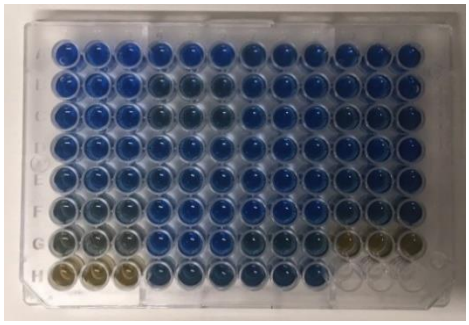
Sammendrag

Til tross for fremskritt innen forebyggende tiltak, så fortsetter karies å være en kilde til dårlig helse hos barn i hele verden. Årsakene til dette er komplekse, men man tror at forskjeller i sammensetning av saliva er en faktor, og dette er et fokusområde for dagens forskning. Denne masteroppgaven beskriver en litteraturgjennomgang av feltet, prosessen for å velge to biomarkører i spytt som har en kjent eller mistenkt rolle i kariesutvikling hos barn, pilotstudier som vurderer pre-analytiske og ELISA-metoder for spyttanalyser, ELISA-evaluering av de to utvalgte biomarkørene i spytt og en tverrsnittstudie for å undersøke assosiasjoner mellom karieserfaring og nivåene til de to biomarkørene i spytt fra barn som har deltatt i en større folkehelsestudie i Norge.

Oppgaven fant at cathelicidin (LL37) og statherin er biomarkører som kan spille en en rolle i kariesetiologi, at nivåene av LL37 og statherin lett kan måles i spytt ved bruk av kommersielle ELISA-kit, og at resultatene i denne studien er sammenliknbare med det som tidligere er publisert. Forekomsten av karieserfaring i denne studiepopulasjonen er i tråd med nasjonale trender og, ved bruk av logistisk regresjon, viste en positiv assosiasjon til nivået av statherin i saliva, men ingen signifikant assosiasjon til nivået av LL37. Ingen av assosiasjonene ble signifikant endret ved å kontrollere for konfounderende faktorer som var identifisert gjennom en DAG-analyse. Denne studien viste også at forholdene mellom spyttvolum og proteinkonsentrasjon er komplekse. Én observasjon var at nivåene av LL37, men ikke av statherin i spytt blir fortynnet når spyttproduksjonen er stor. Betydningen av preanalytisk spyttprøvehåndtering og valgte metoder for normalisering av analysenivå blir også diskutert, i tillegg til fordeler og utfordringer med saliva som prøvemateriale, og utsiktene for fremtidig forskning.

Acknowledgements

With thanks to Hilde Lysvand (Department of Clinical and Molecular Medicine, NTNU) for technical guidance with the ELISA experiments.



Contents

SUMMARY i
1. INTRODUCTION 1
1.1 Teeth and Saliva 1
1.2 Dental Caries and Salivary Components 4
2. RESEARCH QUESTIONS AND APPROACHES USED 8
3. PRELIMINARY INVESTIGATION 9
3.1 Methodology 9
3.2 Results 11
4. MAIN INVESTIGATION 14
4.1 Methodology 14
4.1.1 The TRIP-Tann Data 14
4.1.2 Characterising the saliva samples 16
4.1.3 Testing salivary LL37 and statherin levels 19
4.1.4 Testing for possible associations between the analytes and caries experience 22
4.2 Results 28
4.2.1 Descriptive analysis of the sample population, health data and saliva 28
4.2.2 Determining best pre-analytical sample preparation was important 30
4.2.3 Cathelicindin (LL37) was found in varying levels in saliva 32
4.2.4 Statherin was generally quite abundant in saliva 33
4.2.5 Descriptive analysis of assay results 34
4.2.6 Inferential statistical analyses of analyte levels and caries experience 35
5. DISCUSSION 38
5.1 The two saliva components chosen to analyse were LL37 and statherin 38
5.1.1 Cathelicidin-derived LL37: a direct antimicrobial peptide 38
5.1.2 Statherin: an enamel-maintaining peptide 40
5.2 Choice of pre-analytical and data normalising methods are important 42
5.3 Saliva ELISA 46
5.3.1 The ELISA results from this study are consistent with the published literature 47
5.4 Caries experience associations with LL37 and statherin 48
5.5 Project strengths and weaknesses 50
6. CONCLUSIONS 52
REFERENCES 54
ATTACHMENTS 62

List of Tables and Figures

Table 1: A short-list of saliva components with associations to dental caries demonstrated in more than one published study (those in italics were found to have saliva-validated commercial kits available to order). 11
Table 2: A summary of the findings from literature, catalogue and database searches for the two selected salivary analytes 13
Table 3: Bivariate analyses of possible confounders 25
Table 4: Logistic regression models 27
Table 5: Background table- exposure counts, proportions and/or means for each variable in relation to whole sample and with- and without caries experience subgroups. 28
Table 6: Descriptive analysis of the total protein and ELISA assay results by caries experience in this sample population 34
Table 7: The results of logistic modelling of caries experience on LL37 concentration (ng/mL). 36
Table 8: The results of logistic modelling of caries experience on statherin concentration (µg/mL) 36
Table 9: Human Cathelicidin gene, transcript and products 39
Table 10: Human Statherin gene, transcript and protein details 41

Figure 1: Factors contributing to the development of dental caries (reproduced from Selwitz et al. (2007)) 5
Figure 2: Literature search 9
Figure 3: The number of participants and samples at each stage 15
Figure 4: DAGs of hypothetical causal pathway/networks involving A) LL37, and B) statherin levels and caries experience 24
Figure 5: Plots of the correlations between various saliva measurements: A) Volume versus total protein (TP) concentration; B) Volume of spit by age; C) Volume of spit by sex; D) Volume of spit by collection time 26
Figure 6: The collected volume of the saliva samples in this study 30
Figure 7: A) Thawing of whole saliva at 4 °C, showing sediment already present in the frozen state; B) Re-sedimentation of vortexed whole saliva within 5 minutes at room temperature 31
Figure 8: The total protein concentration of 175 saliva samples 32
Figure 9: LL37 concentration (ng/mL) in 173 saliva samples determined by ELISA. 33
Figure 10: Statherin concentration (ng/mL) in 165 saliva samples determined by ELISA 34
Figure 11: A) LL37 concentration versus dmft/DMFT score, B) Statherin concentration versus dmft/DMFT score 35
Figure 12: TP Concentration by caries experience 37
Figure 13: The sequence of human cathelicidin aligned with consensus sequences 38
Figure 14: Image taken from Engelberg and Landau (2020): electron microscopy of LL37 fibrils interacting with a bacterium (Micrococcus luteus) (scale bar represents 500 nm). 40
Figure 15: The sequence of human statherin aligned with the consensus sequence 41

List of Abbreviations used:

%CV	coefficient of variability (%)
AEP	acquired enamel pellicle
BLAST	Basic Local Alignment Search Tool
BSA	bovine serum albumin
CI	confidence interval
DAG	directed acyclic graph
DMFT	decayed, missing or filled teeth
ELISA	enzyme-linked immunosorbant assay
GCF	gingival crevicular fluid
H0	null hypothesis
HGNC	HUGO Gene Nomenclature Committee
IKOM	Institutt for Klinisk og Molekylær Medisin
IL	Interleukin
LPS	Lipopolysaccharide
NCBI	National Centre for Biotechnology Information
NTNU	Norges teknisk-naturvitenskapelige universitet
OR	odds ratio
PBS	phosphate-buffered saline
PS	passive supernatant
RANTES	Regulated upon Activation, Normal T Cell Expressed and Secreted
RO	reverse osmosis
TkMidt	Kompetansesenteret Tannhelse Midt
TMB	3,3’,5,5’-Tetramethylbenzidine
TNF	Tumour necrosis factor
TP	total protein
TREM	Triggering receptor expressed on myeloid cells
TRIP	Training in Pregnancy
VDR	vitamin D responsive element
WHO	World Health Organization
WSH	whole saliva, homogenate
WSS	whole saliva, supernatant

Key words

caries, early childhood caries, LL37, cathelicidin, statherin, saliva, ELISA

1. INTRODUCTION

Despite advances in preventative care, dental caries continues to be a source of poor health in children around the world, with far-reaching consequences that can extend into adulthood. Saliva plays a role in caries protection, but the full mechanisms by which it does so have not been elucidated.

1.1 Teeth and Saliva

In most vertebrates efficient digestion is a process that begins with the teeth. The ability to grow and maintain healthy teeth evolved in the Devonian Period and has been conserved through the reptilian and mammalian lineages (Vaškaninová & Chen *et al.*, 2020). The mineral composition of teeth enables them to withstand physical and chemical damage, however they are not indestructible and direct assistance from the blood and lymphatic supply is limited by the very nature of the dental structure. The biological fluids of the mouth, namely saliva and mucosal secretions and transudates collectively referred to as "saliva" or spit, have evolved to contribute to tooth homeostasis (de Sousa-Pereira & Amado *et al.*, 2013).

The physical properties of saliva help maintain a moist, neutral pH environment free of food and other debris. Additionally, saliva provides minerals for enamel maintenance and delivers innate and adaptive immune system agents that act as the first line of defence against pathogens that might otherwise damage the teeth and mouth and/or invade further into the body. These latter functions must be carefully regulated so as to maintain a balanced microbiome and avoid autoimmune damage (Hemadi & Huang *et al.*, 2017; Van Nieuw Amerongen & Bolscher *et al.*, 2004).

By volume, the majority of saliva is produced by three pairs of salivary glands- the parotid, submandibular and sublingual glands (Carlson, 2000; Carpenter, 2013). These salivary glands contain specialised secretory cells and have a muscular capsule, which is subject to hormonal and autonomic control. There are also numerous smaller, unencapsulated glands scattered through the lining of the mouth. Together, the glands produce a cocktail of components with multiple biological functions, with some individual components having themselves more than one function (Proctor, 2016). The rate of flow of saliva into the mouth from the salivary glands varies with time of day, conscious state, chewing, the senses of taste and smell and with stress (Carpenter, 2013; Dawes, 1972; Proctor, 2016). The quantity and chemical form of salivary components can also be varied in response to signals at any level from gene expression through to post-secretion proteolysis (Aidoukovitch & Dahl *et al.*, 2020; Helmerhorst & Traboulsi *et al.*, 2010; Jensen & Xu *et al.*, 1995; Proctor, 2016; Vitorino & Barros *et al.*, 2009).

The gingival crevicular fluid (GCF) is present in the small space between the neck of the tooth and the free gingiva surrounding it. It is composed of ions and minerals, by-products of metabolism and digestion, epithelial and inflammatory cells, inflammatory mediators, bacteria and bacterial by-products, and other proteins (Subbarao & Nattuthurai *et al.*, 2019). These components are produced by the mucosal epithelium itself, or arise from the underlying connective tissue and blood supply, passing through the epithelium to the gingival sulcus as a transudate (Subbarao *et al.*, 2019; Sultan & Mali *et al.*, 2017). GCF is typically present in only small amounts but the volume is increased by increased local blood pressure (Subbarao *et al.*, 2019) and its composition is altered in local inflammation (Sultan *et al.*, 2017). The GCF can spill out from the sulcus into the general oral cavity.

The substance that is generally referred to as saliva or spit is thus a combination of that emerging from the major salivary ducts, unencapsulated glands and from the gingival sulci, as well as epithelial cell debris and secretions (from the digestive and respiratory tract epithelia), bacteria (whole and by-products) and the residue of products that have been taken into the mouth (food, drink, tobacco products etc). As a result, it has been estimated that only a quarter of the total protein content of saliva that is typically spat out from the mouth is directly produced by the salivary glands (Amado & Lobo *et al.*, 2010; Vitorino *et al.*, 2009).

While being approximately 99% water, saliva has very different properties to water that ensure it both spreads easily and also clings to oral surfaces. The ion composition is hypotonic, but it is supersaturated in terms of calcium (Carpenter, 2013). By percentage, the main protein components are amylase, mucins, proline-rich peptides, histatins (histidine rich), cystatins (cysteine rich) and statherin. Post-translational modifications including phosphorylations and glycosylation are important for function, the former promoting calcium-binding (e.g. of statherin) and the glycosylations promote the formation of large protein aggregates, which amongst other actions assists bacterial clearance. Mucin, in particular, is a large glycoprotein that readily forms large aggregates and is the main source of saliva viscosity (Carpenter, 2013).

All tooth surfaces, but especially those closest to the tongue, are covered in a thin, organic layer of proteins and glycoproteins called the acquired enamel pellicle (AEP). Salivary molecules that adsorb to the tooth surface and contribute to the AEP include lysozyme, lactotransferrin, statherin, myeloperoxidase and proline rich protein-3. The AEP functions as a kind of semi-permeable shield for the teeth, allowing controlled access of minerals and limiting the colonisation of harmful bacteria (Ventura & Cassiano *et al.*, 2017).

In the healthy mouth, a film of *non-harmful* microbes is allowed to form over the AEP and mucosal surfaces. This biofilm consists of hundreds of species, the precise quantity and identity of which varies with site in the mouth and an individual's age and general health. The most commonly detected species are facultative anaerobes including *Neisseria, Streptococcus, Actinomyces, Veillonella and*

Granulicatella species (plus Bacteroidaceae spp and spirochetes subgingivally). These species exist in homeostasis with each other and symbiosis with the host. Their presence limits the success of more harmful bacterial competitors. The production of acids by some members of this microbiome are balanced by the production of alkali by others within the microbe population and they are typically "neutrophilic microbes", that is, producing only a mild, neutrophilic inflammatory response from the host. The composition of saliva also plays a role in microbe homeostasis, by providing nutrients for the microbes on the one hand, while delivering the innate immunity components that manage them and their competitors on the other hand (Lamont & Koo *et al.*, 2018; Sanz & Beighton *et al.*, 2017).

1.2 Dental Caries and Salivary Components

Dental caries constitutes a non-communicable bacterial disease that leads to localised demineralisation of the dental enamel (the inorganic layer of the tooth) that, without treatment, progresses through to the dentine (the organic layer of the tooth) (Selwitz & Ismail *et al.*, 2007). Dental caries is a common health problem worldwide and there is concern regarding its prevalence in children in particular, and the consequences on their physical, psychological and emotional health (Hemadi *et al.*, 2017; Tao & Jurevic *et al.*, 2005). Childhood caries have also been cited as a strong predictor of adult caries (Broadbent & Page *et al.*, 2013).

The prevalence of caries in children under the age of six has been estimated at approxiately 23% in the USA and 60% in China (Hemadi *et al.*, 2017). In Norway in 2018, 18.7% of 5-year-old children and 39.6%[1] of 12-year-old children were found to have dental caries, with a similar pattern seen across the preceding five years (Statistics Norway, 2020). This high prevalence with wide geographic spread, persisting despite public health initiatives, has led to interest in developing biomarkers for early detection of childhood caries to instigate additional, targeted intervention measures (Hemadi *et al.*, 2017).

For dental caries to develop, teeth must be present in conjunction with several oral environment conditions (Selwitz *et al.*, 2007) (Figure 1). These conditions can include suboptimal salivary volume and composition and the presence of cariogenic bacteria.

[1] Statistics based on national dental service records. Approximately 77% of all 5- and 12-year-olds had visited the publicly funded dental service in 2018.

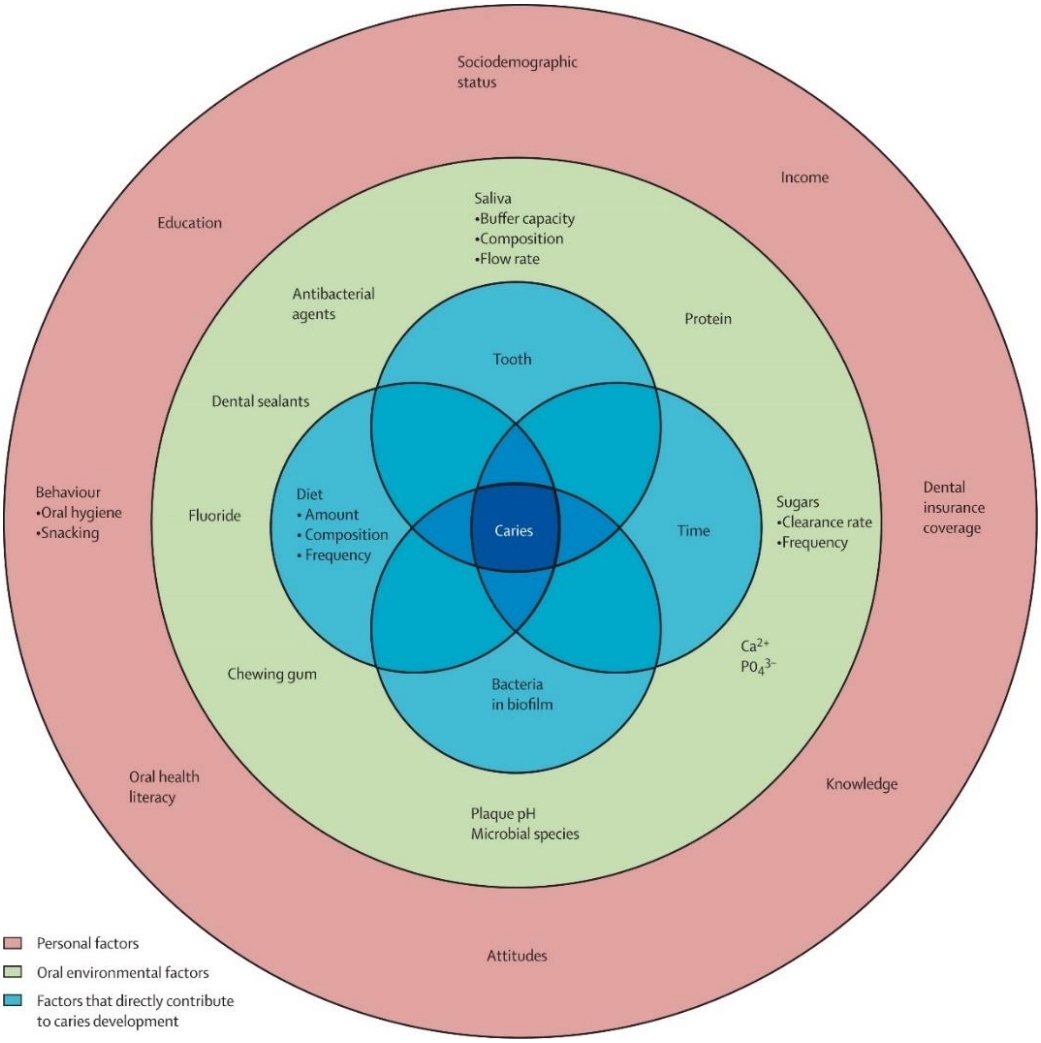


FIGURE 1: FACTORS CONTRIBUTING TO THE DEVELOPMENT OF DENTAL CARIES (REPRODUCED FROM SELWITZ ET AL. (2007))

The main inorganic component of dental enamel is hydroxyapatite, a hardened complex of calcium and phosphorous. It is produced by foetal ameloblasts and, throughout life, demineralisation-remineralisation cycling can occur. Remineralisation is dependent on salivary levels of calcium and phosphorous and the saliva pH (Robinson & Brookes *et al.*, 1998).

Calcium is actively secreted from parotid and mandibular gland acinar cells into the salivary gland duct (Homann & Kinne-Saffran *et al.*, 2006) and its reabsorption, as the saliva passes along the duct is limited and regulated (Bandyopadhyay & Swaim *et al.*, 2012). The resultant high concentrations of calcium ions in the saliva that enter the oral space would ordinarily lead to the formation of calcium crystals, making the ions less available for enamel remineralisation. Components of saliva, most notably statherin, prevent the calcium crystalisation (Helmerhorst *et al.*, 2010). Serum vitamin D metabolites also play a significant role in calcium

homeostasis, through effects on signalling pathways that control calcium-carrying protein gene expression, amongst others (Bikle, 2020). The latter occurs via the Vitamin D Receptor (VDR) which can act as a transcription regulator when bound to its ligand, vitamin D isoform $1,25(OH)_2D_3$ (also known as calcitriol) (Bikle, 2020; Carlberg, 2019). The role of Vitamin D in dental health, through effects on calcium homeostasis amongst others, is the focus of several current research projects (Børsting, 2019; Kühnisch & Thiering *et al.*, 2015; Nørrisgaard & Haubek *et al.*, 2019; Schroth & Lavelle *et al.*, 2014; van der Tas & Elfrink *et al.*, 2018).

A salivary pH of below 5.5 leads to hydroxyapatite destabilisation and demineralisation. This can be reversed if the pH does not remain low and if sufficient calcium and phosphorous (or fluoride) are present at the tooth surface to favour remineralisation. Some fluctuations in pH at the tooth surface are unavoidable and tolerated, however a very low and/or sustained low pH brings a risk of caries.

High dietary sugar intake can lead to pH changes via oral bacteria. The symbiotic bacteria in the healthy dental biofilm, described above, normally have an evolutionary advantage over pathogenic bacteria and do not produce harmful changes in pH (Lamont *et al.*, 2018). However, the addition of sucrose to the environment shifts the balance to one more favourable to sacchrolytic, acidogenic bacteria. The main sacchrolytic and acidogenic species implicated are the Streptococcus species *S. mutans* and *S. sorbrinus* and lactobacilli (Hemadi *et al.*, 2017), although *Bifidobacterium dentium and Scardovia wiggsiae* are also recognised as important in the aetiology of caries, particularly in children (Sanz *et al.*, 2017). Once these species gain some advantage, a positive feedforward loop is initiated that leads to a shift away from the symbiotic homeostasis towards dysbiosis. A new, highly specialised biofilm community becomes established, that perpetuates the acidic environment. Some species also synthesise a new extracellular matrix that gives the community additional self-sustaining properties, through protection from salivary sheer forces and even brushing, protection from antimicrobial peptides and by acting as an alternative food source when dietary sugar levels are insufficient (Lamont *et al.*, 2018; Sanz *et al.*, 2017). At this point, the demineralisation of the enamel outpaces potential remineralisation.

It is hypothesised that, even with moderate amounts of dietary sugar, the growth and proliferation of cariogenic microbes will only proceed in the absence of an appropriate change in the saliva. Saliva can physically clear non-adherent microbes, a property enhanced by mucin and agglutinins (Polley & Louzada *et al.*, 2015; Sanz *et al.*, 2017). Saliva also delivers components of the innate immune system. This includes peptides that have direct antimicrobial activity (for example, the cathelicidin-derived antimicrobial peptide LL-37, the defensins and the histatins (Khurshid & Naseem *et al.*, 2016)); and peptides that act as pro-inflammatory cytokines (for example: TREM1 (S. S. Chen & Wang *et al.*, 2017; Nylund & Ruokonen *et al.*, 2018), the chemokine CCL5/RANTES and interleukins, including IL-8 (Gornowicz & Bielawska *et al.*, 2012)).

Some studies have shown that levels of antimicrobial or proinflammatory salivary components are higher in the presence of caries (Colombo & Ribas *et al.*, 2016; Gyll & Ridell *et al.*, 2018; Kumar & Reddy *et al.*, 2016), as would be expected in a normal physiological response to the presence of pathogenic bacteria and inflammation. Other studies have concluded that the levels of these types of saliva components are inversely correlated to the presence of caries (Davidopoulou & Diza *et al.*, 2012; Tao *et al.*, 2005), suggesting their absence might play a role in caries' aetiology. As with all aspects of immunological physiology, the regulation of these salivary components is complex.

Other authors have attempted to perform systematic reviews of the role, if any, of saliva components in protecting teeth against caries, but their efforts have been limited by the differences in study design, case definition and component analysis methods (for example, Martins & Buczynski *et al.* (2013) and Piekoszewska-Ziętek & Turska-Szybka *et al.* (2019)). Some saliva studies have employed techniques that are difficult to replicate, due to institute-specific materials and/or methods (for example, Jentsch & Beetke *et al.* (2004)).

There remains substantial interest within public health, dental and bioscience research in the role/s of saliva components in caries aetiology (Arias-Bujanda & Regueira-Iglesias *et al.*, 2020; Hemadi *et al.*, 2017; Kumar *et al.*, 2016; Van Nieuw Amerongen *et al.*, 2004) and a lot remains unanswered. Through thoroughly detailed studies that follow standardised procedures for case classification, sample collection and sample pre-analysis and analysis methods, the pieces may start to fit together.

2. RESEARCH QUESTIONS AND APPROACHES USED

Preliminary research questions

Given access to a biobank of saliva samples with accompanying data, which salivary components should be investigated for associations with dental caries? Are there published studies of similar investigations? What methods were used to analyse salivary components and what were the results? Are these methods practical for our biobank samples?

Approach: A review of published literature, databases and supplier catalogues.

Main research questions

- What are the levels of the chosen salivary components in our saliva samples? What impact do pre-analytical methods have?

  Approach: Laboratory experiments based on the preliminary research results using saliva samples in the biobank.

- Are there any associations between the levels of the chosen salivary components and the dental caries experience observed in the sample population? Can we identify any confounders of these association/s?

  Approach: An analysis utilising existing cross-sectional survey health data connected to the saliva biobank, the measurement of salivary component levels from this project, epidemiological methods, models and statistical tests.

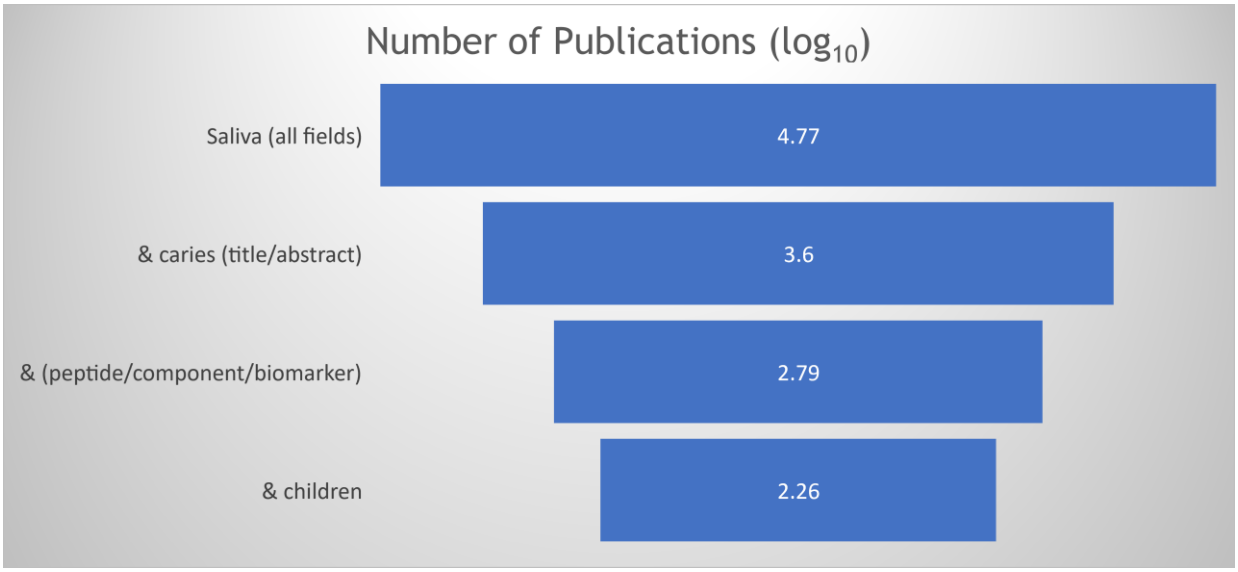
3. PRELIMINARY INVESTIGATION

3.1 Methodology

Selection of which saliva components to assay and which method to use to analyse those salivary components

In order to select saliva components with a possible role in childhood caries aetiology, a literature search was conducted using the PubMed advanced search tool[2] with the following search terms, limited to english language publications (Figure 2).

FIGURE 2: LITERATURE SEARCH



This was then limited to those publications for which we had full PDF access via the NTNU library.

The short-list of publications was reviewed for information on saliva collection and analysis methods, statistically significant saliva component associations with caries, typical concentration ranges for these components in saliva and the possible role of genetic polymorphisms.

Information on genes, transcripts, mature peptide features and variants was obtained from databases provided by HGNC (genenames.org; Braschi & Denny *et al.* (2019)), UNIprot (uniprot.org; TheUniProtConsortium (2018)) and NCBI

[2] https://pubmed.ncbi.nlm.nih.gov/advanced/)

(ncbi.nlm.nih.gov; Resource_Coordinators_NCBI (2016); Altschul & Gish *et al.* (1990)).

Selection of method of analysis

In order to determine the feasibility of analysis methods, the website Biocompare (biocompare.com) and the online catalogs of Gentaur[3] and NordicBiosite[4] were accessed. Information was sought regarding the availability of proteomic kits to assay for any of the saliva components of interest and whether these assays were validated for use with saliva samples and had an assay sensitivity low enough for the anticipated concentration range determined from the literature search. Preference was given to kits that contained all the necessary standards and reagents to perform the assay.

[3] GENTAUR Europe BVBA, Voortstraat 49, 1910 Kampenhout BELGIUM; Tel 0032 16 58 90 45; www.gentaursearch.com

[4] Nordic BioSite AS, Postboks 8007, 4675 Kristiansand, Norway; Tel: 2396 0418; nordicbiosite.com

3.2 Results

Two analytes selected from a short-list of six candidates

The literature search returned 153 full-text articles. Further filtering of the literature, for example limiting it to studies involving healthy seven to nine year old children where the saliva was collected in the same way as this study, was not performed due to the scarcity of such studies in the literature.

Approximately 18 individual saliva components were identified that have been confirmed to be present in spit and have been observed to have a correlation with dental disease in multiple studies (including at least one relating to children) (Table 1). This was further narrowed down to approximately eight candidates once the availability of saliva-validated ELISA kits was explored. Several components including lactotransferrin (Doetzer & Brancher *et al.*, 2015), agglutinin (Polley *et al.*, 2015), proline-rich peptides (Manconi & Castagnola *et al.*, 2016; K. Wang & Wang *et al.*, 2018) and betadefensins (Hatipoğlu & Saydam, 2020) were excluded due to concerns regarding genetic polymorphisms that alter caries risk (Lips & Antunes *et al.*, 2017; Tao *et al.*, 2005) as the suppliers of ELISA kits for these saliva components did not provide evidence regarding variant/isoform specificity. This left LL37, histatin-1, sTREM1, statherin, cystatin-C and IL-8 as strong potential candidates for analysis.

TABLE 1: A SHORT-LIST OF SALIVA COMPONENTS WITH ASSOCIATIONS TO DENTAL CARIES DEMONSTRATED IN MORE THAN ONE PUBLISHED STUDY (THOSE IN ITALICS WERE FOUND TO HAVE SALIVA-VALIDATED COMMERCIAL KITS AVAILABLE TO ORDER).

	Examples	***Selected References***
Direct antimicrobial peptides	*LL37*, *defensin beta-2*, defensin-alpha, *histatin 1* and histatin 5	(Aldred & Hollox *et al.*, 2005; Colombo *et al.*, 2016; Dale & Tao *et al.*, 2006; Davidopoulou *et al.*, 2012; Khurshid *et al.*, 2016; Tao *et al.*, 2005)
Released receptors	*sTREM1*, sCD14	(Bergandi & Defabianis *et al.*, 2007; Nishana & Bhat *et al.*, 2019)
Calcium maintenance and pellicle forming	*Statherin*, Proline-rich basic peptides (1 and/or 3)	(Levine, 2011; Shimotoyodome & Kobayashi *et al.*, 2006; Van Nieuw Amerongen *et al.*, 2004; Vitorino & Lobo *et al.*, 2005; K. Wang *et al.*, 2018)
Cysteine protease inhibitors	*Cystatin C* and S, *Lactotransferrin*	(Jentsch *et al.*, 2004; Sikorska & Mielnik-Blaszczak *et al.*, 2002; Van Nieuw Amerongen *et al.*, 2004; K. Wang *et al.*, 2018)
Cytokines	*IL-8*, IL-6, TNF, CCL5	(Gornowicz *et al.*, 2012; Sharma & Gupta *et al.*, 2017)

After objectively ranking these six candidates based on the literature search, based on the number and robustness of published association/s with caries or dental disease in children, statistically significant or otherwise, the two analytes selected were LL37 (cathelicidin antimicrobial peptide) and statherin. As can be seen in Table 2, reported salivary concentration ranges for the selected analytes were broad. This reflects differences in saliva collection and laboratory analysis methods in addition to any underlying physiological differences.

TABLE 2: A SUMMARY OF THE FINDINGS FROM LITERATURE, CATALOGUE AND DATABASE SEARCHES FOR THE TWO SELECTED SALIVARY ANALYTES

	LL-37	***Statherin***
Uniprot Protein ID	P49913	P02808
Molecular weight (kDa)	18 (19.3 pre-cleaved)	5.4 – 7.3 (depending on isoform, phosphorylation, proteolysis)
Approx. Concentration in saliva (ng/mL)	0.1-1200 (Davidopoulou *et al.*, 2012; Simon-Soro & Sherriff *et al.*, 2018; Tao *et al.*, 2005)	500 - 4000 (Huq & Cross *et al.*, 2007; Pateel & Gunjal *et al.*, 2017)
Mode of action	Antimicrobial peptide with some chemotactic properties (Oppenheim & Yang, 2005; Kun Wang & Zhou *et al.*, 2019)	Maintenance of anti-bacterial physical environment at the enamel and some antimicrobial action (Rudney & Staikov *et al.*, 2009; Shimotoyodome *et al.*, 2006; Vitorino *et al.*, 2005; K. Wang *et al.*, 2018)
Reported association/s with caries in children	Negative association (Davidopoulou *et al.*, 2012; Simon-Soro *et al.*, 2018; Tao *et al.*, 2005), positive association (Colombo *et al.*, 2016; Gyll *et al.*, 2018)	Low saliva statherin with caries (Rudney *et al.*, 2009; Vitorino *et al.*, 2005; K. Wang *et al.*, 2018)
Special sequence and structural features	2 disulfide bonds	High concentration of TYR, PRO and E residues; Hydrophobic N-terminus, -ve and acidic C-terminus
Vulnerability to intrinsic proteolysis	Low	High
ELISA kit	Hycult HK321	Cusabio CSB-EL022817HU
ELISA kit detection range (ng/mL)	0.1-100	78-5000
Possible ELISA issues	Does not discriminate between pro-peptide (intracellular) and active form. More than one sample dilution may be required	Does not discriminate between intact and degraded forms, which may have less activity. More than one sample dilution may be required

See main body text for protein information source details and Supplementary Data 1 for ELISA kit supplier details. Not all references used for salivary concentrations are cited.

4. MAIN INVESTIGATION

4.1 Methodology

4.1.1 The TRIP-Tann Data

This cross-sectional study utilised saliva samples, and dental health and lifestyle data from the participants. All data and saliva samples were collected prior to this project, according to the protocols of the TRIPTann study (REK approval #2015/639/REKSørØst – "The TRIP-Tann Study: Maternal pre-, peri- and postnatal vitamin D levels and dental caries in the primary teeth of children by 7 years of age"). These protocols are detailed in this section and in Supplementary Data 2 in order to provide a thorough background. All data was deidentified and securely stored at Kompetansesenteret Tannhelse Midt (TkMidt). Only data relevant to the current project was accessed. Any saliva that remained after the completion of the current project was returned to TkMidt.

Initially, the Training in Pregnancy (TRIP) project was a randomised-control trial run between 2007 and 2009 at Trondheim and Stavanger University Hospitals which investigated exercise interventions during pregnancy. In 2014-2016, the 855 TRIP project mother-child pairs were invited to participate in a 7-year follow-up of the TRIP study, and TRIPTann, which was a substudy involving the children. This lead to a sampling population for the current project constituting children between the ages of seven to nine years, resident in the Trøndelag and Rogaland regions of Norway, whose mothers were involved in the TRIP study. A dental examination, health and lifestyle data and saliva were collected from these children, who agreed to participate in the TRIPTann study, between May 2016 to August 2017 (n = 176; see Figure 3). Parents of the participating children gave written consent for their participation.

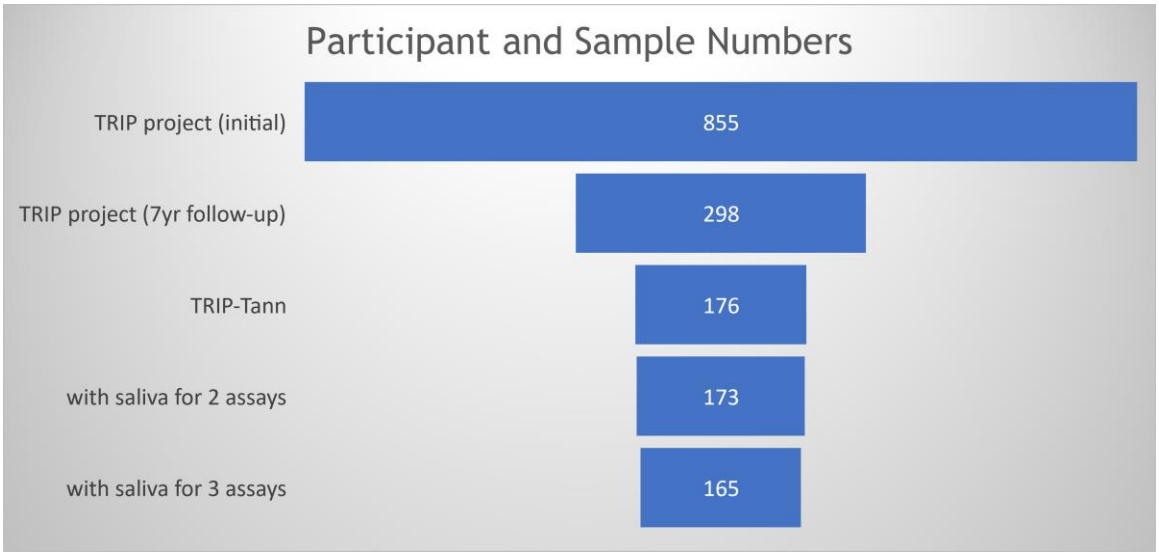


FIGURE 3: THE NUMBER OF PARTICIPANTS AND SAMPLES AT EACH STAGE

A standardised method was followed for the collection of saliva based on guidelines from the University of Oslo, Faculty of Odontology[5], the full details of which are presented in Supplementary Data 2. Saliva was always collected before the dental examination and questionnaire.

Immediately after the saliva was collected, a 0.5 mL aliquot (or the whole sample if less than 0.5 mL in total) was taken and stored at 4°C and the remainder of the sample was immediately placed in a freezer (-80°C). Three samples of saliva were of insufficient volume to allow for subsequent analysis.

The refrigerated aliquot was used for pH and bacterial tests the same day as the sample was collected. Semi-quantitative analysis of *Streptococcus mutans* and *Lactobacilli* levels was performed using a Strip test method (CRT- Bacteria, Ivoclar Vivodent Inc). The procedure involved homogenising the saliva with a disposable pipette before following the kit instructions. This gave a score of 1-4 for each bacterial species.

After the saliva sample was collected, the parent/guardian was asked to complete an electronic diet and lifestyle questionnaire (CheckWare). The questionnaire included variables relating to the child's vitamin supplement intake, frequency of toothbrushing, fluoride intake and diet.

Thereafter, the participant underwent the dental examination. The two dentists carrying out the dental examination were blinded with regards to the results of the survey. They were trained in how to do the dental exams and were calibrated according to a 5-graded caries diagnostic tool (Amarante & Raadal *et al.*, 1998). The two examiners achieved a Cohen's kappa value of 0.95 for intra-examiner reliability and 0.90 for inter-examiner reliability (personal communication, T.

[5] https://www.odont.uio.no/studier/ressurser/kariologi/Kariesutredning/Salivatester/saliva_sekresjonshastighet.html

Børsting). The dental examination recorded plaque and tartar scores, caries scores and scores for enamel defects.

The caries score is based on the recorded number of decayed, missing or filled primary (dmft) and permanent (DMFT) teeth present for each individual according to World Health Organization guidelines (WHO, 1997). This is abbreviated as dmft/DMFT and is considered a measure of the caries experience of the individual up to that point in time. For the descriptive analysis and some statistical analyses, the dmft/DMFT variable was dichotomised such that a score greater than zero was defined as "having caries experience", while a dmft/DMFT score of zero was defined as "not having caries experience".

Descriptive analysis of the TRIP-Tann data

The TRIPTann data set was used to determine the outcome proportions for those with- versus those without caries experience. These were determined for the sample population overall and within variables included in the data set. The categorisation of variables was defined during the TRIPTann study, with the exception of participant age and the saliva collection time variable. Age was rounded to the nearest whole year. Collection time was divided into four categories based on the division of the total range of values into approximately equal-sized intervals of time with similar sized total counts of participants.

All calculations were performed using Excel (Microsoft version 2104). A histogram of the distribution of the collected saliva volume was created using the statistical program R (version 4.0.3).

4.1.2 Characterising the saliva samples

Pilot studies

According to the data, the mean volume of the saliva samples (at the time of collection) was 4.43 mL. A subgroup of saliva samples with a collected volume within 0.25 standard deviations of the mean volume, were identified from the dataset (n= 26). From this subgroup, a random sample was taken for pilot studies by numbering the 26 samples and using a random number generator to choose nine of them. These pilot studies were designed to explore properties of saliva such as viscosity, protein concentration and sedimentation that might affect proteomic analysis, and later to determine dilution factors for the ELISA tests.

Thawing and sedimentation properties

The original saliva samples were contained in 15 mL polypropylene tubes stored at -80 °C at TKMidt for several months before being transferred to a –20 °C freezer at The Institute for Clinical and Molecular Medicine (IKOM), NTNU, for the analysis.

The time taken to fully thaw a sample at 4 °C was recorded for the pilot samples. Representative photos were taken of one sample thawing: at t_0, t_{30} minutes and t_{60} minutes to show the degree of sedimentation that had already occurred before freezing. After thawing of the pilot samples, experiments and observations were carried out on:

- whole saliva vortexed at 2000 g for 5 seconds using a bench-top vortex (WSH)
- the supernatant that forms "passively" in WSH after standing for at least 60 minutes at 4 °C (PS)
- the supernatant obtained after centrifugation of the WSH for 3 minutes at 10,000 g at 4 °C (WSS)

A representative photo was taken of one WSH aliquot after 5 minutes at room temperature to illustrate the natural sedimentation process.

Total protein estimation and sample preparation methods

The pilot studies analysed the total protein concentration of several saliva preparations:

- the 9 pilot samples assayed as WSH, PS and WSS at less than one minute and one hour (at 4°C) after preparation
- one pilot sample as PS and WSS one week (stored -20 °C, thawed in a fridge) after preparation.
- one pilot sample as WSH diluted 1:4 in an ELISA buffer (includes 0.05 % Tween).

All sampling for these assays occurred from the middle of the sample volume.

The Bradford (Coomassie G-250) standard, microplate method was used for the total protein measurements (Thermo Scientific). Bovine serum albumin (BSA) was used as a standard and dilutions of known concentration BSA were prepared (0 – 2000 µg/mL) using filtered (reverse osmosis (RO)) water as the diluent according to the manufacturers protocol. All standard dilutions and samples were assayed in triplicate.

For the standard dilutions and pilot samples alike, 10 µL per sample or standard was added to the microplate wells first, then 250 µL Coomassie reagent was rapidly added to each well. The plate was gently shaken, orbitally, on a microplate shaker for 30 seconds after the addition of the Coomassie reagent and incubated at room

temperature for at least 10 minutes and no more than 30 minutes (being mindful of first and last wells to have Coomassie reagent added). Absorbance at 595 nm was measured using a BMG Labtech Fluostar Optima plate reader, with orbital shaking for one second prior to measurement and gain adjustment.

The myassays.com online tool was used to generate the standard curve equation. Firstly, the mean absorbance of the three 0 µg/mL wells on the plate was calculated. This was subtracted from all the other absorbance measurements on the plate (normalisation). The mean normalised absorbance of each standard dilution triplicate was used to plot a 3-factor polynomial standard curve, as per the manufacturer's recommendation (Thermo Scientific). The myassays.com results included the standard curve equation. The standard curve equation and mean normalised sample absorbance measurements were then used with Excel to calculate the total protein concentration of the samples. Additionally, the intraassay variation coefficient (%CV) of the absorbance measurements was assessed for each set of (normalised) triplicates and the results across the samples and preparation methods was assessed.

Initial sample handling (all samples)

After the pilot experiments, each of the remaining 15 mL sample tubes were thawed for up to 2 hours at 4 °C and then vortexed at 2000 g for 10 seconds using a bench vortex before immediately and rapidly obtaining up to five 250 µL aliquots, depending on the original volume. These WSH aliquots were either kept in the fridge/on ice for immediate use or returned to -20 °C. Approximately one in ten of the 15 mL sample tubes had a sediment that was solidly wedged into the conical tip of the sample tube and resistant to vortexing. This was more common in large (greater than 5 mL) volume samples. These samples required tipping the tube almost horizontally and tapping the tip end sharply a few times before vortexing as above. Samples that contained yellow mucus were not able to be fully homogenised by vortexing. These samples were all of sufficient volume that it was possible to take aliquots from the presumed homogenous, non-mucoid component.

Total protein assay method (all samples)

Total protein estimation was performed on 175 saliva samples (an additional two saliva samples were available that did not have associated dental health data). WSH aliquots were thawed, centrifuged at 10,000 g for 3 minutes at 4 °C and a WSS supernatant obtained. Each sample WSS was assayed in triplicate using the Bradford (Coomassie G-250) standard microplate method and BSA as the standard as described above. The resulting 175 total protein concentrations were plotted as a boxplot using Excel and a histogram using R.

4.1.3 Testing salivary LL37 and statherin levels

ELISA pilot studies

These pilot studies were designed to determine the optimal sample preparation, dilution and plate incubation conditions when using the selected commercial ELISA kits.

Sample preparation and dilutions were determined from the manufacturers' recommendations, the reported concentrations in published literature and the testing range stated for the kits. The Cusabio statherin kit recommended centrifuging the saliva prior to plating, but did not suggest any dilution factor for saliva. The Hycult LL37 kit manual did not specify pre-analysis preparation for saliva but it did recommend diluting saliva at least 5 x in the buffer provided. The dilutions were mixed by adding the sample to the buffer provided with the kit and slowly pipetting up-and-down several times to avoid bubbles. WSH and WSS are defined above.

For the LL37 ELISA pilot: three samples were assayed as WSH and WSS at 1:4 dilution and two of those samples were assayed as WSH and WSS at 1:25 dilution and 1:50 dilution.

For the statherin ELISA pilot: two samples were assayed as WSS at 1:4 dilution and 1:10 dilution.

All sample permutations were assayed in quadruplicate with one pair undergoing the initial incubation as per the manufacturers protocol (1 hour at room temperature for LL37; 2 hours at 37°C for statherin) and another pair incubating overnight at 4 °C. In all cases the test standards were provided in the kits, and were always assayed in duplicate along with the samples. The subsequent ELISA steps followed the manufacturers' protocols (full details in the following section).

ELISA assay methods (all samples)

There were 173 saliva samples with associated health data and sufficient volume to be assayed for at least one of the selected analytes and 165 samples of sufficient volume for both of the analytes.

All the kits employed the sandwich ELISA method and provided all the standards, buffers and reagents required. The kits are listed in Table 2 and further supplier details are included in the supplementary section (Supplementary Data 1).

All kit components were stored, reconstituted and/or diluted as per manufacturers protocol and brought to room temperature before use. The protocols supplied with the ELISA kits were followed except where indicated in italics (changes based on the respective pilot studies).

Saliva aliquots used for the Hycult LL37 ELISA were v*ortexed for 5 seconds at 2000g with a benchtop vortex and then diluted 1:4* with the wash/dilution buffer provided with the kit. Saliva aliquots used for the Cusabio Statherin ELISA assay were *vortexed for 5 seconds at 2000g* with a bench top vortexed before centrifuging for 10 minutes at 4000 g at 2-8 °C and finally *diluting the supernatant 1:4* with the sample diluent provided in the kit.

100 µL of either diluted sample or one of a serial dilution of the provided standards were added to each pre-coated microplate well. Standards were assayed in duplicate for every plate while *samples were generally assessed in single wells*.

For both kits, the first incubation was then carried out *overnight (> 20 hours) at 4°C*.

LL37 kit: After the first incubation, the plate/s were washed to removed non-bound material. This was carried out using a ThermoScientific WellWash Versa platewasher using the provided wash/dilution buffer at a volume of 200 µL per wash for five washes. This was immediately followed by the addition of 100 µL of the anti-LL37, biotinylated, tracer antibody to each well. The plate/s were incubated for one hour at room temperature before a second washing procedure was carried out using the same plate washing procedure as before to remove unbound tracer. Immediately after the second wash, 100 µL of streptavidin-peroxidase conjugate was added to each well and the plate/s incubated for another hour at room temperature before the third and final wash (same procedure as before). Thereafter 100 µL of the 3,3',5,5' – tetramethylbenzidine substrate (TMB) was added and allowed to react (protected from light) with the peroxidase for 20 -30 minutes before the addition of 100 µL of the acidic stop solution to achieve a final, standardized total volume in each well.

Statherin kit: After the first incubation, as much liquid as possible was withdrawn from the wells using a pipette with care taken not to touch the sides or bottom of the wells before tipping the plate/s upside down onto absorbant paper to remove any excess liquid remaining. Without prior washing, 100 µL of anti-statherin,

biotinylated tracer antibody was then added to each well and the plates were covered with an adhesive strip and incubated for one hour at 37 °C. The plate/s were then washed using 200 µL of the provided wash buffer per well and the WellWash automated platewasher for a total of three washes. Immediately after the washing procedure, 100 µL of streptavidin-peroxidase conjugate was added to each well and the plate/s incubated for another hour at 37 °C before the third and final wash (200 µL of wash buffer per well repeated five times). Thereafter, 90 µL of the TMB subtrate was added and allowed to react (protected from light) with the peroxidase for 20 -30 minutes before the addition of 50 µL of the stop solution.

In all cases, after the addition of the stop solution the plates were covered and the solutions in the wells mixed by using an orbital shaker for 30 seconds. Any remaining air bubbles were removed by gently tapping the sides of the plate. The plates were read with a BioTek ELx808 plate reader within 5 - 10 minutes of the addition of the stop solution.

As well as stopping the oxidation of TMB by peroxidase, the acidic stop solution shifts the absorbance spectrum of the substrate from 650 nm (blue) to 450 nm (yellow). Therefore, the absorbance was read at 450 nm.

The absorbances of the standard and sample wells were normalised to the mean absorbance of the blank (0 mg/mL) wells.

Four- point (statherin) or five-point (LL37) logistic regressions were used to produce the standard curves, as per manufacturers' recommendations. These were performed using the online tool myassays.com. Calculation of the analyte concentrations were thereafter also determined using the myassays.com tool. It should be noted that this online tool is dependent on user set-up and as such there can be some discrepancy between assays in terms of the methods used to normalise results. Therefore, quality control checks of the results were performed using Excel to ensure normalisation was carried out appropriately.

Plots of the concentration results were prepared using Excel (boxplots) and R (histograms).

4.1.4 Testing for possible associations between the analytes and caries experience

Descriptive analysis of ELISA assay results

The mean analyte concentrations of the saliva samples, including the concentrations non-normalised, normalised to total protein content and as the product of analyte concentration by spit volume collected, were calculated for those with versus those without caries experience using Excel.

Spearman's Rank Correlation test was performed to assess for any correlation between LL37 and statherin concentrations. The same test was used to assess for correlations between LL37 or statherin concentration and the total protein concentration or volume of the saliva sample. These correlation tests were performed using R.

Inferential statistical analysis

Spearman's Rank Correlation tests to assess for correlations between caries experience as a score and LL37 or statherin concentrations (non-normalised) were performed using R.

Potential associations between caries experience as a binary variable and LL37 or statherin concentration were then analysed by unadjusted binary logistic regression models, also using R (unadjusted models are referred to as Model A).

In addition, binary logistic regression models including additional variables were designed in order to explore and control for possible confounding. Two tools were utilised in order to guide the design of the multivariable models – Directed Acyclic Graphs (DAGs) and statistical tests of bivariate relationships.

The DAG models (Joffe & Gambhir *et al.*, 2012) were prepared separately for LL37 and statherin using the online tool at daggity.net, and were based on the following observations from previously published research (as discussed further in Section 5.1):

- LL37 effects the probability of caries development primarily via its effects on harmful bacteria, but also possibly through other mechanisms, with the level of LL37 in spit possibly being influenced by age, sex, vitamin D supplementation and saliva flow
- statherin effects the probability of caries development through its calcium carrying and bacteria inhibition roles and the level of statherin in spit is possibly influenced by saliva flow rate and sex
- bacterial dysbiosis at the tooth surface is a major determinant of caries development in all cases and it is influenced by the sugar content of the diet

and mechanical forces such as saliva volume and tooth brushing, as discussed in the introduction, with a possible influence from the analytes under investigation
- fluoride intake, either in the form of toothpaste, dental treatments or as supplements, directly influences caries risk (Clark & Keels *et al.*, 2020).

The DAG models (Figure 4) provided further support for sex, age, saliva flow and vitamin D as plausible confounders of the relationship between LL37 and the caries experience outcome, and sex and saliva flow as plausible confounders of the relationship between statherin and the caries experience outcome. Fluoride supplementation could also act a covariate, as could mother's education, toothbrushing habits and dietary sugar intake mediated by oral bacterial dysbiosis.

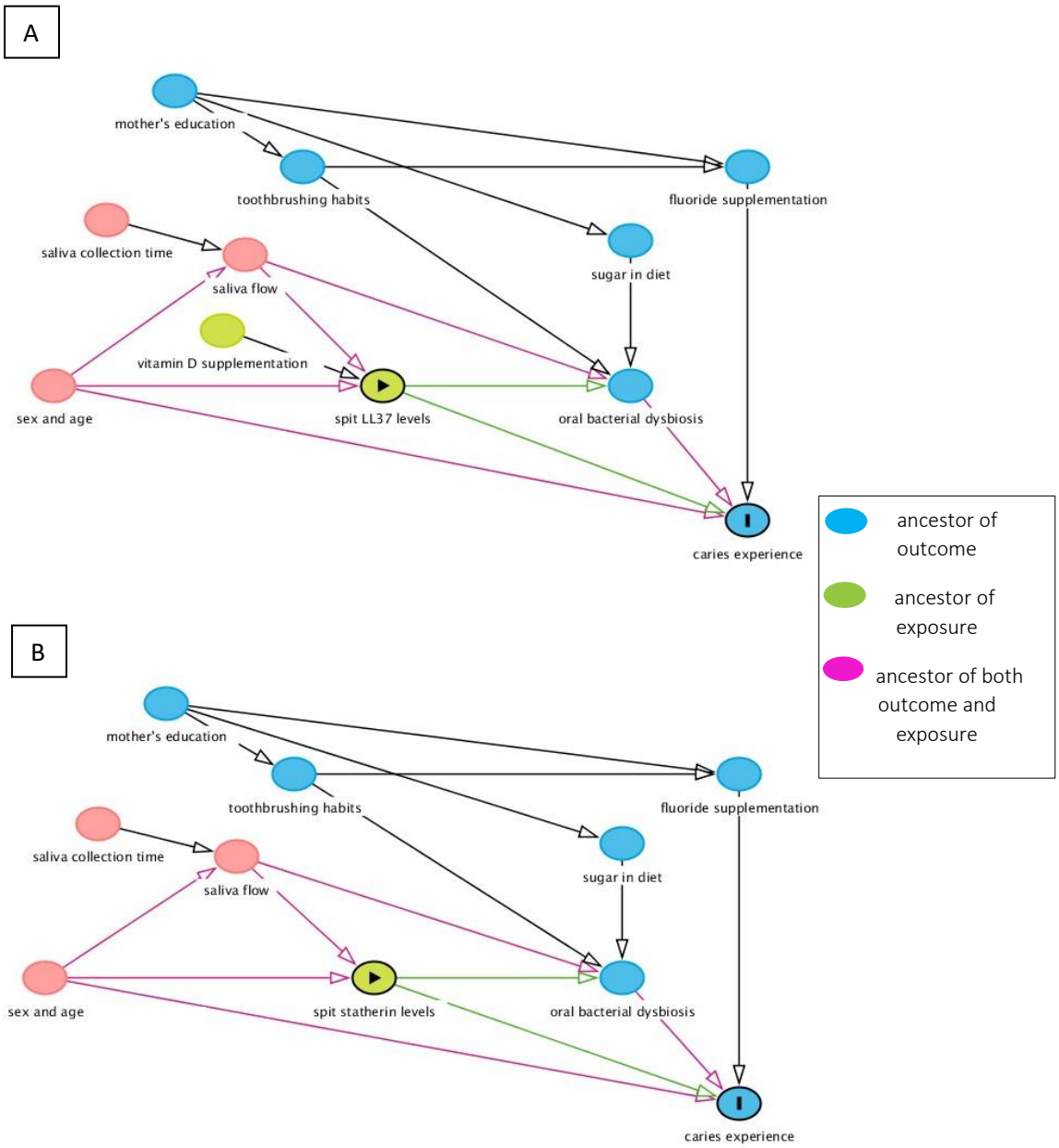


FIGURE 4: DAGS OF HYPOTHETICAL CAUSAL PATHWAY/NETWORKS INVOLVING A) LL37, AND B) STATHERIN LEVELS AND CARIES EXPERIENCE

The potential confounders identified by the DAGs were further explored for evidence of collinearity via plots and statistical tests (both using R) as detailed in Table 3 and Figure 5.

TABLE 3: BIVARIATE ANALYSES OF POSSIBLE CONFOUNDERS

Variable 1	***Variable 2***	***Test of relationship***	***Summary measure (95% conf. interval)***	***p-value****
Volume	Total protein concentration	Spearman's Rank Correlation	ρ = -0.22 (-0.36, -0.07)	0.004
Volume	Age	Spearman's Rank Correlation	ρ = 0.155 (0.005, 0.30)	0.04
Volume	Sex	Mann-Whitney U test	Est. difference = 1.1 mL (0.5, 1.8)	0.0007
Total protein concentration	Age	Spearman's Rank Correlation	No significant correlation	
Total protein concentration	Sex	Mann-Whitney U test	No significant correlation	
Age	Sex	Mann-Whitney U test	No significant correlation	
Collection time	TP concentration	Spearman's Rank Correlation**	No significant correlation	
Collection time	Volume	Spearman's Rank Correlation**	No significant correlation	

* p-values below 0.05 were considered statistically significant
**Also Kruskal-Wallis and Jonckheere-Terpstra tests
Units: volume (mL), total protein concentration (µg/mL), statherin and LL37 concentration (ng/mL), age (years).

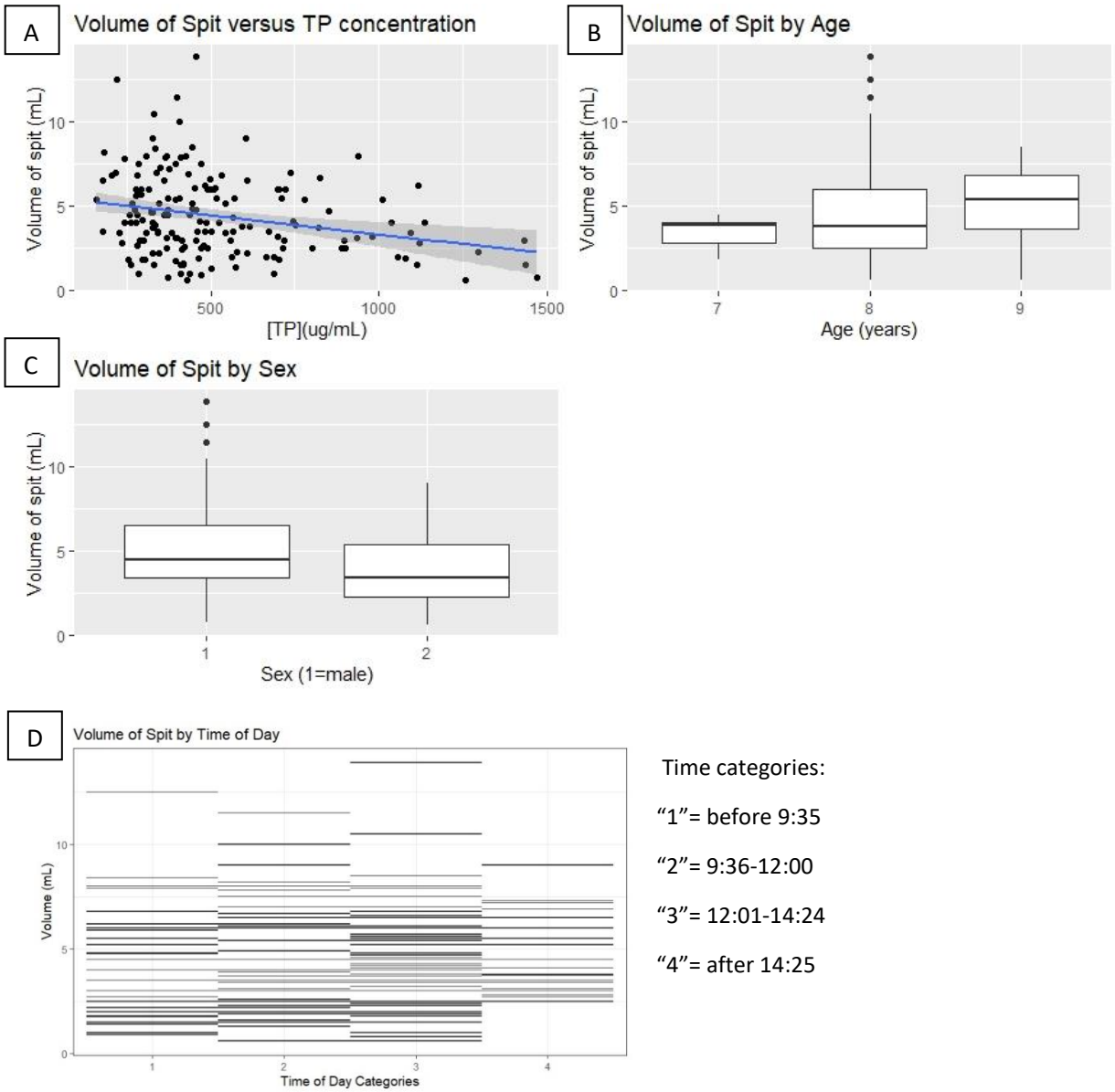


FIGURE 5: PLOTS OF THE CORRELATIONS BETWEEN VARIOUS SALIVA MEASUREMENTS: A) VOLUME VERSUS TOTAL PROTEIN (TP) CONCENTRATION; B) VOLUME OF SPIT BY AGE; C) VOLUME OF SPIT BY SEX; D) VOLUME OF SPIT BY COLLECTION TIME

As can be seen in Table 3, statistically significant ($p < 0.05$) correlations were found between collected saliva volume and TP concentration, and between collected saliva volume and the sex and age of the subject. However, there was no significant correlation between time of day and saliva flow (in contrast to passive flow studies, for example, Dawes (1972)) (Figure 5D).

The observation of these correlations, and the results from the DAG models, lead to the design of the following binary logistic regression multivariable models (Table 4). The use of more than one of the saliva flow variables in a model was avoided (one of sex, TP concentration or volume of spit was included). These multivariate

models were estimated as odds ratios (ORs) and 95% confidence intervals (95% CIs) using R. A p-value less than 0.05 was considered statistically significant.

TABLE 4: LOGISTIC REGRESSION MODELS

LL37	***age***	***vitamin D supplement***	***saliva flow variable***	***Other potential confounders****
Model A				
Model Bi	X	X	sex	
Model Bii	X	X	[TP]	
Model Biii	X	X	volume	
Model Ci	X	X	sex	X
Model Cii	X	X	[TP]	X
Model Ciii	X	X	volume	X

Statherin	***saliva flow variable***	***Other potential confounders****
Model A		
Model Bi	Sex	
Model Bii	[TP]	
Model Biii	Volume	
Model Ci	Sex	X
Model Cii	[TP]	X
Model Ciii	Volume	X

**mother's education (high school ≤13 yrs; university ≤4 yrs; university > 4 yrs), fluoride supplementation (0-5 times/week vs. 6-7 times/week), Lactobacillus level (≤10^5 CFU/mL vs. > 10^5CFU/mL), saliva collection time (before 9:35; between 9:36 and 12:00; between 12:01 and 14:24; after 14:25), toothbrushing habits (≤ once a day vs. ≥ twice a day), dairy dessert servings (never/rarely; 1-3 servings/month; 1-2 servings/week; 3-4 servings/week), sugary beverage servings (never/rarely; 1-3 glasses/month; 1-3 glasses/week; 4-6 glasses/week).*

Finally, further analysis for any correlation between caries experience as a binary variable and total protein concentration was performed (Mann-Whitney U/Wilcoxon rank sum test) and a boxplot of the TP concentrations for each outcome group was generated using R.

4.2 Results

4.2.1 Descriptive analysis of the sample population, health data and saliva

Demographic details of the participants are included in Table 5. For the sample population, and the "with caries experience" and "without caries experience" groups within that population, the exposure counts, proportions or means for the available variables are shown.

TABLE 5: BACKGROUND TABLE- EXPOSURE COUNTS, PROPORTIONS AND/OR MEANS FOR EACH VARIABLE IN RELATION TO WHOLE SAMPLE AND WITH- AND WITHOUT CARIES EXPERIENCE SUBGROUPS.

		ALL	*Without caries experience*	*With caries experience*
		Count	*Count (outcome proportions %)*	
		173	134 (77.5%)	39 (22.5%)
Variables	*Categories*	Count	*Exposure proportions(%)*	
age	7	5	3.0	2.6
	8	145	85.1	79.5
	9	23	11.9	18.0
sex	Male	89	50.8	53.8
	Female	84	49.2	46.2
Vitamin D supplementation	No intake	36	22.4	15.4
	Any intake	137	77.6	84.6
Fluoride supplementation[†]	0-5 times a week	78	41.8	56.4
	6-7 times a week	94	57.5	43.6
Education level (mother)	High school <=13 years	11	6.0	7.7
	University <=4 years	71	44.0	30.8
	University >4 years	91	50	61.5
Toothbrushing habits	≤ once a day	56	33.6	28.2
	≥ twice a day or more	117	66.4	71.8
Lactobacillus level[‡]	$<10^5$ CFU/ml	83	53.0	30.8
	$\geq 10^5$ CFU/ml	90	47.0	69.2

Streptococcus mutans level‡	$<10^5$ CFU/ml	140	83.6	71.8
	$\geq 10^5$ CFU/ml	33	16.4	28.2
Dairy dessert servings	Never/rarely	3	1.5	2.6
	1-3 times a month	92	57.5	38.5
	1-2 times a week	67	35.1	51.3
	3-4 times a week	11	6.0	7.7
Sugary, carbonated drink servings	Never/rarely	40	24.6	18
	1-3 glasses a month	81	45.5	51.3
	1-3 glasses a week	48	27.6	28.2
	4-6 glasses a week	4	2.2	2.6
Saliva collection time	before 9:35	34	17.2	28.2
	9:36 – 12:00	50	27.6	33.3
	12:01 – 14:24	66	39.6	33.3
	after 14:25	23	15.7	5.1
			Means	
		ALL	Without caries experience	With caries experience
	Saliva volume (mL)	4.43	4.44	4.40

† n= 172(data missing from one participant without caries); ‡ Strip test method (CRT-Bacteria, Ivoclar VIvodent Inc

Overall, the subjects were of an approximately equal gender mix. The ages ranged from seven years and two months to nine years and two months of age but the majority took part in the project between their eighth and ninth birthdays. The majority of mothers had some university education. Nearly half (42%) of the children took regular supplementary doses of vitamin D and fluoride, and the majority brushed their teeth twice a day and only consumed high sugar products in moderation (less than once a week). Ethnicity was not observed in this study.

The volumes of saliva collected ranged from 0.1 to 13.9 mL (mean +/- 1 SD = 1.92 – 6.8 mL). A histogram of the collected saliva volumes is shown in Figure 6.

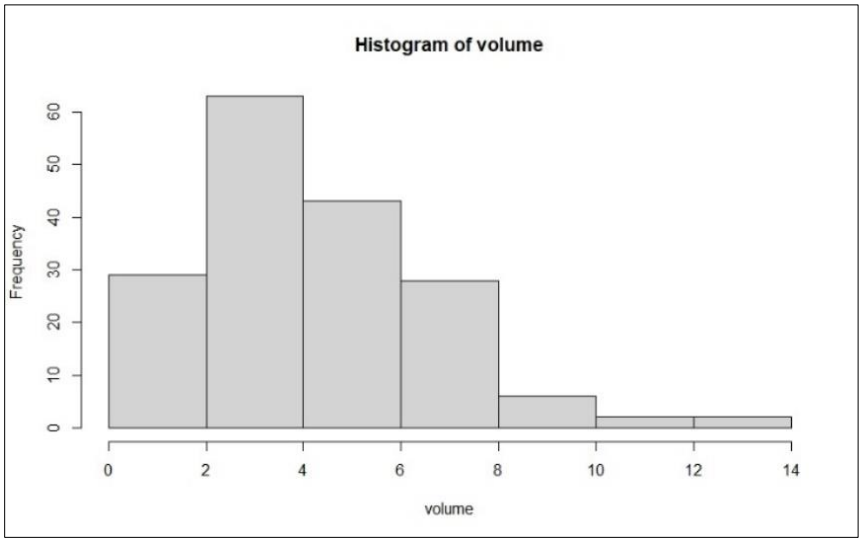


FIGURE 6: THE COLLECTED VOLUME OF THE SALIVA SAMPLES IN THIS STUDY

The proportions across outcome groups did not seem visibly very different for most variables. There may be some increase in bacteria scores for the with-caries group.

The counts of dmft/DMFT scores were as follows: score 0 (n=134), score 1 (n=19), score 2 (n=11), score 3 (n=5), score 4 (n=2), score 5 (n=0), score 6 (n=1) and score 7 (n=1). The mean dmft/DMFT score was 0.44 with a standard deviation of 1.05.

4.2.2 Determining best pre-analytical sample preparation was important

In order to optimise the analyses of the chosen saliva components, efforts were made to first characterize some aspects of the saliva samples through the initial pilot studies.

The time taken to fully thaw at 4 °C ranged from 60 minutes to 90 minutes. This appeared (visually) to roughly correlate to saliva volume. A thin layer, presumably of lipids and surfactants, was visible at the top of most samples once thawed. Some samples (not in the pilot group) contained thick, yellow mucus or blood. By closely observing the samples as they thawed it was evident that whole saliva readily forms a sediment. In all cases, the collected saliva had already formed a sediment before it froze (see Figure 7A) and after re-suspending a sample using a bench top vortex at 2000 g for 10 seconds and dividing into aliquots, a visible sediment reappeared within 5 minutes at room temperature (see Figure 7B). Therefore, it was concluded that whenever working with whole saliva it should be re-vortexed every few minutes to try and prevent sedimentation from adversely affecting precision.

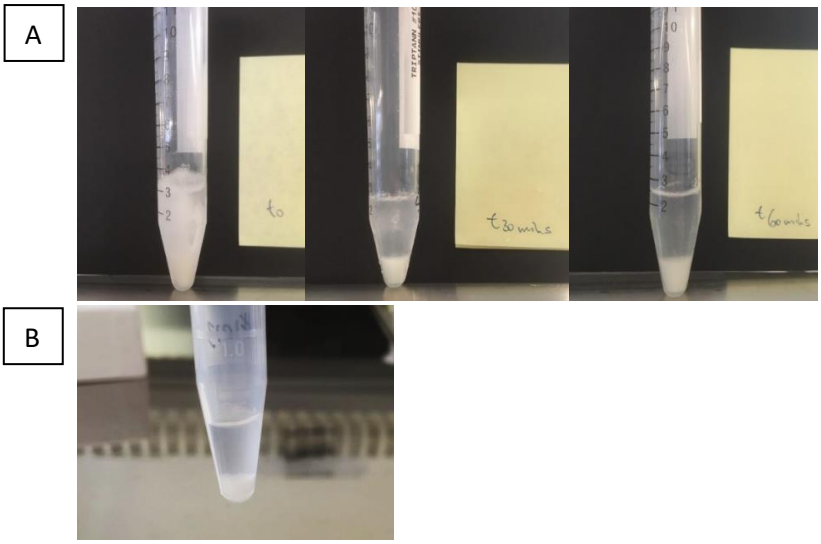


FIGURE 7: A) THAWING OF WHOLE SALIVA AT 4 °C, SHOWING SEDIMENT ALREADY PRESENT IN THE FROZEN STATE; B) RE-SEDIMENTATION OF VORTEXED WHOLE SALIVA WITHIN 5 MINUTES AT ROOM TEMPERATURE

Neither the supernatant removed after an hour of passive WSH sedimentation (PS) or the supernatant obtained after WSH centrifugation (WSS) formed visible sediment within seven days.

Total protein content in different fractions of saliva showed that total protein concentration estimates were usually highest for the vortexed whole saliva (WSH) preparations, but the intraassay precision (%CV) of measurements from WSH was poorer than for "clarified" i.e. centrifuged WSH saliva (WSS). Consistent with the visible sedimentation of WSH, the total protein concentration one hour after vortexing dropped on average 26.4% (range 18 – 54%) (Supplementary Data 3).

While there was no visible additional sedimentation of the supernatants, their total protein estimation when pipetting from the middle of the volume did decrease slowly. There was very little decrease in the first hour (< 3%), however the TP concentration had dropped on average 38% of the original concentration for the PS and 24% for the WSS after seven days (Supplementary Data 3). The variability of total protein estimation of freshly prepared PS was found to be high, especially interassay (results not shown).

Dilution of saliva, in any form, was not necessary to keep the absorbance within the linear range of the assay's standard curve. Attempts to analyse total protein content of whole saliva diluted in ELISA buffer showed that, although visible sedimentation was prevented, the background absorbance was too high to enable accurate colorimetric measurements, presumably due to the detergent content[6].

[6] The exact composition of the dilution buffer is not stated by the kit manufacturers. ELISA buffers are most commonly PBS with 0.05% Tween. Detergents are known to interfere with the Coomassie assay (Thermofisher Scientific Protein Assay Technical Handbook, 2017).

The median salivary total protein concentration of the 175 samples prepared by vortexing and centrifugation was 434.3 µg/mL with a range of 160 to 1471 µg/mL (St. dev. 266.7; Figure 8; Supplementary Data 6).

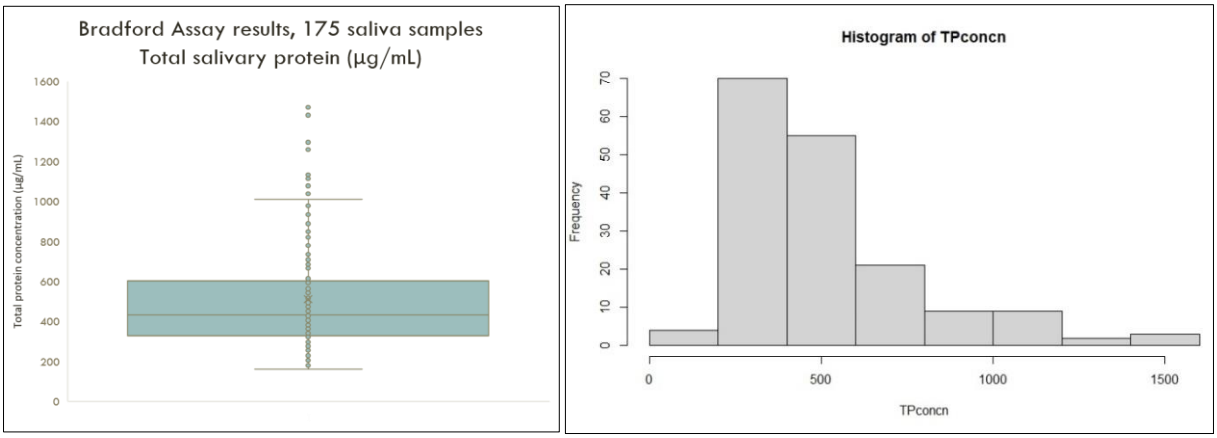


FIGURE 8: THE TOTAL PROTEIN CONCENTRATION OF 175 SALIVA SAMPLES

4.2.3 Cathelicindin (LL37) was found in varying levels in saliva

The pilot studies with the Hycult LL-37 kit found that LL37 was bound up in the sediment and released, at least to some extent, by vortexing as demonstrated by higher concentration results. The 1:4 dilution was sufficient to obtain absorbance readings within the standard range and overnight incubation only increased sensitivity of very dilute preparations, at the expense of precision (Supplementary Data 4). The intraassay variability (%CV) for samples assayed as whole saliva as described was below 8% and generally lower than that for supernatant preparations. Thus, taking together the higher concentration and good precision seen in the pilot studies, it was concluded that it was best to use the vortexed whole saliva diluted 1:4 in the provided buffer for this assay.

The standards for the kit were of high quality, with %CVs consistently below 10% and the standard curve derived using a five-point logistic regression consistently producing an R^2 of above 99.9 %.

Following this procedure, the LL37 concentration was determined for 173 of the saliva samples (Supplementary Data 6). The range was 12.42 – 270.7 ng LL37/mL saliva; median 60.1 ng/mL, mean 72.2 ng/mL with a standard deviation of 46.1 (Figure 9). The LL37 to total protein ratios were 0.01 - 0.54 (median 0.13) ng per µg of total protein content. The concentrations were non-normally distributed (Figure 9).

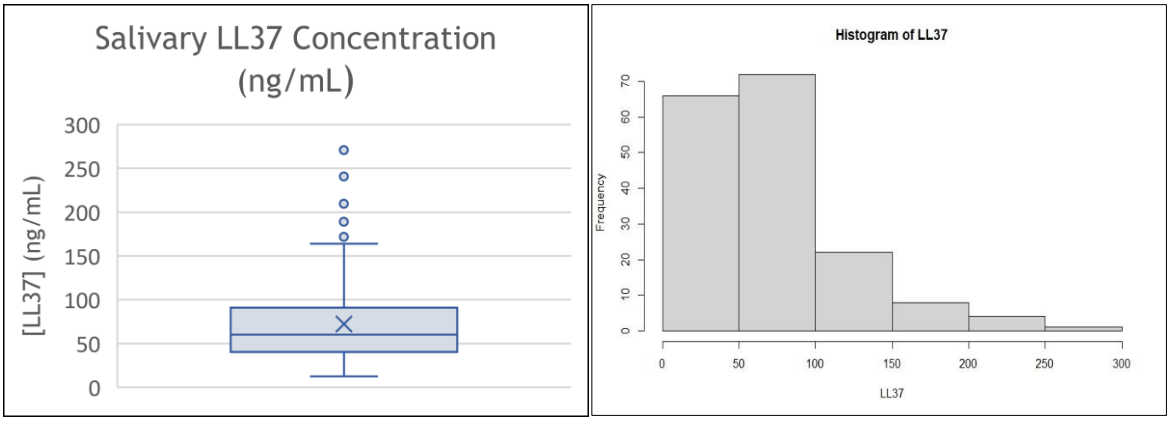


FIGURE 9: LL37 CONCENTRATION (NG/ML) IN 173 SALIVA SAMPLES DETERMINED BY ELISA.

4.2.4 Statherin was generally quite abundant in saliva

No pilot studies into WSH were performed because the manufacturer of the selected ELISA kit (Cusabio) specifies the use of saliva supernatant obtained after centrifugation. The pilot studies focussed on the best dilution and incubation conditions and found that a 1:4 dilution produced absorbance measurements within the standard range and the precision was greater if the first incubation was allowed to occur overnight (>20 hours) at 4 °C rather than two hours at 37 °C as specified in the product manual. The coefficient of variability for samples analysed this way was consistently below 8% (see Supplementary Data 5). The %CVs for the standards was also better when allowed to incubate overnight, although they were still not always all below 10%. The R^2 using the recommended 4-point logistic regression was always above 99%.

Following this procedure, the statherin concentration was determined for 165 saliva samples in total (Supplementary Data 6). Eight samples that had been assayed for LL37 and total protein concentration did not have sufficient volume to also carry out the statherin assay. The range was 581- 9685 ng statherin/mL saliva; median 3784 ng/mL, mean 3766 ng/mL with a standard deviation of 1591 (Figure 10). The statherin to total protein ratios were 1.46 – 27 ng per µg of total protein content. The distribution of concentrations was somewhat skewed to the right (Figure 10).

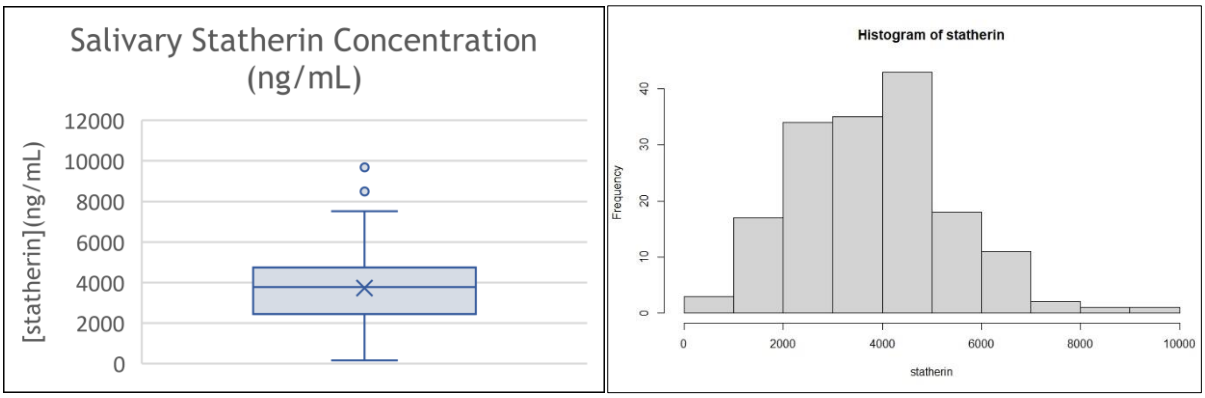


FIGURE 10: STATHERIN CONCENTRATION (NG/ML) IN 165 SALIVA SAMPLES DETERMINED BY ELISA

4.2.5 Descriptive analysis of assay results

The mean analyte concentrations of the saliva samples, with the concentrations non-normalised, normalised to total protein content and as the product of analyte concentration by spit volume collected, for those with versus those without caries experience, are shown in Table 6.

TABLE 6: DESCRIPTIVE ANALYSIS OF THE TOTAL PROTEIN AND ELISA ASSAY RESULTS BY CARIES EXPERIENCE IN THIS SAMPLE POPULATION

	ALL	*Without caries experience*	*With caries experience*
	Count	*Count (outcome proportions %)*	
	173	134 (77.5%)	39 (22.5%)
		Means	
Saliva analysis variable	ALL	Without caries experience	With caries experience
Saliva volume (mL)	4.43	4.44	4.40
Total protein (µg/mL)	512.37	474.85	641.25
LL37 (ng/mL)	72.22	70.17	79.29
LL37 (ng/µg total protein)	0.162	0.17	0.15
LL37 (ng)	302.2	293.73	331.3
Statherin (µg/mL)	3.77▫	3.60	4.33◇
Statherin (ng/µg total protein)	8.68▫	8.80	8.24◇
Statherin (µg)	17.03▫	16.46	18.99◇
			▫ n= 165; ◇ n= 37

LL37 was diluted in large volume samples

The analytes (LL37 and statherin concentration) were found to not be significantly correlated to each other (test statistic S = 616544, p-value = 0.023; rho = 0.1765). LL37 concentration was significantly correlated to TP concentration (test statistic S = 624099, p-value = 0.0002; rho = 0.2777) and had a statistically significant negative correlation with volume (S = 1045145, p-value = 0.005; rho = -0.2112). Statherin was significantly correlated to TP concentration (S = 505800, p-value = 2.13e-05; rho= 0.3244) but not to volume (S = 754594, p-value = 0.92).

4.2.6 Inferential statistical analyses of analyte levels and caries experience

Although a trend towards an inverse relationship between LL37 concentration and caries experience (as a score from 0-7) seemed present in the plot (Figure 11), this was not supported by the Spearman's Rank Correlation test (Spearman's rank correlation statistic S = 789620, sample estimate rho = 0.085, p-value of H_0 = 0.27). There was statistical support at the 5% significance level for a positive correlation between caries experience score and statherin concentration (Spearman's rank correlation statistic S = 599443, rho = 0.20, p-value of H_0= 0.01).

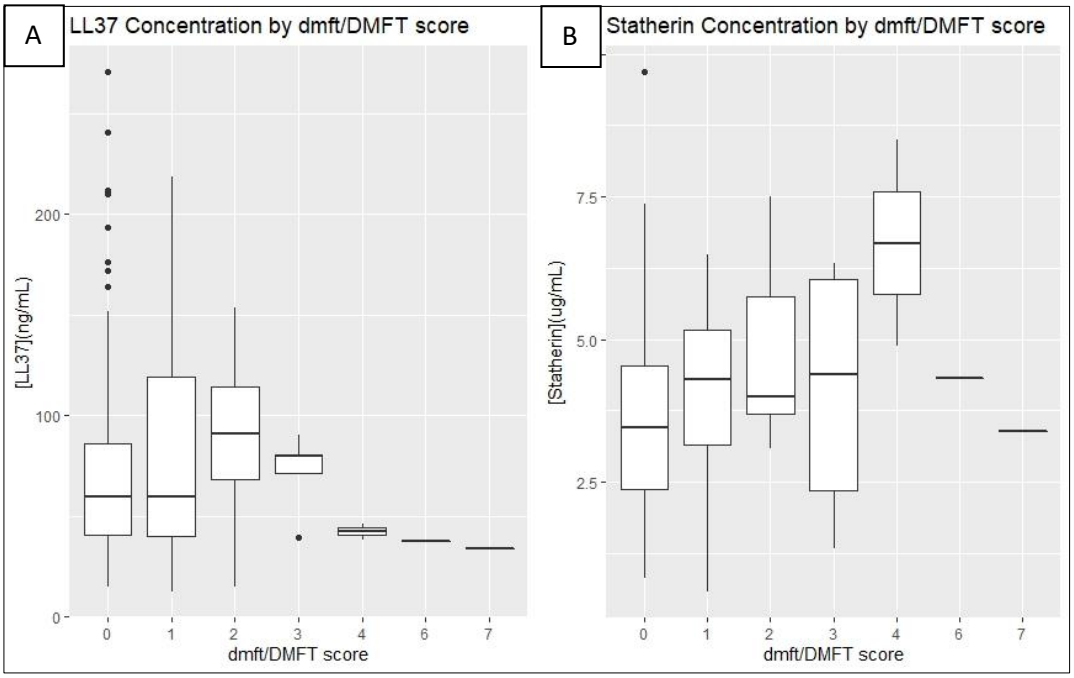


FIGURE 11: A) LL37 CONCENTRATION VERSUS DMFT/DMFT SCORE, B) STATHERIN CONCENTRATION VERSUS DMFT/DMFT SCORE

Table 7 summarises the results of the logistic regression analyses of the dichotomised caries experience outcome dependent on LL37 concentration. This includes the unadjusted model and adjusted models. For the R code and results for the logit modelling described in this section, see Supplementary 7.

TABLE 7: THE RESULTS OF LOGISTIC MODELLING OF CARIES EXPERIENCE ON LL37 CONCENTRATION (NG/ML).

LL37	***OR***	***95% CI***	***p-value of coefficient***
Model A (unadjusted)	1.004	0.996-1.011	0.28
Model Bi	1.005	0.996-1.012	0.22
Model Bii	1.004	0.996-1.011	0.36
Model Biii	1.005	0.997-1.012	0.24
Model Ci	1.003	0.99-1.01	0.44
Model Cii	1.002	0.99-1.01	0.57
Model Ciii	1.003	0.99-1.01	0.43

Refer to Table 4 for details of the variables included in these models
n = 173 (39 with caries, 134 without caries experience)
Model C (i, ii and iii) one participant without caries experience missing the fluoride supplement variable

None of the models showed a significant association between caries experience and LL37 concentration. The p-values were more than 0.05 and the OR estimates were close to one with a narrow 95 % confidence interval straddling 1.0. The estimates were not altered by adjusting the model.

The results of the logistic regression analyses of caries experience dependent on statherin concentration, unadjusted and adjusted models, are summarised in Table 8.

TABLE 8: THE RESULTS OF LOGISTIC MODELLING OF CARIES EXPERIENCE ON STATHERIN CONCENTRATION (MG/ML).

Statherin	***OR***	***95% CI***	***p-value of coefficient***
Model A (unadjusted)	1.33	1.06-1.69	0.017*
Model Bi	1.33	1.06-1.69	0.016*
Model Bii	1.22	0.96-1.57	0.11
Model Biii	1.33	1.06-1.69	0.017*
Model Ci	1.31	1.02-1.72	0.04*
Model Cii	1.19	0.90-1.59	0.22
Model Ciii	1.31	1.02-1.72	0.04*

Refer to Table 4 for details of the variables included in these models
**a p-value less than 5% was considered statistically significant*
n = 165 (37 with caries, 128 without caries experience)
Model C (i, ii and iii) one participant without caries experience missing the fluoride supplement variable

The unadjusted regression (Model A) showed a statistically significant positive correlation between caries experience and statherin concentration. The odds ratio was 1.33 (per µg/mL) (95% CI 1.06 - 1.69).

Controlling for saliva flow (as sex or volume of spit) did not alter the strength of association between statherin and caries experience. The addition of the variables education, fluoride supplementation, Lactobacillus score, toothbrushing habits, diet variables and time of saliva collection (Models Ci and Ciii) also had little effect.

When TP concentration was used as the saliva flow variable for either analyte, the size and significance of the association between the analyte and caries experience was reduced (Models Bii and Cii). For statherin, this was sufficient to remove the significance of the association estimate (p-value >10%).

A statistically significant correlation between total protein concentration and caries experience was observed (Figure 12; Mann-Whitney U test, p-value 0.0085; estimated difference 117.9 µg/mL; 95% CI 29 – 216).

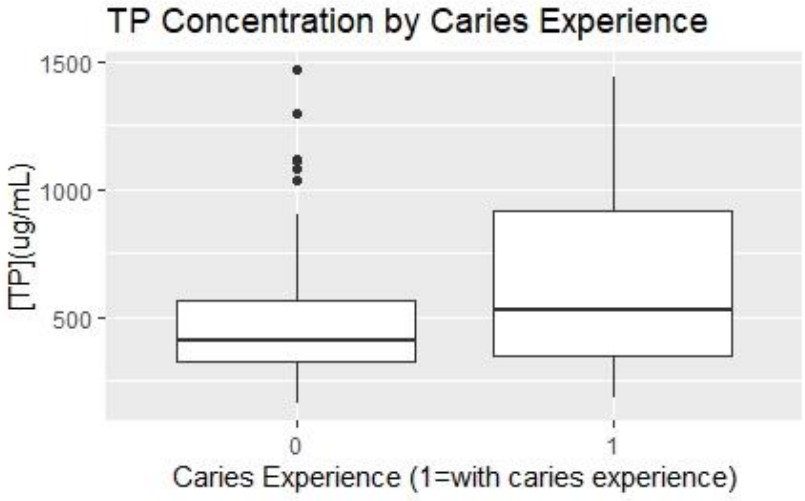


FIGURE 12: TP CONCENTRATION BY CARIES EXPERIENCE

5. DISCUSSION

A review of the literature showed that, despite growing interest in the topic, there exists only a moderate number of published papers detailing investigations into associations between saliva composition and caries in children. The saliva components analysed and methods used varied widely between projects. From an initial list of 18 analyte candidates, several were ruled out due to concerns regarding polymorphisms or a lack of available commercial ELISA kits verified for use with saliva samples.

5.1 The two saliva components chosen to analyse were LL37 and statherin

5.1.1 Cathelicidin-derived LL37: a direct antimicrobial peptide

The antimicrobial peptide LL-37 is produced by the proteolytic cleavage of a precursor- the 170 amino acid long, Human Cathelicidin Antimicrobial Peptide Preproprotein, also known simply as Human Cathelicidin.

The Cathelicidins are a superfamily of peptides, highly conserved through vertebrates (Bals & Wilson, 2003; Peschel & Sahl, 2006). Members of this family possess two peptide domains, namely the N-terminal cathelicidin signal domain (approximately 100 amino acids) and a C-terminal antimicrobial peptide domain (approximately 27 amino acids). Figure 13 shows the Human Cathelicidin (NP_004336.4) sequence aligned with the cathelicidin protein domain consensus sequence in blue (pfam00666) (Zanetti & Gennaro *et al.*, 1995) and the LPS-binding, antimicrobial peptide domain consensus sequence in red (pfam12153)(C. Chen & Brock *et al.*, 1995) using BLAST (Altschul *et al.*, 1990).

```
Human    21   GLVMPLAIIA   QVLSYKEAVL   RAIDGINQRS   SDANLYRLLD   LDPRPTMDGD   70
                           QVLSY+EAVL   RA+D  N++S   S+ANLYRLL+   LDP P  D D
Consensus                  QVLSYREAVL   RAVDQFNEQS   SEANLYRLLE   LDPPPQDDED

Human    71   PDTPKPVSFT   VKETVCPRTT   QQSPEDCDFK   KDGLVKRCMG   TVTLNQARGS   120
              PDTPKPVSF    VKETVCPRTT   QQ PE CDFK   +DGLVKRC+G   TVTLNQAR S
Consensus     PDTPKPVSFR   VKETVCPRTT   QQPPEQCDFK   EDGLVKRCVG   TVTLNQARDS

Human   121   FDISCDKDNK   RFALLGDFFR   KSKEKIGKEF   KRIVQRIKDF   LRNLVPRTES   170
              FDISC++D    R     DF R   K  EKIG      K I Q IKDF     N
Consensus     FDISCNEDQS   R     DFLR   KGGEKIGEKL   KKIGQKIKDF   FQN
```

FIGURE 13: THE SEQUENCE OF HUMAN CATHELICIDIN ALIGNED WITH CONSENSUS SEQUENCES

The antimicrobial peptides that are derived from cathelicidin exert direct antibacterial effects against a broad range of gram positive and negative bacteria, including *S. mutans* (Nilsson, 2020). The cathelicidins have also been shown to bind to the bacterial endotoxin lipopolysaccharide (LPS) and have chemotactic properties (Aidoukovitch *et al.*, 2020; Nilsson, 2020; Peschel & Sahl, 2006; G. Wang & Mishra *et al.*, 2014).

In humans, there is a single cathelicidin gene, the *CAMP* gene on chromosome 3 (Table 9; HGNC Gene ID 1472; OMIM ref 600474; location 3p21.31). There are no documented associations between polymorphisms in *CAMP* and caries risk in modern humans, however there is evidence of historical evolution in *CAMP* (Peschel & Sahl, 2006) and ongoing evolution of commonly encountered bacteria (Phattarataratip & Olson *et al.*, 2011) reflecting the dynamic balance between the pathogens and the AMP.

TABLE 9: HUMAN CATHELICIDIN GENE, TRANSCRIPT AND PRODUCTS

HGNC			NCBI			Translated protein features
Approved gene name	Gene Symbol	Gene ID	Gene ID	Transcript RefSeq	Translated protein RefSeq	
Cathelicidin Antimicrobial Peptide	*CAMP*	1472	820	NM_004345.5	NP_004336.4	170 amino acid long, 19.3kDa pre-proprotein

CAMP is expressed by granulocytes, lymphocytes and epithelial cells of the mucosae (Agerberth & Charo *et al.*, 2000; Aidoukovitch *et al.*, 2020) and by salivary gland cells to a lesser degree (Nilsson, 2020). The mRNA transcript is translated into Human Cathelicidin which then undergoes post-translational modifications including the formation of two disulphide bonds and is trimmed into a 140 amino acid long precursor before being stored in cytoplasmic granules (Lehrer & Ganz, 2002). Upon degranulation, the propeptide hCAP18 become incorporated in bacteriocidal neutrophil extracellular traps (NETs) (Mohanty & Sjögren *et al.*, 2015) and/or is further cleaved by proteases in the oral cavity (Aidoukovitch *et al.*, 2020). This cleavage can occur at a number of sites in the propeptide, producing AMPs of slightly different length. LL-37 is the most common and most studied of the human cathelicidin-derived AMPs (G. Wang *et al.*, 2014).

The direct antimicrobial action of LL37 is through disruption of the bacterial cell membrane (Aidoukovitch *et al.*, 2020). While the precise mechanism has not conclusively been determined as yet, there is evidence of pore formation (G. Wang *et al.*, 2014). Recent studies into the supramolecular structure of LL37 have

revealed a self-assembling, thermostable, hollow tubular structure several micrometres long that co-locates with bacteria and interacts with them in a way that is disturbed by selective mutagenesis (Engelberg & Landau, 2020) (Figure 14).

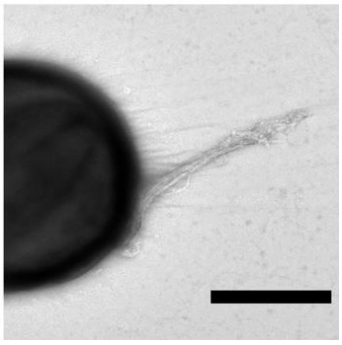


FIGURE 14: IMAGE TAKEN FROM ENGELBERG AND LANDAU (2020): ELECTRON MICROSCOPY OF LL37 FIBRILS INTERACTING WITH A BACTERIUM (MICROCOCCUS LUTEUS) (SCALE BAR REPRESENTS 500 NM).

The chemotactic properties of LL-37 are also of current research interest. There is evidence that it can act as a ligand for a number of receptors including the formyl-peptide receptors of neutrophils and eosinophils, thereby initiating inflammatory cell movement (Tjabringa & Ninaber *et al.*, 2006).

The final activity of LL-37 in human saliva against bacteria is therefore determined by the regulation of gene expression, the subsequent peptide modification, transportation, exocytosis and supramolecular assembly, as well as GCF flow rate and the presence of other saliva components such as proteases, mucins and neutrophils. Additionally, transcription of *CAMP* is regulated by vitamin D and the VDR (Koivisto & Hanel *et al.*, 2020), and cytokines (Agerberth *et al.*, 2000). The activity of the proprotein activating proteases is subject to neurohormonal regulatory mechanisms including the sex hormones (Proctor, 2016) and supramolecular structure formation is determined by the sequence and concentration of LL37 itself (Engelberg & Landau, 2020). The importance of LL37 to human oral health is illustrated by Severe Congenital Neutropaenia (Morbus Kostmann Syndrome) which includes early onset periodontal disease mediated at least in part by an LL37 deficiency (Nilsson, 2020).

5.1.2 Statherin: an enamel-maintaining peptide

Statherin is an important component of saliva and the acquired enamel pellicle. In humans, it is secreted by the parotid and submandibular salivary glands and is typically present in saliva in µg/mL concentrations making it one of the more abundant saliva proteins (Huq *et al.*, 2007). Via charged glutamic acid residues

and a hydrophobic domain, statherin prevents calcium crystals forming in the calcium-supersaturated saliva and thus facilitates the supply of calcium ions needed for enamel remineralisation, which in turn maintains an enamel surface that is smooth and not amenable to colonisation by pathogenic bacteria (Helmerhorst *et al.*, 2010).

Figure 15 shows Human Statherin A aligned with the Statherin family consensus sequence (pfam03875) using BLAST. The first 20 amino acids are considered the signal region, the region in purple is the Hydroxyapatite-interacting domain and the hydrophobic domain is in orange.

```
Human    11  ALMVSMIGAD    SSEEKFLRRI    GRFGYGYGPY    QPVPEQPLYP    QPYQPQYQQY  TF 62
                       D    SSEEKFLRR+     RF  GYGPY    QP PEQPLYP    QPYQP YQQY
Consensus              D    SSEEkflrrl    rrfdGGYGPY    QPFPEQpLYP    QPYQPYYQQY
```

FIGURE 15: THE SEQUENCE OF HUMAN STATHERIN ALIGNED WITH THE CONSENSUS SEQUENCE

Statherin is coded for by the *STATH* gene on chromosome 4 (HGNC Gene ID 11369; OMIM ref. 184470; location 4q13.3). Alternative splicing produces two transcripts and two protein isoforms (see Table 10). There is no evidence to date that polymorphisms in *STATH* or the two isoforms contribute to caries risk (Lips *et al.*, 2017).

TABLE 10: HUMAN STATHERIN GENE, TRANSCRIPT AND PROTEIN DETAILS.

HGNC			NCBI			Uniprot	
Approved name	Gene Symbol	Gene ID	Gene ID	Transcript RefSeq	Translated protein RefSeq	Protein ID	Features
Statherin	*STATH*	11369	6779	NM_003154 679nt	NP_003145.1	P02808-1	Isoform A; 62 aa
				NM_001009181.2 649nt	NP_001009181.1	P02808-2	Isoform B; 52aa

The concentration of statherin in saliva is determined primarily by salivary gland secretion and contraction (Jensen *et al.*, 1995) but is also influenced by underlying health, for example diabetes (Izumi & Zhang *et al.*, 2015).

5.2 Choice of pre-analytical and data normalising methods are important

Pre-analytical handling of saliva samples

After reviewing the published literature, it was noted that the majority of researchers prefer to "clarify" saliva samples before analysis. Sometimes this involves the addition of reagents that initiate mucin precipitation and/or physical filtration but most commonly clarifying is achieved through centrifugation. The speed and time of centrifugation varied markedly. Questions arose, therefore, regarding how important it is to "clarify" saliva before analysis and whether this depends upon the type of assay being considered.

Interactions between saliva components are well documented, and exist in many forms, but of greatest significance are the complexes involving mucin/s (Iontcheva & Oppenheim *et al.*, 1997). The mucins are large, abundant macromolecules that form heterotypic complexes with many saliva components (Felgentreff & Beisswenger *et al.*, 2006; Iontcheva *et al.*, 1997). These mucin aggregates precipitate out of the saliva and can be lost in the sediment after centrifugation (Bucki & Namiot *et al.*, 2008; Felgentreff *et al.*, 2006; Soares & Vitorino *et al.*, 2011). The pilot studies carried out for this project confirmed that aggregation and sedimentation readily occur in saliva and that this property was an important factor to consider during saliva preparation and analysis. The sediment contained proteins including those of interest in this study.

Use of a mucinase to break up the aggregates is not advisable due to the heating required for the enzyme to work. Very high salt or detergent content buffers might likewise free up the saliva components of interest but alter their structure in a manner that affects antibody recognition sites or otherwise affects the ELISA efficiency (Agarkhed & O'Dell *et al.*, 2018).

The presence of debris, and host and microbe cells in saliva sediment is also not disputed. These can impede analysis and vigorous attempts to physically break up the sediment to release bound components could lead to the disruption of cells, releasing proteins that could affect proteomic analysis- for example the cathelicidin pre-proprotein that cross-reacts with the Hycult LL37 ELISA assay- and, significantly, intracellular proteases. Many saliva proteins are susceptible to proteolytic enzymes (Vitorino *et al.*, 2009). In fact, proteolytic cleavage is an important physiological mechanism in saliva, activating or degrading salivary proteins, having pro-inflammatory and anti-inflammatory effects (see Amado *et al.* (2010) and Huq *et al.* (2007) for thorough reviews). LL-37, for example, relies on proteolytic cleavage for physiological activity and possesses multiple sites vulnerable to proteases. Furthermore, the precise composition of proteases in saliva can reflect the inflammatory status in the mouth (Feng & Li *et al.*, 2019), presenting a potential confounder in saliva proteome analysis. The addition of protease inhibitors during this project was carefully considered but eventually not performed. Small, diluted volumes of the thawed sample were to be added to

each assay well, leading to concern that adding more to the already complex sample would introduce an undesirable level of variability and measurement bias in addition to complications related to optimal buffer constitution. Only working with chilled samples, to inhibit enzyme activity, was also not an option as assay kits require all analytes to be at room temperature at the start of analysis.

Thus, there was a conundrum in terms of the best sample preparation method, especially as the samples were already frozen before the start of this project. A compromise between releasing mucin-bound proteins and avoiding cytolysis and the degradation of proteins was sought. An interesting side-note to consider is that the formation of heterotypic complexes *in vivo* is believed to be a strategy to protect proteins against proteolytic enzymes (Iontcheva *et al.*, 1997), and thus the presence of aggregates is a sign of a healthy oral system. Breaking apart aggregates could increase the rate of proteolytic degradation of saliva samples.

A short, low-speed vortexing should not be sufficient to lyse cells in saliva (Yoshizawa & Wong, 2013). Additionally, it was found that the supernatant obtained by centrifuging an aliquot immediately after vortexing has a higher protein content than passively occurring supernatant formed by natural sedimentation over 60 minutes, supporting the hypothesis that vortexing liberates some otherwise soluble proteins from the aggregates in the sediment.

Vortexing was thus used *immediately* before aliquoting and centrifuging. It was also used to resuspend stored aliquots before pipetting out whole saliva to dilute with ELISA buffer (LL37 assay only). The ELISA buffers prevented re-sedimentation. In order to minimize variation between samples introduced through sample freeze-thaw effects and proteolysis, samples used in each assay were represented by identically treated aliquots, typically a freshly thawed aliquot that had been obtained from the original frozen-thawed sample.

In terms of total protein estimation, results were more consistent when using a supernatant obtained after centrifugation. Not only did allowing sedimentation to occur passively reduce total protein estimates, it also reduced intraassay precision in a manner which was highly variable between samples and over time. The best compromise thus seemed to be to obtain total protein concentration estimations by vortexing and immediately centrifuging aliquots and performing a Bradford (Coomassie G-250) microplate assay using the supernatant within 1 hour.

The pre-analytical handling of samples outlined above produced acceptably consistent (<15% CV) total protein measurements for separate aliquots of the same samples, suggesting consistency for other analyses. However, some potential sources of bias remain. Protein degradation differences between samples prior to this project cannot be excluded and the effectiveness of the release of LL37 or other proteins from the aggregates could plausibly be different between samples and was not measured in detail.

Normalising methods

The initial intention was to use saliva total protein content to normalise all analyte levels in order to adjust for physiological dilution. The range of salivary total protein concentrations obtained in this study (160 to 1471 µg/mL) was consistent with some studies that used the Bradford method, for example Walsh & Montague *et al.* (2004) (who concluded that saliva TP concentration range for euhydration in adults was approximately 550 – 950 µg/mL), but lower than ranges reported by studies that used the bichocinic acid (BCA) method, for example Tao *et al*. (2005) (421-7052 µg/mL).

The Bradford (Coomassie G-250) total protein method was chosen for total protein estimation in the present study as the relatively high concentration of protein in saliva obviates the need for a highly sensitive assay. Additionally, this method requires little specialist training or equipment and is tolerant of salt and low concentrations of surfactants, as may be present in saliva samples (Thermofisher Scientific Protein Assay Technical Handbook, 2017). It relies on the binding of the reagent to aromatic and basic amino acid side chains in an acidic environment, provided by the reagent. This binding results in a shift in maximum absorbance of the reagent from 465 nm to 595nm (Bradford, 1976). Compared with the BCA assay, the Bradford assay is subject to more variability through differences in amino-acid composition. The major saliva proteins amylase and the histatins are high in both basic and aromatic side chains. These Coomassie dye-binding residues constitute 25.7% and 45.6% of total amino acid composition respectively, similar to that of the reference protein BSA (25.9%) and higher than the average mammalian protein (22%) (Gaur, 2014). However, mucin and the proline-rich salivary peptides, which are reported to constitute as much as 30 - 70% of total salivary protein (Bennick, 2004; Van Nieuw Amerongen *et al.*, 2004) are low in the necessary amino acids. It is therefore possible that the lower estimates are due to this choice of protein estimation method, although as discussed much of the mucin would be in the sediment and therefore not included in the method used by either this study or Tao *et al*. (2005) (composition calculations derived from Composition & Molecular Weight, proteininformationresource.org; accessed December 2020).

The normalising of saliva component levels to total protein content is a widely used method (for example Devic & Shi *et al.* (2014) and discussed by Pitti & Petrella *et al.* (2019)) but has limitations. There are issues, for example, if analytes of interest differ markedly in terms of their percentage of total protein mass. Statherin is the more abundant of the two analytes in this study, but is still typically reported as no more than 0.01 % of total salivary protein mass (Amado *et al.*, 2010) so this was not considered a risk in this study.

However, in this study we found that there was a correlation between total salivary protein concentration and the caries experience outcome variable, as well as with LL37 and statherin concentrations. Additionally, while spitting out a large volume of saliva was correlated with having an overall more dilute saliva, of the two analytes in question, only LL37 was diluted. This suggests that increased

production of spit by an individual may alter the *relative* abundance of saliva components. In the case of statherin, contraction of the salivary glands might plausibly be resulting in the secretion of statherin at least proportional to water content, an observation supported by previous studies (Jensen *et al.*, 1995), while the increased washing out of gingival sulci with higher volumes of saliva does not recruit sufficient LL37 to compensate for dilution. Since the dilution effects on our analytes of interest are complex, it was concluded that the analytes should not be normalised to total protein concentration. Instead, saliva flow variables were considered as possible confounders in the multivariable analyses.

5.3 Saliva ELISA

Quantitative enzyme-linked proteomic immunoassays were first developed fifty years ago and evolved into the modern enzyme-linked immunosorbent assays (ELISA) techniques employed today in both clinical and research settings (Drijvers & Awan *et al.*, 2017). When compared to other proteomic techniques, such as HPLC-MS or Western Blot, ELISA can be highly sensitive and specific while readily returning quantitative results as well as being tolerant of a wide range of contaminants (R&D Systems ELISA Guide 2021). Provided highly specific antibodies are used, the main disadvantages of ELISA are reduced to the inability to differentiate between different forms of the analyte of interest and that only one analyte can be analysed at a time. Since saliva is a complex material that includes contaminants, suitable commercial ELISA kits could be obtained, and only a small number of saliva components were to be analysed, the advantages of the ELISA technique outweighed the disadvantages in this case.

As discussed in the preceding section, the formation of aggregates and sediment in saliva can present a problem for analysis. This formation of large complexes particularly complicates chromatography and electrophoretic analysis of saliva, and if the complexes are not dissociated, this can lead to measurement error (Iontcheva *et al.*, 1997). This can also be a source of systematic bias, as there is some evidence that these complexes are more likely to affect estimates from "normal" saliva, where antimicrobial peptides are present in larger, more active forms that more readily form complexes, as compared with the fragments often seen in inflamed mouths, as discussed by Vitorino *et al.* (2005).

These interactions can also, in theory, affect ELISA analysis, through the formation of aggregates that precipitate out of the sample supernatant or through the blocking of monoclonal antibody recognition sites. ELISA kits were chosen that had undergone some quality verification for use with saliva samples and thus had hopefully achieved some compromise in terms of reducing aggregation without impacting on antibody binding efficiency.

Our pilot studies confirmed that the buffers provided did indeed reduce sedimentation and/or reduced aggregate formation, as evidenced by good intra-ELISA assay precision when using "non-clarified" saliva for the LL-37 assays, in addition to visual observations. The increased tolerance of the ELISA method for "non-clarified" saliva over other proteomic techniques was therefore an advantage when dealing with an analyte with a tendency to become involved in aggregates.

In this study, the main significant departure from the manufacturers' stated protocols was the analysis of samples in singles rather than duplicates or triplicates in most cases. This is justified in situations where samples are of limited volume and pilot studies indicate reasonable (< 10 %) intraassay coefficient of variability for standards and samples, as was the case in this study.

5.3.1 The ELISA results from this study are consistent with the published literature

The results for the salivary concentrations of LL37 and statherin reported in this study are similar to those in published literature.

The concentration of LL37 was between 12.42 – 270.7 ng of LL37/mL of saliva. This compares with 0.22 – 275 ng LL37/mL saliva reported by Davidopoulou *et al*. (2012). In their study, Davidopoulou *et al.* (2012) also used a Hycult ELISA kit to analyse saliva from children. However, they analysed clarified saliva and included samples from children across a wider age range (2 to 18 years old compared with 7 to 9 years old in this study). They noted an age-dependent effect on LL37 concentration, with the lowest concentrations occurring in children that were younger than those included in the present study. This may explain the difference in lower end of the concentration ranges observed. An age-dependent effect is also supported by the results of Colombo *et al*. (2016) where the median salivary LL37 concentration in children under five years of age was 40 ng/mL compared with the median in the present study of 60 ng/mL. A tendency for higher LL37 concentration with age was observed in this study but was not significant due to the small number of seven year olds in the sample group. The study of Gyll *et al.* (2018) also used the same kit and paraffin tablet-stimulated saliva from children about the same age but, again, clarified the saliva by centrifugation. Their results were so much lower (typically less than 2 ng/mL) that proteolysis or some other artefact is suspected.

When the LL37 results of the present study were normalised to total protein content, the ratios were 0.01 - 0.54 (median 0.13) ng of LL37 per µg of total protein content. This range is lower than that reported by Tao *et al*. (2005) of 0.07 – 25.33 (median 2.29), despite the higher average TP concentrations found in this previous study. Again, this is possibly due to the differences in participant age and assay methods used.

The salivary statherin concentration reported in this study (581- 9685 ng/mL saliva) is similar to that reported by Pateel *et al.* (2017) (500 – 4000 ng/mL), who used the same ELISA kit and centrifuged the saliva samples following the same method. However, the saliva used in their analysis was collected via an absorption method which may explain the difference in range of concentrations found. The participants in the Pateel *et al.* (2017) study were also all adults, however there is little evidence that salivary gland secretion of statherin varies with age (Cabras & Pisano *et al.*, 2009).

5.4 Caries experience associations with LL37 and statherin

The data showed a 22.5% prevalence of caries experience in the sample, which is consistent with the prevalence of caries in five-year olds (18.7%) and 12 year-olds (39.6%) in Norway as a whole (Statistics Norway, 2020).

There was no apparent statistically significant association between LL37 level and caries experience in these analyses. This was the case when treating the outcome as a score[7] or as a dichotomised variable, and with the latter in both unadjusted and adjusted models. This is contrary to the observations of Davidopolou *et al.* (2012) who reported a statistically significant negative association using the same Hycult ELISA kit. The main difference from this study is that Davidopolou *et al.* used clarified saliva, as discussed in Section 5.3.1. If our observation that LL37 is bound up in salivary aggregates is correct, and if the observations of Iontcheva *et al.* (1997) and Vitorino *et al.* (2005) that aggregation is more likely to involve salivary proteins in caries-free mouths are also correct, then the use of WSH instead of WSS should if anything increase the strength of any negative association between LL37 and caries experience. Again, in contrast to the absence of a statistically significant association in this study, Gyll *et al.* (2018) reported a statistically significant positive association of LL37 concentration and caries experience using the same kit and clarified saliva from children. Limitations relating to the definition of a positive outcome remain the most likely cause of these differences in observations, as discussed further in Section 5.5.

Our results suggested a statistically significant positive association between statherin concentration and caries experience in children. We estimated that there is an approximately 1.14 times increase in the probability of a dmft/DMFT score of >0 with each µg/mL increase in statherin concentration as determined by the methods described. The 95% confidence interval for the odds ratio was quite wide with the lower end close to 1.0, possibly reflecting the relatively small sample size and low prevalence of caries in this population. This positive association differs from the negative associations reported by Rudney *et al.* (2009), Vitorino *et al.* (2005) and K. Wang *et al.* (2018), however none of these studies used an ELISA method, and only the K. Wang *et al.* (2018) study involved child participants.

As this is an observational study, it is necessary to consider possible confounding of any associations between the exposure of interest and the outcome of interest by other variables. These were controlled for by multivariable models. The DAG analysis helped guide the design of the multivariable models by clarifying possible pathways between variables while bivariate analysis helped ensure models were as simple as possible. This resulted in models with a minimal adjustment set (one saliva flow variable, +/- age, +/- vitamin D supplementation). A review of the literature lead to the design of multivariable models with a set of additional

[7] It should be noted that the count of observations of dmft/DMFT score of three or more was only nine, making it likely there is low statistical power using the non-dichotomised outcome variable.

variables that have been shown to have some strength as predictors in childhood caries prediction models (Powell, 1998).

There was little effect on the strength of association between either analyte and the caries experience outcome after adjusting for these variables. This suggests that none of the variables considered are significant confounders of the associations. In other words, the statherin concentration in the saliva of children in this age range with caries experience is increased relative to those without caries experience, regardless of gender and the volume of saliva produced (when following the methods outlined). However, not all variance is accounted for and other, unobserved confounders are possible.

When models included total protein concentration, the strength of association between the analytes and caries experience was decreased suggesting the relationship between total protein and the exposure and/or outcome of interest is more complex.

The initial plan to use total protein concentration to normalise both of the analyte concentration variables prior to statistical analysis was rejected based on the initial findings of differences in dilution effects on saliva components (Section 5.2). However, its use in the multivariable model as a covariate may also be flawed. The eventual concentration of protein in saliva spat out by an individual in a five minute period in response to a chewing stimulus is a function of many things. This includes salivary gland protein expression but also factors that influence the contraction of the salivary glands, including the degree of chewing, psychology (possibly including age and gender), time of day, the absorption rate of water from the saliva as it passes through the salivary ducts, and the initial dryness and soluble protein in the mouth. Many of these factors can be affected by caries, due to for example pain, stress and anxiety, a generalised inflammatory state, and increased bacterial numbers and bacterial debris in the mouth. It is plausible that total protein concentration is a bad control for the association of analyte concentration with caries experience as a result of being a dependent variable of the outcome itself (Cinelli, 2020).

Although proteolysis as a result of inflammation, or by bacteria directly, could reduce the amount of analytes detectable by ELISA though disruption of recognition sites, proteolysis could also result in fragments of protective saliva proteins that are more soluble and/or less likely to become involved in aggregates than the intact protein (Vitorino *et al.*, 2005). When analysing saliva supernatant, this could lead to a biased increase in detectable peptide levels in those with bacterial dysbiosis. This, along with increased debris, is a plausible explanation for the association between total protein concentration and caries experience, and is also another plausible explanation for the increase in statherin levels seen in those with caries experience in this study.

5.5 Project strengths and weaknesses

There are considerable challenges when investigating the association of saliva components with caries experience outcomes in children. The contribution of the levels of a single saliva component to the risk of caries may be too low to detect without very large sample sizes and, on a practical level, it can be challenging to obtain saliva samples from a very large number of children in a controlled and regulated way as a result of differences in individual participant psychological resilience and capacity to follow instructions, in addition to consent concerns.

This study involved saliva samples from a reasonably large sample of children, however, the low prevalence of the outcome of interest still limited some statistical analysis.

The outcome measure used was the clinical diagnosis of decayed, missing or filled teeth (dmft/DMFT), also referred to as caries experience. The dmft/DMFT score is generated through dental examination and also considers the normal change in dentition from deciduous to permanent teeth that occurs in the sampling population age group. It is clear that this diagnosis captures active caries activity, recently treated caries and fully resolved caries without discriminating between the three. As such, the levels of any analyte that is involved in one stage of caries more than another could be biased by an unobserved imbalance in the distribution of caries stage in the sample. For example, if low LL37 levels are only associated with early caries development and those observed as having a caries experience within the sample mostly have advanced or fully resolved caries, then the association will be missed. Additionally, if only observed at a single time point and caries development relies on a failure of gingival cells to increase LL37 production at a brief but critical time relative to baseline, for example at the first signals of bacterial dysbiosis, then the association between the low LL37 level and caries will be missed (there is some evidence of just such a phenomenon in Simon-Soro *et al*. (2018)). Thus observations of dmft/DMFT at a single time point may miss important information that could be obtained from repeat observations or a different outcome measure, although the latter would require even larger sample sizes.

As discussed in the introduction, bacterial dysbiosis is required for caries to develop and, once established, it has self-sustaining properties. These could include the ability to suppress or evade the inflammatory responses, for example via the extracellular matrix produced by some of the cariogenic bacteria (Lamont *et al*., 2018), through proteolysis of protective peptides (Vitorino *et al.*, 2005) or acquisition of resistance mutations (Nilsson, 2020). It is therefore plausible that any change in statherin or LL37 levels and/or activity will always be mediated by bacterial factors, including many that were unobserved in this study.

An additional weakness of this project is the extent to which the results can be generalised. As the selection of subjects was not randomised, but rather based on

previously volunteering for a research project and then re-volunteering for a follow-up study, it is not possible to exclude some bias including selection bias. This could include, for example, parents being more likely to volunteer if the family is interested in research and dental health more specifically, possibly due to personal experience with poor dental health. This project could also be affected by reporting bias, since questionnaires by their nature always involve a certain degree of subjectivity or recollection bias. Additionally, the sampling population is unlikely to have included a broad range of ethnicities, genotypes or socioeconomic backgrounds. The ethnicity of participants was not recorded, but it is known that the proportion of all Trøndelag residents with non-norwegian ancestry is less than 10%[8] and the proportion in this sample was anecdotally at or close to zero percent. The mothers were all highly educated and likely only represent high socioeconomic groups. Whilst limiting generalisability, the homogenous sampling population reduced the number of variables that needed to be considered as confounders.

This project had advantages in comparison to many previously published studies which generally had lower sample sizes, did not control for possible confounders (for example Tao *et al.* (2005)), had less transparent participant selection processes (for example Rudney *et al.* (2009) and Tao *et al.* (2005)) or less standardised case definitions (for example Davidopolou *et al.* (2012)). The methods for measuring analyte concentrations were also more detailed in this study and thus more replicable than, for example, Simon-Soro *et al.* (2018) and Davidopolou *et al.* (2012), who did not expressly state whether their ELISA method wholly followed the manufacturer's manual, and Rudney *et al.* (2009) who used a non-ELISA, semi-quantitative proteomic approach. Other studies that followed detailed, standardised procedures for subject selection and sample analysis and controlled for some variables did not define significance levels for their statistical analyses, which is especially important when sample sizes are very small (for example Vitorino *et al*. (2005), n=20).

[8] https://www.ssb.no/befolkning/statistikker/innvbef/arkiv/2012-04-26; accessed 25th April, 2021.

6. CONCLUSIONS

The goals for this project were to determine:

- which salivary components to investigate for associations with early dental caries experience,
- the best way to obtain information about the levels of those saliva components and what the levels of those components of interest were in the biobank samples, and
- if there were any associations between the levels of those saliva components and the dental caries experience in the sample population.

It was found that LL37 and statherin are saliva components plausibly involved in caries aetiology and that the levels of LL37 and statherin can be readily and consistently measured in saliva using commercial ELISA kits, albeit with some deviations from the manufacturers' protocols. The range of concentrations of LL37 and statherin found were similar to those reported in the literature, especially when considering differences in study design. The concentration of statherin was positively associated with caries experience and this association was not influenced by saliva flow. No statistically significant association was found between LL37 concentration and caries experience.

Future possibilities

The finding of a statistically significant, positive association between statherin concentration and caries experience in this study supports a role for statherin in childhood caries. Although it was not found to be statistically significant, there was some trend towards an inverse correlation between LL37 levels and caries experience score. In both cases the results justify further research, ideally with repeated observations over time. In future studies, it would also be preferable to perform some pre-analysis sample preparation immediately after saliva collection if practicable such that the concerns regarding proteolysis and aggregation bias discussed in Sections 5.2 and 5.3 are somewhat addressed.

The findings of this project suggests that, in principle, ELISA of a small number of saliva samples should be technically and economically feasible for childhood caries management in referral dental practices. However, further work is still required with regards to the precise components to analyse, the timing of tests and the thresholds for diagnosis and/or prognosis. The finding of a strong correlation between saliva total protein concentration and caries experience also warrants separate investigation since this has the potential to be even more simple and cheap to measure.

The use of saliva as a diagnostic sample has increased in recent decades, as researchers and companies appreciate the relative ease of sample collection and the range of feasible analytes increases with advances in technology. Private companies specialising in the analysis of saliva have been established, for example Salimetrics (Carlsbad, California), which offers services ranging from saliva collection kits, single cytokine and hormone analyte analysis or analyte panels through to genotyping from cells in the saliva. However, a major issue remains with the comparison of analyte results across test methods.

Saliva is a complex material both in terms of end composition and in its creation. Researchers have used saliva collected directly from salivary gland duct openings in order to reduce some of this complexity (for example the very careful and detailed work of Saitou & Gaylord *et al.* (2020)). Controlling for other factors such as saliva flow rate and aggregation at preanalytical and/or data analysis stage is helpful, but may not be sufficient when there are a high number of unobserved variables.

In terms of understanding pathophysiology, use of salivary gland duct opening collection methods in research may be preferable, however, expectorated saliva (spit) is more practical in studies with large sample sizes, especially when they involve children. Additionally, it will always be necessary to make some reference to the composition of spit in any study relating to dental health, since it is the conditions at the tooth surface that are relevant. As such, the best method of saliva collection and/or preanalytical processing will continue to depend on the primary goals and extra care must therefore be taken when comparing results.

The role of childhood caries as a source of inequality has been highlighted by the World Health Organisation (Phantumvanit & Makino *et al.*, 2018). Research into public health childhood caries intervention strategies has been a part of the WHO Global Oral Health Programme, which is due to publish their report later in 2021. This report is expected to include guidelines regarding caries surveillance and research priorities, which will further influence the direction of future caries research.

REFERENCES

Agarkhed, M., O'Dell, C., Hsieh, M.-C., Zhang, J., Goldstein, J., & Srivastava, A. (2018). Effect of Surfactants on Mechanical, Thermal, and Photostability of a Monoclonal Antibody. *AAPS PharmSciTech, 19*(1), 79-92. doi:10.1208/s12249-017-0845-7

Agerberth, B., Charo, J., Werr, J., Olsson, B., Idali, F., Lindbom, L., . . . Gudmundsson, G. H. (2000). The human antimicrobial and chemotactic peptides LL-37 and α-defensins are expressed by specific lymphocyte and monocyte populations. *Blood, 96*(9), 3086-3093. doi:10.1182/blood.V96.9.3086

Aidoukovitch, A., Dahl, S., Fält, F., Nebel, D., Svensson, D., Tufvesson, E., & Nilsson, B. O. (2020). Antimicrobial peptide LL-37 and its pro-form, hCAP18, in desquamated epithelial cells of human whole saliva. *Eur J Oral Sci, 128*(1), 1-6. doi:10.1111/eos.12664

Aldred, P. M., Hollox, E. J., & Armour, J. A. (2005). Copy number polymorphism and expression level variation of the human alpha-defensin genes DEFA1 and DEFA3. *Hum Mol Genet, 14*(14), 2045-2052. doi:10.1093/hmg/ddi209

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410. doi:10.1016/s0022-2836(05)80360-2

Amado, F., Lobo, M. J., Domingues, P., Duarte, J. A., & Vitorino, R. (2010). Salivary peptidomics. *Expert Rev Proteomics, 7*(5), 709-721. doi:10.1586/epr.10.48

Amarante, E., Raadal, M., & Espelid, I. (1998). Impact of diagnostic criteria on the prevalence of dental caries in Norwegian children aged 5, 12 and 18 years. *Community Dent Oral Epidemiol, 26*(2), 87-94. doi:10.1111/j.1600-0528.1998.tb01933.x

Arias-Bujanda, N., Regueira-Iglesias, A., Balsa-Castro, C., Nibali, L., Donos, N., & Tomás, I. (2020). Accuracy of single molecular biomarkers in saliva for the diagnosis of periodontitis: A systematic review and meta-analysis. *J Clin Periodontol, 47*(1), 2-18. doi:10.1111/jcpe.13202

Bals, R., & Wilson, J. M. (2003). Cathelicidins - a family of multifunctional antimicrobial peptides. *Cellular and Molecular Life Sciences CMLS, 60*(4), 711-720. doi:10.1007/s00018-003-2186-9

Bandyopadhyay, B. C., Swaim, W. D., Sarkar, A., Liu, X., & Ambudkar, I. S. (2012). Extracellular Ca(2+) sensing in salivary ductal cells. *J Biol Chem, 287*(36), 30305-30316. doi:10.1074/jbc.M112.394122

Bennick, A. (2004). Salivary proline-rich proteins. *Molecular and Cellular Biochemistry, 45*, 83-99.

Bergandi, L., Defabianis, P., Re, F., Preti, G., Aldieri, E., Garetto, S., . . . Ghigo, D. (2007). Absence of soluble CD14 in saliva of young patients with dental caries. *Eur J Oral Sci, 115*(2), 93-96. doi:10.1111/j.1600-0722.2007.00437.x

Bikle, D. D. (2020). Vitamin D: Newer Concepts of Its Metabolism and Function at the Basic and Clinical Level. *J Endocr Soc, 4*(2), bvz038. doi:10.1210/jendso/bvz038

Børsting, T., Annemarie Schuller, Paula van Dommelen, Signe Nilssen Stafne, Marit Slåttelid Skeie, Anne B. Skaare, Miriam K. Gustafsson, Unni Syversen, Astrid Kamilla Stunes, Mats Peder Mosti, Siv Mørkved, Kjell Å. Salvesen,, Tone Natland Fagerhaug. (2019). Associations between maternal vitamin D status in second and third trimester of pregnancy and offspring enamel hypomineralisation at 7-9 years: a longitudinal study. *Norsk Epidemiologi, 28*.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem, 72*(1), 248-254. doi:10.1016/0003-2697(76)90527-3

Braschi, B., Denny, P., Gray, K., Jones, T., Seal, R., Tweedie, S., . . . Bruford, E. (2019). Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res, 47*(D1), D786-d792. doi:10.1093/nar/gky930

Broadbent, J. M., Page, L. A. F., Thomson, W. M., & Poulton, R. (2013). Permanent dentition caries through the first half of life. *British Dental Journal, 215*(7), E12-E12. doi:10.1038/sj.bdj.2013.991

Bucki, R., Namiot, D. B., Namiot, Z., Savage, P. B., & Janmey, P. A. (2008). Salivary mucins inhibit antibacterial activity of the cathelicidin-derived LL-37 peptide but not the cationic steroid CSA-13. *J Antimicrob Chemother, 62*(2), 329-335. doi:10.1093/jac/dkn176

Cabras, T., Pisano, E., Boi, R., Olianas, A., Manconi, B., Inzitari, R., . . . Messana, I. (2009). Age-Dependent Modifications of the Human Salivary Secretory Protein Complex. *Journal of proteome research, 8*(8), 4126-4134. doi:10.1021/pr900212u

Carlberg, C. (2019). Nutrigenomics of Vitamin D. *Nutrients, 11*(3). doi:10.3390/nu11030676

Carlson, G. W. (2000). THE SALIVARY GLANDS: Embryology, Anatomy, and Surgical Applications. *Surgical Clinics of North America, 80*(1), 261-273. doi:10.1016/S0039-6109(05)70405-9

Carpenter, G. H. (2013). The secretion, components, and properties of saliva. *Annu Rev Food Sci Technol, 4*, 267-276. doi:10.1146/annurev-food-030212-182700

Chen, C., Brock, R., Luh, F., Chou, P. J., Larrick, J. W., Huang, R. F., & Huang, T. H. (1995). The solution structure of the active domain of CAP18--a lipopolysaccharide binding protein from rabbit leukocytes. *FEBS Lett, 370*(1-2), 46-52. doi:10.1016/0014-5793(95)00792-8

Chen, S. S., Wang, K., Zhao, J., Wu, W. C., Wu, Y. F., & Zhao, L. (2017). Increased expression of triggering receptor expressed on myeloid cells 1 and 2 in inflamed human gingiva. *J Periodontal Res, 52*(3), 512-521. doi:10.1111/jre.12417

Cinelli, C., Forney, A. and Pearl, J. (2020). *A Crash Course in Good and Bad Controls*. Retrieved from https://ssrn.com/abstract=3689437

Clark, M. B., Keels, M. A., & Slayton, R. L. (2020). Fluoride Use in Caries Prevention in the Primary Care Setting. *Pediatrics, 146*(6), e2020034637. doi:10.1542/peds.2020-034637

Colombo, N. H., Ribas, L. F., Pereira, J. A., Kreling, P. F., Kressirer, C. A., Tanner, A. C., & Duque, C. (2016). Antimicrobial peptides in saliva of children with severe early childhood caries. *Arch Oral Biol, 69*, 40-46. doi:10.1016/j.archoralbio.2016.05.009

Dale, B. A., Tao, R., Kimball, J. R., & Jurevic, R. J. (2006). Oral Antimicrobial Peptides and Biological Control of Caries. *BMC Oral Health, 6*(1), S13. doi:10.1186/1472-6831-6-S1-S13

Davidopoulou, S., Diza, E., Menexes, G., & Kalfas, S. (2012). Salivary concentration of the antimicrobial peptide LL-37 in children. *Archives of Oral Biology, 57*(7), 865-869. doi:10.1016/j.archoralbio.2012.01.008

Dawes, C. (1972). Circadian rhythms in human salivary flow rate and composition. *The Journal of physiology, 220*(3), 529-545. doi:10.1113/jphysiol.1972.sp009721

de Sousa-Pereira, P., Amado, F., Abrantes, J., Ferreira, R., Esteves, P. J., & Vitorino, R. (2013). An evolutionary perspective of mammal salivary peptide families: cystatins, histatins, statherin and PRPs. *Arch Oral Biol, 58*(5), 451-458. doi:10.1016/j.archoralbio.2012.12.011

Devic, I., Shi, M., Schubert, M. M., Lloid, M., Izutsu, K. T., Pan, C., . . . Presland, R. B. (2014). Proteomic Analysis of Saliva from Patients with Oral Chronic Graft-Versus-Host Disease. *Biology of Blood and Marrow Transplantation, 20*(7), 1048-1055. doi:10.1016/j.bbmt.2014.03.031

Doetzer, A. D., Brancher, J. A., Pecharki, G. D., Schlipf, N., Werneck, R., Mira, M. T., . . . Trevilatto, P. C. (2015). Lactotransferrin Gene Polymorphism Associated with Caries Experience. *Caries Res, 49*(4), 370-377. doi:10.1159/000366211

Drijvers, J. M., Awan, I. M., Perugino, C. A., Rosenberg, I. M., & Pillai, S. (2017). Chapter 7 - The Enzyme-Linked Immunosorbent Assay: The Application of ELISA in Clinical Research. In M. Jalali, F. Y. L. Saldanha, & M. Jalali (Eds.), *Basic Science Methods for Clinical Researchers* (pp. 119-133). Boston: Academic Press.

Elm, E. v., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ, 335*(7624), 806-808. doi:10.1136/bmj.39335.541782.AD

Engelberg, Y., & Landau, M. (2020). The Human LL-37(17-29) antimicrobial peptide reveals a functional supramolecular structure. *Nature Communications, 11*(1), 3894-3894. doi:10.1038/s41467-020-17736-x

Felgentreff, K., Beisswenger, C., Griese, M., Gulder, T., Bringmann, G., & Bals, R. (2006). The antimicrobial peptide cathelicidin interacts with airway mucus. *Peptides, 27*(12), 3100-3106. doi:10.1016/j.peptides.2006.07.018

Feng, Y., Li, Q., Chen, J., Yi, P., Xu, X., Fan, Y., . . . Zhang, P. (2019). Salivary protease spectrum biomarkers of oral cancer. *Int J Oral Sci, 11*(1), 7. doi:10.1038/s41368-018-0032-z

Gaur, R. K. (2014). Amino acid frequency distribution among eukaryotic proteins. *IIOAB Journal, 5*, 6-11.

Gornowicz, A., Bielawska, A., Bielawski, K., Grabowska, S. Z., Wójcicka, A., Zalewska, M., & Maciorkowska, E. (2012). Pro-inflammatory cytokines in saliva of adolescents with dental caries disease. *Ann Agric Environ Med, 19*(4), 711-716.

Gyll, J., Ridell, K., Ohlund, I., Karlsland Akeson, P., Johansson, I., & Lif Holgerson, P. (2018). Vitamin D status and dental caries in healthy Swedish children. *Nutr J, 17*(1), 11. doi:10.1186/s12937-018-0318-1

Hatipoğlu, Ö., & Saydam, F. (2020). Association between rs11362 polymorphism in the beta-defensin 1 (DEFB1) gene and dental caries: A meta-analysis. *J Oral Biosci, 62*(3), 272-279. doi:10.1016/j.job.2020.06.004

Helmerhorst, E. J., Traboulsi, G., Salih, E., & Oppenheim, F. G. (2010). Mass spectrometric identification of key proteolytic cleavage sites in statherin affecting mineral homeostasis and bacterial binding domains. *Journal of proteome research, 9*(10), 5413-5421. doi:10.1021/pr100653r

Hemadi, A. S., Huang, R., Zhou, Y., & Zou, J. (2017). Salivary proteins and microbiota as biomarkers for early childhood caries risk assessment. *International Journal of Oral Science, 9*(11), e1-e1. doi:10.1038/ijos.2017.35

Homann, V., Kinne-Saffran, E., Arnold, W. H., Gaengler, P., & Kinne, R. K. H. (2006). Calcium transport in human salivary glands: a proposed model of calcium secretion into saliva. *Histochemistry and Cell Biology, 125*(5), 583-591. doi:10.1007/s00418-005-0100-2

Huq, L., Cross, K., Ung, M., Myroforidis, H., Veith, P., Chen, D., . . . Reynolds, E. (2007). A Review of the Salivary Proteome and Peptidome and Saliva-derived Peptide Therapeutics. *International Journal of Peptide Research and Therapeutics, 13*, 547-564. doi:10.1007/s10989-007-9109-9

Iontcheva, I., Oppenheim, F. G., & Troxler, R. F. (1997). Human salivary mucin MG1 selectively forms heterotypic complexes with amylase, proline-rich proteins, statherin, and

histatins. *Journal of dental research, 76*(3), 734-743. doi:10.1177/00220345970760030501

Izumi, M., Zhang, B.-X., Dean, D., Lin, A., Saunders, M., Hazuda, H., & Yeh, C.-K. (2015). Secretion of salivary statherin is compromised in uncontrolled diabetic patients. *BBA Clinical, 89*. doi:10.1016/j.bbacli.2015.01.002

Jensen, J., Xu, T., Lamkin, M. S., Brodin, P., Aars, H., Berg, T., & Oppenheim, F. (1995). Physiological Regulation of the Secretion of Histatins and Statherins in Human Parotid Saliva. *Journal of dental research, 73*, 1811-1817. doi:10.1177/00220345940730120401

Jentsch, H., Beetke, E., & Göcke, R. (2004). Salivary analyses and caries increment over 4 years: an approach by cluster analysis. *Clin Oral Investig, 8*(3), 156-160. doi:10.1007/s00784-004-0263-6

Joffe, M., Gambhir, M., Chadeau-Hyam, M., & Vineis, P. (2012). Causal diagrams in systems epidemiology. *Emerging themes in epidemiology, 9*, 1. doi:10.1186/1742-7622-9-1

Khurshid, Z., Naseem, M., Sheikh, Z., Najeeb, S., Shahab, S., & Zafar, M. S. (2016). Oral antimicrobial peptides: Types and role in the oral cavity. *Saudi Pharm J, 24*(5), 515-524. doi:10.1016/j.jsps.2015.02.015

Koivisto, O., Hanel, A., & Carlberg, C. (2020). Key Vitamin D Target Genes with Functions in the Immune System. *Nutrients, 12*(4). doi:10.3390/nu12041140

Kühnisch, J., Thiering, E., Kratzsch, J., Heinrich-Weltzien, R., Hickel, R., Heinrich, J., . . . group, L. I. s. (2015). Elevated serum 25(OH)-vitamin D levels are negatively correlated with molar-incisor hypomineralization. *Journal of dental research, 94*(2), 381-387. doi:10.1177/0022034514561657

Kumar, N., Reddy, V., Padakandla, P., Togaru, H., Kalagatla, S., & Chandra, S. (2016). Evaluation of chemokines in gingival crevicular fluid in children with dental caries and stainless steel crowns: A clinico-biochemical study. *Journal of Indian Society of Pedodontics and Preventive Dentistry, 34*(3), 273-279. doi:10.4103/0970-4388.186754

Lamont, R. J., Koo, H., & Hajishengallis, G. (2018). The oral microbiota: dynamic communities and host interactions. *Nature Reviews Microbiology, 16*(12), 745-759. doi:10.1038/s41579-018-0089-x

Lehrer, R. I., & Ganz, T. (2002). Cathelicidins: a family of endogenous antimicrobial peptides. *Curr Opin Hematol, 9*(1), 18-22. doi:10.1097/00062752-200201000-00004

Levine, M. (2011). Susceptibility to dental caries and the salivary proline-rich proteins. *International Journal of Dentistry, 2011*, 953412-953412. doi:10.1155/2011/953412

Lips, A., Antunes, L. S., Antunes, L. A., Pintor, A. V. B., Santos, D. A. B. d., Bachinski, R., . . . Alves, G. G. (2017). Salivary protein polymorphisms and risk of dental caries: a systematic review. *Brazilian Oral Research, 31:e41*, 1-12. doi:10.1590/1807-3107bor-2017.vol31.0041

Manconi, B., Castagnola, M., Cabras, T., Olianas, A., Vitali, A., Desiderio, C., . . . Messana, I. (2016). The intriguing heterogeneity of human salivary proline-rich proteins: Short title: Salivary proline-rich protein species. *J Proteomics, 134*, 47-56. doi:10.1016/j.jprot.2015.09.009

Martins, C., Buczynski, A. K., Maia, L. C., Siqueira, W. L., & Castro, G. F. B. d. A. (2013). Salivary proteins as a biomarker for dental caries—A systematic review. *Journal of Dentistry, 41*(1), 2-8. doi:10.1016/j.jdent.2012.10.015

Mohanty, T., Sjögren, J., Kahn, F., Abu-Humaidan, A. A., Fisker, N., Assing, K., . . . Sørensen, O. E. (2015). A novel mechanism for NETosis provides antimicrobial defense at the oral mucosa. *Blood, 126*(18), 2128-2137. doi:10.1182/blood-2015-04-641142

Nilsson, B.-O. (2020). What can we learn about functional importance of human antimicrobial peptide LL-37 in the oral environment from severe congenital neutropenia (Kostmann disease)? *Peptides, 128*, 170311. doi:10.1016/j.peptides.2020.170311

Nishana, E., Bhat, S. S., Sahana, K. S., Hegde, S. K., Bhat, V., & Kalal, B. S. (2019). Estimation of Salivary sCD14 in Children with Early Childhood Caries in Association with Pneumonia. *Rep Biochem Mol Biol, 8*(2), 132-138.

Nørrisgaard, P. E., Haubek, D., Kühnisch, J., Chawes, B. L., Stokholm, J., Bønnelykke, K., & Bisgaard, H. (2019). Association of High-Dose Vitamin D Supplementation During Pregnancy With the Risk of Enamel Defects in Offspring: A 6-Year Follow-up of a Randomized Clinical Trial. *JAMA Pediatrics, 173*(10), 924-930. doi:10.1001/jamapediatrics.2019.2545

Nylund, K. M., Ruokonen, H., Sorsa, T., Heikkinen, A. M., Meurman, J. H., Ortiz, F., . . . Bostanci, N. (2018). Association of the salivary triggering receptor expressed on myeloid cells/its ligand peptidoglycan recognition protein 1 axis with oral inflammation in kidney disease. *J Periodontol, 89*(1), 117-129. doi:10.1902/jop.2017.170218

Oppenheim, J. J., & Yang, D. (2005). Alarmins: chemotactic activators of immune responses. *Current Opinion in Immunology, 17*(4), 359-365. doi:10.1016/j.coi.2005.06.002

Pateel, D. G. S., Gunjal, S., Math, S. Y., Murugeshappa, D. G., & Nair, S. M. (2017). Correlation of Salivary Statherin and Calcium Levels with Dental Calculus Formation: A Preliminary Study. *International Journal of Dentistry, 2017*, 2857629. doi:10.1155/2017/2857629

Peschel, A., & Sahl, H.-G. (2006). The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nature Reviews Microbiology, 4*(7), 529-536. doi:10.1038/nrmicro1441

Phantumvanit, P., Makino, Y., Ogawa, H., Rugg-Gunn, A., Moynihan, P., Petersen, P. E., . . . Ungchusak, C. (2018). WHO Global Consultation on Public Health Intervention against Early Childhood Caries. *Community Dent Oral Epidemiol, 46*(3), 280-287. doi:10.1111/cdoe.12362

Phattarataratip, E., Olson, B., Broffitt, B., Qian, F., Brogden, K. A., Drake, D. R., . . . Banas, J. A. (2011). Streptococcus mutans strains recovered from caries-active or caries-free individuals differ in sensitivity to host antimicrobial peptides. *Molecular Oral Microbiology, 26*(3), 187-199. doi:10.1111/j.2041-1014.2011.00607.x

Piekoszewska-Ziętek, P., Turska-Szybka, A., & Olczak-Kowalczyk, D. (2019). Salivary proteins and peptides in the aetiology of caries in children: Systematic literature review. *Oral Dis, 25*(4), 1048-1056. doi:10.1111/odi.12953

Pitti, E., Petrella, G., Di Marino, S., Summa, V., Perrone, M., D'Ottavio, S., . . . Cicero, D. O. (2019). Salivary Metabolome and Soccer Match: Challenges for Understanding Exercise induced Changes. *Metabolites, 9*(7), 141.

Polley, S., Louzada, S., Forni, D., Sironi, M., Balaskas, T., Hains, D. S., . . . Hollox, E. J. (2015). Evolution of the rapidly mutating human salivary agglutinin gene (DMBT1) and population subsistence strategy. *Proc Natl Acad Sci U S A, 112*(16), 5105-5110. doi:10.1073/pnas.1416531112

Powell, L. V. (1998). Caries prediction: a review of the literature. *Community Dentistry and Oral Epidemiology, 26*(6), 361-371. doi:10.1111/j.1600-0528.1998.tb01974.x

Proctor, G. B. (2016). The physiology of salivary secretion. *Periodontol 2000, 70*(1), 11-25. doi:10.1111/prd.12116

Protein Assay Technical Handbook. (2017). Retrieved from https://www.thermofisher.com/no/en/home/global/forms/life-science/protein-assay-technical-handbook-download.html

R&D Systems ELISA Guide: A Clear and Easy Guide to ELISAs. (2021). Retrieved from https://resources.rndsystems.com/pdfs/brochures/rndsystems_br_elisa_guide_asdfy20-10071.pdf

Resource_Coordinators_NCBI. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research, 44*(D1), D7-D19. doi:10.1093/nar/gkv1290

Robinson, C., Brookes, S. J., Shore, R. C., & Kirkham, J. (1998). The developing enamel matrix: nature and function. *Eur J Oral Sci, 106 Suppl 1*, 282-291. doi:10.1111/j.1600-0722.1998.tb02188.x

Rudney, J. D., Staikov, R. K., & Johnson, J. D. (2009). Potential biomarkers of human salivary function: A modified proteomic approach. *Archives of Oral Biology, 54*(1), 91-100. doi:10.1016/j.archoralbio.2008.08.007

Saitou, M., Gaylord, E. A., Xu, E., May, A. J., Neznanova, L., Nathan, S., . . . Gokcumen, O. (2020). Functional Specialization of Human Salivary Glands and Origins of Proteins Intrinsic to Human Saliva. *Cell Reports, 33*(7). doi:10.1016/j.celrep.2020.108402

Sanz, M., Beighton, D., Curtis, M. A., Cury, J. A., Dige, I., Dommisch, H., . . . Zaura, E. (2017). Role of microbial biofilms in the maintenance of oral health and in the development of dental caries and periodontal diseases. Consensus report of group 1 of the Joint EFP/ORCA workshop on the boundaries between caries and periodontal disease. *J Clin Periodontol, 44 Suppl 18*, S5-s11. doi:10.1111/jcpe.12682

Schroth, R., Lavelle, C., Tate, R., Bruce, S., Billings, R., & Moffatt, M. (2014). Prenatal Vitamin D and Dental Caries in Infants. *Pediatrics, 133*. doi:10.1542/peds.2013-2215

Selwitz, R. H., Ismail, A. I., & Pitts, N. B. (2007). Dental caries. *The Lancet, 369*(9555), 51-59. doi:10.1016/S0140-6736(07)60031-2

Sharma, V., Gupta, N., Srivastava, N., Rana, V., Chandna, P., Yadav, S., & Sharma, A. (2017). Diagnostic potential of inflammatory biomarkers in early childhood caries - A case control study. *Clin Chim Acta, 471*, 158-163. doi:10.1016/j.cca.2017.05.037

Shimotoyodome, A., Kobayashi, H., Tokimitsu, I., Matsukubo, T., & Takaesu, Y. (2006). Statherin and histatin 1 reduce parotid saliva-promoted Streptococcus mutans strain MT8148 adhesion to hydroxyapatite surfaces. *Caries Res, 40*(5), 403-411. doi:10.1159/000094286

Sikorska, M. H. J., Mielnik-Blaszczak, M., & Kapeć, E. (2002). The relationship between the levels of SigA, lactoferrin and α1 proteinase inhibitor in saliva and permanent dentition caries in 15-year-olds. *Oral microbiology and immunology, 17*(5), 272-276. doi:10.1034/j.1399-302X.2002.170502.x

Simon-Soro, A., Sherriff, A., Sadique, S., Ramage, G., Macpherson, L., Mira, A., . . . Malcolm, J. (2018). Combined analysis of the salivary microbiome and host defence peptides predicts dental disease. *Scientific Reports (Nature Publisher Group), 8*, 1-10. doi:10.1038/s41598-018-20085-x

Soares, S., Vitorino, R., Osório, H., Fernandes, A., Venâncio, A., Mateus, N., . . . de Freitas, V. (2011). Reactivity of Human Salivary Proteins Families Toward Food Polyphenols. *J Agric Food Chem, 59*(10), 5535-5547. doi:10.1021/jf104975d

Statistics Norway. (2020). *11959: Tannhelsetilstand og kariesforekomst blant utvalgte aldersgrupper, etter alder (F) 2015 - 2018*. Oslo: Statistics Norway

Subbarao, K. C., Nattuthurai, G. S., Sundararajan, S. K., Sujith, I., Joseph, J., & Syedshah, Y. P. (2019). Gingival Crevicular Fluid: An Overview. *Journal of pharmacy & bioallied sciences, 11*(Suppl 2), S135-S139. doi:10.4103/JPBS.JPBS_56_19

Sultan, Z., Mali, M., Najeeb, S., Naseem, M., & Zafar, M. (2017). Human Gingival Crevicular Fluids (GCF) Proteomics: An Overview. *Dentistry Journal, 5*, 12. doi:10.3390/dj5010012

Tao, R., Jurevic, R. J., Coulton, K. K., Tsutsui, M. T., Roberts, M. C., Kimball, J. R., . . . Dale, B. A. (2005). Salivary antimicrobial peptide expression and dental caries experience in children. *Antimicrob Agents Chemother, 49*(9), 3883-3888. doi:10.1128/aac.49.9.3883-3888.2005

TheUniProtConsortium. (2018). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research, 47*(D1), D506-D515. doi:10.1093/nar/gky1049

Tjabringa, G. S., Ninaber, D. K., Drijfhout, J. W., Rabe, K. F., & Hiemstra, P. S. (2006). Human cathelicidin LL-37 is a chemoattractant for eosinophils and neutrophils that acts via formyl-peptide receptors. *Int Arch Allergy Immunol, 140*(2), 103-112. doi:10.1159/000092305

van der Tas, J., Elfrink, M., Heijboer, A., Rivadeneira, F., Jaddoe, V., Tiemeier, H., . . . Voortman, T. (2018). Foetal, neonatal and child vitamin D status and enamel hypomineralization. *Community Dentistry and Oral Epidemiology, 46*. doi:10.1111/cdoe.12372

Van Nieuw Amerongen, A., Bolscher, J. G., & Veerman, E. C. (2004). Salivary proteins: protective and diagnostic value in cariology? *Caries Res, 38*(3), 247-253. doi:10.1159/000077762

Vaškaninová, V., Chen, D., Tafforeau, P., Johanson, Z., Ekrt, B., Blom, H., & Ahlberg, P. E. (2020). Marginal dentition and multiple dermal jawbones as the ancestral condition of jawed vertebrates. *Science (American Association for the Advancement of Science), 369*(6500), 211-216. doi:10.1126/science.aaz9431

Ventura, T. M. d. S., Cassiano, L. d. P. S., Souza e Silva, C. M. d., Taira, E. A., Leite, A. d. L., Rios, D., & Buzalaf, M. A. R. (2017). The proteomic profile of the acquired enamel pellicle according to its location in the dental arches. *Archives of Oral Biology, 79*, 20-29. doi:10.1016/j.archoralbio.2017.03.001

Vitorino, R., Barros, A., Caseiro, A., Domingues, P., Duarte, J., & Amado, F. (2009). Towards defining the whole salivar peptidome. *Proteomics. Clinical applications, 3*, 528-540. doi:10.1002/prca.200800183

Vitorino, R., Lobo, M. J. C., Duarte, J. R., Ferrer-Correia, A. J., Domingues, P. M., & Amado, F. M. L. (2005). The role of salivary peptides in dental caries. *Biomedical Chromatography, 19*(3), 214-222. doi:10.1002/bmc.438

Walsh, N. P., Montague, J. C., Callow, N., & Rowlands, A. V. (2004). Saliva flow rate, total protein concentration and osmolality as potential markers of whole body hydration status during progressive acute dehydration in humans. *Arch Oral Biol, 49*(2), 149-154. doi:10.1016/j.archoralbio.2003.08.001

Wang, G., Mishra, B., Epand, R. F., & Epand, R. M. (2014). High-quality 3D structures shine light on antibacterial, anti-biofilm and antiviral activities of human cathelicidin LL-37 and its fragments. *Biochim Biophys Acta, 1838*(9), 2160-2172. doi:10.1016/j.bbamem.2014.01.016

Wang, K., Wang, Y., Wang, X., Ren, Q., Han, S., Ding, L., . . . Zhang, L. (2018). Comparative salivary proteomics analysis of children with and without dental caries using the iTRAQ/MRM approach. *J Transl Med, 16*(1), 11. doi:10.1186/s12967-018-1388-8

Wang, K., Zhou, X., Li, W., & Zhang, L. (2019). Human salivary proteins and their peptidomimetics: Values of function, early diagnosis, and therapeutic potential in combating dental caries. *Archives of Oral Biology, 99*, 31-42. doi:10.1016/j.archoralbio.2018.12.009

WHO. (1997). *Oral health surveys : basic methods*(4th ed. ed.), World Health Organization.

Yoshizawa, J. M., & Wong, D. T. W. (2013). Salivary microRNAs and oral cancer detection. *Methods in molecular biology (Clifton, N.J.), 936*, 313-324. doi:10.1007/978-1-62703-083-0_24

Zanetti, M., Gennaro, R., & Romeo, D. (1995). Cathelicidins: a novel protein family with a common proregion and a variable C-terminal antimicrobial domain. *FEBS Lett, 374*(1), 1-5. doi:10.1016/0014-5793(95)01050-o

ATTACHMENTS

Supplementary 1: ELISA kit supplier details.
Supplementary 2: Translation of triptann protocol spytprøvetaking
Supplementary 3: Total protein assay pilot results
Supplementary 4: LL37 ELISA assay pilot results
Supplementary 5: Statherin ELISA pilot studies results
Supplementary 6: Raw data, all assays
Supplementary 7: Coding from R
Supplementary 8: Strobe guidelines checklist

SUPPLEMENTARY 1: ELISA KIT SUPPLIER DETAILS.

LL37 ELISA kit

Manufacturer: Hycult Biotech, Frontstraat 2a, 5405, PB Uden, The Netherlands

Hycult catalogue no. HK321

Supplier: Nordic Biosite Nordic BioSite AS, Arbins Gate 2, 0253 Oslo, Norway

Statherin ELISA kit

Manufacturer: Cusabio Technology LLc, 7707 Fannin St., Ste 200-V126, Houston, TX 77054, USA

Cusabio catalogue no. CSB-EL022817HU

Supplier: Nordic Biosite Nordic BioSite AS, Arbins Gate 2, 0253 Oslo, Norway

SUPPLEMENTARY 2: TRANSLATION OF TRIPTANN PROTOCOL SPYTPRØVETAKING

Protokoll TRIP-tann klinisk undersøkelse: Spytprøvetaking (personal communication- T. Børsting, accessed 2021)

Based on: University of Oslo. (2010). Clinical routines - Cariology. https://www.odont.uio.no/studier/ressurser/kariologi/Kariesutredning/Salivatester/saliva_sekresjonshastighet.html

The saliva test must be performed before all the other examinations to get the most accurate sample of the participant's normal saliva production and saliva content.

General instructions for collecting saliva samples:

- The patient is placed in a bright, comfortable, temperate room in a chair with a straight back

- The patient should sit still for approx. 5 min before the sample is taken

- The patient should not have eaten, drunk (anything other than water) or had anything in their mouth in the last hour before the examination

- The patient should not use rinse aid or brush their teeth on the same day as the examination takes place

- They must also not have taken other medications than usual the night before or the same day as the examination takes place or have performed hard physical exercise

- All medications used must be noted on the examination form

- The tube into which the saliva sample is taken must be weighed with a lid before the sample is taken, and weight is written on the tube and in an electronic form for the saliva samples. This is done for each sample since the collection tubes may have different weights.

Instructions for chewing stimulated saliva sample:

- Write down in the examination form whether the participant has fasted for 1 hour before the test and if not, how long it has been since they ate / drank something other than water, and whether they have brushed their teeth and / or used mouthwash before they arrived and if so how long ago last.

- Write down if the participant has taken any medication in the last 24 hours before the saliva test.

- The participant should then chew on a paraffin tablet for half a minute so that it becomes soft, saliva is swallowed.

- The participant is then asked to chew at normal frequency and pressure (as fast and as hard as when the participant eats), start timing the saliva sample from this point.

- Chew a little on one side, then on the other side (change regularly).

- At the same time, the participant should spit regularly in a 15ml collection tube (which was weighed before the examination) with a glass funnel on top to catch the saliva, ingestion should be avoided.

- Saliva is collected for 5 minutes. It is ok if it is not exactly 5 min or a little over, as long as you write down on the form exactly how long the saliva was collected in minutes and seconds, plus time of day the saliva sample was taken.

SUPPLEMENTARY 3: TOTAL PROTEIN ASSAY PILOT RESULTS

	Mean [TP] (range)	**Mean intraassay %CV (range)**	**Change in [TP] after 1 hour (9 samples)**	**Change in [TP] after 7 days (1 sample)**	**Interassay %CV**
WSH	550.65 (327 – 912)	4.69 (2.55 – 8.62)	18 – 54% decrease	*ND (expect same as PS)*	*ND*
WSS	434.63 (299 – 746)	4.51 (0.11 – 7.7)	*< 1% change*	Approx. 24% decrease in TP	17.8%
PS	425.33 (237 – 720)	6.05 (3.74 – 9.88)	Approx. 3% decrease	Approx. 38% decrease in TP	*ND*
	Total protein concentration (µg/mL) using Bradford (Coomassie G-250) method WSH = whole vortexed saliva WSS = the supernatant after centrifuging WSH PS = the passively occurring supernatant Results from eight samples except where indicated. *ND* = not determined				

SUPPLEMENTARY 4: LL37 ELISA ASSAY PILOT RESULTS

A) LL37 concentrations in two saliva samples tested as whole, vortexed saliva (WSH) and supernatant (incubated 1hr room temperature, dilution 1:4)

	Preparation	LL37 concentration (ng/mL)
Sample 1	WSH	36.53
	Supernatant	23.32
Sample2	WSH	37.84
	Supernatant	4.84

B) LL37 ELISA pilot studies raw data showing mean %CVs with varying preparation, incubation and dilution conditions. Results summarised from two to three experiments, each in duplicate or triplicate (%CV > 15 in italics).

WSH	Supernatant	1hr	24hr	1:4	1:25	1:50	%CVs
x		x		x			2.73, 4.24, 5.51
	x	x		x			5.34, 6.39, *80.55*
x		x			x		*17.44, 18.41*
	x	x			x		2.05, *65.12*
x		x				x	2.38, *15.86*
	x	x				x	3.12, *29.64*
x			x		x		2.55, *16.49*
	x		x		x		6.96, *24.9*
x			x			x	2.72, *50.64*
	x		x			x	7.86, *28.72*

SUPPLEMENTARY 5: STATHERIN ELISA PILOT STUDIES RESULTS

Statherin ELISA pilot studies raw data showing mean %CVs with varying incubation and dilution conditions. Results summarised from two experiments, each in duplicate or triplicate (%CV > 15 in italics)

2hr	24hr	1:4	1:10	%CVs
x		x		*17, 61*
x			x	*48, 49*
	x	x		4.73, 8.06
	x		x	10.4, 10.9

SUPPLEMENTARY 6: RAW DATA, ALL ASSAYS

Total protein concentrations (Coomassie G-250; µg/mL), LL37 concentrations (ELISA; ng/mL) and statherin concentrations (ELISA; ng/mL) (ND = not determined).

#	[TP]	[LL37]	[Statherin]	#	[TP]	[LL37]	[Statherin]
1	326.0	39.7	6332	**89**	702.0	91.5	4288
2	745.0	86.2	4314	**90**	421.3	143.9	3517
3	593.0	48.9	5740	**91**	571.4	37.4	4040
4	280.0	17.7	6366	**92**	702.0	52.8	3678
5	428.0	43.5	ND	**93**	850.0	38.4	8492
6	434.0	55.0	2355	**94**	321.4	92.3	3109
7	721.0	57.6	6647	**95**	1437.4	19.0	6493
8	482.3	189.0	3989	**96**	340.5	95.1	4068
9	332.0	32.8	4275	**97**	406.2	43.6	2798
10	271.0	34.5	1898	**98**	566.0	41.3	3216
11	1093.5	59.5	2083	**99**	525.7	158.8	ND
12	452.0	77.2	1068	**100**	436.3	50.2	3479
13	372.6	70.9	3350	**101**	501.5	270.7	4728
14	443.9	32.2	1611	**102**	506.8	218.5	3536
15	393.1	16.8	1824	**103**	467.5	209.7	3412
16	463.0	124.7	3926	**104**	735.0	40.6	5575
17	214.6	87.1	1060	**105**	567.0	14.7	1480
18	183.0	28.8	4124	**106**	372.1	53.7	1361
19	1260.0	126.0	ND	**107**	606.1	110.1	2066
20	325.0	59.7	6139	**108**	276.0	12.4	3024
21	305.6	71.4	1327	**109**	481.0	19.8	2459
22	709.0	34.7	9685	**110**	716.2	120.4	6564
23	322.6	85.3	4851	**111**	262.8	48.6	2103
24	325.3	44.3	3224	**112**	666.0	26.6	4306
25	780.4	49.7	4530	**113**	476.9	80.5	4393
26	470.2	60.1	ND	**114**	280.0	16.3	4107
27	712.4	34.9	1776	**115**	821.8	103.3	5991
28	346.3	44.9	1889	**116**	443.4	103.4	3726
29	1012.1	90.9	4076	**117**	488.0	176.4	4173
30	1120.1	65.0	3847	**118**	329.0	54.1	3631
31	579.0	84.6	5267	**119**	413.1	39.8	2141
32	398.0	35.4	580.8	**120**	455.9	45.7	4642
33	471.0	53.8	3926	**121**	604.1	193.8	2892
34	350.5	62.3	5248	**122**	494.0	70.3	3176
35	400.0	48.1	4426	**123**	609.1	116.2	6098
36	373.0	37.6	4314	**124**	500.1	130.4	5219

37	375.8	100.7	2826	***125***	1114.0	97.0	2011
38	229.0	62.5	2822	***126***	673.1	98.8	5856
39	825.8	38.6	3201	***127***	299.6	33.1	2817
40	302.0	46.8	2355	***128***	259.1	64.2	3612
41	516.0	59.7	5223	***129***	314.0	46.1	4882
42	362.0	90.9	1495	***130***	395.8	47.5	4575
43	889.0	79.9	4594	***131***	532.4	79.0	5045
44	343.0	40.7	4068	***132***	565.0	151.8	6924
45	344.6	70.3	2949	***133***	500.5	107.7	2099
46	901.7	69.6	4060	***134***	1432.4	93.4	7510
47	1078.8	211.7	4658	***135***	425.0	29.4	6535
48	686.0	76.9	4132	***136***	383.0	35.8	6749
49	511.4	108.5	4011	***137***	556.2	122.2	4489
50	1055.0	44.6	4658	***138***	937.0	80.1	6049
51	409.0	65.8	4100	***139***	242.0	14.8	3280
52	342.0	28.9	2941	***140***	434.6	104.9	5209
53	336.7	63.6	5084	***141***	261.3	107.5	4767
54	443.0	135.9	5238	***142***	1038.8	45.8	4426
55	276.0	37.4	3271	***143***	1134.9	33.8	3389
56	325.3	115.8	4155	***144***	1115.9	37.3	2172
57	367.0	92.9	2838	***145***	499.0	69.4	3784
58	395.1	87.9	4824	***146***	366.0	35.8	1242
59	934.6	87.2	3082	***147***	286.0	69.1	817.4
60	331.7	104.0	2422	***148***	295.0	44.0	1007
61	327.0	20.2	4243	***149***	489.0	61.3	3374
62	159.7	47.9	2225	***150***	285.0	88.3	2387
63	542.7	50.5	4844	***151***	326.2	46.1	7383
64	1296.3	38.1	3194	***152***	206.0	25.5	2956
65	1471.0	69.1	ND	***153***	255.0	75.6	1857
66	264.0	29.7	4235	***154***	369.0	39.7	2379
67	895.0	124.5	5142	***155***	800.0	64.7	2268
68	454.0	50.1	4859	***156***	282.0	23.4	2091
69	179.0	28.3	3334	***157***	333.0	53.2	ND
70	454.0	52.0	2180	***158***	697.0	71.7	3145
71	417.0	53.4	4770	***159***	613.9	240.8	2172
72	978.8	59.5	5215	***160***	289.0	52.6	2883
73	393.2	82.4	3289	***161***	754.0	47.6	4757
74	486.9	73.0	5180	***162***	544.0	41.7	4126
75	179.0	81.1	4826	***163***	296.3	60.3	2544
76	409.0	45.1	3429	***164***	364.5	59.5	3840
77	411.3	123.1	ND	***165***	572.6	153.4	3916
78	293.4	65.4	3384	***166***	307.6	164.1	3270
79	472.5	171.9	4326	***167***	406.0	52.6	5199
80	370.0	61.6	1190	***168***	699.0	68.0	5061
81	232.9	17.8	1842	***169***	292.0	20.0	817.4

82	220.0	21.8	2666	***170***	416.1	123.7	4193
83	277.2	90.2	2356	***171***	686.0	82.5	3697
84	412.1	60.0	5450	***172***	361.3	26.1	1545
85	282.0	31.0	2234	***173***	563.7	109.0	ND
86	244.0	69.3	4135	***174****	516.1	ND	ND
87	368.7	40.4	2150	***175****	395	ND	ND
88	458.0	85.9	2628	* same population, no health data			

SUPPLEMENTARY 7: CODING FROM R

```
#LL37 models
> #model A
> LL37modelA.glm<-glm(dmftfact~LL37concn, data=data_df, family = "binomial")
> summary(LL37modelA.glm)

Call:
  glm(formula = dmftfact ~ LL37concn, family = "binomial", data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-0.9958  -0.7162  -0.6811  -0.6455   1.8378

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.534581   0.338863  -4.529 5.94e-06 ***
LL37concn    0.004031   0.003723   1.083    0.279
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 184.66  on 172  degrees of freedom
Residual deviance: 183.53  on 171  degrees of freedom
AIC: 187.53

Number of Fisher Scoring iterations: 4
> exp(cbind(OR = coef(LL37modelA.glm), confint(LL37modelA.glm)))

> #model B1
> LL37modelB1.glm<-glm(dmftfact~LL37concn+agefact+sexfact+vitDfact, data=data_df,
family = "binomial")
> summary(LL37modelB1.glm)

Call:
  glm(formula = dmftfact ~ LL37concn + agefact + sexfact + vitDfact,
      family = "binomial", data = data_df)
```

```
Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.0822  -0.7360  -0.6696  -0.5356   2.0322

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.847528   1.182993  -1.562    0.118
LL37concn    0.004650   0.003795   1.225    0.220
agefact8    -0.075992   1.166154  -0.065    0.948
agefact9     0.409759   1.232674   0.332    0.740
sexfact2    -0.180427   0.374352  -0.482    0.630
vitDfact1    0.436664   0.498067   0.877    0.381

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 184.66  on 172  degrees of freedom
Residual deviance: 181.42  on 167  degrees of freedom
AIC: 193.42

Number of Fisher Scoring iterations: 4
> exp(cbind(OR = coef(LL37modelB1.glm), confint(LL37modelB1.glm)))

> #modelB2
> LL37modelB2.glm<-glm(dmftfact~LL37concn+agefact+data_TPconcn+vitDfact,
data=data_df, family = "binomial")
> summary(LL37modelB2.glm)

Call:
  glm(formula = dmftfact ~ LL37concn + agefact + TPconcn +
        vitDfact, family = "binomial", data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.4739  -0.6967  -0.5896  -0.4224   2.1527

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.8236362  1.2548003  -2.250  0.02443 *
```

```
LL37concn         0.0035885  0.0039314   0.913  0.36136
agefact8         -0.3067262  1.1898805  -0.258  0.79658
agefact9          0.2453511  1.2625236   0.194  0.84591
TPconcn           0.0020972  0.0006558   3.198  0.00139 **
vitDfact1         0.4719725  0.5184668   0.910  0.36265
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 184.66  on 172  degrees of freedom
Residual deviance: 171.27  on 167  degrees of freedom
AIC: 183.27

Number of Fisher Scoring iterations: 4
> exp(cbind(OR = coef(LL37modelB2.glm), confint(LL37modelB2.glm)))

> #model B3
> LL37modelB3.glm<-glm(dmftfact~LL37concn+agefact+volumespit+vitDfact,
data=data_df, family = "binomial")
> summary(LL37modelB3.glm)

Call:
  glm(formula = dmftfact ~ LL37concn + agefact + volumespit + vitDfact,
      family = "binomial", data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.0403  -0.7365  -0.6711  -0.5384   1.9869

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.852493   1.224440  -1.513    0.130
LL37concn    0.004548   0.003843   1.184    0.237
agefact8    -0.164622   1.158447  -0.142    0.887
agefact9     0.322484   1.231776   0.262    0.793
volumespit   0.004613   0.079766   0.058    0.954
vitDfact1    0.426748   0.498082   0.857    0.392
```

```
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 184.66  on 172  degrees of freedom
Residual deviance: 181.65  on 167  degrees of freedom
AIC: 193.65

Number of Fisher Scoring iterations: 4
> exp(cbind(OR = coef(LL37modelB3.glm), confint(LL37modelB3.glm)))

> #model C1
> LL37modelC1.glm<-
glm(dmftfact~LL37concn+agefact+sexfact+vitDfact+fluoridecatsfact+lactofact+toothbru
shingcatsfact+educationfact+DairyDessertcatsfact+SugaryDrinkcatsfact+saltime4catsfa
ct, data=data_df, family = "binomial")
> summary(LL37modelC1.glm)

Call:
  glm(formula = dmftfact ~ LL37concn + agefact + sexfact + vitDfact +
        fluoridecatsfact + lactofact + toothbrushingcatsfact + educationfact +
        DairyDessertcatsfact + SugaryDrinkcatsfact + saltime4catsfact,
      family = "binomial", data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.5548  -0.7033  -0.5062  -0.2709   2.4310

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              0.623902   2.051596   0.304    0.761
LL37concn                0.003262   0.004179   0.780    0.435
agefact8                -0.773498   1.244189  -0.622    0.534
agefact9                -0.295569   1.344383  -0.220    0.826
sexfact2                -0.252717   0.410660  -0.615    0.538
vitDfact1                0.514815   0.530000   0.971    0.331
fluoridecatsfact2       -0.600223   0.428658  -1.400    0.161
lactofact1               0.819474   0.452241   1.812    0.070 .
toothbrushingcatsfact2   0.419024   0.471476   0.889    0.374
educationfact2          -0.657804   0.831837  -0.791    0.429
educationfact3          -0.176292   0.814467  -0.216    0.829
DairyDessertcatsfact1   -2.006478   1.376753  -1.457    0.145
```

```
DairyDessertcatsfact2  -1.158496   1.370731  -0.845    0.398
DairyDessertcatsfact3  -1.087174   1.544402  -0.704    0.481
SugaryDrinkcatsfact1    0.251446   0.557421   0.451    0.652
SugaryDrinkcatsfact2    0.547512   0.621834   0.880    0.379
SugaryDrinkcatsfact3    0.375289   1.354419   0.277    0.782
saltime4catsfact2      -0.506962   0.551457  -0.919    0.358
saltime4catsfact3      -0.770169   0.550315  -1.400    0.162
saltime4catsfact4      -1.388121   0.917171  -1.513    0.130
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 184.15  on 171  degrees of freedom
Residual deviance: 161.11  on 152  degrees of freedom
(1 observation deleted due to missingness)
AIC: 201.11

Number of Fisher Scoring iterations: 5
> exp(cbind(OR = coef(LL37modelC1.glm), confint(LL37modelC1.glm)))

> #model C2
> LL37modelC2.glm<-
glm(dmftfact~LL37concn+agefact+vitDfact+TPconcn+fluoridecatsfact+lactofact+toothbru
shingcatsfact+educationfact+DairyDessertcatsfact+SugaryDrinkcatsfact+saltime4catsfa
ct, data=data_df, family = "binomial")
> summary(LL37modelC2.glm)

Call:
  glm(formula = dmftfact ~ LL37concn + agefact + vitDfact + TPconcn +
        fluoridecatsfact + lactofact + toothbrushingcatsfact + educationfact +
        DairyDessertcatsfact + SugaryDrinkcatsfact + saltime4catsfact,
      family = "binomial", data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.3958  -0.6948  -0.4495  -0.1825   2.7055

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)             -0.0076556  2.0667650  -0.004  0.99704
LL37concn                0.0024195  0.0042434   0.570  0.56856
agefact8                -1.0731218  1.2604932  -0.851  0.39457
agefact9                -0.4814401  1.3554362  -0.355  0.72245
vitDfact1                0.4662853  0.5460151   0.854  0.39312
TPconcn                  0.0023104  0.0007739   2.985  0.00283 **
fluoridecatsfact2       -0.4985402  0.4376320  -1.139  0.25463
lactofact1               0.8907698  0.4714339   1.889  0.05883 .
toothbrushingcatsfact2   0.5310218  0.5081349   1.045  0.29600
educationfact2          -0.8931479  0.8519540  -1.048  0.29448
educationfact3          -0.1899884  0.8274170  -0.230  0.81839
DairyDessertcatsfact1   -2.2337147  1.3794984  -1.619  0.10540
DairyDessertcatsfact2   -1.5792257  1.3830642  -1.142  0.25352
DairyDessertcatsfact3   -1.3738762  1.5673522  -0.877  0.38073
SugaryDrinkcatsfact1     0.1919130  0.5680159   0.338  0.73546
SugaryDrinkcatsfact2     0.3172222  0.6496163   0.488  0.62532
SugaryDrinkcatsfact3     0.6056305  1.3710206   0.442  0.65868
saltime4catsfact2       -0.6759214  0.5806160  -1.164  0.24437
saltime4catsfact3       -0.8319749  0.5711483  -1.457  0.14521
saltime4catsfact4       -1.6205357  0.9538331  -1.699  0.08932 .
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 184.15  on 171  degrees of freedom
Residual deviance: 151.74  on 152  degrees of freedom
(1 observation deleted due to missingness)
AIC: 191.74

Number of Fisher Scoring iterations: 5
> exp(cbind(OR = coef(LL37modelC2.glm), confint(LL37modelC2.glm)))


> #model C3
> LL37modelC3.glm<-
glm(dmftfact~LL37concn+agefact+volumespit+vitDfact+fluoridecatsfact+lactofact+tooth
brushingcatsfact+educationfact+DairyDessertcatsfact+SugaryDrinkcatsfact+saltime4cat
sfact, data=data_df, family = "binomial")
> summary(LL37modelC3.glm)
```

```
Call:
  glm(formula = dmftfact ~ LL37concn + agefact + volumespit + vitDfact +
        fluoridecatsfact + lactofact + toothbrushingcatsfact + educationfact +
        DairyDessertcatsfact + SugaryDrinkcatsfact + saltime4catsfact,
      family = "binomial", data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.4995  -0.6823  -0.5075  -0.2815   2.4780

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              0.460503   2.090395   0.220   0.8256
LL37concn                0.003305   0.004222   0.783   0.4337
agefact8                -0.918540   1.239041  -0.741   0.4585
agefact9                -0.485173   1.348418  -0.360   0.7190
volumespit               0.024056   0.088314   0.272   0.7853
vitDfact1                0.524642   0.529255   0.991   0.3215
fluoridecatsfact2       -0.618268   0.431712  -1.432   0.1521
lactofact1               0.813358   0.452543   1.797   0.0723 .
toothbrushingcatsfact2   0.431839   0.471630   0.916   0.3599
educationfact2          -0.604870   0.858358  -0.705   0.4810
educationfact3          -0.157241   0.829483  -0.190   0.8496
DairyDessertcatsfact1   -1.984878   1.369755  -1.449   0.1473
DairyDessertcatsfact2   -1.167105   1.366276  -0.854   0.3930
DairyDessertcatsfact3   -1.075317   1.540255  -0.698   0.4851
SugaryDrinkcatsfact1     0.312500   0.557387   0.561   0.5750
SugaryDrinkcatsfact2     0.617190   0.624608   0.988   0.3231
SugaryDrinkcatsfact3     0.493301   1.374828   0.359   0.7197
saltime4catsfact2       -0.535754   0.549751  -0.975   0.3298
saltime4catsfact3       -0.774370   0.551019  -1.405   0.1599
saltime4catsfact4       -1.435477   0.912519  -1.573   0.1157
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 184.15  on 171  degrees of freedom
```

```
Residual deviance: 161.42  on 152  degrees of freedom
(1 observation deleted due to missingness)
AIC: 201.42

Number of Fisher Scoring iterations: 5
> exp(cbind(OR = coef(LL37modelC3.glm), confint(LL37modelC3.glm)))

> #statherin models
> #model A
> stathmodelA.glm<-glm(dmftfact~stathconcn, data=data_df, family = "binomial")
> summary(stathmodelA.glm)

Call:
  glm(formula = dmftfact ~ stathconcn, family = "binomial", data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.3471  -0.7381  -0.6263  -0.4973   2.1477

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3667     0.5263  -4.497 6.89e-06 ***
 stathconcn    0.2847     0.1190   2.392   0.0168 *
  ---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 175.63  on 164  degrees of freedom
Residual deviance: 169.71  on 163  degrees of freedom
(8 observations deleted due to missingness)
AIC: 173.71

Number of Fisher Scoring iterations: 4
> exp(cbind(OR = coef(stathmodelA.glm), confint(stathmodelA.glm)))

> #model B1
> stathmodelB1.glm<-glm(dmftfact~stathconcn+sexfact, data=data_df, family =
"binomial")
```

```
> summary(stathmodelB1.glm)

Call:
  glm(formula = dmftfact ~ stathconcn + sexfact, family = "binomial",
      data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.3186  -0.7431  -0.6215  -0.4991   2.1226

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3076     0.5506  -4.191 2.78e-05 ***
 stathconcn    0.2856     0.1189   2.403   0.0163 *
 sexfact2     -0.1322     0.3817  -0.346   0.7291
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 175.63  on 164  degrees of freedom
Residual deviance: 169.59  on 162  degrees of freedom
(8 observations deleted due to missingness)
AIC: 175.59

Number of Fisher Scoring iterations: 4
> exp(cbind(OR = coef(stathmodelB1.glm), confint(stathmodelB1.glm)))

> #modelB2
> stathmodelB2.glm<-glm(dmftfact~stathconcn+TPconcn, data=data_df, family =
"binomial")
> summary(stathmodelB2.glm)

Call:
  glm(formula = dmftfact ~ stathconcn + TPconcn, family = "binomial",
      data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
```

```
-1.2760  -0.7067  -0.5730  -0.4494   2.2439

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.0506171  0.6147630  -4.962 6.97e-07 ***
stathconcn   0.2005589  0.1252359   1.601  0.10928
TPconcn      0.0018859  0.0007245   2.603  0.00925 **
  ---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 175.63  on 164  degrees of freedom
Residual deviance: 162.92  on 162  degrees of freedom
(8 observations deleted due to missingness)
AIC: 168.92

Number of Fisher Scoring iterations: 4
> exp(cbind(OR = coef(stathmodelB2.glm), confint(stathmodelB2.glm)))

> #model B3
> stathmodelB3.glm<-glm(dmftfact~stathconcn+volumespit, data=data_df, family =
"binomial")
> summary(stathmodelB3.glm)

Call:
  glm(formula = dmftfact ~ stathconcn + volumespit, family = "binomial",
      data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.3452  -0.7375  -0.6254  -0.4974   2.1454

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.351055   0.658730  -3.569 0.000358 ***
stathconcn   0.284487   0.119092   2.389 0.016903 *
volumespit  -0.003275   0.083179  -0.039 0.968593
---
```

```
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 175.63  on 164  degrees of freedom
Residual deviance: 169.71  on 162  degrees of freedom
(8 observations deleted due to missingness)
AIC: 175.71

Number of Fisher Scoring iterations: 4
> exp(cbind(OR = coef(stathmodelB3.glm), confint(stathmodelB3.glm)))

> #model C1
> stathmodelC1.glm<-
glm(dmftfact~stathconcn+sexfact+fluoridecatsfact+lactofact+toothbrushingcatsfact+ed
ucationfact+DairyDessertcatsfact+SugaryDrinkcatsfact+saltime4catsfact,
data=data_df, family = "binomial")
> summary(stathmodelC1.glm)

Call:
  glm(formula = dmftfact ~ stathconcn + sexfact + fluoridecatsfact +
        lactofact + toothbrushingcatsfact + educationfact + DairyDessertcatsfact +
        SugaryDrinkcatsfact + saltime4catsfact, family = "binomial",
      data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.6571  -0.6886  -0.4855  -0.1939   2.7542

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              -0.9929     1.7839  -0.557   0.5778
stathconcn                0.2718     0.1332   2.041   0.0412 *
sexfact2                 -0.1003     0.4304  -0.233   0.8158
fluoridecatsfact2        -0.8245     0.4499  -1.833   0.0668 .
lactofact1                0.9789     0.4650   2.105   0.0353 *
toothbrushingcatsfact2    0.7613     0.5111   1.490   0.1364
educationfact2           -0.6046     0.8355  -0.724   0.4692
educationfact3           -0.1557     0.7952  -0.196   0.8448
DairyDessertcatsfact1    -1.7412     1.4231  -1.223   0.2211
```

```
DairyDessertcatsfact2   -1.0095     1.4293  -0.706   0.4800
DairyDessertcatsfact3   -0.8062     1.5850  -0.509   0.6110
SugaryDrinkcatsfact1     0.1636     0.5725   0.286   0.7751
SugaryDrinkcatsfact2     0.7342     0.6383   1.150   0.2500
SugaryDrinkcatsfact3     1.2134     1.4487   0.838   0.4022
saltime4catsfact2       -0.4491     0.5679  -0.791   0.4290
saltime4catsfact3       -0.9604     0.5748  -1.671   0.0947 .
saltime4catsfact4       -1.5141     0.9335  -1.622   0.1048
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 175.12  on 163  degrees of freedom
Residual deviance: 147.08  on 147  degrees of freedom
(9 observations deleted due to missingness)
AIC: 181.08

Number of Fisher Scoring iterations: 5
> exp(cbind(OR = coef(stathmodelC1.glm), confint(stathmodelC1.glm)))

> #model C2
> stathmodelC2.glm<-
glm(dmftfact~stathconcn+TPconcn+fluoridecatsfact+lactofact+toothbrushingcatsfact+ed
ucationfact+DairyDessertcatsfact+SugaryDrinkcatsfact+saltime4catsfact,
data=data_df, family = "binomial")
> summary(stathmodelC2.glm)

Call:
  glm(formula = dmftfact ~ stathconcn + TPconcn + fluoridecatsfact +
        lactofact + toothbrushingcatsfact + educationfact + DairyDessertcatsfact +
        SugaryDrinkcatsfact + saltime4catsfact, family = "binomial",
      data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.6383  -0.7051  -0.4330  -0.1523   2.8411

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)            -1.4128119  1.7669950  -0.800   0.4240
stathconcn              0.1759268  0.1442201   1.220   0.2225
TPconcn                 0.0020104  0.0008813   2.281   0.0225 *
fluoridecatsfact2      -0.6998524  0.4588196  -1.525   0.1272
lactofact1              0.9871715  0.4732841   2.086   0.0370 *
toothbrushingcatsfact2  0.8304635  0.5358597   1.550   0.1212
educationfact2         -0.8057627  0.8554773  -0.942   0.3463
educationfact3         -0.1415368  0.8110413  -0.175   0.8615
DairyDessertcatsfact1  -1.9327891  1.4187782  -1.362   0.1731
DairyDessertcatsfact2  -1.3383468  1.4315037  -0.935   0.3498
DairyDessertcatsfact3  -1.0627879  1.6057153  -0.662   0.5080
SugaryDrinkcatsfact1    0.1343959  0.5733000   0.234   0.8147
SugaryDrinkcatsfact2    0.5146859  0.6558275   0.785   0.4326
SugaryDrinkcatsfact3    1.5623413  1.4483673   1.079   0.2807
saltime4catsfact2      -0.6149028  0.5910970  -1.040   0.2982
saltime4catsfact3      -0.9865991  0.5910601  -1.669   0.0951 .
saltime4catsfact4      -1.7337769  0.9731304  -1.782   0.0748 .
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 175.12  on 163  degrees of freedom
Residual deviance: 141.65  on 147  degrees of freedom
(9 observations deleted due to missingness)
AIC: 175.65

Number of Fisher Scoring iterations: 5
> exp(cbind(OR = coef(stathmodelC2.glm), confint(stathmodelC2.glm)))


> #model C3
> stathmodelC3.glm<-
glm(dmftfact~stathconcn+volumespit+fluoridecatsfact+lactofact+toothbrushingcatsfact
+educationfact+DairyDessertcatsfact+SugaryDrinkcatsfact+saltime4catsfact,
data=data_df, family = "binomial")
> summary(stathmodelC3.glm)


Call:
  glm(formula = dmftfact ~ stathconcn + volumespit + fluoridecatsfact +
        lactofact + toothbrushingcatsfact + educationfact + DairyDessertcatsfact +
```

```
        SugaryDrinkcatsfact + saltime4catsfact, family = "binomial",
      data = data_df)

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.6261  -0.6906  -0.4806  -0.1896   2.7798

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)             -1.11975    1.83866  -0.609   0.5425
stathconcn               0.27331    0.13301   2.055   0.0399 *
volumespit               0.01034    0.09088   0.114   0.9094
fluoridecatsfact2       -0.83128    0.45318  -1.834   0.0666 .
lactofact1               0.97340    0.46441   2.096   0.0361 *
toothbrushingcatsfact2   0.77128    0.50950   1.514   0.1301
educationfact2          -0.58747    0.86043  -0.683   0.4948
educationfact3          -0.14927    0.81018  -0.184   0.8538
DairyDessertcatsfact1   -1.73406    1.42349  -1.218   0.2232
DairyDessertcatsfact2   -1.01514    1.43146  -0.709   0.4782
DairyDessertcatsfact3   -0.79836    1.58645  -0.503   0.6148
SugaryDrinkcatsfact1     0.18910    0.57199   0.331   0.7409
SugaryDrinkcatsfact2     0.76204    0.63841   1.194   0.2326
SugaryDrinkcatsfact3     1.26524    1.44232   0.877   0.3804
saltime4catsfact2       -0.46263    0.56499  -0.819   0.4129
saltime4catsfact3       -0.96490    0.57555  -1.676   0.0936 .
saltime4catsfact4       -1.55230    0.92083  -1.686   0.0918 .
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 175.12  on 163  degrees of freedom
Residual deviance: 147.12  on 147  degrees of freedom
(9 observations deleted due to missingness)
AIC: 181.12

Number of Fisher Scoring iterations: 5
> exp(cbind(OR = coef(stathmodelC3.glm), confint(stathmodelC3.glm)))
```

SUPPLEMENTARY 8: STROBE GUIDELINES CHECKLIST

(Elm & Altman *et al.*, 2007)

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page numbers
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	vi
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	vi
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	1
Objectives	3	State specific objectives, including any prespecified hypotheses	8
Methods			
Study design	4	Present key elements of study design early in the paper	9, 14
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	14
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	14
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	15, 22
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	20
Bias	9	Describe any efforts to address potential sources of bias	22
Study size	10	Explain how the study size was arrived at	15
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	16
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	22
		(*b*) Describe any methods used to examine subgroups and interactions	
		(*c*) Explain how missing data were addressed	
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	

Continued on next page

			Page numbers
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	15
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	28
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	28
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	36
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	37
Discussion			
Key results	18	Summarise key results with reference to study objectives	52
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	50
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	52
Generalisability	21	Discuss the generalisability (external validity) of the study results	50
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.