

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Evaluation of a national complex oral health improvement programme: a population data linkage cohort study
<b>AUTHORS</b>	Kidd, Jamie; McMahon, Alex; Sherriff, Andrea; Gnich, Wendy; Mahmoud, Ahmed; Macpherson, Lorna; Conway, D

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Leonard Crocombe Centre for Rural Health, University of Tasmania, Australia
<b>REVIEW RETURNED</b>	29-Mar-2020

<b>GENERAL COMMENTS</b>	This is a very well written and presented paper and is of importance. We can but hope that the Scottish Government has a re-think about fluoridation of the public water supply.
-------------------------	--

<b>REVIEWER</b>	Bradley Christian La Trobe University, Australia
<b>REVIEW RETURNED</b>	11-Apr-2020

<b>GENERAL COMMENTS</b>	<p><b>General comments</b></p> <p>This study, which is part of the Childsmile programme's evaluation plan, used data linkage to investigate: (1) the programme's reach and (2) impact on child caries experience.</p> <p>This is a very important evaluation piece and has the potential to inform oral health promotion at the global level.</p> <p>In general, the paper is reads ok. However, there are issues in the way the data analysis, results, and discussion sections are reported. Improvements are needed in the use of scientific language, particularly related to the reporting of the data analysis section, the statistics and the related analysis results.</p> <p>This is a cross-section analysis of data. As such the term effectiveness should be completely avoided, as effectiveness of an intervention is in most cases demonstrated by comparison to a matched group. With a cross-section study design, all you can demonstrate are associations between exposures and an outcome.</p> <p>The discussion section has no discussion of the results for the programme's reach and possible reasons for the low reach for some interventions. For example, it appears that the reach for the FV intervention was low – the majority only received 1 application during the study period.</p> <p>The paper needs major revision.</p>
-------------------------	--

## Specific comments

### 1. Abstract

1.1 Design section: The study design should be stated as cross-sectional. The word 'longitudinal' should be taken out from this section. It is misleading as it gives the impression that longitudinal analyses were conducted on this cohort.

1.2 Primary Outcome Measure: Isn't 'reach' a primary outcome too?

1.3 Results: Please check all reported percentages for accuracy. For example, 30% of 50,379 is 15,114 and not 15,032 as reported in this section of the paper. If there is some reason for the discrepancies, please state them.

1.4 Results, Sentence 2: Please state the reach upfront for the targeted interventions (FVA and DHSW). For the targeted interventions the reach should be reported for the particular targeted population and not the whole cohort.

1.5 Results, last sentence: What does 'less clear' and 'unchanged' mean? This is the results section and the estimates should be reported as is.

The simple way to state this result given your cross-section study design is: 'DHSW and FVAs were not associated with caries experience in this study'

1.6 Conclusion: The first sentence is not a conclusion and should be removed. The way the conclusion is currently written is a clear example of the inappropriate use of scientific/statistical language that is observed throughout the paper. In the first part of sentence 2, "...strongest reductions in risk..." – this should be re-worded to state that caries experience was significantly lower among children exposed to the two (name them) interventions.

The second part of the same sentence: The term 'effective' cannot be used as there was no comparison group, and this was a cross-section study design. It should be re-worded to read something like: "..., with supervised nursery toothbrushing having the greatest impact among children in areas of high deprivation."

The word 'effect' should be replaced with 'impact' throughout the paper.

### 2. Introduction

2.1 Intro, 1<sup>st</sup> paragraph, 2<sup>nd</sup> sentence: Delete the word 'tooth decay'. It is a scientific paper. Also, in this sentence it is not clear how caries can peak in 1- to 4-year-olds? It should peak at a narrower age range, for example at 5 years of age or 4-5 years of age, which makes more sense than 1-4-year-olds.

2.2 Intro, 1<sup>st</sup> paragraph, last sentence: Please state the exact prevalence or prevalence range, rather than "...over 50%..." which could be anything.

2.3 Intro, 2<sup>nd</sup> paragraph: 'Proportionate universal approach' is introduced here. It needs a reference and a bit of a rationale and background for its use.

2.4 Intro, 4<sup>th</sup> paragraph: The aim of the study was not to develop a cohort. This should be re-worded to read like 'Here, we developed a cohort using data linkage of routine administrative data: to....'

### **3. Methods**

3.1 Databases section: The first sentence needs to be revised to clearly demonstrate that 4 of 5 databases used were part of/used for the Childsmile programme. It currently reads as though five unrelated databases were linked.

3.2 Cohort inclusion criteria: This section is not at all clear and needs a re-write.

3.3 Data management: "To assess data completeness and linkage success, the linked cohort data were compared with appropriate published reports" - compared against what variables?

3.4 Outcome and intervention data definitions: No outcome measures reported here for 'programme reach' and the rationale for using them. Also, dmf is an index to quantify and record caries experience, not to diagnose/define. The BASCD criteria was used to define. Please revise to reflect this.

3.5 Statistical analyses: This section definitely needs a re-write for clarity. This section is the heart of the paper and is very under-developed. While the key elements are there it comes across as all over the place. The language around the use of the statistics needs to improve. For example, "Logistic regression was used throughout to model caries experience" – you need to explain why logistic regression, you need to define the outcome variable and how it was coded for the analysis etc. Detail is required in this section. I suggest you structure this section as: cohort description, analysis of aim/objective 1: programme's reach and then, the analysis of aim/objective 2: intervention impact on caries experience.

3.6 Statistical analyses: In the first sentence, you state that SIMD is a continuous variable. This is not correct. SIMD is an ordinal categorical variable.

3.7 Statistical analyses: Unadjusted odds ratios – Again, you have to explain your rationale for using and reporting unadjusted odds ratios. In this study, the use of unadjusted odds ratio does not seem appropriate. In a classic regression analysis, you would use unadjusted regression estimates as part of the model building exercise, i.e., to understand which variables were associated with the outcome of interest; and then use this information to inform which variables to include in the adjusted/final model. Where as here, it appears that the unadjusted serves no purpose. I suggest remove from the paper and only report on Model 1 and Model 2. Also, use one term for consistency - unadjusted or crude.

### **4. Results**

	<p>4.1 The first paragraph can be removed as it is not a result.</p> <p>4.2 All the reported percentages need to be verified for accuracy. The numbers seem to be a bit off. For example, in the paper it is reported that 30% of 50379 is 15032. However, <math>50379 \times 0.30 = 15114</math>.</p> <p>4.3 Table 1 – The odds ratios do not serve any purpose here and should be removed, particularly since this a descriptive table. Also, in these tables, it is more useful to show the column percentages. Readers want to know the distribution of caries by SES, age and sex (not the distribution of caries within each level of the categorial variables).</p> <p>4.4 Table 1: 4-year-olds are included here. Weren't children in the cohort 5 and older in 2014/15? Or do you mean that for the caries analysis you included 5-year-olds?</p> <p>4.5 Paragraphs 3: A table to show all this information would be useful as it is a primary outcome of the study. These paragraphs need revision to focus on: for the two targeted interventions, please only report the most important results for those targeted groups. The reported OR of 0.67 is of no value, as this was a targeted intervention and it would be expected that the programme's reach would be lower among the least deprived. The use of some OR's in this paper appears to be without rationale for their use.</p> <p>4.5 Paragraphs 4: You need to explain/interpret at least some of the ORs in the correct scientific manner, as there are no tables for results paragraphs 3 &amp; 4 to assist the reader. For example, "There was a high level of reach across the population for the (universal) nursery supervised toothbrushing intervention (89% n=44868). The decreasing trend by SIMD was considerably weaker (OR=0.75; 95% 0.73, 0.77)..." – you can write these two sentences in terms of the odds ratio and the reference.</p> <p>4.6 Table 2: The column 'No Caries Experience' is redundant information and can be deleted. Again column % would be more useful. The 'Unadjusted ORs' is of no value and can also be removed from this table.</p> <p>4.6: The results section on interventions and caries experience is too long. Report only on the key findings from model 2 – the full model with all variables.</p> <p><b>5. Discussion</b></p> <p>5.1 Programme reach is the first aim/objective, yet there is nothing in this section about reach. The results of the programme's impact on caries needs to be discussed in the context of the programme's reach. For example, your finding of 49% of the cohort receiving only 1 fluoride varnish application (low reach), in the study period, could be a key reason why this targeted intervention did not have an impact on caries experience.</p> <p>Programme reach needs to be thoroughly discussed in this paper as it provides insight into the realities of implementing population level interventions, that program planners/policy makers can use to improve all aspects of programme efficiencies and effectiveness.</p> <p>5.2 Sentence 1: Were these routine administrative datasets?</p>
--	--

	<p>Weren't 4 datasets part of the Childsmile programme that were linked with 1 administrative database? The 4 datasets that were part of the Childsmile programme were probably setup at the start for evaluation and research. Hence, they are probably not as routine as what occurs in other countries.</p> <p>5.3 Sentence 2: Is there a reference for this sentence? This sentence is about the Programme's fidelity. How do you know it was delivered as planned? Was there an evaluation of implementation process that showed that the programme was delivered as planned?</p> <p>5.4 "...there was insufficient evidence for an independent effect of the nursery fluoride varnish application intervention...." - Comments in relation to this statement:</p> <p>5.4.1 The 'effect' should be replaced with 'association' to read as: "...there was insufficient evidence for an independent <del>effect</del> association of the nursery fluoride varnish application intervention with caries experience."</p> <p>5.4.2 There is nothing further in the discussion as to explain why you may have observed this result.</p> <p>5.4.3 How does this finding compare with the Protecting Teeth@3 Study?</p> <p>5.5 Discussion, paragraph 2, "The results support.....": This is clearly the strongest finding and the discussion around this result needs to be beefed up. Why this reduction? What were the estimates of caries experience (dmf scores in the least deprived VS the most deprived)? How does this compare to the literature? Why do you think this intervention was successful? What were the enablers and barriers? What was the policy context that made this possible?</p> <p>5.6 Discussion, paragraph 2, last sentence: "One possible explanation for this is that children living in the most affluent areas are more likely to already be regularly toothbrushing at home <b>and as a result may have had lower caries experience to begin with.</b>"</p> <p>5.7 Discussion, paragraph 3: The section about the FVA intervention should be a separate paragraph. It is strange to call a 2013 review 'historic'. Again, you need to discuss possible reasons for the FV not working in your population. Did it not reach enough of the population? Did the children get at least 2 FV applications in a 12-month period etc?</p>
--	--

<b>REVIEWER</b>	Sharat Chandra Pani Schulich School of Medicine and Dentistry, University of Western Ontario
<b>REVIEW RETURNED</b>	23-Jun-2020

<b>GENERAL COMMENTS</b>	<p>Abstract</p> <p>Participants: The term approximately 5 years is not acceptable. I realize that the WHO recommends 5 years as a target age, but the authors should either mention mean age or age range in brackets to elaborate on this.</p> <p>Primary outcome measure: Missing is not a valid expression of</p>
-------------------------	--

	<p>primary decay. The authors need to clarify if this was extracted due to caries: the deft?  Conclusion: delete “This is the first population-wide data-linkage cohort study to evaluate a complex public health programme” this is not a finding of the study and is not needed here</p> <p>Strengths and limitations  Rephrase the second point. There is no need to mention the systematic reviews here. Focus instead on the evidence and the evaluation of the evidence in this study.  Are the authors suggesting there are no limitations? Use of records created by multiple dentists? Calibrations? Please mention the limitations here</p> <p>Introduction  Pg4 line 57: Please be specific about the cohort measured. I would recommend the use of the same terminology for age in the abstract introduction and methodology</p> <p>Methods  Page 5 line 30: Dental examination involved... please stick to past tense through the methodology section  Patient identity protection: I was unable to find any mention if the records were de-identified? Or if there were any measures taken to protect identifiable datasets</p> <p>Outcome measures  Lines 54-55: Please clarify the term “missing” – was this all missing teeth or only teeth with a record of extraction due to dental caries. Were there exclusion criteria (congenitally missing teeth). Though teeth missing due to reasons other than dental carries is unlikely , it is important to clarify this.  What is the level of variation/calibration for dental caries in the NDIP database. Are practitioners routinely trained/calibrated for the recording of dental caries. This is a point that needs to be mentioned if present and/or addressed in the discussion if not present.</p> <p>Results  Acceptably presented – the models are clear</p> <p>Discussion  The discussion needs to discuss the limitations of the dataset. The use of population based data has limitations. While the discussion makes valid points it is not complete unless it addresses those limitations</p>
--	---

## VERSION 1 – AUTHOR RESPONSE

### Reply to Reviewer 1

**This is a very well written and presented paper and is of importance. We can but hope that the Scottish Government has a re-think about fluoridation of the public water supply.**

Thank you- we appreciate your feedback.

### Reply to Reviewer 2

**investigate: (1) the programme’s reach and (2) impact on child caries experience.**

**This is a very important evaluation piece and has the potential to inform oral health promotion at the global level.**

**In general, the paper is reads ok. However, there are issues in the way the data analysis, results, and discussion sections are reported. Improvements are needed in the use of scientific language, particularly related to the reporting of the data analysis section, the statistics and the related analysis results.**

**This is a cross-section analysis of data. As such the term effectiveness should be completely avoided, as effectiveness of an intervention is in most cases demonstrated by comparison to a matched group. With a cross-section study design, all you can demonstrate are associations between exposures and an outcome.**

**The discussion section has no discussion of the results for the programme's reach and possible reasons for the low reach for some interventions. For example, it appears that the reach for the FV intervention was low – the majority only received 1 application during the study period.**

**The paper needs major revision.**

We would like to thank Reviewer 2 for recognising the importance of this paper and for their useful and detailed comments. Please see below for a point by point response.

## **Specific comments**

### **(1) Abstract**

**(1.1) Design section: The study design should be stated as cross-sectional. The word 'longitudinal' should be taken out from this section. It is misleading as it gives the impression that longitudinal analyses were conducted on this cohort.**

The study is genuinely longitudinal in nature. The cohort was based on all children who had a valid "caries" outcome in Primary 1 (first year of primary school) in the school year 2014/15. The records of these children (approximately 83% of the whole population) were then linked to the various administrative and Childsmile datasets that collected "intervention/exposure" data prospectively in the years from birth up to the primary endpoint for each child. All intervention/exposure data were ascertained prior to the primary endpoint.

We have stated the time stamp for interventions in the second paragraph of the introduction section (e.g. "nursery (kindergarten) fluoride varnish applications (targeted to children from the of age three-years" and we have now explained the longitudinal nature off this study in more detail in the Methods section:

"We derived appropriate categories for each of the four Childsmile interventions. The number of times (from birth to outcome) a targeted family received a DHSW contact (DHSW Contacts), the number of times a targeted child received a nursery FVA between the age of three-years and outcome (Number of Nursery FVA), and the number of Primary Care Dental Practice Visits interventions a child received between birth and the outcome were calculated. Children in the cohort who were not enrolled at a nursery targeted for the FVA intervention or were not identified by a health visitor as requiring a DHSW Contact were categorised as 'Not targeted' for these interventions. Nursery Supervised Toothbrushing participation was captured using the parent/carer annual consent forms – categorised as the number of years the child was consented to participate in toothbrushing prior to the cohort outcome endpoint".

**(1.2) Primary Outcome Measure: Isn't 'reach' a primary outcome too?**

We have two main outcome measures in this study, and you are correct to say that we have inadvertently omitted “reach” as an outcome in the abstract. We have amended this in the abstract.

“Reach of the programme defined as the percentage of children receiving each intervention at least once by SIMD fifth”.

**(1.3) Results: Please check all reported percentages for accuracy. For example, 30% of 50,379 is 15,114 and not 15,032 as reported in this section of the paper. If there is some reason for the discrepancies, please state them.**

There are no journal guidelines on decimal places so we had rounded percentages up throughout the paper to whole numbers (e.g. 29.8% to 30%) but as a result have lost accuracy, therefore we have taken your comment on board, and for consistency and clarity have reverted to one decimal place throughout the paper.

**(1.4) Results, Sentence 2: Please state the reach upfront for the targeted interventions (FVA and DHSW). For the targeted interventions the reach should be reported for the particular targeted population and not the whole cohort.**

Thank you for your suggestion. We have included the reach for the FVA targeted intervention in the Results section of the abstract and have chosen to summarise reach in the most versus least deprived areas. The Childsmile programme targeting principle is to deliver interventions in a proportionate universal approach, with greatest delivery of the targeted interventions to children from the most deprived areas. Due to the implementation of the FVA programme in nurseries being slightly different across different health boards in Scotland, targeting was based on different criteria, and some health boards autonomously chose not to target at all. We therefore felt it important to look across the SIMD distribution at the proportion of children reached, as well as within the most deprived areas. Targeting of the DHSW intervention was not based on area-based deprivation alone, and was often determined on a judgement made by a health visitor based on the family's need. Therefore, there could legitimately be children/families targeted that did not live in areas of high deprivation, however, there was the expectation that there should be a general trend in reach of the DHSW intervention towards reaching children from the more deprived areas. We have added this information to the results section of the abstract in response to comments below.

“The targeted interventions strongly favoured children from the most deprived areas: DHSW contacts (SIMD1:29.5% vs SIMD5:7.7%), nursery FVAs (SIMD1:75.2% vs SIMD5:23.2%)”.

**(1.5) Results, last sentence: What does 'less clear' and 'unchanged' mean? This is the results section and the estimates should be reported as is. The simple way to state this result given your cross-section study design is: 'DHSW and FVAs were not associated with caries experience in this study'**

We used the wording “less clear” for DHSW intervention as it was associated with a reduction in caries experience for children with just one contact but not for those with two or more contacts. We agree to alter the text for the FVA association, which we agree with, but wish to retain the original text for DHSW, due to the nature of the association described above.

“The findings were less clear for DHSW contacts. Nursery FVAs were not independently associated with caries experience”.

**(1.6) Conclusion: The first sentence is not a conclusion and should be removed.**

We have amended as suggested.



**The way the conclusion is currently written is a clear example of the inappropriate use of scientific/statistical language that is observed throughout the paper. In the first part of sentence 2, “..strongest reductions in risk...” – this should be re-worded to state that caries experience was significantly lower among children exposed to the two (name them) interventions.**

We have modified the language throughout in relation to the term risk and have focused on the strength of association measured by the odds ratio, however we have avoided focusing on significant differences between findings. The conclusion paragraph to the abstract now reads:

“The universal interventions, nursery toothbrushing and regular dental practice visits, were independently and most strongly associated with reduced odds of caries experience in the cohort, with nursery toothbrushing having the greatest impact among children in areas of high deprivation.”

**(1.7) The second part of the same sentence: The term 'effective' cannot be used as there was no comparison group, and this was a cross-section study design. It should be re-worded to read something like: "..., with supervised nursery toothbrushing having the greatest impact among children in areas of high deprivation." The word 'effect' should be replaced with 'impact' throughout the paper.**

We have modified as suggested (See change in comment 1.6). We also refer back to our response to comment 1.1 regarding the study design. We have considered the use of “effective” and concluded we would be more comfortable with “impact”, however with the longitudinal design (exposures pre-date outcome) and the inclusion of “zero” categories in each exposure for comparison, we feel confident in our conclusions. We have changed “effect” to “impact” throughout the paper where appropriate.

## **(2) Introduction**

**(2.1) Intro, 1st paragraph, 2nd sentence: Delete the word ‘tooth decay’. It is a scientific paper. Also, in this sentence it is not clear how caries can peak in 1- to 4-year-olds? It should peak at a narrower age range, for example at 5 years of age or 4-5 years of age, which makes more sense than 1-4-year-olds.**

We use the term “tooth decay” in brackets after “dental caries” for the benefit of the non-dental audience in BMJ Open, and is often used in scientific articles (see for example: the 2019 Lancet article: Peres MA, Macpherson LMD, Weyant RJ, *et al.* Oral diseases: a global public health challenge. *Lancet* 2019;**394**:249-60.[doi:10.1016/S0140-6736\(19\)31146-8](https://doi.org/10.1016/S0140-6736(19)31146-8)

We do agree that we have chosen the wrong wording with regards to the peaking of caries. What we actually meant was that the prevalence of caries in children is highest in those under 5 years of age, again as per the Lancet article which we have referenced. We have amended the text accordingly:

“Untreated dental caries (tooth decay) of the deciduous teeth affects 8% of the global child population, with greatest prevalence in those under five-years of age.”

**(2.2) Intro, 1st paragraph, last sentence: Please state the exact prevalence or prevalence range, rather than “..over 50%...” which could be anything.**

We have amended as suggested and updated to reported level of 60% .

**(2.3) Intro, 2nd paragraph: ‘Proportionate universal approach’ is introduced here. It needs a reference and a bit of a rationale and background for its use.**

The rationale and background was described in detail in our own paper which we had cited. We have added the original reference (Marmot M. Fair Society, Healthy Lives: The Marmot Review. London: University College London, 2010) along with some information on the rationale and background for its use.

“It follows a proportionate universal approach – delivering both universal interventions to all children and additional targeted interventions focussed on children predicted to be at higher risk of dental caries from the most socioeconomically deprived backgrounds, with the twin aims of improving child oral health and reducing associated inequalities in the population[9,10]”.

**(2.4) Intro, 4th paragraph: The aim of the study was not to develop a cohort. This should be reworded to read like 'Here, we developed a cohort using data linkage of routine administrative data: to....'**

We have reworded as per your suggestion.

### **(3) Methods**

**3.1 Databases section: The first sentence needs to be revised to clearly demonstrate that 4 of 5 databases used were part of/used for the Childsmile programme. It currently reads as though five unrelated databases were linked.**

We agree this could be clearer and therefore have added 'Childsmile' to the start of 3 of the 5 databases. We also added additional wording to the MIDAS dataset description to make it clearer that this a national dataset that collects all child and adult primary care dental treatments and registrations, at the population level.

“...(MIDAS)[19] – collated information on all child and adult primary care dental practice appointments and treatments in Scotland (including Childsmile practice prevention items).”

It should be noted that these Childsmile databases are not bespoke research databases, rather they are databases used by the front line healthcare teams for recording and monitoring their service delivery. They are all used by different workforce groups for different components of the programme – they are not related or linked for operational purposes, or designed for primary research purposes.

**(3.2) Cohort inclusion criteria: This section is not at all clear and needs a re-write.**

We have amended this section as suggested:

“Our longitudinal cohort included all children in P1 at local authority schools in the 2014/15 school year (July 2014 - June 2015) who underwent a NDIP dental inspections and were aged between four- and six-years-of-age and whose record could be reliably linked across datasets. Details of the linkage procedure can be found elsewhere[22]. This cohort was initially seeded via probability matching[23] with the Community Health Index (CHI) which is NHS Scotland’s unique patient identifier number. As the CHI number is held on all the other national level health datasets in NSS, we then linked the children in our cohort to their corresponding records in the Childsmile intervention datasets”.

**(3.3) Data management: “To assess data completeness and linkage success, the linked cohort data were compared with appropriate published reports” - compared against what variables?**

We have amended to clarify the variables that were compared in our linked dataset with the published reports using the single database.

“To assess data completeness and linkage success, where possible, the total number of children in the linked cohort receiving each intervention type and the outcome of their NDIP inspection were compared with appropriate published reports which had been based on single databases [24,25]”.

**(3.4) Outcome and intervention data definitions: No outcome measures reported here for ‘programme reach’ and the rationale for using them. Also, dmft is an index to quantify and record caries experience, not to diagnose/define. The BASCD criteria was used to define. Please revise to reflect this.**

As per the comment in the abstract, we have added detail on the reach of the programme outcome. In relation to caries outcome as mentioned in the section on the National Dental Inspection Programme (NDIP) database every year all children (approx. 5 year olds) in their first year of local authority schools receive a basic dental inspection (examination) this is undertaken by trained and standardised (not calibrated) examiners. This examination records the presence or absence of obvious caries experience (caries into dentine) and also missing (extracted due to caries) and filled teeth – although does not record a formal DMFT. Obvious caries into dentine is defined using the BASCD criteria. The section has been modified as:

“The reach of each of the programmes interventions was measured descriptively by the proportion of the child population receiving each intervention on at least one occasion or having consented to nursery supervised toothbrushing by SIMD deprivation fifth.

The impact of the interventions on dental caries (defined as “caries experience” throughout) was measured by the presence or absence of obvious caries experience which was determined clinically by the presence of decay (caries into dentine), missing (extracted due to decay), or filled deciduous teeth – following recognised criteria[26], although due to the nature of the basic NDIP dental inspection being undertaken (rather than a detailed epidemiological assessment) a dmft score was not available. This outcome measure was available in all children in the cohort for the school year 2014/15 from the NDIP database[25]”.

**(3.5) Statistical analyses: This section definitely needs a re-write for clarity. This section is the heart of the paper and is very under-developed. While the key elements are there it comes across as all over the place. The language around the use of the statistics needs to improve. For example, “Logistic regression was used throughout to model caries experience” – you need to explain why logistic regression, you need to define the outcome variable and how it was coded for the analysis etc. Detail is required in this section. I suggest you structure this section as: cohort description, analysis of aim/objective 1: programme’s reach and then, the analysis of aim/objective 2: intervention impact on caries experience.**

We have taken onboard your comments for this section of the paper and have structured it as suggested. Logistic regression is the standard analysis method for binary outcome (0,1) data, and it would not be standard practice in such a journal to have to justify its use. We do, however agree that the outcome (caries experience) should be explicitly defined in this section, and that the description of the analysis of reach and caries experience outcomes should be separated, which is included in the response to point 3.4.

We have altered the wording in the first sentence of the second paragraph to:

“Caries Experience: Logistic regression was used throughout to model the binary endpoint caries experience and a series of steps were taken in the modelling process”.

**(3.6) Statistical analyses: In the first sentence, you state that SIMD is a continuous variable. This is not correct. SIMD is an ordinal categorical variable.**

While we agree that SIMD is an ordinal categorical variable, and have defined it as such, for this part of the analysis we treated SIMD as a continuous variable to test whether there was an increasing or

decreasing trend in those children with at least one dose of a component across the deprivation groups. as this is considered a better way of examining trend than the Cochrane Armitage test.

“Programme Reach: Differences in the reach (gradient across SIMD groups) of each intervention by area-based deprivation (SIMD) was tested using logistic regression of reach with SIMD fifths treated as a continuous variable. This provides the odds ratio for “reach” according to a one unit change in the SIMD indicating whether there was a significant increasing or decreasing trend in those children with at least one dose of a component across the deprivation groups”.

**(3.7) Statistical analyses: Unadjusted odds ratios – Again, you have to explain your rationale for using and reporting unadjusted odds ratios. In this study, the use of unadjusted odds ratio does not seem appropriate. In a classic regression analysis, you would use unadjusted regression estimates as part of the model building exercise, i.e., to understand which variables were associated with the outcome of interest; and then use this information to inform which variables to include in the adjusted/final model. Where as here, it appears that the unadjusted serves no purpose. I suggest remove from the paper and only report on Model 1 and Model 2. Also, use one term for consistency - unadjusted or crude.**

It is true that unadjusted estimates are part of any model building exercise, however, “change in estimate” is a core concept in epidemiology. The change in estimate demonstrates that confounding had happened moving from unadjusted to adjusted. We believe it is important to show any attenuation (or otherwise) from unadjusted > to Model one > to Model two.

We agree that for consistency, we should use the one term throughout and have opted for unadjusted.

#### **(4) Results**

**(4.1) The first paragraph can be removed as it is not a result.**

We have moved the first two sentences from this paragraph the to methods section under the data management header and deleted the last sentence.

**(4.2) All the reported percentages need to be verified for accuracy. The numbers seem to be a bit off. For example, in the paper it is reported that 30% of 50379 is 15032. However,  $50379 \times 0.30 = 15114$ .**

We have amended all percentages in the paper as per our response to comment 1.3.

**(4.3) Table 1 – The odds ratios do not serve any purpose here and should be removed, particularly since this a descriptive table. Also, in these tables, it is more useful to show the column percentages. Readers want to know the distribution of caries by SES, age and sex (not the distribution of caries within each level of the categorial variables).**

We agree with the reviewer that the column percentages would be informative and have produced a new table “Supplementary Table 1: Distribution of Caries by Potential Confounders and Exposure to each Childsmile Intervention” and added the text “The distribution of caries experience by both the potential confounders and exposure levels of each Childsmile intervention are reported in Supplementary Table 1” to the first paragraph of the results section to reflect this. We have also renumbered the supplementary tables appropriately.

The odds ratios demonstrate that the bar has been raised from potential confounders to possible/likely confounders. By definition, the confounder has to be associated with an end point.

The odds-ratios for SIMD are of particular importance as it demonstrates the stark inequalities in oral

health in the cohort.

We have also redesigned Table 1 in light of the comments here and renamed the table to provide a better description of what it is contained within this table to “Table 1: Unadjusted and Adjusted Odds Ratios and 95% CIs for Caries Experience According to Potential Confounders.”

**(4.4) Table 1: 4-year-olds are included here. Weren't children in the cohort 5 and older in 2014/15? Or do you mean that for the caries analysis you included 5-year-olds?**

The NDIP inspection where caries experience is measured is undertaken in the first year of a child's primary schooling. Entry to primary school is based on your birthday falling somewhere within a calendar year, and therefore children in Primary 1 may be as young as 4 and as old as 6 years of age. In addition to this some children can be slightly older than 6 years of age if they have been held back one year at nursery, for example. For the purposes of our analyses we decided to include children aged 4-6 years inclusive. We have now described the age distribution more clearly in our results.

“The majority (n=43165, 85.7%) of children in the cohort were five-years-old (mean age = 5.5, S.D. = 0.3)...”.

**(4.5) Paragraphs 3: A table to show all this information would be useful as it is a primary outcome of the study. These paragraphs need revision to focus on: for the two targeted interventions, please only report the most important results for those targeted groups. The reported OR of 0.67 is of no value, as this was a targeted intervention and it would be expected that the programme's reach would be lower among the least deprived. The use of some OR's in this paper appears to be without rationale for their use.**

Thank you for your comment .We have included a new table entitled 'Supplementary Table 2: Association Between SIMD and the Reach of Each Childsmile Intervention' showing this information as suggested and added the text “...and the ORs [95% CI] for reach (gradient) of each intervention by SIMD are presented in Supplementary Table 2” to the first sentence of the first paragraph under the sub-header: Reach of the programme according to area-based deprivation.

Combined with Figure 2, the text in para 3 and the new table, we believe we are providing all the information required to the reader to show that the targeted interventions have a steeper gradient across SIMD than the universal interventions. We have also added additional text to aid the readers interpretation of the results in this section:

“across the SIMD distribution (most to least deprived) (OR for slope=...”.

As per our response to comment 1.4, It is not possible to define a targeted population for the DHSW intervention for example because this intervention is targeted to children and their parents/carers in greatest need as identified by health visitors and although the assumption is that children with the greatest need are more likely to be living in the areas of highest deprivation, this is not to say that children across the deprivation gradient would not also be targeted. This is why we conducted these tests and wish to present these results within the text. In addition, we have added this explanation to the Discussion (as per point 5.1).

**(4.5) Paragraphs 4: You need to explain/interpret at least some of the ORs in the correct scientific manner, as there are no tables for results paragraphs 3 & 4 to assist the reader. For example, “There was a high level of reach across the population for the (universal) nursery supervised toothbrushing intervention (89% n=44868). The decreasing trend by SIMD was considerably weaker (OR=0.75; 95% 0.73, 0.77)...” – you can write these two sentences in terms of the odds ratio and the reference.**

We hope that our explanation above (point 3.4) as to how we analysed the reach of the interventions

by SIMD helps the reviewer (and our readers) with these two paragraphs. The odds ratio from the logistic regression of reach (“at least one contact of an intervention”) by SIMD (treated as a continuous variable going from most deprived to least deprived) represents the decrease in odds of reach for every unit increase in SIMD (eg SIMD 1- SIMD 2 or equally SIMD 4 to SIMD 5). This is the standard interpretation of an odds ratio using a continuous exposure variable.

**(4.6) Table 2: The column ‘No Caries Experience’ is redundant information and can be deleted. Again column % would be more useful. The ‘Unadjusted ORs’ is of no value and can also be removed from this table.**

We agree that the ‘No Caries Experience’ is redundant and have deleted. Row percentage is useful as the ORs are calculated using these. We have added column percentages to Supplementary Table 1 as per point 4.3.

As noted earlier, we believe it is important to show any attenuation from unadjusted > to model one > to model two, and would respectfully argue to keep this column.

**(4.7): The results section on interventions and caries experience is too long. Report only on the key findings from model 2 – the full model with all variables.**

We have only reported on Model 2 results. Please refer to the second sentence of the paragraph under the sub-header Impact of the interventions on caries experience.

“The main results, adjusted for confounders (age, sex, SIMD) and all other interventions, are described here (Model 2).”

As this is an evaluation of an entire public health programme, we consider it important to report on the impact from all of the four interventions. We also consider it important to describe very briefly whether the Model 2 effects for each intervention have attenuated (or strengthened) after adjusting for the effect of the other interventions.

In order to give the section a bit of structure we have added subheadings and are still well within our word count limit:

Reach of the programme according to area-based deprivation  
Impact of the interventions on caries experience  
DHSW contacts intervention  
Nursery FVA intervention  
Primary care dental practice intervention  
Nursery supervised toothbrushing intervention

## **(5) Discussion**

**(5.1) Programme reach is the first aim/objective, yet there is nothing in this section about reach. The results of the programme’s impact on caries needs to be discussed in the context of the programme’s reach.**

**For example, your finding of 49% of the cohort receiving only 1 fluoride varnish application (low reach), in the study period, could be a key reason why this targeted intervention did not have an impact on caries experience.**

**Programme reach needs to be thoroughly discussed in this paper as it provides insight into the realities of implementing population level interventions, that program planners/policy makers can use to improve all aspects of programme efficiencies and effectiveness.**

We have expanded on this in the Discussion:

“The four Childsmile interventions examined here are largely being delivered as envisaged in the Childsmile strategy[30,31] with respect to their differing targeted and universal aims. This demonstrates a good example of proportionate universalism, where the intensity of interventions across the socioeconomic gradient is proportionate to need. There was near universal coverage observed for the nursery supervised toothbrushing intervention in keeping with findings that nearly all the nurseries nationally (establishment-level) in 2015 were participating in the programme[24]. There were no socioeconomic inequalities observed with the reach of the primary care dental practice intervention which may in part be explained by findings that DHSWs were effective at getting targeted children from more deprived areas into a dental practice earlier than expected[32]. The Childsmile programme health boards implemented the level of targeting in the fluoride varnish intervention in nurseries in slightly different ways[24]. It was therefore important, at the national population-level, to assess the proportion reached across the SIMD distribution, as well as focusing on the most deprived areas. Targeting of the DHSW intervention was often determined on a judgement made by a health visitor based on an individual family’s need. Therefore, there could be children/families targeted that did not live in areas of high deprivation. Nevertheless, there was an expectation that there should be a general trend in reach of the DHSW intervention towards reaching children from the more deprived areas. However, with only 30% of children from the most deprived areas receiving the DHSW intervention, there is room for improving the targeting approach in the programme”.

The finding that we reported that 49% of the cohort “had at least one nursery fluoride varnish application”. Of the cohort, only 9% received only one fluoride varnish application. As can be ascertained from Table 2, of the 31,581 children targeted for this intervention, there were similar numbers of children and similar rates of caries experience within each category suggesting that the reach of this intervention was not influencing its impact.

**(5.2) Sentence 1: Were these routine administrative datasets? Weren't 4 datasets part of the Childsmile programme that were linked with 1 administrative database? The 4 datasets that were part of the Childsmile programme were probably setup at the start for evaluation and research. Hence, they are probably not as routine as what occurs in other countries.**

This was also raised earlier (point 3.1). The primary purpose of the three Childsmile datasets are for the operationalisation and day to day administration of these nationally implemented interventions, with the evaluation and research forming a secondary purpose of these data. This is emphasised further in that these data are administered and analysed for operational purposes by the NHS and not by researchers associated with the programme’s monitoring and evaluation. These Childsmile datasets are now stored in NHS Scotland information services along with both the NDIP and MIDAS datasets are routine NHS Scotland administrative health datasets.

**(5.3) Sentence 2: Is there a reference for this sentence? This sentence is about the Programme’s fidelity. How do you know it was delivered as planned? Was there an evaluation of implementation process that showed that the programme was delivered as planned?**

We agree, we have added the references below, and have modified this whole paragraph as per comments above (5.1). We have added references.

Macpherson L, Ball G, Brewster L, *et al.* Childsmile: the national child oral health improvement programme in Scotland. Part 1: establishment and development. *Br Dent J* 2010;**209**:73–78. <https://doi.org/10.1038/sj.bdj.2010.62>

Turner S, Brewster L, Kidd J, *et al.* Childsmile: the national child oral health improvement programme in Scotland. Part 2: monitoring and delivery. *Br Dent J* 2010;**209**, 79–83. <https://doi.org/10.1038/sj.bdj.2010.629>

**(5.4) “...there was insufficient evidence for an independent effect of the nursery fluoride varnish application intervention....” - Comments in relation to this statement:**

**(5.4.1) The ‘effect’ should be replaced with ‘association’ to read as: “...there was insufficient evidence for an independent effect association of the nursery fluoride varnish application intervention with caries experience.”**

As suggested, we have removed term ‘effect’ from throughout the paper where appropriate and amended the wording of this section.

“...there was insufficient evidence for an independent association of the nursery fluoride varnish application intervention with caries experience.”

**(5.4.2) There is nothing further in the discussion as to explain why you may have observed this result.**

There was further discussion as to why we observed this result. In the third paragraph of the discussion we state “Furthermore, there was very little evidence that fluoride varnish applications within the nursery setting reduced odds of caries experience after adjustment for the other three interventions. Although systematic reviews of fluoride varnish show a clear caries preventive effect in children[16], a more recent review is beginning to cast doubt over fluoride varnish effectiveness and cost-effectiveness[37]. As the impact attenuated following adjustment with the other interventions, it is plausible that there was little to no benefit for receiving fluoride varnish over and above the almost universal coverage and caries preventive impact of nursery supervised toothbrushing, or the other interventions, particularly for those living in the most deprived fifth”.

**(5.4.3) How does this finding compare with the Protecting Teeth@3 Study?**

Results from our Protecting Teeth@3 study are yet to be published.

**(5.5) Discussion, paragraph 2, “The results support.....”: This is clearly the strongest finding and the discussion around this result needs to be beefed up. Why this reduction? What were the estimates of caries experience (dmf scores in the least deprived VS the most deprived)? How does this compare to the literature? Why do you think this intervention was successful? What were the enablers and barriers? What was the policy context that made this possible?**

We agree this is the strongest finding, this is discussed in relation to the literature and explained in the subsequent paragraph, we have combined these paragraphs to make it clearer. As mentioned above we do not have dmf scores. The policy context of delivering the supervised toothbrushing programme is in our earlier ecological study to which we refer, and the enablers and barriers is beyond the scope of this present study.

**(5.6) Discussion, paragraph 2, last sentence: “One possible explanation for this is that children living in the most affluent areas are more likely to already be regularly toothbrushing at home and as a result may have had lower caries experience to begin with.”**

We have amended this last section as per your suggestion and chosen the wording:

“...and as a result may have been at a lower risk of caries experience to begin with.”

**(5.7) Discussion, paragraph 3: The section about the FVA intervention should be a separate paragraph. It is strange to call a 2013 review ‘historic’. Again, you need to discuss possible reasons for the FV not working in your population. Did it not reach enough of the population? Did the children get at least 2 FV applications in a 12-month period etc?**



We have removed the term 'historic'. In the paper, we discuss the reason that the FVAs were not working in our population was most likely due to the almost universal coverage of the supervised toothbrushing programme. Furthermore, both Table 2 and Supplementary Table 4 show there no difference in the association between FVA and caries experience regardless of how frequently this intervention was applied across the targeted cohort and therefore does not merit further discussion. We did not look at the specific timings of the varnish but the programme policy was for children to receive two varnishes per school year.

Due to a limitation in the number of paragraphs permitted in the discussion section, we are unable to split this paragraph.

### **Reply to Reviewer 3**

#### **(1) Abstract**

**(1.1) Participants: The term approximately 5 years is not acceptable. I realize that the WHO recommends 5 years as a target age, but the authors should either mention mean age or age range in brackets to elaborate on this.**

Agreed . The majority (n=43165, 85.7%) of children in the cohort were five-years-old but the age range was 4- to 6-years of age inclusive. The amended text is:

“50,379 children (mean age=5.5years,S.D.=0.3) attending local authority schools (2014/15)”.

**(1.2) Primary outcome measure: Missing is not a valid expression of primary decay. The authors need to clarify if this was extracted due to caries: the deft?**

Thank you, we agree that this needed further clarity as per here and point 4.1. This examination records the presence or absence of obvious caries experience (caries into dentine) and also missing (extracted due to caries) and filled teeth – although does not record a formal DMFT. Obvious caries into dentine is defined using the BASCD criteria. This has been detailed in the text and modified in the abstract as:

“Obvious dental caries experience (presence/absence) defined as the presence of decay (into dentine), missing (extracted) due to decay, or filled deciduous teeth”.

**(1.3) Conclusion: delete “This is the first population-wide data-linkage cohort study to evaluate a complex public health programme” this is not a finding of the study and is not needed here**

We have deleted as suggested.

**(1.4) Strengths and limitation: Rephrase the second point. There is no need to mention the systematic reviews here. Focus instead on the evidence and the evaluation of the evidence in this study.**

We have removed the reference to Cochrane review, however, have retained the general thrust of this point as we feel evaluating the complex multicomponent aspect of our study is a strength.

“There is evidence of Cochrane systematic reviews demonstrate effectiveness of oral health improvement interventions for children including fluoride toothpaste and professionally applied fluoride varnish, however the evidence of combining these into a complex oral health improvement programme delivered via a proportionate universal approach has not previously been evaluated”.

**(1.5) Strengths and limitations: Are the authors suggesting there are no limitations? Use of records created by multiple dentists? Calibrations? Please mention the limitations here.**

The limitations of this study are mainly associated with the use of routine administrative data, which has built in compromises around the standardisation of data collection. We have amended this section.

- The study utilises routine administrative data, which has some limitations in the variables available, including a lack of information on intermediate individual behaviours.
- The outcome data available, the presence or absence of obvious dental caries experience, collected by trained and standardised dental inspection teams and available at the population level, shows a high level of agreement with detailed dmft scores collected by calibrated dental inspection teams on a much smaller sample of children.
- The study strengths are in the robust data linkage approach, where there were no concerns about the quality and completeness of the data linkage, resulting in a cohort with population-wide coverage of outcome and intervention data.

**(2) Introduction**

**(2.1) Pg4 line 57: Please be specific about the cohort measured. I would recommend the use of the same terminology for age in the abstract introduction and methodology.**

Thank you, we agree we needed to be more specific. We have chosen to add age range in the introduction as we just describing the NDIP programme as this state but have used mean age in the abstract and when presenting the results.

**(3) Methods**

**(3.1) Page 5 line 30: Dental examination involved... please stick to past tense through the methodology section.**

We had deliberated over tense but as the dental inspection and indeed the databases are ongoing and not historical points of data collection, but we have now chosen past tense based on your recommendation.

**(3.2) Patient identity protection: I was unable to find any mention if the records were de-identified? Or if there were any measures taken to protect identifiable datasets.**

Thank you for raising this point. Yes, all records for analysis were anonymised (pseudonymised). A lot of time and work was undertaken to be granted access to these datasets with data governance and data protection at the core of this. We have added a paragraph at the start of the data management section which describes the pseudonymisation of the cohort.

“Prior to obtaining the datasets for our study, NSS removed the personal identifiable variables (CHI number, forename, surname and home postcode) from the datasets. The CHI numbers were replaced with study-specific pseudo anonymised IDs that allowed all the records belonging to an individual across all the datasets to be linked without the need for personal identifiers”.

**(4) Outcome Methods**

**(4.1) Lines 54-55: Please clarify the term “missing” – was this all missing teeth or only teeth with a record of extraction due to dental caries. Were there exclusion criteria (congenitally**

**missing teeth). Though teeth missing due to reasons other than dental carries is unlikely, it is important to clarify this.**

We agree that we had to clarify that missing teeth were only those that were extracted due to dental caries. We clarified this in more detail in response to reviewer 2 comment 3.4.

**(4.2) What is the level of variation/calibration for dental caries in the NDIP database. Are practitioners routinely trained/calibrated for the recording of dental caries. This is a point that needs to be mentioned if present and/or addressed in the discussion if not present.**

We have clarified in the text that dentists involved in the basic NDIP inspection programme are “trained and standardised”, but they are not calibrated. This limitation has also been added to the Discussion (see response to point 6 below).

Methods section, Databases subleader, second last sentence:

“The dental inspection involved a simple assessment of the mouth and teeth of each child undertaken by trained and standardised primary care dental teams within primary schools. Dental caries experience of the deciduous dentition was recorded[20]”.

**(5) Results: Acceptably presented – the models are clear.**

Thank you.

**(6) Discussion**

**(6.1) The discussion needs to discuss the limitations of the dataset. The use of population based data has limitations. While the discussion makes valid points it is not complete unless it addresses those limitations.**

Thank you we agree that population data has its limitations and have expanded the discussion to further reflect this.

“Our study used routine administrative databases, the limitations of which are recognised [42] as they are established for other purposes, and therefore the variables available are more limited. These are more than offset by the large population coverage, and in our case because all the datasets had robust quality and completeness procedures. The NDIP basic inspection data had good population coverage providing presence or absence of caries experience collected by trained and standardised examiners, this has less detail than detailed epidemiological inspection data which includes dmft scores collected by calibrated examiners – although these data would only be available on a small sample (20%) of children. The NDIP reports show high level of agreement between the basic and detailed inspection caries prevalence data [20]. Moreover, the linkage process with intervention datasets was robust with a high linkage rate, which did not exclude many records from those expected in published reports providing a cohort representative of the population[25]”.

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Bradley Christian La Trobe University, Australia
<b>REVIEW RETURNED</b>	04-Oct-2020
<b>GENERAL COMMENTS</b>	Excellent revisions. The paper reads so much better now. Well done!! I still don't agree with the way some of the statistics is written/presented but everyone has a different approach and as long as there is sound rationale for using a particular approach it is all

	good. I look forward to hearing about future successes of the Childsmile program!
<b>REVIEWER</b>	Sharat Pani University of Western Ontario, Canada
<b>REVIEW RETURNED</b>	06-Oct-2020
<b>GENERAL COMMENTS</b>	The authors have addressed all concerns raised by me in the first round of review.