Author’s response to reviews

Title: Outcome measures for oral health based on clinical assessments and claims data. Feasibility evaluation in practice

Authors:

Riët Hummel (g.i.hummel@acta.nl)
Josef (J.J.M.) Bruers (j.bruers@acta.nl)
Onno (O.P.) van der Galiën (onno.van.der.galien@zilverenkruis.nl)
Wil (W.J.M.) van der Sanden (wil.vandersanden@radboudumc.nl)
Geert (G.J.M.G.) van der Heijden (geert.vander.heijden@acta.nl)

Version: 1 Date: 29 Apr 2017

Author’s response to reviews:


Dear Editor,

We are grateful for the critical comments and constructive and insightful responses on our manuscript from you and the reviewers. On the following pages we reply to each of the comments, one by one, and we have modified our manuscript accordingly. We feel that this has helped to improve the manuscript. We sincerely hope that you and the reviewers appreciate our rebuttal and agree that we have complied with the suggestions for improvement.

Major changes in response to the comments are:

• The title. ‘Oral healthcare outcome measures for children, adolescents and adults. A proof of practice study’ has been changed into ‘Outcome measures for oral health based on clinical assessments and claims data. Feasibility evaluation in practice.’ With this adjustment we intend to clarify the scope of the project and to explain the term ‘proof of practice study’.
• ‘Proof of practice’ is replaced by ‘feasibility evaluation in practice’ throughout the whole manuscript.

• We deleted less relevant details to shorten the article and make it easier for the reader to follow the text.

• We deleted tables 8 and 9 and additional file 1 from the manuscript. They contained details that were less relevant.

• We deleted parts containing repetition of earlier parts.

• We replaced information in parts where readers expect to find the information.

• The manuscript was read by a native speaker. His remarks were used in the revision.

In our reply to the reviewers comments we have included page and line numbers which refer to the page and line numbers in the submitted version without track changes. The other submitted version shows the revisions marked by track changes. In this version, not all revisions in our use of the English language are shown. We can send these changes separately, if requested. Figure 1 is also revised and resubmitted. We would like to withdraw additional file 1.

We look forward to hearing from you regarding our submission and would be glad to respond to any further questions and comments that you may have.

Also on behalf of my co-authors.

Yours sincerely,

Riët Hummel
Reviewer 1 - Maha El Tantawi

The point of the study is important and sheds light on a much needed area where dentistry has lagged behind the medical profession for some years. The manuscript includes lots of details that sometimes make it difficult for a reader to follow. It needs to be revised and to be made more concise by avoiding repetition and by placing the information where readers expect to find it so that the rationale for an indicator is placed in the Methods section, for example, rather than in the Results as is currently the case. Please notice the following:

1.1. The manuscript includes lots of details that sometimes make it difficult for a reader to follow.

We appreciate this critical comment. Accordingly we have deleted some details that we feel were not essential or of lesser relevance for understanding the study.

1.2. It needs to be revised and to be made more concise by avoiding repetition and by placing the information where readers expect to find it so that the rationale for an indicator is placed in the Methods section, for example, rather than in the Results as is currently the case.

We critically reread the text searching for repetitions and deleted these, and moved information to more appropriate sections. We have moved the rationale for the measures from the results section to the methods section, page 5, line 11.

1.3. The authors need to explain the concept/meaning of proof of practice they used in the title and in several areas in the manuscript.
To avoid any confusion on meaning we have changed ‘proof of practice’ into ‘evaluation of feasibility in practice’ in both the title and throughout the manuscript. That is, we have evaluated and tested the feasibility of measures in practice.

1.4. The title, aim and several areas indicate that the focus is on developing indicators of quality of care. The authors totally omit all patient-based and non-clinical indicators such as pain, quality of life, satisfaction with care and professionals, length to getting an appointment, ... The dimensions of quality of health care services includes many other aspects. It is up to the authors to select which of these they address. However, the title and aim as well as other areas need to reflect this.

We acknowledge that quality of care includes clinical as well as non-clinical measures. Our focus is on clinical measures from clinical assessments and claims data. To emphasize the focus from the beginning we changed the title into ‘Outcome measures for oral health based on clinical assessments and claims data’. We also added the clinical and non-clinical aspects of quality of care to the introduction section, page 3, line 9.

1.5. Since the authors used clinical indicators, why were not indicators of oral health status used by others included/ referred to in Introduction or Discussion? Example are those used by the Association for State and Territorial Dental Directors in the US of percentage of children with decay, with untreated caries or with sealant and so on. The objective differs but the authors already referred to indicators developed/ used for other purposes.

We have chosen not to list all kind of indicators used by others, but to refer to current initiatives since it is not currently routine to collect data to improve quality of oral health care, as described by Baâdoudi et al. However, as stated in the methods section (page 4, line 16), we used clinical measures described by others in the development of our measures. Where measures derived from other sources or initiatives, this was mentioned in figure 1. Measures used by others relating to our measures are described in the discussion session on page 18, lines 1 and 8.
1.6. If the aim is to assess the quality of care provided so that 3rd party such as insurance organizations can evaluate the care their clients receive, how can clinical assessment be used as source for outcome measures? How would this be applied? Is it suggested that these items be added to the clinical records of patients so that they are later extracted for evaluation? It would not be very helpful if dentists are asked to generate data for the purpose of evaluating the quality of care they provide. This would introduce bias. One of the indicators differed when clinically assessed and when the records were used as source of data. A potential explanation for this is examiner bias.

Our purpose has not been to identify possible bias or provide indicators for normative evaluation, but we aim to provide measures that can help to inform discussion and reflection by practitioners on quality of care. Moreover, we have found that clinical audits have been shown to be strong procedures in the evaluation of quality of care (discussion section, page 19, line 25)

The difference in the distribution of risk categories for dental caries between the two data sources was not caused by bias. The main reason was that these indicators measure different stages in the caries process. The clinically assessed risk categories not only included carious lesions with an indication for restorative treatment, but also included active initial lesions without an indication for restorative treatment (yet). Whereas the risk categories based on claims data only included lesions that were filled (methods section, page 5, lines 14-21).

1.7. The authors refer to the controlling for the effect of some factor in one part of the manuscript. This requires multivariable analysis which would be most useful to assess the several factors included (diabetes, smoking, socioeconomic, age and provider). The methods of analysis described refer only to bivariate analysis. This raises the chance of type 1 error and increases the risk of spurious conclusions. If multivariable methods were used, they need to clearly shown and explained.

This is intended and has been designed as a descriptive study in which we mainly evaluated the feasibility of procedures in practice. This study has neither been designed nor is meant to search for causal or a predictive associations. As such multivariable methods were not used as they were beyond the objective of this study.
Moreover the limited number of practitioners and patients would not allow to test any (causal or non-causal) associations.

1.8. Some indicators are built on very few observations (table 8, n=34) and for others, the number isn't clearly defined (table 9). It is recommended that the indicators that were not assessed be removed and may be they can be just mentioned in the methods since data doesn't support or refute their merit.

We appreciate this suggestion by the reviewer, and accordingly removed tables 8 and 9 to avoid possible confusion. We now only mention that these indicators are deleted due to the low number of observations (results section, page 12, line 19)

1.9. Some tables contain only percentages while others contain numbers and percentages. Pls add numbers to all so that the reader can see the sample size used in assessing each indicator.

Only table 9 did not include numbers. This table is deleted. All other tables contain numbers in the patient characteristics section.

1.10. There is mention in the manuscript to reliability of the indicators. How was this assessed?

Reliability has not been formally tested but has been judged and taken at face value. The next step will be formal psychometric reproducibility evaluation. This is added to the discussion section (page 17, line 10).
1.11. The aim definitely needs revision. Oral health cannot be assessed by clinical examinations performed by clinicians in their practices because of the high variability. They cannot also be used as the gold standard for claims or records data to validate them for the same reason. The authors must justify the inclusion of this data. If the aim is restricted to developing indicators of quality of care from records (claims and clinical), this may provide more focus and would be more useful.

The aim is restricted to developing measures from clinical assessments and claims records. This includes data from clinical assessments during consultation (for example assessing tooth wear), data from the dental files from GDPs and claims records from the insurance company. This focus is clarified in the title.

We did not revise the aim any further, because first of all, we have designed this study as a descriptive evaluation of the feasibility of the measures. And secondly, it has not been our intention to evaluate diagnostic accuracy, nor have we designed this study for such purposes. Measures using claims records or clinical records are proxies for oral health (abstract, page 2 line 25) and not intended as diagnostic tools.

Variability does not render a measure invalid. Of course, standardization of procedures and large numbers of observations are necessary to reduce variability as much as possible. But assessments and examinations show variability by definition, as do claims data. This does not imply they cannot be used for comparisons and discussions for care improvement.

1.12. The number of GPs needs to be explained. Are these 3-5 practitioners/dentists or practices with potentially several dentists per practice? Their number is too small to assess the ability of the indicators to discriminate between dentists. In comparing values across patients’ groups, we know from the literature that patients from low SES are expected to have poor oral health. When data supports this, we can infer that the indicator discriminates between people based on SES. How can this be applied to dentists. At best, the indicators would support the presence of variation among dentists in treatment decisions and care process. This is already supported by huge literature. The claim that the indicators, therefore, have discriminant validity is questionable.
All participating dental practices were group practices with several dentists per practice. Clinically assessed data were derived from the participating GDPs, but data from records were potentially from other dentists working in the dental practices. We added a description of the dental practices to the methods section (page 7, line 6).

The SES of the patients of 3 of the participating dental practices was comparable (added on page 7, line 10). The SES of the other dental practice was much lower and they had very different scores on the FM-score for example. However, SES was not the only explanation for this difference. When we compared the scores for the various SES groups between dental practices, the FM-score was still much higher in the dental practice with the low SES population. From discussions with the GDPs it became apparent that this practice had the policy to fill lesions in an earlier stage of the caries process than the other dental practices in the expectation to prevent endodontic treatments. (page 16, line 13).

Insight in these variations among dentists in treatment decisions and care process is important when comparing oral health outcomes. Feedback information on the delivery process and discussions on treatment decisions are important conditions for quality improvement.

Reviewer 2.

General comments:

This is an interesting study and it is very important too, as it could help dentists to review their practice and the outcome of the treatment performed by them. In another dimension, it could help dentists to improve clinical documentation and software developers to improvise their existing clinical management platforms.

The limitations and the future perspectives are well discussed.

The paper is a bit long and there are too many tables.
I have the following minor comments:

2.1. The paper is a bit long and there are too many tables.

We critically reread the manuscript and removed some probably unnecessary details and we have removed tables 8 and 9.

2.2. The term “distribution risk categories dental caries” could be written as “distribution of risk categories for dental caries” throughout the manuscript.

Thank you for the suggestion. We changed ‘distribution risk categories dental caries’ in ‘distribution of risk categories for dental caries’.

2.3. The caries risk categories used in the manuscript could be changed to: Low, medium, high, and very high or Very low, low, high, and very high.

We changed the names of the categories in low, decreased, increased and high. These names reflect the original names described in the reference quite well.

2.4. The word “filling” that has been used in several places of the manuscript could be changed to “restoration”.

We changed ‘filling’ into ‘restoration’ as supposed. Filling is now only used in context of the FM-score: filled-and-missing score, as it is derived from the DMF-index that also uses ‘filled’. 
2.5. The word “re-restoration” could be changed to “repeat restoration”

We consider re-restoration an appropriate term, which also is used in several articles in the PubMed database. We therefore have not followed this suggestion.

Remarks in pdf

2.6. Title: pls define/ explain this term/ concept

‘Proof of practice’ means feasibility evaluation in practice. We changed ‘proof of practice’ throughout the whole manuscript into ‘feasibility evaluation in practice’.

2.7. Abstract, line 9, page 1. responsiveness isn't included in the data, pls remove

Responsiveness has been removed from the abstract.

2.8. Abstract, line 11, page 1. Is this prospective or retrospective?

Clinically assessed data during consultation were prospective, the other data were collected retrospectively. We added the word ‘prospective’: ‘Data sources were claims records from Achmea, clinical records from the dental practices, and prospective, pre-determined clinical assessment data obtained during routine consultations’. (Abstract, page 2, line 10)
2.9. Abstract, line 24, page 1. The evaluated measures ‘restoration free time’, ‘distribution of risk categories for dental caries’, ‘filled and missing score’ and ‘retreatment after filling’, were considered relevant measures for oral health status: for oral health status or oral health care?

These measures serve as a proxy for oral health status. We added the words ‘proxy for’ (abstract, page 2, line 25). Oral health was approached conceptually by measuring delivered oral health services that are considered as related to oral health outcomes.

2.10. Abstract, line 24, page 1. They may serve as process measures and, as such, improve transparency on oral health services delivery and can be related to oral health outcomes: are they measures of process or of outcome? if oral health is process, what would the outcome be for health care services?

These are measures of health care services delivery which are considered to be related to oral health outcomes. For example: a new restoration due to caries is a healthcare service delivery that is related to oral health, because it can be seen as a proxy for caries progression. We clarified this by changing the sentence into: ‘As such, they improve the transparency of oral health services delivery that can be related to oral health outcomes, and with time may serve to improve these oral health outcomes.’ (Abstract, page 2, line 25).

2.11. Introduction section, line 23, page 3. Complied lists of: Compiled?

This was a mistake, we meant compiled indeed. This is corrected in the manuscript.
Yes, we mean ‘its’. The sentence is removed from the manuscript since it was considered an unnecessary detail.

The focus of this project was on measures derived from clinical data from dental practices and claims data. The manuscript title is altered in a way that reflects that only these measures are included: measures based on clinical assessment and claims data. See also comment 1.4.

There is no reference for this score, this measure was developed by the study team. There are described measures based on DMF, like the mean DMFT-score per dental practice. Measures on DMFT are not feasible in the Netherlands as DMFT is not measured by GDPs. Apart from that, DMFS and DMFT reflect what happened in the past. A goal of oral healthcare is to maintain or improve oral health. Therefore an increment in DMFT or DMFS is more relevant. Since we used claims records we had to approach this by claimed restorations and extractions.

‘As sample size approximation we used a rule of thumb of 60 patients as the minimum number of patients to be analyzed per measure and an additional 20 patients for each case mix factor we measured.’ What is the reference for this rule of thumb?
Achmea applied the minimum number of patients per measure on 60 in all projects in its program ‘Quality of care’. The first part of the sentence is changed into: ‘We set the minimum number of patients to be analyzed per measure at 60; measures with less than 60 patients per dental practice were excluded.’ (methods section page 10, line 10) The second part of the sentence is removed, in general the numbers were too low for multivariate analysis.

2.16. Methods section, line 10, page 8. ‘Measures with less than 60 patients per dental practice were excluded.’ Such as what?

The measures ‘percentage 18 year-olds with no tooth decay’ and ‘improvement in gingival bleeding index at reassessment’ were excluded for this reason.

We clarified this in the results section (page 12, lines 19-22) by adding that these measures had less than 60 patients per dental practice and these measures were therefore excluded. ‘During the data collection the available data for the selected measures ‘percentage 18-year-olds with no tooth decay’ and ‘improvement in gingival bleeding index at reassessment’ were for less than 60 patients per dental practice and therefore were too limited to provide accurate estimates per practice. These measures were excluded.’

2.17. Methods data analysis. Why were the 2 softwares used? SPSS can do all of that and so can SAS.

Achmea cannot use SPSS due to the large amount of data. SAS was used for all analysis performed by data analysts from Achmea. The statistical tests were performed at ACTA using SPSS.

2.18. Methods section, line 4, page 9. ‘Validity and reliability of the data were assessed during a meeting with the participating GDPs by showing the data and asking them whether they recognized the data.’ Is there a reference that this method can be used to assess reliability?
There is no reference that this method can be used to assess reliability. We took this one at face value. The next step will be psychometric reproducibility. This is added to the discussion session, page 17, line 10.

2.19. Results section, line 26, page 9. In general, there is confusion between whether the point of the study is to develop indicators of oral health or indicators of quality of care. What seems to be the case is that indicators of oral health were the main focus all through with some exceptions.

Our aim was to develop outcome measures for oral health. These measures can be used to improve quality of care. As stated in the introduction section, lines 7, page 3, quality of care can be explored using data which describes links between the healthcare services provided and improved oral health outcomes. We see the confusion and made some adjustments in the first paragraph of the introduction section (page 3, line 5): ‘The aim of oral healthcare is to maintain or improve oral health for individuals and populations [1]. The degree to which the likelihood of this goal is increased by the delivered oral healthcare services is regarded as quality of care [2]. Quality of care can therefore be explored using data which describes links between the healthcare services provided and improved oral health outcomes. These may comprise both clinical and non-clinical, patient-derived outcomes.’

2.20. Results section, line 7, page 10. ‘Measures for oral health were..’ This is not the same as quality of care.

We agree. The main focus is on outcome measures for oral health. See also comment 2.19.

2.21. Results section, line 25, page 10 up to line 18, page 12. The rationale of the measures. This belongs to the methods NOT the results.

We agree and replaced this rationales of the measures in the methods section, page 5, line 11.
2.22. Results section, line 2, page 11. ‘Based on the survival curves we looked at the restoration free cumulative incidence of seven year olds.’ So these children were cared for by the practitioners since year 0?

These children were at least cared for by the GDP since they were three years old. One of the inclusion criteria was that there were no carious lesions during the first visit and the first half year after the first visit (table 1, page 29). So there were no children included with dental caries before the GDPs provided oral health care.

2.23. Results section, line 2, page 11. ‘There were differences between the dental practices.’ Are these dental practices with many dentists or dental practitioners? Do the tables compare 4 dental practitioners or 4 practices that include 8 or 10 or 20 dentists, group practices?

The measure ‘Restoration-free time’ is based on patient records. As mentioned in comment 1.12 these data compared group practices. We have added a description of the participating dental practices in the methods section, page 7, line 6.

2.24. Results section, line 11, page 11. ‘the four risk categories (low, lower, higher, high).’ Can these be replaced by ranks? like category 1 to 4 or A to D so that readers can readily know which is worst and which is best?

We changed the names in low, decreased, increased and high. We think this covers the definitions as described in tables 2a and 2b the best.

2.25. Results section, line 6, page 12. ‘we do not consider this difference as clinically relevant.’ What difference would be considered clinically relevant? based on what? doesnt this belong in Disc?
We do agree that this does not belong in the results section and removed it.

2.26. Results section, line 10, page 12. ‘when the scores were stratified by socioeconomic groups, the FM-score for dental practice #1 was higher in all groups (not presented). Data per SES categorie not presented. Why not?

We chose not to present the results per case mix group to limit the length of the tables. Due to the generally small numbers of patients per case mix group, we could not compare the results per case mix group per dental practice, but only compared aggregated case mix group. We described the results for the testing of the aggregated case mix groups in order to explore the influence of case mix factors on the results.

2.27. Results section, line 8, page 13. ‘The data from the measure ‘time to first restoration’ were automatically extracted from the clinical registration from the dental practices.’ Is this the same as clinical records or clinical assessment?

This is the same as clinical records. We changed ‘clinical registration’ in ‘clinical records’.

2.28. Results section, line 6, page 15. ‘The difference in FM-score remained after correction for age…’ Where is this correction? I have not seen any multivariable analysis correcting/adjusting for things. Did i miss it?

Multivariable methods were not used, see rebuttal on comment 1.7.

‘Correction for age’ is not the right expression. We changed this sentence in:
‘The socioeconomic status of the patients in dental practice #1 was lower than in the other dental practices, but when the scores were compared between the same socioeconomic groups in the other dental practice, the FM-score for dental practice #1 was found to be higher in all groups.’ (Results section, page 13, line 24)

2.29. Table 4, page 34. ‘Retreatment after filling’. All DPs were shown in the other tables. Why not here?

Results for all DPs were used to test the effect of case mix groups on overall results. Case mix was assumed not to have an influence on this measure and therefore not tested nor shown.

2.30. Table 5, page 36. ‘Proof of practice measures’. Where is the last one? Gingival bleeding?

The two excluded measures ’percentage 18 year-olds with no tooth decay’ and ‘improvement in gingival bleeding index at reassessment’ are not described in table 5. See also comment 2.16.

2.31. Table 5, page 36. How was reliability evaluated?

There was several validity mentions but not reliability.

Reliability was evaluated on face value. See also comment 1.10 and 2.18.
2.32. Table 5, page 36. ‘Proof of practice measures’. face validity. Why would the same measure obtained by different methods have diff face validity evaluations?

These measures have different evaluations of the degree in which they reflect aspects of quality of care. The distribution of risk categories for dental caries based on claims records is only a proxy caries progression as claims data cannot distinguish fillings on behalf of dental caries from fillings on behalf of trauma. The distribution of risk categories for dental caries based on clinical assessment is more precise, only caries lesions are taken into account.

Reviewer 3 - Hermann Agis

In their study entitled "Oral healthcare outcome measures for children, adolescents and adults. A proof of practice study" the authors established measures to evaluate oral health. The performance was evaluated based on insurance company data and general dental practices data. Based on their search and evaluation they established eight measures. The authors conclude that 'restoration free time', 'distribution risk categories dental caries', 'filled and missing score' and 'retreatment after restoration' are feasible to describe the oral health status. While the study is interesting with regard to the public health area and the insurance sector there are several issues that need to be resolved and thus major revision is required.

Major Issues:

3.1. While the study is indeed important for the public health and the insurance sector. The scientific impact of the study is not totally clear. The study was conducted in the Netherlands and it is known that there are big differences in the healthcare and reporting systems within Europe and beyond. It remains unclear if and how these findings translate to an international community. For example can is this approach applicable for USA, Switzerland, Germany, and Australia or are does every system need to establish a distinct approach?
Thank you for pointing this out. We added remarks on international applicability in the discussion section, page 20, line 6. This approach is applicable in all countries were patients regularly visit the same dental practice. In situations where fee per item is not used, the data can be collected from clinical records. Clinical assessment during consultation is possible in every country. It might be necessary to make some slight modifications to the measures.

3.2. How can this approach be used in a scientific setting to improve the quality of clinical studies? What are the limitations and what further research has to be done?

We see the point, but improving the quality of clinical studies is out of scope for our paper. Our intention is to develop indicators for oral healthcare delivery improvement and improving healthcare research in practice. The limitations and further research have been described in the discussion section, page 17, line 9 to page 19, line 14. We also included a remark that we did not intend to use measures for normative purposes, but our intention was to inform and help GDPs in reflections and discussions about their oral healthcare deliveries (discussion section, page 21, line 1).

3.3. It is not clear how that 'restoration free time', 'distribution risk categories dental caries', 'filled and missing score' and 'retreatment after restoration' can be used as quality scores? Is it mainly influenced by the patient's behavior, the DMD's approach to prevention and treatment, or the socioeconomically background of the patient?

In the discussion section we described ‘Information on these measures provides relevant, valid and important feedback for GDPs which has the potential to improve quality in these aspects of oral healthcare’ (page 17, line 7) and ‘Further research is also required for the relationship between the outcomes, patient population and treatment approach of GDPs before they can be considered as measures that deliver improved health outcomes.’ (line 11). For the interpretation of the results patient characteristics like socioeconomic background should be taken into account as well as the GDPs’ approach to prevention and the GDPs’ restorative treatment threshold. Our intention is to measure the influence of GDPs’ treatment approaches on oral health outcomes.
3.4. Overall the approach is descriptive and it remains unclear how to improve and fine tune the system based on these measures. Furthermore, the clinical situation of these criteria cannot be improved retrospectively. Even if a patient has received fillings and was subjected to retreatment after restoration the oral health associated live quality can improve by education of the patients by the DMD and behavioral changes preventing further fillings, extractions and other treatments. How is this reflected in the model?

Improvement of the system can be achieved by providing GDPs feedback information wherein their results are compared to the results of GDPs/dental practices with similar patient populations and taken the preventive policy of the GDPs into account. Indeed the clinical situation cannot be approved retrospectively, but the measures are meant for improvement in the future. We performed baseline measurements which allowed GDPs to reflect on their treatment approach (preventive as well as restorative). By discussing the outcomes and treatment approaches quality improvement is stimulated what can result in preventing further fillings, extractions, etc. This is added to the discussion section, page 20, line 2. First of all fine tuning is required for the measures itself and the effect of case mix factors and treatment policies on the results of the measures. After that we can build a system of quality improvement based on these measures.

3.5. Measures can induce a behavioral change to improve based on the outcomes of these measures. When including these measures to assess the treatment quality is there a possibility of inducting the selection bios for DMD with regard to their patients? Would it encourage treatment approaches in a direction that improve the measures rather than the oral health associated life quality? Or would it be a counterpart to the mentioned treatment bios which is introduced by the "fee per item" approach.

Whether there is a possibility of inducting selection of patients depends on how these measures are used. When the measures are looked upon as performance indicators and are coupled to financial incentives, selection bias is a risk.

When, for example, a norm is set on a maximum percentage of re-restorations within a certain period, there is a change that the measure encourages a delay in necessary treatment. To observe unwanted treatment approaches to get better results for the measures, it is advisable to add patient perceptions to the system.

On the other hand ‘over treatment’, resulting from a potentially perverse incentive in a system with fee per item, might be prevented with measures based on records.
At this stage our measures are meant for reflection and discussions on quality of care on behalf of internal quality policies of GDPs. These measures can be used for patients and policy makers after further development.

Minor issues:

3.6. The title should be changed in "Oral healthcare outcome measures for children, adolescents and adults in the Netherlands - A proof of practice study" and the entire manuscript should be clearly revised to reflect this.

We changed the title in ‘Outcome measures for oral health based on clinical assessments and claims data. Feasibility evaluation in practice’ as mentioned in comments 1.4, 1.11 and 2.13. We did not add ‘in the Netherlands’ since the methods are internationally applicable as stated in comment 3.1.

3.7. The authors should revise the manuscript with the help of a scientifically trained native speaker?

We asked an experienced researcher who is native English, to help us with the revision of the manuscript concerning the use of scientific English. His remarks are incorporated in the manuscript.

3.8. The way of data presentation should be refined. 9 Tables is quite intense for the reader.

Thank you for this suggestion. We deleted tables 8 and 9.