Article:
Larvin, H orcid.org/0000-0001-7263-4182, Kang, J orcid.org/0000-0002-2770-1099, Aggarwal, VR orcid.org/0000-0003-0838-9682 et al. (2 more authors) (2021) Multimorbid disease trajectories for people with periodontitis. Journal of Clinical Periodontology, 48 (12). pp. 1587-1596. ISSN 0303-6979

https://doi.org/10.1111/jcpe.13536

Title

Multimorbid disease trajectories for people with periodontitis

Authors

Harriet Larvin[1], Jing Kang[2], Vishal. R. Aggarwal[1], Sue Pavitt[1], Jianhua Wu[1,3]

1. School of Dentistry, University of Leeds, Leeds, UK.
2. Oral Biology, School of Dentistry, University of Leeds, Leeds, UK
3. Leeds Institute for Data Analytics, University of Leeds, Leeds, UK

Correspondence to:

Jianhua Wu

Worsley Building, Level 6

Clarendon Way, University of Leeds

Leeds, UK, LS2 9LU

j.h.wu@leeds.ac.uk

Tel: +44 113 343 3431

Abstract:

Aim

Periodontitis is a multifactorial condition linked to increased risk of systemic diseases. This study aimed to identify disease trajectories of people with periodontitis using the process mining technique as a heuristic approach.

Material and Methods

A total of 188,863 participants from the UK Biobank cohort were included. Self-reported oral health indicators (bleeding gums, painful gums, loose teeth) were surrogates for periodontitis at baseline. Systemic disease diagnoses and dates formed the process mining event log. Relative risk (RR) of systemic diseases, disease trajectories and Cox proportional hazard ratio (HR) models for mortality were compared to age and sex-matched controls who did not report history of periodontitis.

Results

Participants with loose teeth had shorter median time to most systemic diseases and crude RR were increased for several diseases including cardiovascular disease (crude RR: 1.15, 95% CI: 1.03-1.28), hypertension (crude RR: 1.14, 95% CI: 1.05-1.24) and depression (crude RR: 1.33, 95% CI: 1.09-1.61). Participants with loose teeth had increased RR for 20 disease

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jcpe.13536

trajectories, though these were not significant after adjustments. Participants with bleeding/painful gums had similar disease trajectories compared to matched controls.

Conclusions

Self-reported periodontitis may be associated with early and frequent multimorbidity development, though further evidence is required to confirm this hypothesis. People with periodontitis should be informed of the risks of disease progression and be targeted in prevention initiatives.

Word count: 209

Keywords: Periodontitis, Tooth Loss, Noncommunicable Diseases, Multimorbidity, Cohort Studies

Clinical Relevance

Scientific Rationale for Study

Evidence from longitudinal cohort studies have previously demonstrated that periodontitis increases the risk of several subsequent systemic conditions. However, little is known regarding the full disease trajectory of periodontitis and interactions with subsequent multimorbidities.

Principle Findings

Through the use of the novel process mining method, we have been able demonstrate increased risk of a range of systemic diseases disease trajectories in people with self-reported bleeding gums, painful gums and loose teeth.

Practical Implications

The general population should be informed of the risks of possible systemic disease progression and multimorbidities associated to periodontitis.

Introduction

Periodontitis is the sixth most prevalent condition worldwide and severe periodontitis affects approximately 10% of the population [1]. Progression of poor oral health to periodontitis is multifactorial, with risk factors including age, sex, poor oral hygiene and smoking [2-4]. Amongst the same adult population, systemic diseases including cardiovascular disease (CVD), hypertension, diabetes, and inflammatory disease are also highly prevalent [5]. Robust evidence has shown that periodontitis is associated increased risk of a number of systemic diseases [6-8]. This may be due to shared common risk factors between conditions, or periodontitis may be a risk factor for developing systemic diseases [9]. Longitudinal studies have demonstrated that periodontitis increases the risk of subsequent systemic conditions including CVD, diabetes and cancer [10-12]. However, the full disease trajectory of periodontitis and its interactions with subsequent multimorbidities are yet to be quantified.

A disease trajectory refers to the probable pathway of a patient with a given condition may have, incorporating time-dependent sequential diagnoses [13]. Traditional epidemiological approaches employ pairwise methods to identify conditions associated with temporal disease progression [13, 14]. Exposure and outcome conditions must be predefined in pairs in such studies therefore a tangible disease trajectory cannot be investigated. As a heuristic method, process mining examines datasets as a whole by analysing an event log containing patient

diagnoses and dates as cases. Using this data mining approach, it is feasible to map common temporal disease trajectories end-to-end and examine direction of multimorbidities [15].

In this analysis, we aim to apply the novel process mining technique to identify the disease trajectories of periodontitis using linked systemic disease outcomes within the UK Biobank cohort. Furthermore, we will quantify the incidence rates and subsequent mortality for common disease trajectories in participants with periodontitis versus healthy controls.

Methods

Study Design

This is a prospective cohort study using 40-69 years old participants in the UK Biobank dataset spanning between 2006 to 2020.

Database

Data from the UK Biobank was used for this analysis. Ethical approval for UK Biobank was received by the North West Multi-centre Research Ethics Committee. UK Biobank approved the present study under protocol reference: 54633. Informed consent was obtained by UK Biobank and participants are able to withdraw at any time [16]. The UK Biobank is a national biomedical database containing information on 500,000 participants aged 40 – 69 years. Participants were invited to enrol in the study by mail and attend assessment centres where data from physical assessment and self-reported questionnaires were collated [16]. The core dataset is linked to multisource datasets including hospital episode statistics (HES), primary care records from general practice, national cancer registries, and civil registration of death. Participants were assessed and followed up at multiple time points, with further emergency or routine health visits and new disease diagnoses available through linked databases from healthcare providers.

Demographic and biomarker information including age, sex, ethnicity, household income, body mass index (BMI), C-reactive protein (CRP) level, blood pressure and heart rate was collected at UK Biobank assessment centres. History of smoking was dichotomous and derived according to self-report of current or ex-smoking status. Date of death was extracted from the linked civil registration of death.

Self-reported oral health indicators (bleeding gums, painful gums, loose teeth) were recorded during the baseline visit to UK Biobank assessment centre when participants were asked: “Do

you have any of the following?" Presence of loose teeth was a probable indicator of periodontitis (severe disease); bleeding gums are present across the periodontitis spectrum including gingivitis and mild to moderate cases of periodontitis [17]. Painful gums are associated with moderate to severe periodontitis but can also be associated with other oral conditions. In cases of multiple responses, we utilised the most severe indicator as the surrogate for periodontitis. Participants who did not report any of the oral health indicators were assumed to have normal periodontal health and classified as healthy controls. Information regarding tooth loss was not available.

Study Sample

We included all participants recruited between 2006 and 2010 with available linked follow up HES data, primary care records, cancer registry information and death records and who did not withdraw from the study before 1st August 2020. Study censorship was either on the date of 1st August 2020, or the date a participant died. We excluded participants who were diagnosed with systemic diseases at baseline (n = 307,170) or did not report their oral health status (n = 6,472) (Figure 1).

Systemic Diseases

We defined systemic diseases using validated ICD-10 and Read (v2) codes developed within CALIBER code lists [18]. We aggregated systemic diseases into well-known conditions: CVD; hypertension; cancers; depression; and metabolic, neurological, inflammatory (including autoimmune diseases), respiratory, renal and liver diseases (Supplement Table 1). For each participant, their follow up systemic diseases diagnosis and dates were extracted as events and validated from the linked HES, primary care records, or within core dataset variables derived from the cancer registry.

Statistical Analysis

Baseline characteristics are presented as means and standard deviations (SD) for continuous variables, and frequency (percentage) for categorical variables. We compared the participants with each of the oral health indicators (bleeding gums, painful gums or loose teeth) to those without self-reported history of these indicators as healthy controls. To ensure a fair comparison, we matched the participants from each indicator with healthy controls by age and sex separately according to the maximum possible matching frequency (1:2 in bleeding gums; 1:20 in painful gums and loose teeth) (Figure 1).

Event logs were created for each participant so that process mining methods could be applied. Each participant had at least two events within the log: baseline (recruitment date) and censor (date of last follow up/death date). If a predefined systemic condition happened between baseline and censor dates, an event was created for this diagnosis with the accompanying diagnosis date. In cases of recurrent diagnoses, only the first instance was extracted. Concurrent diagnoses were accounted for according to the diagnosis order in HES. Where the source of concurrent diagnoses was from GP records or cancer registry, cancer and hypertension diagnoses were placed first given the order of common patient pathways. The process mining technique was applied to the event log from participants within each oral health indicator and their corresponding matched controls. The temporal pathways of common disease trajectories (with minimum frequency of over 0.1%) for each indicator were presented in a process map. The incident risk of each disease trajectory was compared by oral health indicator and corresponding healthy controls with crude and adjusted relative risk (RR) with 95% confidence interval.

Cox proportional hazard models quantified mortality within a disease trajectory, adjusted for demographics, deprivation, and smoking. We assessed the proportional hazard assumptions using Schoenfeld residuals tests. Covariates were incorporated in the adjusted models based on associations observed in previous literature [2-4], with continuous variables included as such in the models.

Assuming missing at random, we replaced missing data through multiple imputation with chained equations with generation of 10 datasets. The missing value was predicted through predictive meaning matching method. When fitting Poisson or Cox models, each model was fitted to each of the imputed datasets and Rubin's rule was applied to combine the coefficients [19]. A sensitivity analysis was performed for complete cases only, all other tables in the main manuscript and supplement utilise the imputed dataset. To account for inflammation at baseline, we further added CRP in cubic spline as a covariate in the models.

Data management, analysis and process mining was conducted in R version (4.0.0) [20]. Statistical significance level was set as p-value < 0.05. This report conforms to STROBE guidelines.

Results

Overall, 188,863 participants were included in the analysis. 161,812 (85.7%) participants were identified as healthy controls, while 3,225 (1.7%) participants had painful gums, 19,930 (10.6%) had bleeding gums and 3,896 (2.1%) participants reported loose teeth (Table 1). On average participants with loose teeth were older than those with other oral health indicators (mean = 56.49 years, standard deviation [SD] = 7.70 years). There were fewer females with loose teeth (47.1%) compared to those with painful gums, bleeding gums and healthy controls (62.9%, 62.4% and 54.0%, respectively). A larger proportion of participants with loose teeth had an average household income of less than £18,000 per annum (28.1%) compared to remaining oral health indicators (21.1%, 15.0% and 15.3%) (Table 1). For information on matched controls see Supplement Table 2.

Participants with loose teeth

Process maps for disease trajectories were presented for participants with loose teeth and the matched healthy controls (Figure 2). Compared with the matched healthy controls, there was higher incidence of total systemic disease conditions including CVD (crude relative risk[RR]: 1.15, 95% CI: 1.03 - 1.28), hypertension (crude RR: 1.14, 95% CI: 1.05 - 1.24), respiratory disease (crude RR: 1.63, 95% CI: 1.40 - 1.91) and depression (crude RR: 1.33 95% CI: 1.09 - 1.61) for participants with loose teeth during follow up (Supplement Table 3). Median time to diagnostic events including cancer, CVD, renal disease, depression and inflammatory disease was also shorter in participants with loose teeth (Figure 2).

Overall, 20 disease trajectories had higher crude incidence risk for participants with loose teeth compared to matched healthy controls, however after adjustments several were not significant. *Baseline → hypertension → censor* (adjusted RR: 1.09, 95% CI: 0.97 – 1.23), *baseline → metabolic → censor* (adjusted RR: 1.14, 95% CI: 0.98 – 1.33) and *baseline → CVD → censor* (adjusted RR: 1.07, 95% CI: 0.89 – 1.08) were the most common disease trajectories of participants with loose teeth. The RR of most disease trajectories were not significant after adjustments, however the risk of *baseline → hypertension → respiratory → censor* increased more than two-fold after adjusting for covariates (adjusted RR: 2.08, 95% CI: 1.21 – 3.58) (Table 2). A total of 15 disease trajectories involved CVD and hypertension diagnostic events (Figure 3).

In general, risk of mortality was increased for disease trajectories in participants with loose teeth (Figure 3). In particular, the risk of mortality was increased by 50% for *baseline → CVD → censor* (adjusted HR 1.57, 95% CI: 0.95 - 2.62) and by more than two-fold for *baseline → metabolic → censor* (adjusted HR 2.03, 95% CI: 0.70 - 5.87). Risk of mortality was significantly increased in participants that did not develop systemic disease during follow up (adjusted HR 1.40, 95% CI: 1.11 - 1.77), though risk of mortality was not increased for participants who developed only hypertension before censorship (adjusted HR 0.94, 95% CI: 0.51 - 1.73) (Table 2).

Participants with bleeding and painful gums

Process maps for disease trajectories were presented for participants with bleeding and painful gums and the respectively matched healthy controls (Supplement Figures 1 and 2). Compared with the matched healthy controls, there were higher incidence of individual systemic conditions including CVD (crude RR: 1.29, 95% CI: 1.14 - 1.47) and depression (crude RR: 1.34, 95% CI: 1.09 - 1.64) for participants with painful gums during follow up (Supplement Table 3). For participants with bleeding gums, there were higher incidence of individual systemic conditions including hypertension (crude RR: 1.08, 95% CI: 1.02 - 1.14), and liver disease (crude RR: 1.66, 95% CI: 1.19 - 2.33) (Supplement Table 3). Median time to diagnostic events was comparable between participants with bleeding and painful gums (Supplement Figures 1 and 2).

Compared with respective healthy controls, participants with bleeding gums demonstrated increased risk for *baseline → metabolic→ CVD → censor* after adjustment for covariates (adjusted RR: 1.75, 95% CI: 1.14 - 2.70). Participants with painful gums had increased risk for *baseline → CVD → censor* (adjusted RR: 1.29, 95% CI: 1.06 - 1.58) and *baseline → metabolic → censor* (adjusted RR: 1.20, 95% CI: 1.02 - 1.42).

Risk of mortality was particularly high for participants with bleeding gums for *baseline → hypertension → cancer → censor* (adjusted HR: 5.54, 95% CI: 1.33 - 23.08), though this estimate was limited by small sample size (n = 32). The risk of mortality in participants with painful gums was increased for and *baseline → CVD → censor* (adjusted HR: 2.32, 95% CI: 1.29 - 4.16) (Supplement Tables 4 and 5).

Sensitivity Analyses

There was missing data in the following variables: ethnicity (0.3%), household income (12.9%), BMI (0.4%), systolic and diastolic blood pressure readings (6.4%), heart rate (91.4%), CRP level (5.8%), history of smoking (31.8%). Sensitivity analysis showed similar results for complete cases (Supplement Table 6). Adjusting for CRP in cubic spline also had similar overall effect to disease trajectories in relative risk and mortality in participants with loose teeth (Supplement Table 7).

Discussion

Using the novel process mining technique, our study has revealed common disease trajectories of participants with self-reported bleeding gums, painful gums and loose teeth as surrogates for periodontitis. The results showed that participants with loose teeth (indicative of severe periodontitis) took shorter time to develop multimorbidity than matched healthy controls and were at increased risk for the majority of identified disease trajectories, however this was not significant after adjustment for key confounders. Participants with bleeding gums or painful gums had similar disease trajectories when compared to their respective healthy controls.

Our study took a systematic approach to map out the disease trajectories for participants without history of systemic disease using three self-reported oral health indicators as surrogates for periodontitis. Importantly, loose teeth - indicative of more advanced periodontitis, showed the strongest risk to developing multimorbid disease trajectories. We demonstrated augmented risks of systemic disease in participants with loose teeth, however the key strength for our study was that it was not limited to a single predefined disease outcome. Through process mining we were able to incorporate a range of systemic conditions associated with periodontitis and examine the multimorbid development sequence for each individual. As participants were otherwise free from systemic disease at baseline, it was not appropriate to adjust for co-morbidities. Our results are consistent with findings from previous longitudinal studies investigating periodontitis and subsequent systemic conditions and suggest that severe periodontitis may be a risk factor to subsequent systemic disease development. For example, increased risk of metabolic disease in periodontitis has previously been demonstrated (adjusted HR: 1.69, 95% CI: 1.06 – 2.69) [12], while risk of mortality from respiratory disease is also increased (adjusted HR: 1.69, 95% CI: 1.06 – 2.69) [21]. It should be noted that estimated risks in the present study did not demonstrate significance according to confidence intervals, this is likely due to the modest sample sizes within each of

the identified trajectories. Process mining applied to a larger dataset would provide more precise estimates.

Previous findings suggest conflicting estimates of risk for hypertension and CVD in periodontitis. We identified increased risk of disease trajectories which frequently included hypertension and CVD in older aged participants with loose teeth. Estimates for the risk of hypertension in people with periodontitis is controversial [22, 23]. Findings are similarly variable in studies that estimate risk of CVD [11, 24, 25]. A possible explanation could be divergence in demographics across populations (i.e., age and deprivation). There are indications that periodontitis severity may affect disease progression which could also explain the discrepancies in estimates across studies [8]. In contrast to previous studies, our study excluded participants with any systemic disease at baseline. While this impacted statistical significance due to a reduced number of participants within some trajectories, confounding by existing comorbidities was minimised.

We investigated the disease trajectories of participants with self-reported loose teeth, painful gums and bleeding gums as separate surrogates for periodontitis, as such there were some distinct differences in our findings across oral health indicators. More disease trajectories had increased risk for participants with loose teeth compared to those with bleeding gums and painful gums. There was also a larger proportion of disease trajectories with augmented risk of mortality in participants with loose teeth. Discrepancies across oral health indicators could be explained by differences in periodontitis severity, represented by the surrogates individually. Loose teeth are caused by alveolar bone loss as a result of severe or chronic periodontitis, bleeding gums are usually a sign of gingivitis or mild periodontitis and painful gums can be an indicator of moderate to severe periodontitis or other oral conditions entirely [26]. Differences in the grade of periodontitis could explain the distinct increases to risk estimates in participants across the periodontitis indicators.

Our study found an increased risk of mortality across several disease trajectories in participants with oral health conditions compared to healthy controls. In particular, the mortality risk was significantly increased even in participants with loose teeth who did not develop subsequent systemic conditions, and in participants with loose teeth who developed CVD before censorship. This is consistent with a previous report that severe periodontitis may increase risk of mortality from CVD (HR = 1.35, 95% CI = 1.03 – 1.77) [25]. We

previously demonstrated that self-reported oral health responses for periodontitis in the UK Biobank cohort is associated to increased risk of mortality following COVID-19 infection [27]. While the latter study investigated a communicable outcome, the implication remains that increased risk of mortality could be caused by the chronic inflammation that ensues from untreated periodontitis, causing severe disease progression and poorer outcomes [9].

To our knowledge this study is the first of its kind to assess disease trajectories of periodontitis, as such there are several strengths to note. As a large cohort linked to health outcomes, the UK Biobank cohort enabled comprehensive investigation of disease trajectories of periodontitis using self-reported oral health responses. Furthermore, our study excluded participants with history of systemic disease at baseline which reduced the risk of confounding from comorbidities. Process mining techniques have previously demonstrated their utility in identifying disease trajectories using large electronic health record datasets [15, 28]. While there have been recommendations for the use of process mining in improving dental patient pathways [29], the utility in identifying disease trajectories of oral conditions has not previously been explored. Process mining enabled the first measurement of the temporal progression from periodontitis and subsequent systemic disease diagnostic events. The findings of such analyses can provide a basis for future research in investigating the biological mechanisms that underpin untreated disease interactions and multimorbidity.

We observed augmented associations of systemic disease in participants with loose teeth as an indicator of more severe periodontitis compared to minimal associations in those with self-reported bleeding gums. ICD-10 codes for periodontitis were not available and the study was limited to self-reported responses for oral health status. As these findings align with previous literature on severe periodontitis-systemic associations [8] this could advocate the self-reported measures as valid indicators of periodontitis in this study [30]. However, a previous study of oral health and gastrointestinal cancer in the UK Biobank cohort suggested that self-reported measures of oral health may not be direct indicators of periodontitis and could also be associated with trauma or ageing [31]. Validity of self-reported indicators of periodontitis can be population-dependent [32, 33] and sensitivity/specificity of self-reported periodontitis is suboptimal [34]. Given that the periodontitis indicators have not been validated in the UK biobank population, the limitations of the self-reported responses should be considered in the interpretation of findings. In smokers, periodontitis can silently progress without signs of bleeding gums, while bleeding gums can be caused by vigorous brushing. Use of a clinical

periodontitis classification would reduce false positives and alleviate potential bias of the self-reported responses.

There were some differences in significance levels of crude and adjusted risks and estimates were subject to high levels variability. While a probable cause is the modest numbers of cases within each trajectory, it is also possible that there are interactions between risk factors that affect an individual's subsequent risk of adverse systemic outcomes, and that the covariates in adjusted models do not eliminate all potential confounding. In particular, smoking is highly prevalent in people with periodontitis and while the adjusted models adjusted for smoking status, it is possible residual bias may inflate the associations between periodontitis and systemic disease [35]. Our study also included CRP level as a possible non-linear covariate to represent baseline inflammation and account for confounding from interactions of undiagnosed chronic diseases. It could be argued that CRP level may fluctuate after baseline and prior to any actual disease diagnosis, or a condition may go undiagnosed altogether, potentially impacting on the diagnosis event timings and pathways observed in the present study. Cubic spline exploration of other risk factors not investigated in the present study may also have a non-linear association [36]. In light of this, the findings should be interpreted cautiously and used as a basis for further investigation of the individual characteristics and biomarkers that underpin the association of periodontitis and subsequent systemic disease.

Participants with systemic disease at baseline were excluded from our study, therefore, we were not able to investigate the impact of periodontitis with comorbidities on subsequent disease trajectories. Another implication of our study exclusion criteria was the relatively small number of participants within each oral health indicator for analysis and therefore some risk estimates may not be robust; small numbers of participants in some disease trajectories would have had some effect on significance testing.

In conclusion, our findings suggest that people with loose teeth may be more likely to develop multimorbidities sooner and more frequently, which in turn may increase the risk of mortality; however further evidence is required to confirm this hypothesis. Use of a clinical periodontitis classification in future studies would eliminate the potential biases and allow for more robust conclusions. The common disease trajectories identified in this study could provide a basis for further research on oral and systemic disease interactions. Clinically,

people with severe periodontitis should be informed of the potential risks of development of multimorbidity, with targeted early prevention initiatives is emphasised.

Funding

Harriet Larvin was supported by the Hopper Scholarship at the University of Leeds. The research is supported by the National Institute for Health Research (NIHR) infrastructure at Leeds. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgements

This research has been conducted using data from UK Biobank (www.ukbiobank.ac.uk), a major biomedical database under protocol reference: 54633.

Author Contributions

HL: Contributed to conception, data analysis and interpretation, and drafted the manuscript.
JK: Contributed to interpretation, and critically revised the manuscript
VRA: Contributed to interpretation, and critically revised the manuscript
SP: Contributed to interpretation, and critically revised the manuscript
JW: Contributed to conception, study design, data acquisition and interpretation, and critically revised the manuscript
All authors gave their final approval and agree to be accountable for all aspects of the work.

Data Availability

The dataset analysed for this study can be found in the UK Biobank data repository (http://biobank.ndph.ox.ac.uk/showcase/) which is accessible for researchers with an approved protocol.

Figure 1 Flow chart demonstrating reasons for exclusion from study and number of participants included in the analysis.
Key: cardiovascular disease (CVD), periodontitis (PD).

Figure 2 Process maps depicting a minimum 0.1% of disease trajectories of **a)** age and sex matched healthy controls to participants with **b)** loose teeth. Arrows show trajectory direction and median time in months to diagnosis in flow. Nodes demonstrate relative number of cases per diagnosis and flow.
Key: cardiovascular disease (CVD)

Figure 3 Frequent disease trajectories (minimum rate 0.1%) in participants with **a)** loose teeth, **b)** painful gums and **c)** bleeding gums compared to corresponding age and sex matched healthy controls alongside the crude relative risk (RR) of trajectory and hazard ratio (HR) for mortality.
Key: cardiovascular disease (CVD), hazard ratio (HR), relative risk (RR).
(-) Insufficient numbers in trajectory to report rate.

Table 1 Summary table of UK Biobank participants stratified by oral health indicator.
Key: body mass index (BMI), beats per minute (BPM), c-reactive protein (CRP), number of participants (n), millimetres of mercury (mmHg), periodontitis (PD), standard deviation (SD) **Note:** means and percentages are calculated for variables excluding missing data. There was missing data in the following variables: ethnicity (0.3%), household income (12.9%), BMI (0.4%), systolic and diastolic blood pressure readings (6.4%), heart rate (91.4%), CRP level (5.8%), history of smoking (31.8%).

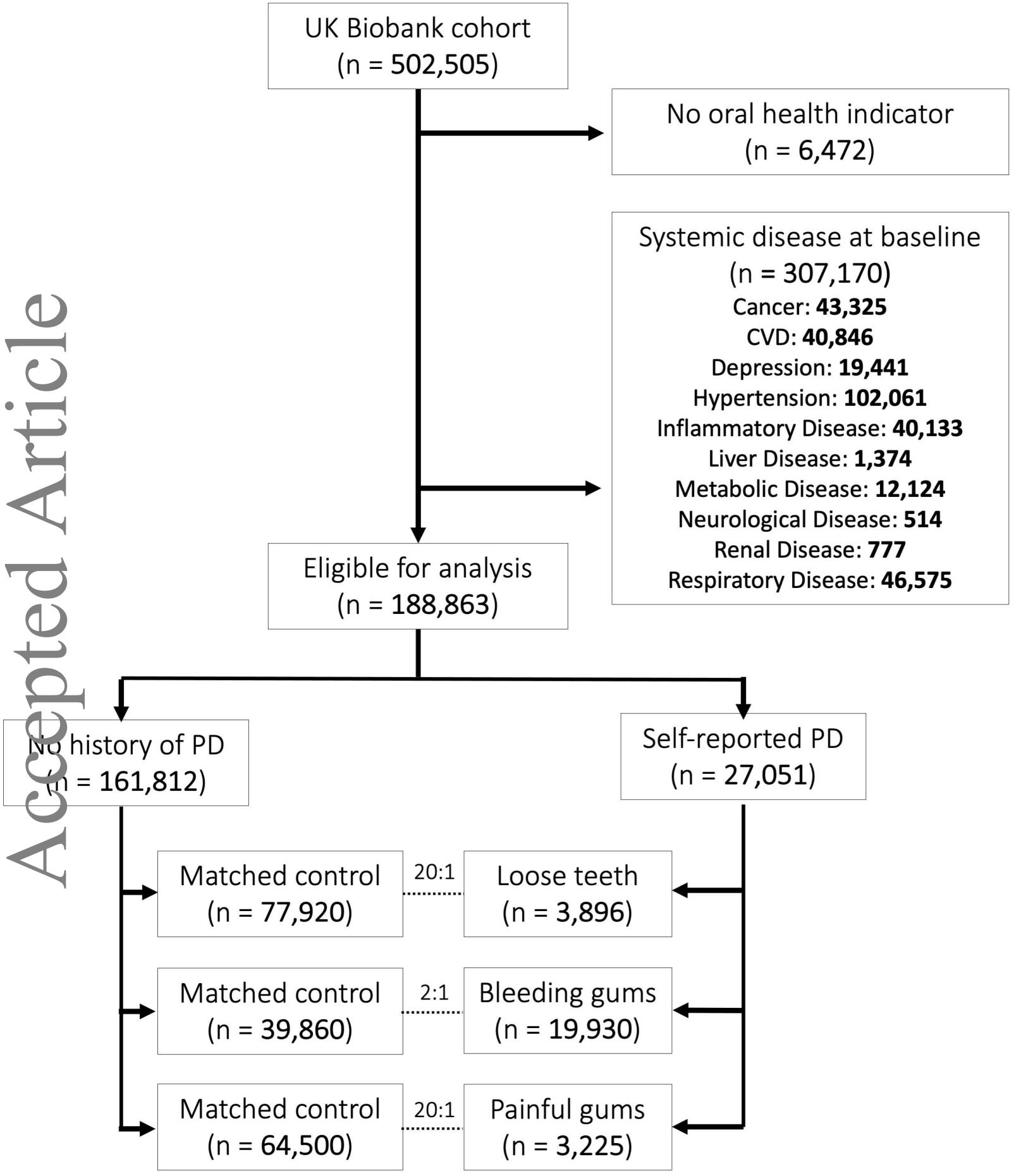
Table 2 Crude and adjusted relative risk (RR) and hazard ratio (HR) for the most frequent trajectories observed in participants with loose teeth compared to age and sex matched healthy controls.
Key: c-reactive protein (CRP), confidence interval (CI), cardiovascular disease (CVD), hazard ratio (HR), number of participants (n), relative risk (RR).
(*) Crude rate reported due to insufficient numbers for adjusted model to converge.
(-) Insufficient numbers in trajectory to report crude or adjusted rate.
Note: adjusted for BMI, CRP level, sex, ethnicity, household income and smoking history.

References

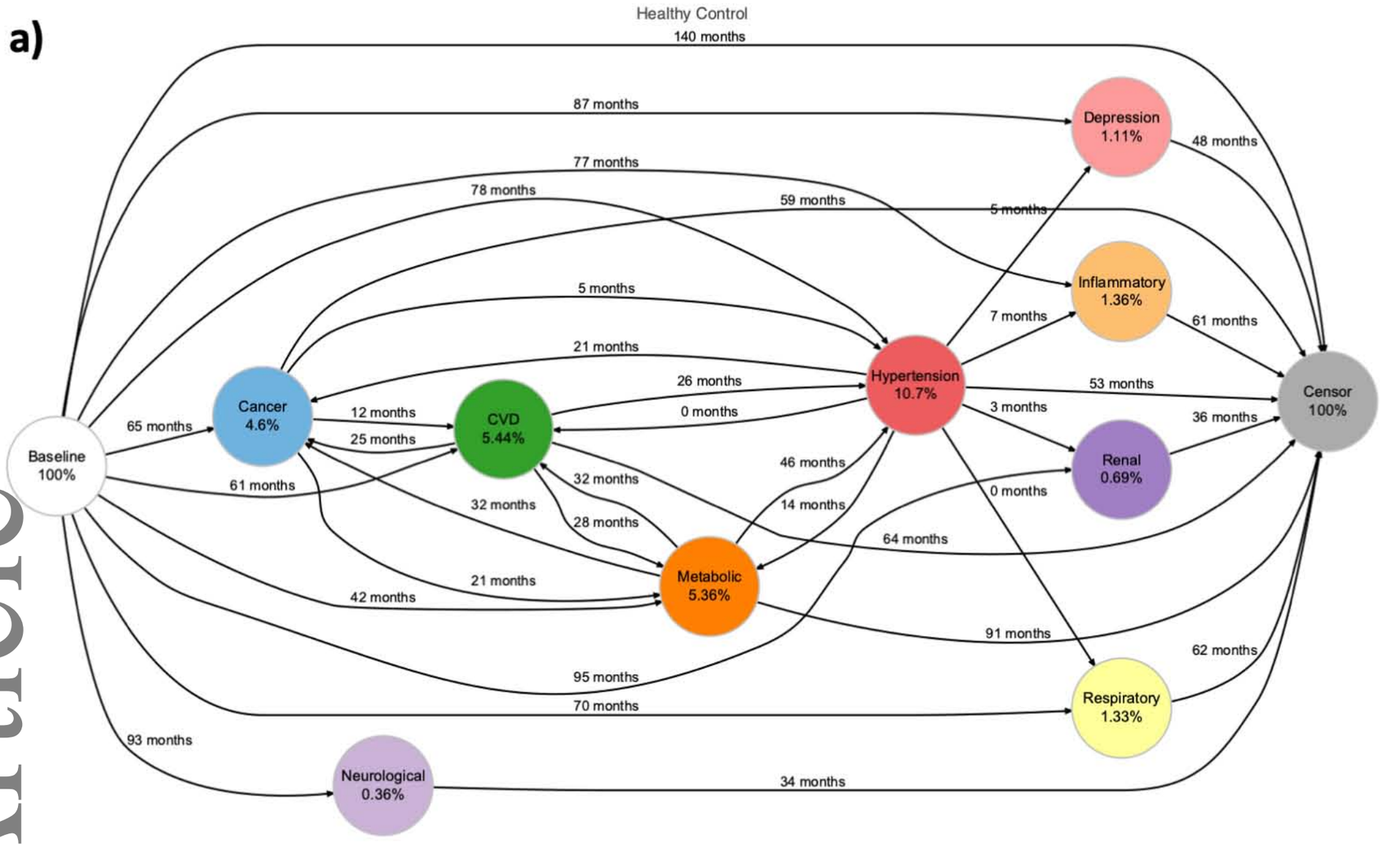
1. Kassebaum, N.J., et al., *Global burden of severe periodontitis in 1990-2010: a systematic review and meta-regression.* J Dent Res, 2014. **93**(11): p. 1045-53.
2. De Angelis, F., et al., *Influence of the oral status on cardiovascular diseases in an older Italian population.* Int J Immunopathol Pharmacol, 2018. **32**: p. 394632017751786.
3. Grover, V., et al., *The Gender Bender effect in Periodontal Immune Response.* Endocr Metab Immune Disord Drug Targets, 2016. **16**(1): p. 12-20.
4. Kinane, D.F. and I.G. Chestnutt, *Smoking and periodontal disease.* Crit Rev Oral Biol Med, 2000. **11**(3): p. 356-65.

5. GBD 2019 Diseases and Injuries Collaborators, *Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019.* Lancet, 2020. **396**(10258): p. 1204-1222.
6. Aguilera, E.M., et al., *Periodontitis is associated with hypertension: a systematic review and meta-analysis.* Cardiovascular Research, 2019.
7. Borgnakke, W.S., et al., *Effect of periodontal disease on diabetes: systematic review of epidemiologic observational evidence.* J Clin Periodontol, 2013. **40 Suppl 14**: p. S135-52.
8. Larvin, H., et al., *Risk of incident cardiovascular disease in people with periodontal disease: A systematic review and meta-analysis.* Clin Exp Dent Res, 2020.
9. Esteves-Lima, R.P., et al., *Association between periodontitis and serum c-reactive protein levels.* J Clin Exp Dent, 2020. **12**(9): p. e838-e843.
10. Abnet, C.C., et al., *Tooth loss is associated with increased risk of total death and death from upper gastrointestinal cancer, heart disease, and stroke in a Chinese population-based cohort.* Int J Epidemiol, 2005. **34**(2): p. 467-74.
11. Winning, L., et al., *Periodontitis and risk of prevalent and incident coronary heart disease events.* J Clin Periodontol, 2020. **47**(12): p. 1446-1456.
12. Winning, L., et al., *Periodontitis and incident type 2 diabetes: a prospective cohort study.* J Clin Periodontol, 2017. **44**(3): p. 266-274.
13. Jensen, A.B., et al., *Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients.* Nature Communications, 2014. **5**(1): p. 4022.
14. Giannoula, A., et al., *Identifying temporal patterns in patient disease trajectories using dynamic time warping: A population-based study.* Scientific Reports, 2018. **8**(1): p. 4216.
15. Kusuma, G., et al., *Process Mining of Disease Trajectories: A Feasibility Study*. 2020. 705-712.
16. UK Biobank. *Protocol for a large-scale prospective epidemiological resource.* 2006; Available from: www.ukbiobank.ac.uk/resources/.
17. Dietrich, T., et al., *Periodontal diagnosis in the context of the 2017 classification system of periodontal diseases and conditions - implementation in clinical practice.* Br Dent J, 2019. **226**(1): p. 16-22.
18. Denaxas, S.C., et al., *Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER).* Int J Epidemiol, 2012. **41**(6): p. 1625-38.
19. Rubin, D.B., *Multiple imputation for nonresponse in surveys*. Vol. 81. 2004: John Wiley & Sons.
20. R Core Team, *R: A language and environment for statistical computing*, in *R Foundation for Statistical Computing*. 2017: Vienna, Austria
21. Qian, Y., et al., *Periodontitis increases the risk of respiratory disease mortality in older patients.* Experimental Gerontology, 2020. **133**: p. 110878.
22. Kawabata, Y., et al., *Relationship Between Prehypertension/Hypertension and Periodontal Disease: A Prospective Cohort Study.* Am J Hypertens, 2016. **29**(3): p. 388-96.
23. Rivas-Tumanyan, S., et al., *Periodontal disease and incidence of hypertension in the health professionals follow-up study.* American Journal of Hypertension, 2012. **25**(7): p. 770-6.

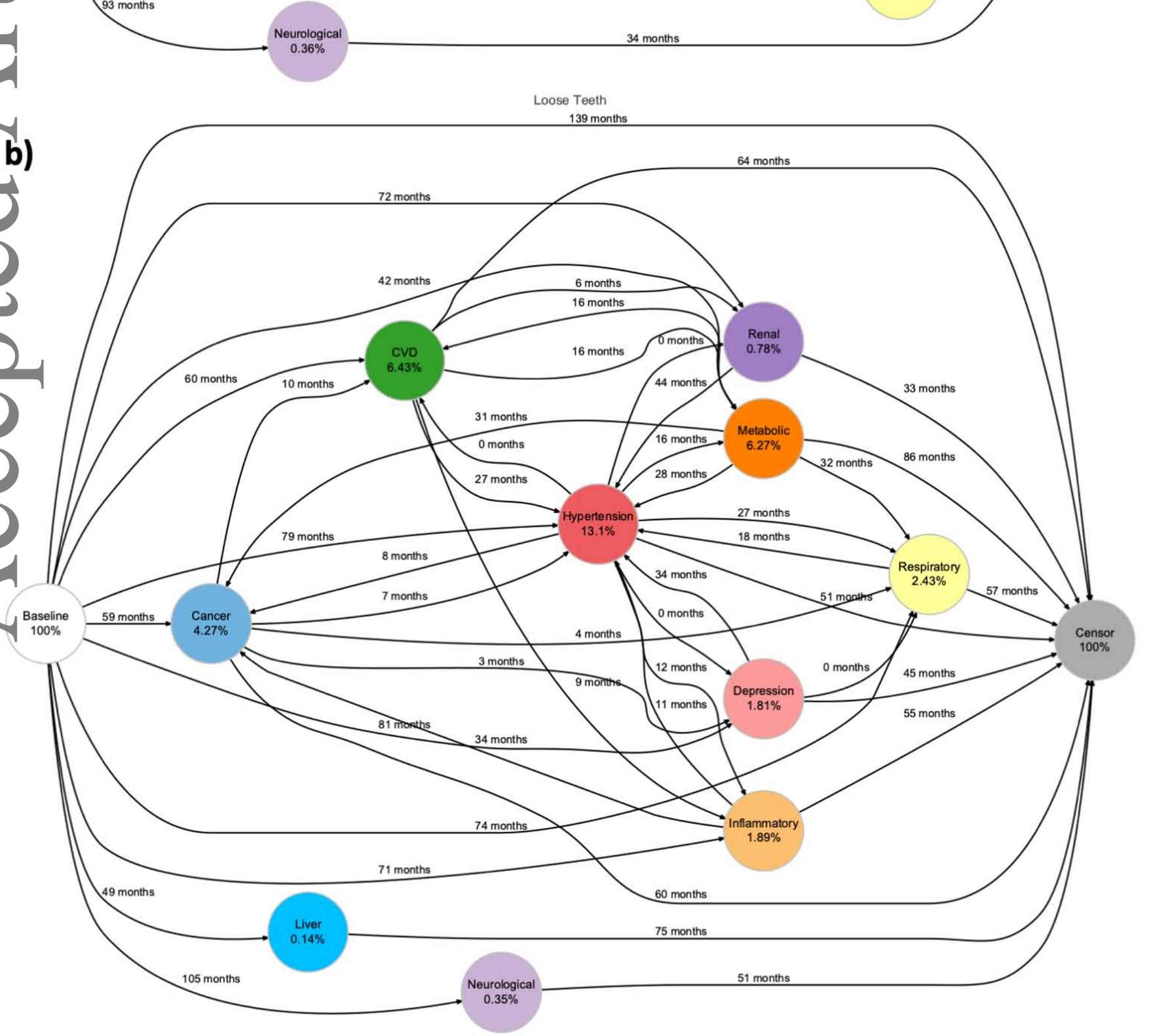
24. Cho, H.J., et al., *Severe Periodontal Disease Increases Acute Myocardial Infarction and Stroke: A 10-Year Retrospective Follow-up Study.* J Dent Res, 2021: p. 22034520986097.
25. Tu, Y.K., et al., *Associations between tooth loss and mortality patterns in the Glasgow Alumni Cohort.* Heart, 2007. **93**(9): p. 1098-103.
26. Coventry, J., et al., *ABC of oral health: periodontal disease.* BMJ (Clinical research ed.), 2000. **321**(7252): p. 36-39.
27. Larvin, H., et al., *The Impact of Periodontal Disease on Hospital Admission and Mortality During COVID-19 Pandemic.* Frontiers in Medicine, 2020. **7**: p. 861.
28. Pebesma, J., et al., *Clustering Cardiovascular Risk Trajectories of Patients with Type 2 Diabetes Using Process Mining.* 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019: p. 341-344.
29. Fox, F., et al., *A Data Quality Framework for Process Mining of Electronic Health Record Data.* 2018 IEEE International Conference on Healthcare Informatics (ICHI), 2018: p. 12-21.
30. Abbood, H.M., et al., *Validity of Self-Reported Periodontal Disease: A Systematic Review and Meta-Analysis.* J Periodontol, 2016. **87**(12): p. 1474-1483.
31. Jordão, H.W., et al., *The association between self-reported poor oral health and gastrointestinal cancer risk in the UK Biobank: A large prospective cohort study.* United European Gastroenterology Journal, 2019. **7**(9): p. 1241-1249.
32. Blicher, B., K. Joshipura, and P. Eke, *Validation of self-reported periodontal disease: a systematic review.* J Dent Res, 2005. **84**(10): p. 881-90.
33. Joshipura, K.J., et al., *Validity of a self-reported periodontal disease measure.* J Public Health Dent, 1996. **56**(4): p. 205-12.
34. Montero, E., et al., *Validation of self-reported measures of periodontitis in a Spanish Population.* Journal of Periodontal Research, 2020. **55**(3): p. 400-409.
35. Hujoel, P.P., et al., *Periodontitis–systemic disease associations in the presence of smoking – causal or coincidental?* Periodontology 2000, 2002. **30**(1): p. 51-60.
36. Hammami, I., B. Lacey, and S. Lewington, *3028The burden of hypertension and associated risk for cardiovascular mortality in the UK biobank.* European Heart Journal, 2018. **39**(suppl_1).



a)



b)



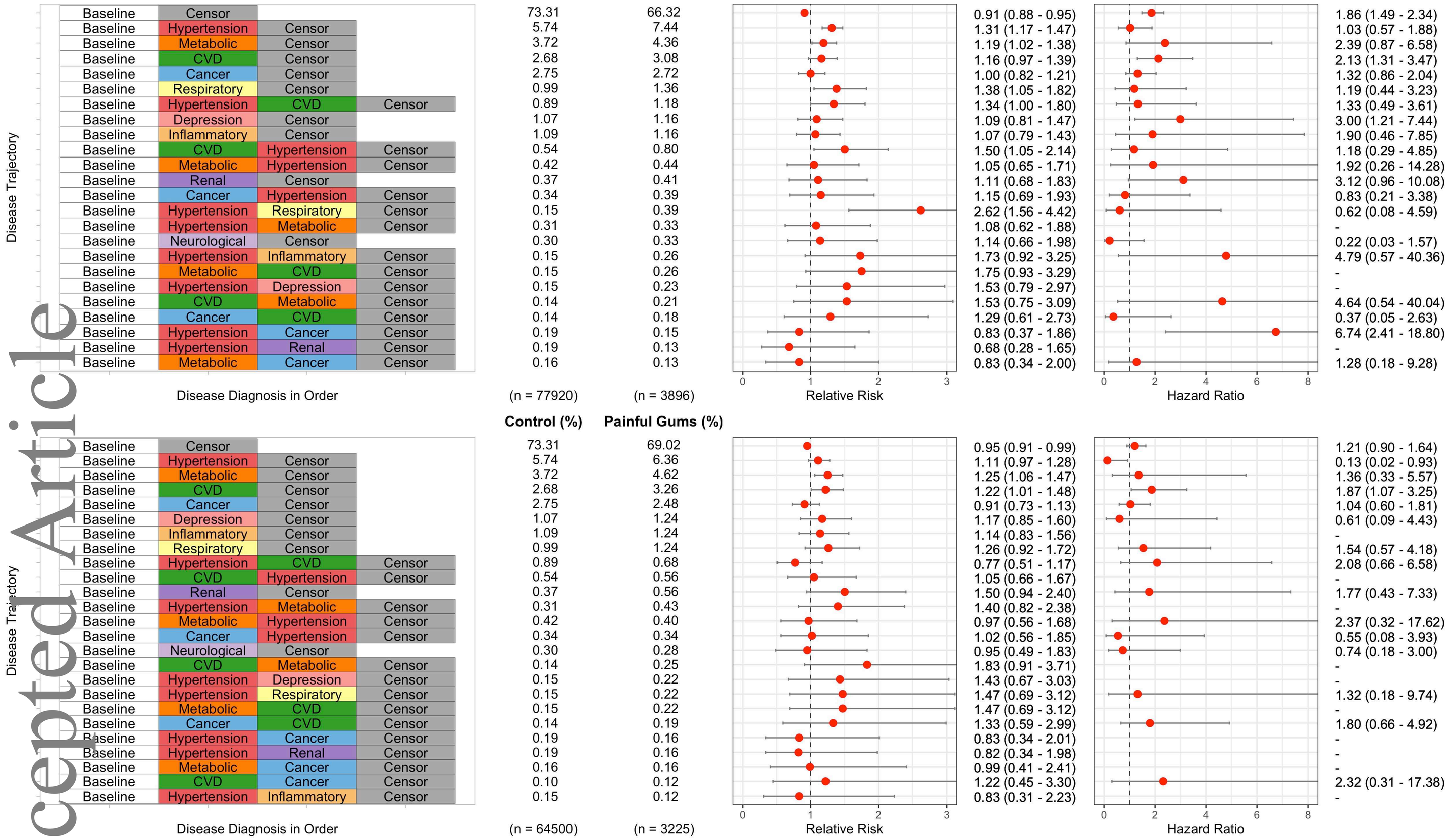


Table 1 Summary table of UK Biobank participants stratified by oral health indicator.

	Overall	Healthy Control	Painful Gums	Bleeding Gums	Loose Teeth	P-value
N, total	188,863 (100.0)	161,812 (85.7)	3,225 (1.7)	19,930 (10.6)	3,896 (2.1)	
Age at Baseline	54.12 (8.09)	54.29 (8.12)	53.68 (7.93)	52.31 (7.59)	56.49 (7.70)	< 0.001
Sex, female	103647 (54.9)	87350 (54.0)	2030 (62.9)	12432 (62.4)	1835 (47.1)	< 0.001
Ethnicity, white	177144 (94.1)	152771 (94.7)	2774 (86.2)	18243 (91.8)	3356 (86.4)	< 0.001
Household Income/ £						< 0.001
Less than 18,000	25731 (15.6)	21609 (15.3)	584 (21.1)	2627 (15.0)	911 (28.1)	
18,000 to 30,999	38229 (23.2)	32560 (23.1)	715 (25.9)	4039 (23.1)	915 (28.2)	
31,000 to 51,999	47166 (28.7)	40422 (28.7)	728 (26.4)	5233 (29.9)	783 (24.1)	
52,000 to 100,000	41656 (25.3)	36008 (25.5)	605 (21.9)	4503 (25.7)	540 (16.7)	
Greater than 100,000	11762 (7.1)	10442 (7.4)	130 (4.7)	1096 (6.3)	94 (2.9)	
BMI/ kg/m^2	26.36 (4.15)	26.30 (4.10)	26.40 (4.38)	26.72 (4.37)	26.95 (4.46)	< 0.001
Systolic Blood Pressure / mmHg	135.94 (18.61)	136.17 (18.65)	133.28 (17.93)	133.96 (18.03)	138.80 (19.35)	< 0.001
Diastolic Blood Pressure / mmHg	80.91 (10.30)	80.93 (10.30)	79.76 (10.15)	80.74 (10.26)	81.75 (10.56)	< 0.001
Heart Rate/ BPM	61.76 (10.18)	61.71 (10.19)	61.29 (9.92)	62.31 (10.12)	61.71 (10.34)	0.012
CRP level	2.05 (3.51)	2.02 (3.48)	2.28 (3.71)	2.12 (3.43)	2.74 (4.49)	< 0.001
Smoker, current or historic	107076 (83.2)	90867 (82.6)	1923 (86.7)	11477 (84.9)	2809 (93.4)	< 0.001

Key: body mass index (BMI), beats per minute (BPM), c-reactive protein (CRP), number of participants (n), millimetres of mercury (mmHg), periodontal disease (PD), standard deviation (SD) **Note:** means and (SD) are reported for continuous variables, n and percentages are reported for categorical variables. Estimates are calculated for variables excluding missing data.

Table 2 Crude and adjusted relative risk (RR) and hazard ratio (HR) for the most frequent trajectories observed in participants with loose teeth compared to age and sex matched healthy controls.

Disease Trajectory	Frequency (n)		Relative Risk (95% CI)		Hazard Ratio (95% CI)	
	Healthy Control	Loose Teeth	Crude	Adjusted	Crude	Adjusted
baseline → censor	57126	2584	0.91 (0.88 - 0.95)	0.96 (0.92 - 1.00)	1.86 (1.49 - 2.34)	1.40 (1.11 - 1.77)
baseline → hypertension → censor	4475	290	1.31 (1.17 - 1.47)	1.09 (0.97 - 1.23)	1.03 (0.57 - 1.88)	0.94 (0.51 - 1.73)
baseline → metabolic → censor	2895	170	1.19 (1.02 - 1.38)	1.14 (0.98 - 1.33)	2.39 (0.87 - 6.58)	2.03 (0.70 - 5.87)
baseline → CVD → censor	2088	120	1.16 (0.97 - 1.39)	1.07 (0.89 - 1.29)	2.13 (1.31 - 3.47)	1.57 (0.95 - 2.62)
baseline → cancer → censor	2144	106	1.00 (0.82 - 1.21)	0.89 (0.73 - 1.08)	1.32 (0.86 - 2.04)	1.09 (0.70 - 1.70)
baseline → respiratory → censor	774	53	1.38 (1.05 - 1.82)	1.12 (0.85 - 1.48)	1.19 (0.44 - 3.23)	1.12 (0.40 - 3.13)
baseline → hypertension → CVD → censor	694	46	1.34 (1.00 - 1.80)	1.08 (0.80 - 1.46)	1.33 (0.49 - 3.61)	0.95 (0.33 - 2.74)
baseline → depression → censor	832	45	1.09 (0.81 - 1.47)	1.17 (0.86 - 1.58)	3.00 (1.21 - 7.44)	2.46 (0.94 - 6.40)
baseline → inflammatory → censor	853	45	1.07 (0.79 - 1.43)	1.02 (0.76 - 1.38)	1.90 (0.46 - 7.85)	1.81 (0.42 - 7.81)
baseline → CVD → hypertension → censor	418	31	1.50 (1.05 - 2.14)	1.11 (0.77 - 1.60)	1.18 (0.29 - 4.85)	1.22 (0.28 - 5.26)
baseline → metabolic → hypertension → censor	326	17	1.05 (0.65 - 1.71)	0.88 (0.54 - 1.43)	1.92 (0.26 - 14.28)	2.60 (0.32 - 21.37)
baseline → renal → censor	291	16	1.11 (0.68 - 1.83)	0.98 (0.59 - 1.62)	3.12 (0.96 - 10.08)	1.91 (0.56 - 6.55)
baseline → cancer → hypertension → censor	263	15	1.15 (0.69 - 1.93)	0.91 (0.54 - 1.53)	0.83 (0.21 - 3.38)	0.86 (0.21 - 3.56)
baseline → hypertension → respiratory → censor	116	15	2.62 (1.56 - 4.42)	2.08 (1.21 - 3.58)	0.62 (0.08 - 4.59)	0.29 (0.03 - 2.93)
baseline → hypertension → metabolic → censor	243	13	1.08 (0.62 - 1.88)	0.92 (0.53 - 1.62)	-	-
baseline → neurological → censor	231	13	1.14 (0.66 - 1.98)	0.99 (0.56 - 1.72)	0.22 (0.03 - 1.57)	0.15 (0.02 - 1.10)
baseline → hypertension → inflammatory → censor	117	10	1.73 (0.92 - 3.25)	1.44 (0.76 - 2.75)	4.79 (0.57 - 40.36)	4.31 (0.37 - 49.63)
baseline → metabolic → CVD →	116	10	1.75	1.33	-	-

censor			(0.93 - 3.29)	(0.69 - 2.54)		
baseline → hypertension → depression → censor	119	9	1.53 (0.79 - 2.97)	1.37 (0.70 - 2.71)	-	-
baseline → CVD → metabolic → censor	106	8	1.53 (0.75 - 3.09)	1.42 (0.69 - 2.93)	4.64 (0.54 - 40.04)	*
baseline → cancer → CVD → censor	110	7	1.29 (0.61 - 2.73)	1.06 (0.50 - 2.27)	0.37 (0.05 - 2.63)	0.27 (0.04 - 2.03)
baseline → hypertension → cancer → censor	146	6	0.83 (0.37 - 1.86)	0.68 (0.30 - 1.55)	6.74 (2.41 - 18.80)	3.80 (1.20 - 12.03)
baseline → hypertension → renal → censor	148	5	0.68 (0.28 - 1.65)	0.47 (0.19 - 1.14)	-	-
baseline → metabolic → cancer → censor	122	5	0.83 (0.34 - 2.00)	0.76 (0.31 - 1.86)	1.28 (0.18 - 9.28)	1.76 (0.22 - 13.90)

Key: c-reactive protein (CRP), confidence interval (CI), cardiovascular disease (CVD), hazard ratio (HR), number of participants (n), relative risk (RR).

(*) Crude rate reported due to insufficient numbers for adjusted model to converge.

(-) Insufficient numbers in trajectory to report crude or adjusted rate.

Note: adjusted for BMI, CRP level, sex, ethnicity, household income and smoking history.