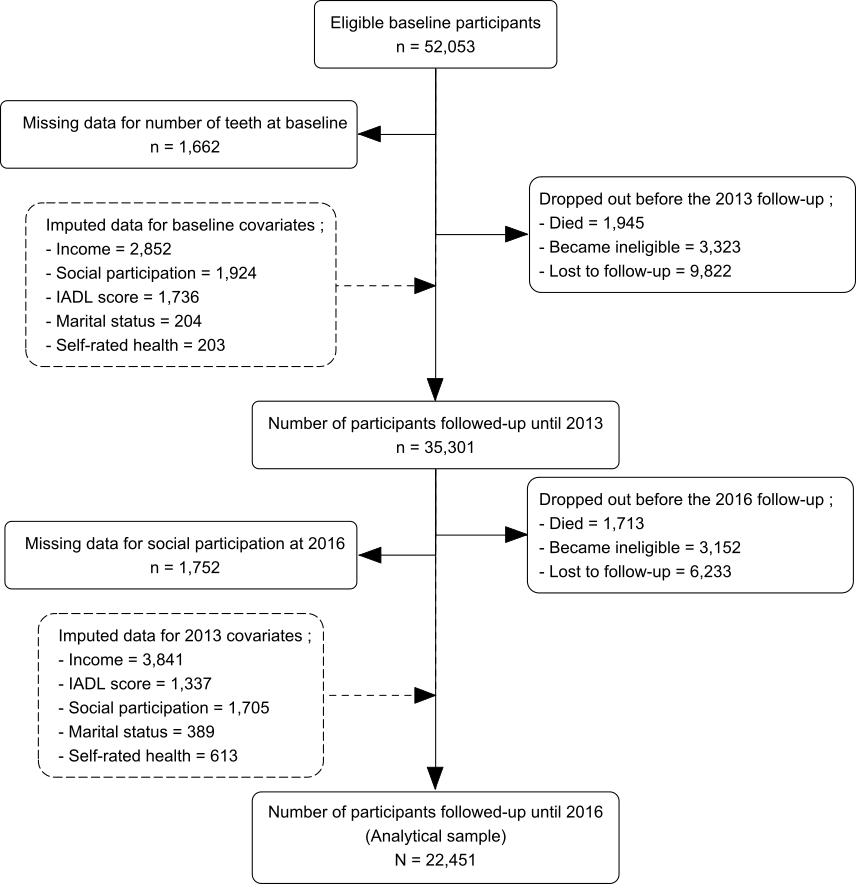
Effect of dental status on social participation: using a doubly robust estimator

Upul Cooray, Jun Aida, Georgios Tsakos,…

Last compiled on 21 June, 2021 at 00:12

#### Data

Figure 1: Selection of analytical sample



We used data from the Japan Gerontological Evaluation Study (JAGES) (Kondo et al. 2018). JAGES is an on-going nationwide cohort study for older adults aged 65 years or more living in Japan who are not eligible for long-term care insurance. In this study, data from the 2010 survey as the baseline and two subsequent follow-up surveys (2013 and 2016) were used. We identified 25,865 functionally independent individuals who responded to all three waves of JAGES. After excluding participants with missing information for social participation in 2016 (n=1,752) and number of teeth in 2010 (n=1,662), a total of 22,451 participants were included in the analyses. The selection of the analytical sample is illustrated in Figure 1.

#### Outcome variable

Social participation in 2016 was the outcome in this study. JAGES recorded the frequency of participation (“nearly every day,” “twice or thrice a week,” “once a week,” “once or twice a month,” “a few times/year,” “never”) for various social activities. We measured social participation by assessing the frequency of participation in any of the following activities, i.e., hobby groups, sports clubs, senior citizens’ clubs, residence groups, or volunteer groups. Participation in any of the aforementioned activities once a month or more often (vs. less than once a month or never) was defined as indicative of social participation (1= participation, 0= non-participation).

#### Exposure

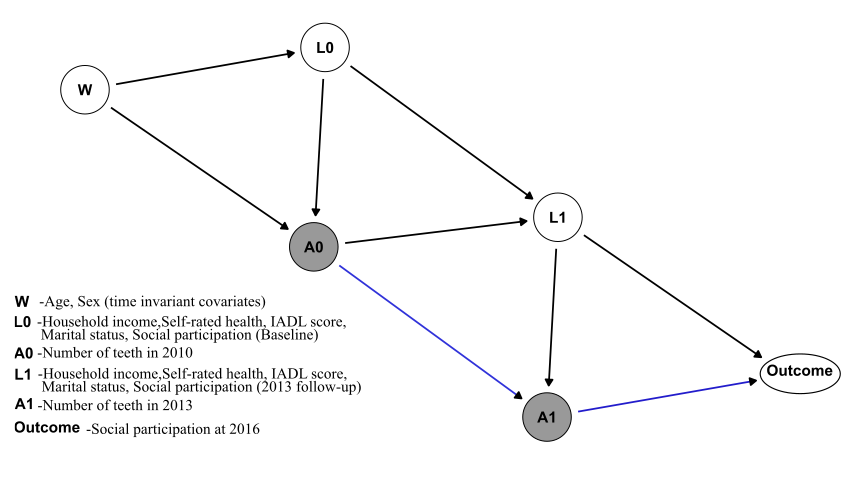
The number of remaining natural teeth at the time of the surveys in 2010 and 2013 were used as a time varying exposure in our analyses. Self-reported number of teeth were recorded using the response to the question “How many natural teeth to you presently have?” (Instructions: include dental implants and capped/crowned teeth as “natural teeth”). Participants’ responses were measured within four categories (i.e. ≥20 teeth/ 10-19 teeth/ 1-9 teeth/ no teeth).

#### Covariates

We controlled our analyses for both time-invariant and time-variant covariates as the number of teeth was assessed as a time varying exposure. Age (65- xx) and sex (male/female) measured in 2010 were treated as time-invariant covariates. Equalised annual household income (million yen), self-rated health (very good/ good/ not good/ bad), instrumental activities of daily living (IADL) score (0-13), and marital status (married/ single, widowed or divorced) were included as time-variant covariates. Furthermore, and the social participation measured in baseline and in 2013 (“nearly every day,” “twice or thrice a week,” “once a week,” “once or twice a month,” “a few times/year,” “never”)

#### Statistical analysis

Figure 2: DAG



Hypothesised temporal associations between study variables are shown in the directed acyclic graph (Figure 2). A descriptive analysis was performed to identify the characteristics of the participants stratified by exposure levels (number of teeth categories). Then, we used doubly-robust targeted maximum likelihood estimation (TMLE) (Schuler and Rose 2016) to estimate the counterfactual prevalence of social participation under hypothetical intervention policies related to number of teeth. Specifically, hypothetical intervention policies included were;

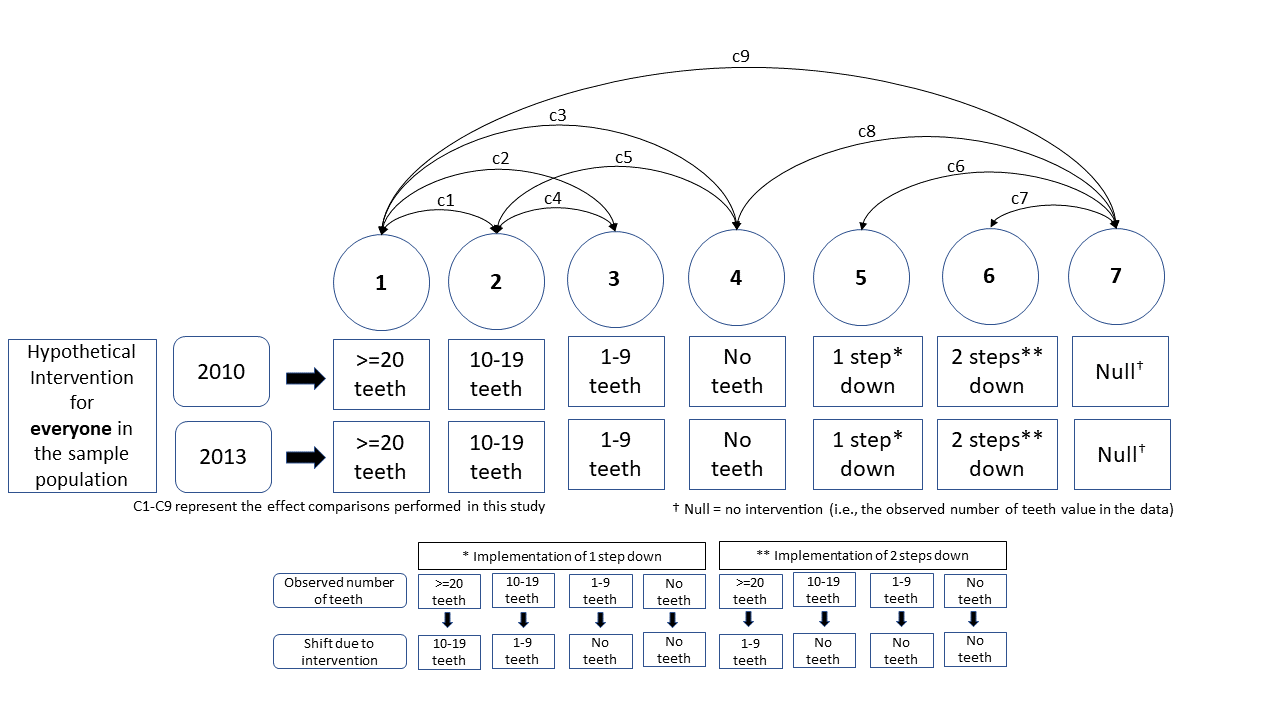
1. “all participants have ≥20 teeth in 2010 (baseline) and in 2013,”
2. “all participants have 10-19 teeth in 2010 and in 2013,”
3. “all participants have 1-10 teeth in 2010 and in 2013,”
4. “all participants are edentate in 2010 and in 2013,”
5. “participants deteriorate by one category in terms of number of teeth between 2010 and 2013 (see Figure 3 )”
6. “participants deteriorate by two categories in terms of number of teeth between 2010 and 2013 (see Figure 3 )”
7. “no intervention” (estimation with originally observed data in 2010 and in 2013).

Finally, we used counterfactual estimates related to the aforementioned different scenarios to calculate the additive treatment effects (ATEs) and the odds ratios (ORs) and their 95% confidence intervals (CIs) for social participation related to the different interventions through the following comparisons:

1. “Having ≥20 teeth at 2010 and 2013” vs “having 10-19 teeth at 2010 and 2013,”
2. “Having ≥20 teeth at 2010 and 2013” vs “having 1-9 teeth at 2010 and 2013,”
3. “Having ≥20 teeth at 2010 and 2013” vs “being edentulous at 2010 and 2013,”
4. “Having 10-19 teeth at 2010 and 2013” vs “having 1-9 teeth at 2010 and 2013,”
5. “Having 10-19 teeth at 2010 and 2013” vs “being edentulous s at 2010 and 2013,”
6. “no intervention” vs “deteriorate by one number of teeth category at 2010 and 2013,”
7. “no intervention” vs “deteriorate by two number of teeth categories at 2010 and 2013,”
8. “no intervention” vs “having ≥20 teeth at 2010 and 2013,” and
9. “no intervention” vs “being edentulous at 2010 and 2013.”

Implemented interventions and their comparisons are graphically illustrated in the Figure 3. All the estimates were controlled for time-variant and time-invariant covariates, and the social participation at baseline and 2013 (Figure 2).

Figure 3: Illustration of implemented hypothetical interventions



The TMLE estimator enabled unbiased estimation of the counterfactual outcomes and their contrasts (Van Der Laan and Rubin 2006; Schuler and Rose 2016) . This was achieved by estimating the probability of the exposure conditional on covariates (exposure model), and the conditional probability of outcome given exposure and covariates (outcome model). Unbiased estimations are obtained if, either the exposure model or the outcome model was consistently estimated (doubly-robust) (Laan and Gruber 2012). To increase the likelihood of correct model specification, the use of machine learning algorithms are typically recommended (Rose and Rizopoulos 2019; Schomaker et al. 2019). Therefore, we used SuperLearner, an ensemble method that uses weighted combinations of multiple machine learning algorithms to ensure robust specifications of exposure and outcome models (Laan et al. 2007). Generalized linear models, extreme gradient boosting models, and neural net were used as candidate algorithms within the SuperLearner (Venables and Ripley 2002; Chen and Guestrin 2016). We applied chained equations multiple imputation for the missing data in covariates generating five imputed datasets. The main analysis was performed using each imputed dataset and Rubin’s rules were used to combine the results (Rubin 2004).

TMLE method was implemented using tmtp and SuperLearner R packages (Williams and Díaz 2020). For multiple imputation mice R package was used. Reproducible codes to generate main result are provide in Appendix xx. Sensitivity analyses were conducted to assess the robustness of our main analysis. First, a complete case analysis was conducted without multiple imputation to assess the validity of the imputation procedure (an analysis of distribution of missing data is reported in Appendix xx). Second, we conducted the analysis without SuperLearner, only using generalised linear models to specify models. R codes used for all the analyses can be found in [<https://github.com/upulcooray/socialParticipation>](https://github.com/upulcooray/social_participation). All analyses were conducted in R studio using R version 4.0.5 for Windows x64.

### Results

Baseline characteristics of the sample stratified by the outcome variable are presented in Table 1. In the 2016 follow-up, 11,762 people reported social participation at least once a month. Older age, male sex, lower income, lower IADL score, being edentulous or having lower number of teeth, and lower frequency of social participation at baseline were all associated with non-participation in social activities in 2016.

Table 1: Baseline characteristics of the study population stratified by sex

|  |  |  |
| --- | --- | --- |
|  | Social-participation in 2016 (outcome) | |
| Characteristic1 | No N = 106891 | Yes N = 117621 |
| **Age (range 65yrs - 99yrs)** | 72.7 (5.3) | 71.8 (4.6) |
| **Annual household income (million yen)** | 2.4 (1.5) | 2.6 (1.6) |
| Missing observations (n) | 1,533 | 1,319 |
| **IADL score (range 0-13)** | 11.6 (1.6) | 12.3 (1.2) |
| Missing observations (n) | 937 | 799 |
| **Sex** |  |  |
| Male | 5,331 (49.9%) | 4,868 (41.4%) |
| Female | 5,358 (50.1%) | 6,894 (58.6%) |
| **Number of teeth (baseline)** |  |  |
| Edentulous | 1,237 (11.6%) | 829 (7.0%) |
| 1-9 teeth | 2,702 (25.3%) | 2,360 (20.1%) |
| 10-19 teeth | 2,915 (27.3%) | 3,223 (27.4%) |
| >= 20 teeth | 3,835 (35.9%) | 5,350 (45.5%) |
| **Social particiapation (baseline)** |  |  |
| Everyday | 141 (1.3%) | 891 (7.6%) |
| 2-3 times a week | 446 (4.2%) | 3,287 (27.9%) |
| Once a week | 700 (6.5%) | 2,710 (23.0%) |
| 1-2 times a month | 1,420 (13.3%) | 2,425 (20.6%) |
| Few times a year | 2,867 (26.8%) | 1,127 (9.6%) |
| Never | 3,778 (35.3%) | 735 (6.2%) |
| Missing (imputed) | 1,337 (12.5%) | 587 (5.0%) |
| **Self-rated health** |  |  |
| Very good | 1,155 (10.8%) | 1,927 (16.4%) |
| Good | 7,682 (71.9%) | 8,672 (73.7%) |
| Fair | 1,558 (14.6%) | 999 (8.5%) |
| Poor | 187 (1.7%) | 68 (0.6%) |
| Missing (imputed) | 107 (1.0%) | 96 (0.8%) |
| **Marital status** |  |  |
| Widowed,divorced, or unmarried | 2,557 (23.9%) | 2,736 (23.3%) |
| Married | 8,021 (75.0%) | 8,933 (75.9%) |
| Missing (imputed) | 111 (1.0%) | 93 (0.8%) |
| 1Mean (SD) for continuous variables; Frequency (%) for categorical variables | | |

Table 2 shows the results from the comparison of TMLE estimates related to different longitudinal intervention (exposure) regimes (figure 3). Having a relatively lower number of teeth negatively affected social participation during the six-year follow-up irrespective of the reference number of teeth category. After adjusting for age, sex, time varying confounders such as annual household income, SRH, IADL score, marital status, and social participation (baseline and 2013), consistent exposure to edentulousness had 18% (OR= 0.82 & 95%CI= 0.72,0.92) lower likelihood of social participation compared to having >=20 teeth (figure 3, c3). When compared to having 10-19 teeth throughout the study period, being edentulous lowered the likelihood of social participation by 11% (OR= 0.89 & 95%CI= 0.77,1.01) (figure 3, cXX). Our results also showed that if all the participants remained edentulous throughout the study period a 12% (OR= 0.88 & 95%CI= 0.78,0.97) reduction in social participation would have been expected (i.e., compared to their actual dental status ) (figure 3, cXX).

We found that if the everyone in our population sample was categorized into the next group with fewer number of teeth at both time points (i.e. in 2010 and 2013), this would lower the odds of social participation by 7% (OR= 0.93; 95%CI= 0.87,0.99) (figure 3, cXX). Similarly, reducing two number of teeth categories at each time point was associated with 10% (OR= 0.9 & 95%CI= 0.78,1.02) reduction in the odds of social participation (figure 3, cXX). Furthermore, a significant positive effect on social participation (OR= 1.07 & 95%CI= 1.04,1.1) was found if all the individuals in our observed sample improved their dental status to have >=20 teeth at baseline and in 2013 (i.e., compared to their actual dental status) (figure 3, cXX). Additive treatment effects for all compared hypothetical interventions are reported in table 2. The highest ATE of negative 5% (95% CI= -7.5%, -2.3%) was observed when continuous edentulousness was compared against continuously having >=20 teeth.

Table 2: Results of TMLE analyses

| Contrast of TMLE estimates | ATE | (95% CI) | OR | (95%CI) |
| --- | --- | --- | --- | --- |
| >=20 teeth vs 10-19 teeth | -0.019 | (-0.039, 0.001) | 0.926 | (0.847, 1.006) |
| >=20 teeth vs 1-9 teeth | -0.031 | (-0.053, -0.009) | 0.883 | (0.797, 0.969) |
| >=20 teeth vs Edentulous | -0.049 | (-0.075, -0.023) | 0.822 | (0.719, 0.925) |
| Observed vs Edentulous | -0.033 | (-0.057, -0.009) | 0.877 | (0.783, 0.97) |
| Observed vs >=20 teeth | 0.016 | (0.009, 0.024) | 1.067 | (1.036, 1.097) |
| Observed vs one category down | -0.018 | (-0.034, -0.002) | 0.931 | (0.868, 0.994) |
| Observed vs two categories down | -0.026 | (-0.056, 0.003) | 0.901 | (0.783, 1.018) |
| 10-19 teeth vs 1-9 teeth | -0.012 | (-0.038, 0.015) | 0.953 | (0.848, 1.059) |
| 10-19 teeth vs Edentulous | -0.030 | (-0.06, 0) | 0.887 | (0.768, 1.007) |

Results of the complete case analysis were consistent with the pattern of the main results. However, the 95% CIs were wider than in the results of main analysis (Appendix xx Figure xx). Similarly, the analysis which was conducted without SuperLearner obtained consistent results with slightly wider 95% CIs (Appendix xx Figure xx).

### References

1. Chen T, Guestrin C. 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA: Association for Computing Machinery.
2. Kondo K, Rosenberg M, WHO. 2018. Advancing universal health coverage through knowledge translation for healthy ageing: Lessons learnt from the japan gerontological evaluation study. World Health Organization.
3. Laan MJ van der, Gruber S. 2012. Targeted minimum loss based estimation of causal effects of multiple time point interventions. The international journal of biostatistics. 8(1).
4. Laan MJ van der, Polley EC, Hubbard AE. 2007. Super learner. Statistical Applications in Genetics and Molecular Biology. 6(1).
5. Rose S, Rizopoulos D. 2019. Machine learning for causal inference in biostatistics. Biostatistics. 21(2):336–338.
6. Rubin DB. 2004. Multiple imputation for nonresponse in surveys. John Wiley & Sons.
7. Schomaker M, Luque-Fernandez MA, Leroy V, Davies MA. 2019. Using longitudinal targeted maximum likelihood estimation in complex settings with dynamic interventions. Statistics in Medicine. 38(24):4888–4911.
8. Schuler MS, Rose S. 2016. Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies. American Journal of Epidemiology. 185(1):65–73.
9. Van Der Laan MJ, Rubin D. 2006. Targeted maximum likelihood learning. The international journal of biostatistics. 2(1).
10. Venables WN, Ripley BD. 2002. Modern applied statistics with s. Fourth. New York: Springer.
11. Williams NT, Díaz I. 2020. Lmtp: Non-parametric Causal Effects of Feasible Interventions Based on Modified Treatment Policies.