

Over-the-counter bite splints. A randomized controlled trial of compliance and efficacy

Running title: OTC appliance efficacy and compliance

Geoffrey Gerstner^{1*}, Wei Yao², Krishnapriya Siripurapu¹, Hadel Aljanabi¹, Ann Decker³, David Ludkin³, Rachel Sinacola³, Katherine Frimenko², Kathryn Callaghan¹, Sean Penoyer¹, Claire Tewksbury¹

¹Department of Biologic and Materials Sciences and Prosthodontics, University of Michigan School of Dentistry, USA

²University of Michigan School of Dentistry, USA

³Department of Periodontics and Oral Medicine, University of Michigan School of Dentistry, USA

***Correspondence:** Department of Biologic and Materials Sciences and Prosthodontics, 1011 N. University Ave. School of Dentistry, Ann Arbor, MI 48109-1078, 734-647-3069 (office); 734-763-3453 (FAX), geger@umich.edu

Acknowledgments

We wish to thank Rajpreet Kaur, Aekta Patel, Abdulkareem Alshehri, Ashley Muehleise, Grace Smith, Vinnie Gupta, Hanna Kwon, Christina Bender, Siddharth Acharya, Kathryn Thompson, and Hina Jaffer for help with data collection and processing. Appreciation goes to Jeffrey Stanley, MD, for consulting assistance with medications. This study was supported by a grant from Delta Dental Foundation of Michigan (G014798) to G Gerstner.

This is the author manuscript accepted for publication and has 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: [10.1002/cre2.315](https://doi.org/10.1002/cre2.315)

Conflict of Interests

The authors have no conflicts to report.

Abstract

Background: Occlusal splints are often used to curb the impacts of sleep bruxism (SB) on the dentition, and over-the-counter (OTC) options are becoming increasingly popular. OTC splints are usually fabricated at home by patients, but not routinely evaluated by dental professionals. It is unclear how OTC splints compare with more traditional splints that receive dental oversight.

Objectives: The present randomized controlled study tested how an OTC splint compared with a gold standard bite splint in terms of patient compliance (primary outcome) and efficacy (secondary outcomes).

Methods: Sixty-seven subjects were randomly assigned to receive either the OTC (SOVA, N=35) splint or the gold standard “Michigan” bite splint (MI, N=32), with 61 completing the study (SOVA, N=30; MI, N=31). OTC-splint subjects were required to fabricate their splints to clinically acceptable standards. Both groups wore the splints nightly for four months. Compliance was measured via daily diary. Efficacy outcomes evaluated stability, retention, periodontal health, night-time rhythmic masticatory muscle activity (RMMA), and material wear.

Results: OTC-splint subjects had difficulty fabricating splints to clinically acceptable standards. The number of night-time RMMA bursts was significantly greater for the OTC splint group. Compliance and all other efficacy measurements were not significantly different between-groups.

Conclusion: The results support the potential use of OTC splints for curbing the impacts of SB. However, the results strongly suggest that dentists should be actively engaged in overseeing patients’ use of self-fabricated appliances. This clinical trial is registered at ClinicalTrials.gov, Identifier number NCT02340663.

Key words: sleep bruxism, occlusal splints, treatment adherence, clinical efficacy

Author Manuscript

INTRODUCTION

Bite splints or occlusal appliances are often used to reduce tooth wear caused by sleep bruxism (SB), clenching and grinding ¹. They serve an important purpose, as tooth wear is a prevalent condition, which increases with age ¹. Terminating splint use can cause both temporomandibular disorder (TMD) and sleep bruxism (SB) symptom exacerbation ².

Three commonly recognized appliances exist ³: (1) an over-the-counter (OTC), non-adjustable type, (2) an OTC, intra-orally-formed or 'boil-and-bite' type, and (3) custom appliances. OTC devices are typically used without dental supervision, whereas the custom appliances involve dental intervention. Studies suggest that custom hard acrylic or boil-and-bite splints have nearly equal efficacy ⁴; however, evidence suggests that hard acrylic splints are superior to soft or repositioning appliances in managing TMD pain ⁵.

Custom splints are usually expensive, and replacements are not typically covered by insurance. This is problematic, as splints can wear out with time ⁶, despite long-term splint use being necessary in bruxers ². OTC splints are becoming more popular, as they are inexpensive and convenient. However, few studies have evaluated OTC appliance efficacy and compliance. This is important, as it is unclear whether private practitioners evaluate OTC appliances routinely with their patients. This study's objective and purpose was to compare a specific custom appliance, the "Michigan bite splint", to an OTC boil-and-bite appliance for compliance and measures of efficacy. By necessity, this was a short-term trial; however, we attempted to evaluate many issues in that time interval. This is an FDA-registered RCT, publically visible at: <https://clinicaltrials.gov/ct2/show/NCT02340663?cond=bruxism&draw=4&rank=38>).

METHODS

Participants

This randomized controlled study was approved by the University of Michigan Medical IRB (HUM00085489). Figure 1A shows recruitment and retention numbers. Subjects were recruited via announcements posted throughout the University of Michigan School of Dentistry. Screening occurred in the principal investigator's (PI, author GEG) laboratory between 21 July 2015 and 2 Sept 2016 and involved 103 candidates, 36 of whom did not meet inclusion-exclusion criteria, decided not to participate, or did not return calls. Subject losses occurred after randomized allocation (3 subjects) or between wk 1 and the study's conclusion (3 subjects). Thus, 30 SOVA and 31 MI subjects completed the study. This was considered sufficient, based upon a power analysis using SB data in ⁷. Using mean bruxing episodes per hour with an occlusal splint=3.97 vs. with a palatal splint=4.45, a SD=0.63, an $\alpha=0.05$ and a $\beta=0.8$, we concluded a study would be sufficiently powered with 28 subjects group⁻¹.

Inclusion criteria were: (1) ≥ 18 years old, (2) clinical signs of dental wear, (3) self-report of nocturnal tooth grinding noises, (4) self-report of a bruxing diagnosis by a family dentist, (5) absence of dental and medical conditions, including periodontal disease, cardiovascular disease, sleep apnea, movement, neurologic and sleep disorders, (6) full dentition sans 3rd molars, (7) no active orthodontics nor removable prostheses, (8) no medications known to have movement disorder or sleep disturbance side effects, (9) ability to follow instructions, (10) no jaw function limitations, and (11) ability to report to the clinical laboratory at appointed times. Presence of TM joint noises or myalgia was permitted; however, joint arthritides were not.

Consented subjects (App 1, Fig. 1B) were randomly assigned to either an OTC 'boil-and-bite' appliance group (SOVA, Akervall Technologies, Saline, MI, Fig. 2A), or an acrylic occlusal appliance group, Michigan (MI) appliance (Fig. 2B) using a stratified randomization procedure ⁸, run in Excel (version 2010, Microsoft) with custom algorithms created by the PI. Groups were matched for gender and presence/absence of TMD signs/symptoms using the Diagnostic Criteria for TMD (TMD-RDC) ⁹, with the TMD-RDC performed by the PI or one other clinician, both of whom had extensive calibrated training in its use. Other instruments used included the Jaw Function Limitation Scale (JFLS) ¹⁰, TMD Pain Screener ¹¹, Measure of Symptoms Sleep Scale (MOS) ¹², Perceived Stress Scale (PSS) ¹³, and Oral Behaviors Checklist (OBC) ¹⁴. All clinical components of the study were performed either in the Implant Clinic or the PI's clinical research laboratory, both in the University of Michigan's School of Dentistry.

Baseline intra- and extra-oral exams performed by the PI evaluated dental and medical health. Alginate impressions were taken and poured in dental stone. The stone models were used to fabricate the MI appliance or to assess SOVA appliance fabrication (below).

Many of the methods described, below, appear in summary form in Table 1.

Bruxism grading

We used the international consensus criteria for grading bruxism in subjects ¹⁵. Question 1 of the OBC was used to define self-reported bruxism. Bruxing signs were clinically evaluated, and confirmed on mandibular stereolithography (stl) models (Fig. 3A) (True Definition Scanner, 3M, St. Paul, MN, USA). Scores of no wear (0), wear into enamel (1), and wear into dentin (2) were assigned to each tooth, based upon severest detected wear on each tooth, and subjects

were scored using the median value for the mandible. Scoring and grading were done by an investigator blinded to group assignments.

Splint delivery

SOVA subjects fabricated SOVA devices in the clinic during appointment 2 (Figure 1B) using manufacturer instructions. If they felt compelled to call the company's help line, they instead asked the PI, who was versed in the company's help line scripts. Otherwise, subjects fabricated splints without oversight. Once fabricated, splints were assessed for critical errors (see footnote, Table 1, for description of error categories). Critical errors required subjects to refabricate splints to acceptable standards. Total refabrications, total time spent fabricating, and types of critical errors were scored.

The MI appliance was fabricated in a professional dental laboratory routinely used by the UM School of Dentistry for fabrication of bite splints. The MI appliance was fabricated by the lab between appointments 1 and 2 (Figure 1B). At delivery, MI splints were adjusted for fit and to establish bilateral occlusal contacts.

Compliance (Primary Outcome)

Daily diaries were filled out at home over the 4-mo study period. Instructions stressed that the diaries be filled out daily. Subjects were instructed to account accurately for nights the splints were not worn. Compliance was defined by the total number of nights of splint wear and the percent of total nights worn (Table 1). Diaries were assessed at wk 1 to determine whether subjects were correctly reporting compliance and to re-educate as necessary. The null hypothesis was that no significant differences in number of nights of splint wear nor in percent of nights worn would exist between groups.

Surveys

SOVA subjects completed a survey covering ease of fabrication. At study's end, all subjects completed a survey on user satisfaction. Standard surveys used at appointments 1 and 4 to assess changes over time included the Oral Health Impact Profile (OHIP) ¹⁶, Tampa Scale for Kinesiophobia for TMD (TSK) ¹⁷, and TMD Pain Screener (PS) ¹¹. No hypothesis was formalized for ease of fabrication; however, we anticipated that SOVA subjects would report that splints were easy to fabricate. Null hypotheses for other surveys were: (1) User satisfaction with splints will not be significantly different between subject groups. (2) There will be no significant between-group differences in responses on the three standard surveys at baseline nor at study's end, nor will there be differences in the survey responses through time.

Efficacy (Secondary Outcomes)

Efficacy was defined by stability, retention, periodontal health, and estimated RMMA (Table 1). Stability was tested at appointment 4 (Fig. 1B) to assess splint dislodgement on the maxilla. Five tasks were performed in the following order (Fig. 4A): clench; grind to the left, right and anteriorly; move the jaw to the left, forward, around to the right, and back to a biting position (Border); tap; open wide. Each task was performed five times in a row (trial, Fig. 4A), and two such trials were performed, resulting in ten observations per task. A >10-s rest period occurred between tasks. The null hypothesis was that there would be no significant between-group differences in splint displacements (measured in mm in three-dimensional space) caused by the tasks.

Stability was monitored using a magnetometer motion analysis system and 1.8 micro sensors attached to the chin and splint (Fig. 4B, Liberty, Polhemus, Colchester, Vermont, USA).

If the splint was dislodged during tasks, the splint sensor recorded this in three dimensions. The maximum change in this distance per task was used to assess stability.

Subjects were asked to bite hard during tasks, and bite force estimates were made to confirm this. Prior to performing tasks, subjects bit on a custom bite plate (Figure 4C) connected to a force sensor (Kistler 9203 force sensor \pm 500 N, Kistler 5010 Dual Mode Amplifier); the design was based on ¹⁸, with modifications to standardize the bite position to the first molar region. While subjects bit on the transducer, surface electromyography (EMG) was sampled from the masseter muscles bilaterally at 10Khz and filtered with a 20 – 500 Hz bandpass filter (Ag/AgCl electrodes, MVAP-II, MVAP Medical Supplies, Newbury Park, CA; Powerlab 8/35, Octal Bio Amp, and LabChart 8, AD Instruments, Colorado Springs, CO).

Subjects bit with mild, medium, and maximum forces, twice with the splint in and twice with the splint out. The bite plate thickness was standardized to 20 mm. A 10-s rest period occurred between bites. An example of one such trial is shown in Figure 4D.

The filtered right masseter EMG bursts were used to extract root mean square (RMS), the mean and median power frequencies, and peak amplitude, expressed as a percent of the maximum peak amplitude (LabChart 8, AD Instruments, Colorado Springs, CO). A step-wise linear regression, with a procedure to eliminate variables expressing high collinearity, was used to create an equation to estimate bite forces. The first 'splint-in' bite force trial was used as the training set; the remaining three trials served as test sets.

Test-set results (Fig. 4E) were evaluated for precision and accuracy with Lin's concordance coefficient, ρ_c ¹⁹ and the bias correction factor, C_B ²⁰, respectively. Results indicated moderately strong precision ($\rho_c=0.9013$, 95% CI=0.89 – 0.91) and high accuracy

$C_B=0.9979$), suggesting that averaging bite force estimates would improve precision; hence, mean estimates were used in analyses.

Retention was assessed during appointments 3 and 4 (Fig. 1B, Table 1) via self-report while subjects performed eight tasks: forced blowing; lip and cheek shearing movements to attempt to unseat the splint; coughing; yawning; smiling; swallowing; sucking and puckering; maximum openings. Each task was repeated five times. Between tasks, subjects bit gently to determine whether splint reseating occurred, in which cases subjects reported a positive score. Total positive observations were tallied (maximum score=40, 8 tasks x 5 replicates per task). The null hypothesis was that no significant between-group differences would exist in number of reported splint reseatings following tasks at either appointment 3 or 4.

Tissue health was assessed using the Rustogi modification of the Navy plaque index (PI) ²¹ and the modified gingival index (MGI) ^{22,23} (Fig. 5A, Table 1). Assessments were done by calibrated periodontal residents in the UM School of Dentistry, who were blinded to group assignments. Each resident assessed equal numbers of subjects from each group to minimize scoring biases. Periodontal data were taken at appointments 2, 3 and 4 (Fig. 1B). PI and MGI scores were calculated separately for upper and lower arches and also for the buccal and lingual of the upper and lower arches. The null hypothesis was that there would be no significant between-group differences in PI or MGI, nor would there be time-dependent effects.

Finally, RMMA was estimated (BioRadio Recording Unit, Great Lakes NeuroTechnologies, Valley View, OH) in a home sleep study performed at appointment 4 (Fig. 1B, Table 1). Bipolar EMG electrodes (Ag/AgCl, MVAP-II, MVAP Medical Supplies, Newbury Park, CA) were placed bilaterally on the masseter and thyroideus muscles (Fig. 6 A,B). A

ground electrode was placed on the mastoid process opposite the subject's sleeping side preference (Fig. 6A). Body movements and positions were also sampled. Audio was used to assist with identifying bruxing events and to identify other nocturnal noises, e.g., coughing, talking, etc. (Fig. 6C). The monitor was worn on an upper arm of the subject's choosing.

Data were recorded at 1kHz, filtered (EMG, 20 – 500 Hz bandpass; body movements, 20 Hz low pass), and evaluated in 10-s epochs by investigators trained and calibrated to analyze recordings using published criteria²⁴. Scorers were blinded to subject group assignment, and they received equal numbers of subjects from each group to reduce rater bias.

Gold standard evaluation of SB uses polysomnography (PSG) and audio-video capture¹⁵; however, home monitoring is an acceptable alternative²⁵, despite moderate false positive rates²⁶. Also, RMMA is not necessarily SB, and without EEG and EOG leads, recording time is not differentiable from sleep time. Thus, this study, *sensu stricto*, reports RMMA during total recording time; however, to reduce the false positive rate, three trained investigators used the monitor at home and performed the following tasks while in bed.

Oral behavior artifacts: swallow; lick lips; speak; moan; snore; cough; yawn; open wide; laugh; RMMA without tooth contact; rapid light tooth tapping. Each task was repeated five times.

Body movement artifacts: rise on elbows; scratch arms; reposition legs; move legs rhythmically; move from prone to left and right sides, and from supine to left and right sides; move head from left to right cheek and reverse while prone. Each task was repeated five times alone, then in conjunction with bruxing-like RMMA, below.

Bruxing-like RMMA: clench for 3 s (tonic clench); clench rhythmically five times at 1 Hz (phasic clench); grind rhythmically five times at 1 Hz to the right and to the left (phasic grind).

Each RMMA was repeated five times without and then with the body movement artifacts.

Investigators were trained to avoid movements other than those called for by each task.

Recorded data were processed and scored similar to data obtained from subjects. All scored observations were matched to their true task categories for cross-validation.

Scored EMG bursts from each investigators' data set were subjected to a step-wise discriminant analysis. Each investigator's resulting functions were cross-validated on the data from the other two investigators. We determined that true positive and negative rates could be reduced if we required agreement from two raters' equations; this generated a true positive rate of 88% and true negative rate of 91%. Thus, positive categorization by at least two raters' equations was used to define estimated RMMA in subject data. The null hypothesis was that no between-group differences would exist in RMMA, tested as both number of EMG bursts and as number of EMG episodes.

To assess splint surface wear, polyvinyl-siloxane (PVS) impressions of the splint were obtained during appointments 2 and 4. Impressions were poured in stone (Diamond Die Scan Stone, Hi-Tech Dental Products, Greenback, TN) and digitized (3M True Definition Scanner). Mesh files (Figure 3B) were uploaded into 3D Slicer (Version 4.8.1 www.slicer.org,²⁷) and registered (CMFreq extension of 3D Slicer). Two regions of interest (ROI) were registered, one around 1st molar-splint contacts, and one around canine-splint contacts (Fig. 3B, rectangles). The left side was analyzed unless clearer signs of wear occurred on the right side.

The Model-to-Model Distance extension in 3D Slicer was used to calculate the signed closest point distances between the aligned surfaces of the baseline and 4-mo-old stl models. The average distance (in mm) was calculated at two specific sites within each ROI, one site

where tooth-splint contact occurred (contact sites) and an adjacent site where no tooth-splint contact could occur (control sites) (Fig. 3B). Means for control and test sites were obtained using areas with 5-voxel radii (Pick 'n Paint extension in 3D Slicer). Mean distances between splint surfaces were calculated (Mesh Statistics extension in 3D Slicer). Estimated wear, w , was calculated as $w = d_{\text{control}} - d_{\text{test}}$, where d_{test} = mean between-model distance at the tooth contact site, and d_{control} = mean between-model distance at the control site (Table 1). The null hypothesis was that no differences in splint material wear will exist between subject groups.

Statistical tests

Continuous data were evaluated for normality using Q-Q plots, and tests for skewness and kurtosis. If normality assumptions were violated, data were transformed. If normality was not achieved through transformation, we used non-parametric tests. General linear models (GLM) were used for tests of within- and between-group differences. A repeated measures design was used for data collected multiple times during the study. Pearson's product moment was used on normally-distributed data.

For non-normally-distributed data and non-continuous data, the Mann-Whitney U test was used for comparisons between groups. For tests involving more than two categories, Kruskal-Wallis or Friedman tests were used. The Wilcoxon or Friedman test was used for paired tests. For correlation analyses of non-continuous data, Spearman's rank correlation test was used.

Because the study's aim was to determine whether the SOVA splint was similar to the MI splint, we employed a liberal, non-corrected $P < 0.05$, except where GLM *post-hoc* tests were performed, in which case, a Bonferroni correction was used. Effect sizes and 95% CI were

calculated for primary and secondary outcomes (www.psychmetrica.de/effect_size.html, www.effect-size-calculator.herokuapp.com/#form4).

RESULTS

Table 2 shows initial and final enrollment demographics. There were no significant differences between the two groups. No between-group differences existed in self-report of SB (MWU=417, P=0.462). All subjects reported current SB noises and showed clinical signs of wear (5 had scores of 1; 5 had scores of 1.5; 51 had scores of 2). Twenty-five subjects (11 SOVA, 14 MI) reported bruxing 4 – 7 nights / wk, 26 subjects (12 SOVA, 14 MI) reported bruxing 1 – 3 nights / wk, and nine subjects (6 SOVA, 3 MI) reported 1 – 3 nights per month. (Data for one remaining SOVA subject spanned multiple categories.)

No significant differences in tooth wear existed between groups (MWU=415, P=0.537). Sixty subjects had wear facets on all posterior teeth; one remaining case (SOVA) had wear facets into enamel on all molars. A trend existed between clinical wear assessment and bruxing habit self-report scores (Spearman's $\rho=0.248$, $p=0.059$, $df=59$).

No significant between-group differences existed for TMD by category (Table 3). Also, there were no significant between-group differences in overbite, overjet, maximum pain-free opening, maximum voluntary opening, maximum opening with passive stretch, maximum protrusion or maximum left or right laterotrusions (Table 3).

Baseline results for JFLS ¹⁰, PSS ¹³, OBC ¹⁴, and MOS ¹² revealed no significant between-group differences (Table 3). Three surveys, the OHIP (broken into subcategorized as per ¹⁶), TSK ¹⁷, and TMDPS ¹¹ were filled out at appointments 1 and 4 (Initial, End, respectively,

in Table 3). There were no differences between groups, nor through time, nor were there interactive effects in survey results.

Twenty SOVA subjects (66.7%) fabricated splints without asking for help, eight (26.7%) asked for help once, and two (6.7%) asked twice. Twenty-eight (93.3%) subjects remolded splints prior to clinical inspection. Mean (SD) fabrication time was 14.0 (11.6) minutes.

Twenty-seven (90%) splints had critical errors. Poor splint alignment (Fig. 2C) on the occlusal table occurred in 26 (86.7%) splints. Other errors included over-stretched material (N=8, 26.7%, Fig. 2D), inadequate or excessive posterior coverage on the posterior-most molars (N=19, 63.3%, Fig. 2E), or excessive mandibular occlusal indentations (N=6, 20%, Fig. 2F).

Table 4 shows results for user satisfaction, composed of four items based upon a five-point Likert scale; 1=Strongly Disagree and 5=Strongly Agree. No significant differences existed between groups.

Table 4 also shows stability test results. An example of a stability test is shown in Fig. 4A. The table reports maximum dislodgements in mm by task, expressed as group means and 1 SD. Movement data for one SOVA subject was corrupted and not included. No significant between-group differences existed. Opening wide produced the greatest dislodgement for both groups. Rest resulted in minimal dislodgement. No significant between-group differences occurred (see also Table 4 footnotes for further results).

Bite force estimates during stability tasks are shown in Figure 4 F,G. Horizontal error bars show significance at the $P < 0.05$ level. Estimated bite forces were higher among SOVA than MI subjects ($F[1,118.6]=12.604$, $P=0.001$, partial $\eta^2=0.096$, 95% CI=0.019 – 0.20; Fig. 4F).

Clenching produced the highest forces, whereas border and grinding produced lower forces for both groups (Fig. 4G). No interactive effects were found.

Retention results, Table 4, are the number of dislodgements (column 1) reported by number of subjects, N, (columns 2,3) . The maximum number of dislodgements that could have occurred=40, but no splint had >6 dislodgements. No significant between-group differences existed at wk 1 or mo 4, and results did not change significantly over the 4-month study (Wk 1 vs. Mo 4 in table).

Finally, Table 4 shows estimated RMMA by group. Variables are defined using nomenclature from ²⁴. Three MI and 5 SOVA subjects' data were not analyzed due to recording errors. Total recording time (mean, SD) for both groups was 6.3 (1.6) hr for MI, 6.5 (1.4) hr for SOVA, $t=-0.4$, $P=0.70$. Blinded raters participated in data analysis; rater effects were not significant (Wilk's lambda=0.582, $F[15,33]=1.583$, $P=0.113$, partial $\eta^2=0.418$, 95% CI=0-0.414). A MANOVA indicated no significant effects due to splint group (Wilk's lambda=0.688, $F[15,33]=0.997$, $P=0.481$, partial $\eta^2=0.312$, 95% CI not calculable). However, note that SOVA had significantly more Bh^{-1} and more BE^{-1} than MI. There were no significant differences for other SB variables between groups.

Figure 7 shows compliance results, both as the number of nights worn (A) and as percent of total days the splints were possessed (B). There were no significant differences in compliance (for number of nights worn, $t=1.201$, $df=62$, $P=0.233$, $d_{\text{cochran}}=0.308$, 95% CI=-0.20 – 0.81; for percent nights worn, $t=0.602$, $df=62$, $P=0.549$, $d_{\text{cochran}}=0.154$, 95% CI=-0.34 – 0.66). Note that these analyses included data for subjects that dropped out before appointment 4, because we deemed compliance to begin with splint delivery. Reasons for reduced compliance

included forgetting to wear the splint, misplacing splint, and/or forgetting to take the splint on vacations (7 SOVA; 7 MI). Issues with splint acceptance, e.g., changes to the bite or splint being too tight, represented the remaining reasons (2 SOVA; 8 MI).

Figure 5 shows results for MGI (B) and PI (C), at baseline, wk-1 and mo-4. MI data are filled bars, SOVA data are open bars. No significant differences existed between subject groups for either index, (for MGI, $F[1,53]=1.213$, $P=0.276$, partial $\eta^2=0.022$, 95% CI=0–0.15; for PI, $F[1,53]=0.426$, $P=0.517$ partial $\eta^2=0.008$, 95% CI undefined). There were significant differences across raters (for MGI, $F[3,53]=163.7$, $P<0.001$, partial $\eta^2=0.903$, 95% CI=0.84 – 0.93; for PI, $F[3,53]=2.943$, $P=0.041$, partial $\eta^2=0.143$, 95% CI=0-0.28), but there were no interactions between rater and subject group (for MGI, $F[3,53]=1.037$, $P=0.384$, partial $\eta^2=0.055$, 95% CI=0 – 0.16; for PI, $F[3,53]=0.904$, $P=0.446$, partial $\eta^2=0.048$, 95% CI undefined).

Because all splints were maxillary, we hypothesized splints might impact maxillary periodontal health. Therefore, we evaluated within-subject between-arch effects. For MGI scores, there were no significant differences across the four areas, viz., BU, LU, BL, LL, ($F[3,53]=2.516$, $P=0.060$, partial $\eta^2=0.125$, 95% CI=0 – 0.26), nor did any of the four areas show significant between-group differences ($F[3,53]=0.309$, $P=0.819$, partial $\eta^2=0.005$, 95% CI undefined). There were no significant changes over the 4 mo time period in any of the four areas between groups ($F[3,174]=2.073$, $P=0.106$, partial $\eta^2=0.035$, 95% CI=0 – 0.09).

However, PI scores differed significantly within subjects across the four areas ($F[3,174]=17.15$, $P<0.005$, partial $\eta^2=0.228$, 95% CI=0.12 – 0.32). The PI was significantly lower for LU than for the other three areas ($P<0.05$). Although the full model suggested significant differences between treatment groups by area ($F[3,174]=3.759$, $P=0.016$, partial $\eta^2=0.061$, 95%

CI=0.003 – 0.13), pairwise comparisons did not reveal any significant differences ($P>0.098$). No significant differences existed between groups by area through time ($F[3,174]=0.587$, $P=0.589$, partial $\eta^2=0.010$, 95% CI undefined). Finally, no significant between-group differences existed at baseline (appointment 2) for measures of either MGI ($U= 447$, $P=0.799$, $d_{\text{Cohen}}=0.067$, 95% CI=-0.44 – 0.57) or PI ($U=433.5$, $P=0.654$, $d_{\text{Cohen}}=0.117$, 95% CI=-0.39 – 0.62).

Figure 3B shows areas where splint wear was sampled; the means (1SD) wear, in mm, on the splints over the 4-mo time period are shown in Figure 3C. There were no significant differences between groups ($F[1,57]=2.173$, $P=0.146$, partial $\eta^2=0.037$, 95% CI=0 – 0.17).

Finally, subjects were allowed to write open-ended responses to the question, “What do you like best about your splint?” The main responses among MI splint subjects had to do with: (1) fit and comfort ($N=19/31$), (2) protection of teeth from bruxism ($N=9/31$), and (3) help with muscle relaxation ($N=3/31$). The main responses among SOVA subjects had to do with: (1) fit and comfort ($N=20/30$), (2) the splint’s small size ($N=7/30$), (3) protection of teeth from bruxism ($N=3/30$), (4) help with muscle relaxation ($N=2/30$), and (5) affordability ($N=1/30$).

DISCUSSION

Ideally, participants randomly assigned to groups are well-matched in RCT. We carefully matched groups based on gender and presence/absence of TMD signs/symptoms. Fortuitously, the groups were also closely matched across a number of other factors, including ethnicity, clinically-observed tooth wear severity, self-reported bruxing nights / wk, and data from several surveys including the JFLS-20, OBC, PSS, MOS, OHIP, TSK for TMD. There were also no significant between-group differences for more detailed TMD findings as well (Table 3).

Perhaps the most significant finding was the inability of subjects to form the OTC splint according to instructions. Only 4 / 31 SOVA splints were clinically acceptable. All subjects were ultimately helped to fabricate clinically acceptable SOVA splints, but this would not occur routinely. It is highly likely that virtually any OTC appliances currently in use are being improperly fabricated. We strongly recommend that dental professionals play pro-active, engaged roles with their patients who possess OTC appliances. It is noteworthy that, based on our findings, the SOVA splint is now available only through dentists and not available OTC.

This study was prompted by the fact that insurers often cover one splint per lifetime; however, rarely do splints last a lifetime. Because severe bruxers are more likely to require replacements, the unfortunate consequence is that patients who need the benefits the most stand to pay the most out-of-pocket. This suggests why OTC devices are becoming increasingly popular.

Clinical studies of obstructive sleep apnea (OSA) appliances emphasize the need to evaluate compliance and efficacy in terms of mean disease alleviation²⁸ or effectiveness²⁹. One treatment may be more efficacious but have poorer compliance than another treatment.

But what defines efficacy in a SB oral appliance? The etiology and pathophysiology of SB are multifactorial and complex^{30,31}. Furthermore, oral appliances do not inhibit SB^{32,33}. In this study, we operationally defined efficacy in terms of retention and stability, periodontal health, nocturnal EMG architecture, and structural integrity, i.e., material wear. Periodontal problems were considered, due to the perforated design of the SOVA appliance.

The results found that efficacy was similar between the two devices in terms of stability, retention, periodontal health, and splint surface wear. On the other hand, there were

significantly more Bh^{-1} and BE^{-1} in the SOVA group compared with the MI group, suggesting that the SOVA splint may exacerbate SB. Whether this is clinically significant remains unclear.

There are shortcomings with performing SB sleep studies with only EMG monitoring, cf. ³⁴⁻³⁶. On the other hand, EMG monitoring of SB can be reliable ³⁷⁻⁴¹. Given the precedent established in the literature, we would argue that the EMG results were worth reporting, but with emphasis on cautionary interpretation. Although we used audio as a means to improve our assessments, we found that, without video, sounds often remained ambiguous.

Of particular interest is the variable Bh^{-1} , which was significantly different between groups, with mean (SD) of 18.8 (SD=17.3) vs. 31.2 (SD=4.24), for the MI vs. SOVA group, respectively (Table 2). Doering, et al. report 34.2 (SD=10.6) Bh^{-1} in a SB population ⁴², which matches closely with our results. However, our results for Eh^{-1} (Table 2) was considerably higher than reported in either the Doering, et al. study (5.6 (SD=1.3)) or Mayer, et al., viz., 2-4 for mild to moderate SB and >4 for severe SB ⁴³. This suggests that our Eh^{-1} values are over-estimates as per ³⁴. Whatever the case, our results suggest that something importantly different in SB architecture associated with SOVA vs. MI splint wear may be occurring. This might be partly due to the softer material and larger occlusal contact areas provided by the SOVA vs. MI splint. Further work needs to be done, ideally with PSG, to assess whether SOVA vs. MI splints differentially impact SB architecture.

We developed a few novel methods to help evaluate splints, including motion analysis of stability, estimating bite forces during stability tests, and assessment of material wear. Motion analysis has an established record in the oral motor literature ⁴⁴⁻⁴⁸, and so we are reasonably confident that our stability results are objective and accurate. Similarly, previous bite force

estimates using EMG data have used variables similar to ours ^{49,50}. Thus, the stability results are probably reasonable.

A digital method of evaluating splint wear was developed by Koriath, et al. ⁶ similar to ours. The wear seen in the previous study was also low. It is possible that the 4-mo time period does not provide sufficient time to identify wear, let alone estimate wear rates. Longer-term studies will likely be more revealing.

We measured compliance through self-report and found no significant between-group differences. Evidence suggests that self-report is a reasonable compliance measure, even when compared to embedded micro-sensor methods ²⁸. Based on patient satisfaction, there did not seem to be any differences between groups, and reasons for not wearing splints were also similar between groups. Thus, the similar scores for both satisfaction and compliance suggest no significant differences in splint preference.

Several study limitations existed. Firstly, the study was not double-blinded. This would have been difficult to do, given the distinct fabrication methods and appearances of each splint (Fig. 2). Obviously splint fabrication assessment could not be blinded either. However, single blinding was done for virtually all other study aspects, e.g., stability, retention, periodontal assessment, EMG analysis, compliance tabulation, and statistical analyses. Also, where multiple investigators were involved, we allotted equal numbers of subjects from each group to each investigator, thereby reducing the impact of investigator biases on results.

Another study limitation was our use of subjects with 'probable' as opposed to 'definite' diagnoses of SB ¹⁵. Given that dentists do not routinely obtain sleep studies on patients with SB, dentists usually treat patients with 'probable' SB diagnoses anyway, and in this respect our

subjects probably represent the 'typical' population treated by dentists for SB. We recognize that this limitation may be a reason for the lack of observed splint surface wear. On the other hand, results of the nocturnal EMG study suggest active SB habits in our subjects.

Other study limitations include the small sample sizes and the abbreviated time period over which the study was performed. Many of the effect sizes demonstrated fairly large confidence intervals, which is likely due to the small study size. Future, larger projects may consider cross-over designs, inclusion of PSG, longer-term splint wear, and use of compliance sensors, among other things.

In conclusion, because OTC splints were difficult to construct, we highly recommend that OTC splints be monitored by dentists. We recognize the need for inexpensive alternatives. Ideally, a large RCT would demonstrate definitely whether an OTC splint would be a legitimate solution. This would potentially provide an important measure of external validity. It is unlikely that such a large RCT will occur in the near future. Hence, a practical approach would be for dentists to be vigilant with OTC splint fabrication and use, since this appears to be a practical alternative for at least the near future.

REFERENCES

1. Cunha-Cruz J, Pashova H, Packard JD, Zhou L, Hilton TJ. Tooth wear: prevalence and associated factors in general practice patients. *Community Dent Oral Epidemiol.* 2010;38(3):228-234.
2. Rehm DDS, Mainieri VC, Saueressig AC, et al. Effects of the bite splint 15-day treatment termination in patients with temporomandibular disorder with a clinical history of sleep bruxism: a longitudinal single-cohort study. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology.* 2012;114(6):740-748.
3. Maeda Y, Kumamoto D, Yagi K, Ikebe K. Effectiveness and fabrication of mouthguards. *Dental Traumatology.* 2009;25:556-564.
4. Klasser GD, Greene CS, Lavigne GJ. Oral appliances and the management of sleep bruxism in adults: A century of clinical applications and search for mechanisms. *International Journal of Prosthodontics.* 2010;23:453-462.
5. Friction J, Look JO, Wright E, et al. Systematic review and meta-analysis of randomized controlled trials evaluating intraoral orthopedic appliances for temporomandibular disorders. *Journal of Orofacial Pain.* 2010;24:237-254.
6. Koriath TW, Bohlig KG, Anderson GC. Digital assessment of occlusal wear patterns on occlusal stabilization splints: a pilot study. *Journal of Prosthetic Dentistry.* 1998;80(2):209-213.
7. Huynh N, Manzini C, Rompré P, Lavigne GJ. Weighing the potential effectiveness of various treatments for sleep bruxism. *Journal of the Canadian Dental Association.* 2007;73(8):727-730b.
8. Suresh KP. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences.* 2011;4(1):8-11.
9. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. *Journal of Craniomandibular Disorders.* 1992;6(4):301-355.

10. Ohrbach R, Larsson P, List T. The jaw functional limitation scale: development, reliability, and validity of 8-item and 20-item versions. *Journal of Orofacial Pain*. 2008;22:219-230.
11. Gonzalez YM, Schiffman E, Gordon SM, et al. Development of a brief and effective temporomandibular disorder pain screening questionnaire: reliability and validity. *Journal of the American Dental Association*. 2011;142:1183-1191.
12. Spritzer KL, Hays RD. *MOS sleep scale: A manual for use and scoring, Version 1.0*. Los Angeles, CA2003.
13. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of Health and Social Behavior*. 1983;24:385-396.
14. Markiewicz MR, Ohrbach R, McCall WD. Oral behaviors checklist: Reliability of performance in targeted waking-state behaviors. *Journal of Orofacial Pain*. 2006;20(4):306-316.
15. Lobbezoo F, Ahlberg J, Glaros AG, et al. Bruxism defined and graded: an international consensus. *Journal of Oral Rehabilitation*. 2013;40(1):2-4.
16. Slade G, Spencer J. Development and evaluation of the Oral Health Impact Profile. *Community Dental Health*. 1994;11:3-11.
17. Visscher CM, Ohrbach R, van Wijk AJ, Wilkosz M, Naeije M. The Tampa Scale for Kinesiophobia for Temporomandibular Disorders (TSK-TMD). *Pain*. 2010;150(3):492-500.
18. Herrel A, Spithoven I, van Damme R, De Vree F. Sexual dimorphism of head size in *Gallotia galloti*: testing the niche divergence hypothesis by functional analyses. *Functional Ecology*. 1999;13:289-297.
19. Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255-268.
20. McBride GB. *A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient*. Hamilton, New Zealand: Ministry of Health;2005.
21. Rustogi KN, Curtis JP, Volpe AR, Kemp JH, McCool JJ, Korn LR. Refinement of the modified Navy plaque index to increase plaque scoring efficiency in gumline and interproximal tooth area. *Journal of Clinical Dentistry*. 1992;3(SupplC):C9-C12.

22. Lobene RR, Weatherford T, Ross NM, Lamm RA, Menaker L. A modified gingival index for use in clinical trials. *Clinical Preventive Dentistry*. 1986;8:3-6.
23. Loe H, Silness J. Periodontal disease in pregnancy (I). Prevalence and severity. *Acta Odontologica Scandinavica*. 1963;21:533-551.
24. Carra MC, Huynh N, Lavigne G. Sleep bruxism: A comprehensive overview for the dental clinician interested in sleep medicine. *Dental Clinics of North America* 2012;56:387-413.
25. Ahlberg K, Savolainen A, Paju S, et al. Bruxism and sleep efficiency measured at home with wireless devices. *Journal of Oral Rehabilitation*. 2008;35(8):567-571.
26. Carra MC, Huynh D, Lavigne GJ. Diagnostic accuracy of sleep bruxism scoring in absence of audio-video recording: a pilot study. *Sleep Breath*. 2015;19:183-190.
27. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*. 2012;30(9):1323-1341.
28. Vanderveken OM, Dieltjens M, Wouters K, De Backer WA, Van de Heyning PH, Braem MJ. Objective measurement of compliance during oral appliance therapy for sleep-disordered breathing. *Thorax*. 2013;68(1):91-96.
29. Sutherland K, Phillips CL, Cistulli PA. Efficacy versus effectiveness in the treatment of obstructive sleep apnea: CPAP and oral appliances. *Journal of Dental Sleep Medicine*. 2015;2(4):175-181.
30. Bader G, Lavigne G. Sleep bruxism; an overview of an oromandibular sleep movement disorder. *Sleep Medicine Reviews*. 2000;4(1):27-43.
31. Manfredini D, Winocur E, Guarda-Nardini L, Paesani D, Lobbezoo F. Epidemiology of bruxism in adults: A systematic review of the literature. *Journal of Orofacial Pain*. 2013;27(2):99-110.
32. Harada T, Ichiki R, Tsukiyama Y, Koyano K. The effect of oral splint devices on sleep bruxism: a 6-week observation with an ambulatory electromyographic recording device. *J Oral Rehabil*. 2006;33(7):482-488.
33. Jagger R. The effectiveness of occlusal splints for sleep bruxism. *Evidence-based dentistry*. 2008;9(1):23.

34. Manfredini D, Ahlberg J, Castroflorio T, Poggio CE, Guarda-Nardini L, Lobbezoo F. Diagnostic accuracy of portable instrumental devices to measure sleep bruxism: a systematic literature review of polysomnographic studies. *Journal of Oral Rehabilitation*. 2014;41(11):836-842.
35. Castroflorio T, Bargellini A, Rossini G, Cugliari G, Deregibus A, Manfredini D. Agreement between clinical and portable EMG/ECG diagnosis of sleep bruxism. *Journal of Oral Rehabilitation*. 2015;42(10):759-764.
36. Yachida W, Arima T, Castrillon EE, Baad-Hansen L, Ohata N, Svensson P. Diagnostic validity of self-reported measures of sleep bruxism using an ambulatory single-channel EMG device. *J Prosthodont Res*. 2016.
37. Minakuchi H, Sakaguchi C, Hara ES, et al. Multiple sleep bruxism data collected using a self-contained EMG detector/analyzer system in asymptomatic healthy subjects. *Sleep Breath*. 2012;16:1069-1072.
38. Castroflorio T, Deregibus A, Bargellini A, Debernardi C, Manfredini D. Detection of sleep bruxism: comparison between an electromyographic and electrocardiographic portable holter and polysomnography. *J Oral Rehabil*. 2014;41(3):163-169.
39. Gallo LM, Lavigne G, Rompre P, Palla S. Reliability of scoring EMG orofacial events: polysomnography compared with ambulatory recordings. *Journal of Sleep Research*. 1997;6(4):259-263.
40. Haketa T, Baba K, Akishige S, Fueki K, Kino K, Ohyama T. Utility and validity of a new EMG-based bruxism detection system. *Int J Prosthodont*. 2003;16(4):422-428.
41. Stuginski-Barbosa J, Porporatti AL, Costa YM, Svensson P, Conti PCR. Diagnostic validity of the use of a portable single-channel electromyography device for sleep bruxism. *Sleep and Breathing*. 2016;20(2):695-702.
42. Doering S, Boeckmann JA, Hugger S, Young P. Ambulatory polysomnography for the assessment of sleep bruxism. *Journal of Oral Rehabilitation*. 2008;35:572-576.
43. Mayer P, Heinzer R, Lavigne G. Sleep Bruxism in Respiratory Medicine Practice. *Chest*. 2016;149(1):262-271.
44. Tanaka Y, Yamada T, Maeda Y, Ikebe K. Markerless three-dimensional tracking of masticatory movement. *J Biomech*. 2016;49(3):442-449.

45. Wilson A, Luck P, Woods C, Foegeding EA, Morgenstern M. Comparison of jaw tracking by single video camera with 3D electromagnetic system. *Journal of Food Engineering*. 2016;190:22-33.
46. Gerstner GE, Lafia C, Lin D. Predicting masticatory jaw movements from chin movements using multivariate linear methods. *Journal of Biomechanics*. 2005;38(10):1991-1999.
47. Gerstner GE, Marchi F, Haerian H. Relationship between anteroposterior maxillomandibular morphology and masticatory jaw movement patterns. *American Journal of Orthodontics & Dentofacial Orthopedics*. 1999;115(3):258-266.
48. Gerstner GE, Parekh VV. Evidence of sex-specific differences in masticatory jaw movement patterns. *Journal of Dental Research*. 1997;76(3):796-806.
49. Gonzalez Y, Iwasaki LR, McCall Jr WD, Ohrbach R, Lozier E, Nickel JC. Reliability of EMG activity versus bite-force from human masticatory muscles. *European Journal of Oral Sciences*. 2011;119(3):219-224.
50. Van Eijden TMGJ, Brugman P, Weijs WA, Oosting J. Coactivation of jaw muscles: recruitment order and level as a function of bite force. *Journal of Biomechanics*. 1990;23(5):475-485.

Table 1. Summary of Methods†

Assessment	Variable	How acquired	When	Metric
Bruxism grading	1. Bruxism frequency over 1-month time period (nights / week) 2. Tooth wear	1. OBC, question 1 2. Intra-oral exam and inspection of mandibular stl model	Screening	Per Ref. 15, viz., 1. Ordinal Likert scale 2. Ordinal: 0 = no wear; 1 = wear into enamel; 2 = wear into dentine
Splint fabrication	± Fabrication errors	Examination of splint presented to PI as completed by subject	Appt 2	Categorical: Presence / Absence of an error; SOVA splints only
Compliance	1. Number of nights worn 2. % total nights worn	Daily diary	Appt 4	1. Numeric 2. Percent

Ease of fabrication	Responses from SOVA subjects	Questionnaire	Appt 4	Ordinal Likert scale
User satisfaction	Responses from all subjects	Questionnaire	Appt 4	Ordinal Likert scale
OHIP, TSK, PS	Responses from all subjects	Questionnaires	Appts 1, 4	See Refs. 11, 16, 17 Combinations of ordinal Likert scale and categorical data.
Stability	Displacement of splint on dentition	2 x 5 trials of each jaw-movement task, EMG and jaw movement sensors	Appt 4	Millimeters
Retention	Self-report of number of splint dislodgements	5 trials of 8 orofacial movement tasks, self-report	Appts 3, 4	Numeric, maximum of 40
Tissue health	1. Number of tooth surface areas with visible plaque 2. Severity of gingivitis	1. Intra-oral exam with disclosing solution 2. Intra-oral exam of marginal gingiva	Appts 2, 3, 4	Per Refs, 21-23. 1. Numeric ratio 2. Mean ordinal grading
RMMA	1. Number of EMG bursts / hr 2. Number of EMG sequences / hr	Home Sleep Test	Appt 4	Quantitative ratios
Splint surface wear	Change in surface	Software algorithms applied to stl files	Appts 2, 4	Difference between stl models (appt 4 minus appt 2) in millimeters

† Key to abbreviations: Appt, appointment; EMG, electromyography; OBC, Oral Behavior Checklist; OHIP, Oral Health Impact Profile; PI, principal investigator; PS, TMD Pain Screener; RMMA, rhythmic masticatory muscle activity; TSK, Tampa Scale for Kinesiophobia for TMD.

‡ Error categories included: (1) water bath temperature not correct, (2) splint not thoroughly warmed, (3) incisal bite not on anterior bar pad of splint blank, (4) lack of snugness against palate (> 1 mm gap), (5) lack of snugness against facial/buccal tooth surfaces (< 1 mm), (6) lack of sufficient material coverage on facials of anterior teeth, i.e., flange is short of gum line, (7) material folded over on itself, (8) marks from lower dentition excessive, (9) material overstretched, i.e., major axis $>$ twice length of minor axis of perforations, (10) maximum intercuspation not even in clench, (11) splint falls off cast when inverted or shaken, (12) material orientation issues, viz., asymmetry/rotation, translated laterally or anteroposteriorly on occlusal surfaces, (13) distal of posterior-most tooth not adequately covered, (14) posterior flange of splint extends onto soft tissues. Note: Because error categories were determined *a priori*, not all error categories were actually observed in the sample. Moreover, no additional error categories were observed or added *post hoc*.

Table 2. Demographics

Initial enrollment	SOVA	MI	Total
N	35	32	67
Age	25.5 (3.30)	25.5 (3.82)	25.5 (3.53)
Gender (F:M)	19:16	15:17	34:33
Ethnicity			
Amerind	0	1	1
Asian	7	5	12
Native Hawaiian; Pacific Islander	0	0	0
Black / African American	2	2	4
White / Caucasian	26	23	49
> 1 race	0	1	1
Unknown/unreported	0	0	0
Hispanic / Latin	3	1	4
Not Hispanic / Latin	0	0	0
Unknown / not reported	0	0	0
Final enrollment			
N	30	31	61
Age	25.0 (2.91)	25.1 (3.25)	25.0 (3.06)
Gender (F:M)	16:14	15:16	31:30
Ethnicity			
Amerind	0	1	1
Asian	6	4	10
Native Hawaiian; Pacific Islander	0	0	0
Black / African American	2	2	4
White / Caucasian	22	23	45
> 1 race	0	1	1
Unknown/unreported	0	0	0
Hispanic / Latin	2	1	3
Not Hispanic / Latin	0	0	0
Unknown / not reported	0	0	0

Age: Initial enrollment, P=0.932; final enrollment, P=0.903.

Gender: Initial enrollment, P=0.547; final enrollment P=0.702.

Ethnicity: Initial enrollment, P=0.545; final enrollment, P=0.302.

Table 3. TMD, periodontal, and survey results.

TMD	SOVA	MI	Statistical Test
	<u>N</u>	<u>N</u>	<u>MWU, P</u>
No symptoms	16	12	397, 0.26
Myalgia	10	5	416, 0.32
Headache due to myalgia	9	8	453, 0.82
Arthralgia	0	1	465, 1.0
DD with reduction	11	15	364, 0.083
DD with reduction, intermittent locking	0	2	435, 0.16
DD No reduction, limited opening	0	0	--
DD No reduction, no limited opening	0	0	--
	<u>Mean (SD)</u>	<u>Mean (SD)</u>	<u>F(1,59), P</u>
Pain-Free Opening	50.2 (6.4)	51.7 (8.8)	0.6, 0.4
Voluntary Opening	55.1 (5.6)	57.2 (7.7)	1.5, 0.2
Open with passive stretch	57.2 (5.1)	58.8 (7.2)	1.0, 0.3
Protrusion	8.4 2.3	8.8 2.4	0.6, 0.4
Left laterotrusion	9.7 2.2	10.1 2.2	0.6, 0.4
Right laterotrusion	9.2 1.8	9.6 2.0	0.7, 0.4
Survey Results			
JFLS	<u>Mean (SD)</u>	<u>Mean (SD)</u>	<u>MWU, P</u>
Masticatory function	0.61 (0.95)	0.59 (0.78)	457.5, 0.91
Mobility	0.41 (0.93)	0.42 (0.81)	427.5, 0.69
Verbal and communication	0.05 (0.19)	0.03 (0.11)	462.5, 0.95
Total	0.35 (0.51)	0.34 (0.45)	438.5, 0.86
MOS			
Sleep Disturbance	27.4 (21.21)	24.7 (20.43)	426, 0.57
Snoring	19.3 (20.67)	18.7 (27.29)	423.5, 0.52
Sleep short of breath or headache	5.3 (10.42)	9.7 (18.53)	420.5, 0.41
Sleep adequacy	48.3 (23.21)	46.1 (19.44)	440, 0.72
Sleep Somnolence	36.7 (16.68)	40.0 (21.84)	436, 0.67
Sleep Problems Index 1	30.4 (15.38)	32.3 (12.86)	414.5, 0.47
Sleep Problems Index 2	32.4 (15.52)	34.0 (13.19)	420.5, 0.52
Optimum Sleep	0.6 (0.50)	0.6 (0.51)	441, 0.69
PSS	13.37 (6.89)	12.23 (6.38)	491, 0.51

OBC			
Raw Sum	26.7 (7.01)	30.8 (8.44)	328.5, 0.05
Sum of Positive Scores	13.37 (2.85)	14.23 (2.65)	381, 0.22
OHIP			
			<u>F(1,59), P[†]</u>
Initial Functional Limitation Scale	5.4 (1.1)	5.5 (0.5)	Tx: 0.16, 0.69
End Functional Limitation Scale	5.6 (0.4)	5.3 (1.1)	Time: 0.006, 0.96
			Tx*Time: 2.0, 0.16
Initial Physical Pain Scale	4.7 (1.0)	4.9 (0.6)	Tx: 0.47, 0.49
End Physical Pain Scale	4.9 (0.6)	4.9 (1.0)	Time: 0.18, 0.68
			Tx*Time: 0.84, 0.36
Initial Psychological Discomfort Scale	5.3 (1.1)	5.5 (0.7)	Tx: 0.10, 0.76
End Psychological Discomfort Scale	5.5 (0.5)	5.4 (1.1)	Time: 0.07, 0.79
			Tx*Time: 0.99, 0.33
Initial Physical Disability Scale	5.5 (1.1)	5.8 (0.3)	Tx: 0.01, 0.93
End Physical Disability Scale	5.9 (0.3)	5.6 (1.1)	Time: 0.37, 0.54
			Tx*Time: 2.3, 0.14
Initial Psychological Disability Scale	5.3 (1.1)	5.7 (0.4)	Tx: 1.0, 0.31
End Psychological Disability Scale	5.5 (0.6)	5.5 (1.0)	Time: 0.08, 0.78
			Tx*Time: 2.8, 0.10
Initial Social Disability Scale	5.6 (1.1)	5.9 (0.3)	Tx: 0.17, 0.68
End Soccial Disability Scale	5.9 (0.3)	5.7 (0.9)	Time: 0.01, 0.92
			Tx*Time: 3.1, 0.08
Initial Handicap Scale	5.7 (1.1)	5.9 (0.2)	Tx: 0.03, 0.85
End Handicap Scale	5.9 (0.2)	5.7 (1.1)	Time: 0.01, 0.93
			Tx*Time: 2.9, 0.10
Initial Total Scale	5.3 (1.0)	5.6 (0.3)	Tx: 0.09, 0.76
End Total Scale	5.6 (0.3)	5.4 (1.0)	Time: 0.07, 0.79
			Tx*Time: 2.3, 0.14
Initial TSK	34.2 (6.2)	33.5 (6.3)	Tx: 0.30, 0.58
End TSK	33.4 (4.7)	32.8 (5.3)	Time: 0.90, 0.35
			Tx*Time: 0.001, 0.98
<u>K-W, P</u>			

Initial TMDPS	2.0 (2.21)	2.4 (2.00)	Initial: 0.764, 0.39
End TMDPS	1.8 (1.87)	2.1 (1.88)	End: 0.49, 0.49
			<i>Friedman, P</i>
			Time: 1.95, 0.39

Key: K-W, Kruskal-Wallis; MWU, Mann-Whitney U; P, P values; SD, standard deviation; Tx, treatment
† Results are effects due to treatment (tx), time, and interactions (treatment * time)

Table 4. Subject satisfaction, stability, retention and estimated RMMA activity results

Independent variable	MI	SOVA	Test	Effect size	95% CI
<u>Satisfaction</u>					
	<u>Mean (SD)</u>	<u>Mean (SD)</u>	<u>MWU</u>		
The splint fits well	4.3 (0.8)	4.1 (1.2)	444.5	0.20 [‡]	-0.31-0.70
I do not use the splint nightly	1.8 (1.2)	2.2 (1.4)	389	0.31 [‡]	-0.20-0.81
I use an additional splint	1.5 (1.1)	1.6 (1.3)	442.5	0.08 [‡]	-0.42-0.59
The splint helps my bruxism	4.0 (0.9)	3.7 (1.0)	377.5	0.32 [‡]	-0.19-0.82
<u>Stability</u>[†]					
	<u>Mean (SD)</u>	<u>Mean (SD)</u>			
Clench	0.71 (1.58)	0.72 (1.94)			
Grind	0.78 (0.70)	0.83 (1.66)			
Border	1.12 (1.33)	0.63 (1.24)			
Tap	0.30 (0.36)	0.33 (0.41)			
Open	1.60 (1.30)	1.16 (0.94)			
Rest	0.16 (0.18)	0.22 (0.23)			
Total	0.78 (1.14)	0.65 (1.16)	F(1,58) 1.13	0.03 [§]	0-0.13
<u>Retention</u>					
	<u>N</u>	<u>N</u>	<u>MWU</u>		
Wk 1					
0	23	21	391	0.28 [§]	-0.22-0.78
1	7	7			
2	1	2			
3	0	0			
4	0	0			
>4	0	0			
Mo 4					
0	25	21	386	0.30 [§]	-0.21-0.80
1	1	2			
2	3	0			
3	1	0			
4	0	1			
>4	1	6			
Wk 1 vs. Mo 4	31	30	Z -0.629	0.16 [§]	-0.34-0.66
<u>Estimated RMMA</u>					
	<u>Mean (SD)</u>	<u>Mean (SD)</u>	<u>F(1,51)</u>		
N _B	120.1 (124.0)	187.3 (147.2)	2.67	0.05 [§]	0-0.20
N _E	58.9 (47.0)	82.3 (51.3)	1.89	0.04 [§]	0-0.17
Bh ⁻¹	18.8 (17.3)	31.2 (21.9)	4.24*	0.08 [§]	0-0.24
Eh ⁻¹	9.4 (6.2)	13.7 (7.8)	3.32	0.06 [§]	0-0.21
BE ⁻¹	1.7 (0.58)	2.2 (0.89)	4.90*	0.09 [§]	0-0.25
T _B	1.1 (0.52)	1.0 (0.33)	0.33	0.01 [§]	--¶
T _E	2.3 (1.45)	2.9 (1.43)	1.47	0.03 [§]	0-0.16

N _{PE}	6.5 (9.1)	12.2 (12.0)	2.81	0.05 [§]	0-0.20
N _{TE}	9.5 (9.0)	10.1 (9.4)	0.01	0.00 [§]	-- [¶]
N _{ME}	8.3 (10.4)	11.9 (10.8)	0.99	0.99 [§]	-- [¶]

[†]Differences, in mm, across tasks were significant, $F(5,290)=52.389$, $p<0.001$, partial $\eta^2=0.475$ (95%CI:0.39-0.53); no task*group interactions $F(5,290)=1.486$, $P=0.194$, partial $\eta^2=0.025$ (95%CI:0-0.054).

[‡] d_{Cohan} effect size

[§]Partial η^2 effect size

[¶]CI undefined (not calculable)

* $P<0.05$

Key: CI, confidence interval; SD, standard deviation; MWU, Mann-Whitney U; RMMA, rhythmic masticatory muscle activity; N_B, number of EMG bursts; N_E, Number of episodes; Bh⁻¹, Bursts per hour; Eh⁻¹, Episodes per hour; BE⁻¹, Bursts per episode; T_B, Burst duration; T_E, Episode duration; N_{PE}, Number of phasic episodes; N_{TE}, Number of tonic episodes; N_{ME}, Number of mixed episodes; z, Wilcoxon z-score.

Figure Legends

Figure 1. A. Summary of participant numbers at selective time points during study. See Table 2 for demographic details. B. Study sequence. Subjects meeting selection criteria were involved in four appointments. Time periods over which each appointment occurred are indicated. Primary outcome (Diary-based compliance, underlined) was evaluated after app 4. Secondary outcomes (*DMFS*, italicized) were evaluated during specific appointments, 2 – 4, as indicated. Key: app, appointment; IC, informed consent; DMFS, decayed missing and filled surfaces; Perio, evaluation of gingiva and plaque indices; HST, home sleep testing (see text).

Figure 2. Occlusal views of the SOVA splint (A), and MI splint (B). Note the circular perforations in the SOVA splint. C-F. Common SOVA splint fabrication errors (see text). C. Rotation of splint. Perforations should be on buccal tooth surfaces, whereas non-perforated portion of splint should be on occlusal surface. D. Stretched material, evidenced by distorted perforation sizes (arrows). E. Excessive material distal to 2nd molar. F. Excessive occlusal indentations from biting into splint during fabrication, despite instructions from manufacturer telling subjects not to bite into splint during fabrication.

Figure 3. A. Mandibular stl models were used for tooth wear grading, here showing examples of wear into enamel (e), which received scores of 1, and wear into dentin (d), which received scores of 2. B. Splint stl model used for studies of material wear over the 4-mo time period. Rectangles show regions of interest used in assessments (see text). C. Mean and 1 SD (error

bars) of splint material wear for the MI vs SOVA group. Ordinate is distance (mm) reported as negative values to represent material loss.

Figure 4. Stability testing methods. A. Example of a trial showing chin (upper trace) and splint (lower trace) movements. Abscissa is time (min:sec); ordinate is distance (shown in cm). Splint data were sampled during the time windows corresponding to each task identified by chin movements. Arrows identify time periods used to construct data for the “Rest” category (see text and Table 4). Letters L, R, A over the Grind trial indicate left, right, and anterior grinding, respectively. Border refers to a task where subjects’ swept the jaw out to the facial border of the splint by moving the jaw first left laterally, then anteriorly, then right laterally and back to a rest position. B. Close-up of a microsensors used for tracking chin and splint movements; outputs from two such sensors resulted in the time series shown in A. C. Picture of the bite plate used to sample bite force data. The shown perforated thermoplastic was used to provide contact with 1st molar regions bilaterally, with and without bite splints in place. With thermoplastic in place, vertical dimension was 20 mm and was not adjusted for trials with and without the splints in place. D. Example of a bite force trial. Upper trace is right masseter EMG; lower trace is output from the force transducer. Rectangles partition trials into four replicates, each of which involved mild, moderate and high bite forces in sequence. The four replicates were: splint in, splint out (tooth trial), splint out, splint in, in that order for all subjects. E. Scatter plot of actual bite force (abscissa) against EMG-based estimate of bite force (ordinate). Both axes are in kg. Plotted data are from the final three trials, one with splints in place and two trials with splints removed,

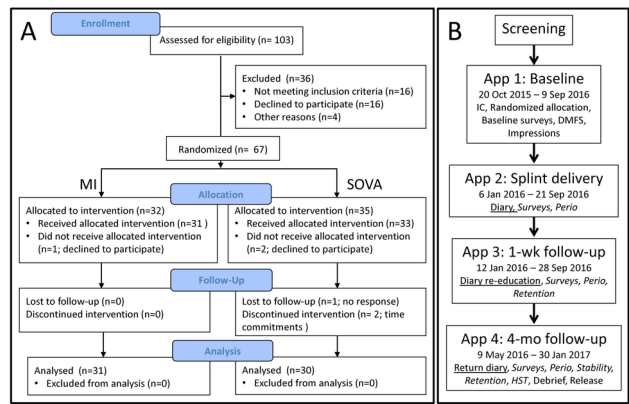
with each trial including mild, moderate and high bite forces as shown in D. F, G. Mean (SD) bite force estimates by group and task. Ordinate in both plots is estimated bite force (kg). Horizontal bars indicate pairwise comparisons that were statistically significant, Bonferonni-corrected at the $P < 0.05$ level. Note that left and right lateral components of the Grind task were pooled to create a Lateral Grind category whereas the anterior component of Grind is Protrusive Grind (see also Table 4, which reports splint displacements in mm during the tasks). The lateral and protrusive components are separated in order to report results for 'roll' and 'pitch' degree-of-freedom dislodgements independently.

Figure 5. A. Page from laboratory manual used by periodontal investigators to measure gingivitis (MGI) and plaque index (PI). Page included instructions and scoring charts as shown. B. Mean (1SD) marginal gingiva indices (MGI) for the Michigan group (filled) and SOVA group (unfilled) at baseline, after 1 wk and 4 mo with splints. Upper = upper arch; lower = lower arch. C. Mean (1SD) plaque indices (PI); format similar to MGI plot.

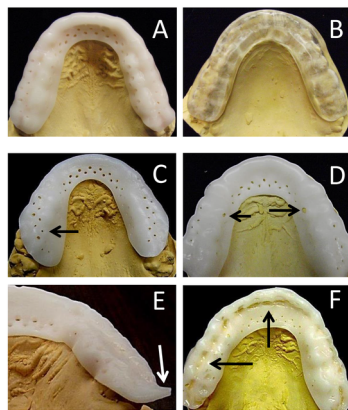
Figure 6. Top row shows instructions from the laboratory manual used for calibrated placement of masseter (A) and thyroideus (B) EMG electrodes. Also shown is the audio monitor (C) used to monitor room and subject sounds, see text. Bottom row (D) shows an example of ~ 15 s from a subject's home sleep study. Top trace is right masseter, middle trace is right thyroideus and bottom three traces are body movements. Note the sequence of 3 masseter bursts,

representing an RMMA sequence. Note that this sequence was associated with thyroideus activity and body movements.

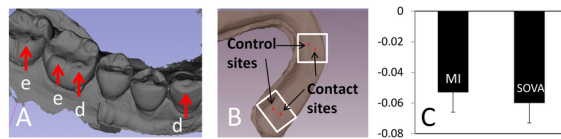
Figure 7. Compliance results plotted as total nights of splint wear (A), and as percent total days splint was in possession of subject (B). Histograms are means with 1 SD error bars for the MI vs. SOVA groups.



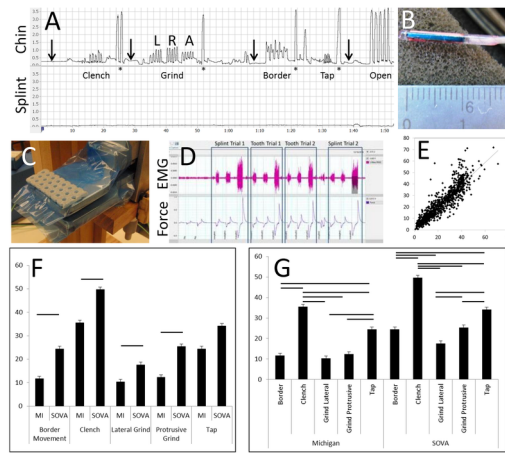
CRE2_315_cre2.20190261-File002.tiff



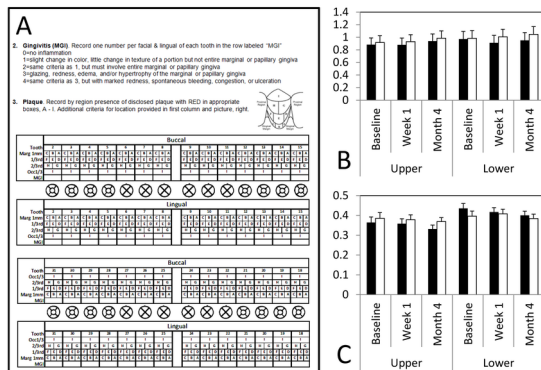
CRE2_315_cre2.20190261-File003.tiff



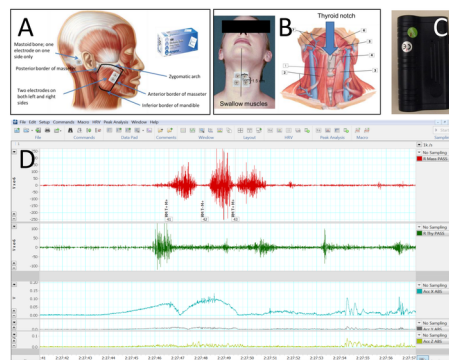
CRE2_315_cre2.20190261-File004.tiff



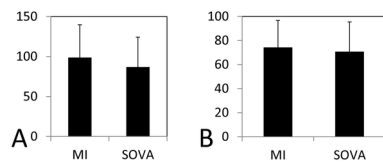
CRE2_315_cre2.20190261-File005.tiff



CRE2_315_cre2.20190261-File006.tiff



CRE2_315_cre2.20190261-File007.tiff



CRE2_315_cre2.20190261-File008.tiff



CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	3
	2b	Specific objectives or hypotheses	3
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	3-4, Figure 1A
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	na
Participants	4a	Eligibility criteria for participants	4
	4b	Settings and locations where the data were collected	4
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	5-10, Figure 1B
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	5-10, Figure 1B
	6b	Any changes to trial outcomes after the trial commenced, with reasons	na
Sample size	7a	How sample size was determined	4
	7b	When applicable, explanation of any interim analyses and stopping guidelines	na
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	4
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	4
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	4
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	4
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	5, 7, 8, 16

		assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	Figure 1
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	10
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	10
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	Table 1
	13b	For each group, losses and exclusions after randomisation, together with reasons	Table 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	Figure 1B
	14b	Why the trial ended or was stopped	3
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Table 1
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	10 - 12
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	12-14, Table 2
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	na
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	Fig. 4C,D
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	na
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	16, 17
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	17
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	17
Other information			
Registration	23	Registration number and name of trial registry	3
Protocol	24	Where the full trial protocol can be accessed, if available	3
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	1

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming; for those and for up to date references relevant to this checklist, see www.consort-statement.org.