ORIGINAL ARTICLE

A crossover randomised controlled trial of oral mandibular advancement devices for obstructive sleep apnoea-hypopnoea (TOMADO)

Timothy G Quinnell,[1] Maxine Bennett,[2] Jake Jordan,[3] Abigail L Clutterbuck-James,[1] Michael G Davies,[1] Ian E Smith,[1] Nicholas Oscroft,[1] Marcus A Pittman,[1] Malcolm Cameron,[4] Rebecca Chadwick,[1] Mary J Morrell,[5] Matthew J Glover,[3] Julia A Fox-Rushby,[3] Linda D Sharples[1,2,6]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2014-205464).

[1]Respiratory Support and Sleep Centre, Papworth Hospital NHS Foundation Trust, Cambridge, UK
[2]Medical Research Council Biostatistics Unit, Institute of Public Health, University Forvie Site, Cambridge, UK
[3]Health Economics Research Group, Brunel University, Uxbridge, Middlesex, UK
[4]Department of Oral and Maxillofacial Surgery, Addenbrooke's NHS Foundation Trust, Cambridge, UK
[5]National Heart & Lung Institute, Royal Brompton Campus, Imperial College, London, UK
[6]Clinical Trials Research Unit, University of Leeds, Leeds, UK

Correspondence to
Dr Timothy George Quinnell, Respiratory Support and Sleep Centre, Papworth Hospital NHS Foundation Trust, Papworth Everard, Cambridge CB23 3RE, UK; Tim.Quinnell@papworth.nhs.uk

Received 24 March 2014
Revised 6 June 2014
Accepted 25 June 2014

To cite: Quinnell TG, Bennett M, Jordan J, *et al*. *Thorax* Published Online First: [*please include* Day Month Year] doi:10.1136/thoraxjnl-2014-205464

ABSTRACT

Rationale Mandibular advancement devices (MADs) are used to treat obstructive sleep apnoea-hypopnoea syndrome (OSAHS) but evidence is lacking regarding their clinical and cost-effectiveness in less severe disease.

Objectives To compare clinical- and cost-effectiveness of a range of MADs against no treatment in mild to moderate OSAHS.

Measurements and methods This open-label, randomised, controlled, crossover trial was undertaken at a UK sleep centre. Adults with Apnoea-Hypopnoea Index (AHI) 5–<30/h and Epworth Sleepiness Scale (ESS) score ≥9 underwent 6 weeks of treatment with three non-adjustable MADs: self-moulded (SleepPro 1; SP1); semi-bespoke (SleepPro 2; SP2); fully-bespoke MAD (bMAD); and 4 weeks no treatment. Primary outcome was AHI scored by a polysomnographer blinded to treatment. Secondary outcomes included ESS, quality of life, resource use and cost.

Main results 90 patients were randomised and 83 were analysed. All devices reduced AHI compared with no treatment by 26% (95% CI 11% to 38%, p=0.001) for SP1, 33% (95% CI 24% to 41%) for SP2 and 36% (95% CI 24% to 45%, p<0.001) for bMAD. ESS was 1.51 (95% CI 0.73 to 2.29, p<0.001, SP1) to 2.37 (95% CI 1.53 to 3.22, p<0.001, bMAD) lower than no treatment (p<0.001 for all). Compliance was lower for SP1, which was the least preferred treatment at trial exit. All devices were cost-effective compared with no treatment at a £20 000/quality-adjusted life year (QALY) threshold. SP2 was the most cost-effective up to £39 800/QALY.

Conclusions Non-adjustable MADs achieve clinically important improvements in mild to moderate OSAHS and are cost-effective. Of those trialled, the semi-bespoke MAD is an appropriate first choice.

Trial registration number ISRCTN02309506.

Key messages

What is the key question?

- Are mandibular advancement devices (MADs) clinically- and cost-effective compared with no treatment in mild to moderate obstructive sleep apnoea-hypopnoea syndrome (OSAHS), and does the degree of MAD sophistication influence outcomes?

What is the bottom line?

- Clinical and cost-effectiveness analyses suggest that semi-bespoke non-adjustable devices should be offered as first line treatment for mild OSAHS and as an alternative to CPAP in moderate disease, whereas dentally-fitted bespoke devices should be reserved for those who cannot produce the mould for, or tolerate, a semi-bespoke device; and while adjustable MADs offer some advantages, their precise role and cost-effectiveness still need to be established.

Why read on?

- This is the first comprehensive randomised controlled trial to evaluate both the clinical and cost-effectiveness of MADs for the treatment of mild to moderate OSAHS; the results and their implications for clinical practice are discussed.

INTRODUCTION

Obstructive sleep apnoea-hypopnoea (OSAH) involves repeated collapse of the pharyngeal airway during sleep, causing oxygen desaturations and brief arousals. OSAH syndrome (OSAHS) incorporates excessive daytime sleepiness (EDS),[1] affecting 2%–7% of adults.[2]

There is a causal link with hypertension[3] and cardiovascular risk is increased 2.5-fold,[4] with a reported 6% increase in stroke risk per unit increase in Apnoea-Hypopnoea Index (AHI/hour).[5] Road traffic accident risk is two to three times higher[6] and health-related quality of life (HRQoL) is impaired.[7] Healthcare usage is almost doubled in OSAHS, with the main determinants of increased cost being cardiovascular disease and psychoactive medication.[8]

Weight loss sometimes cures OSAHS but CPAP therapy gives immediate control of obstructive events. Improvement in EDS usually follows, with added benefits to driving safety and HRQoL.[9] Meta-analyses have shown that CPAP reduces mean blood pressure by around 2 mm Hg[10] and observational data have suggested cardiovascular risk reduction.[11]

CPAP is recommended by the National Institute for Health and Care Excellence as clinically and cost-effective for moderate to severe OSAHS.[12] Benefits are less certain in milder disease, although there is some evidence of improvement in functional outcomes and daytime sleepiness.[13 14] Intolerance of CPAP is common, affecting 46%–83% of patients across the disease spectrum.[15] Effective alternatives to CPAP are therefore needed.

Intraoral mandibular advancement devices (MADs) protrude the mandible and tongue to maintain upper airway patency during sleep. Meta-analyses suggest MADs are beneficial in OSAHS, although trial and device heterogeneity complicate interpretation.[16] Most comparative studies show CPAP is better at controlling respiratory events but both treatments improve sleepiness equally, possibly due to CPAP intolerance.[17] MADs are effective compared with sham MADs in reducing AHI and improving sleepiness. However, adverse sham effects may exaggerate treatment benefits of active devices by undermining sleep quality without reducing respiratory events,[16] although there is evidence to the contrary in more severe OSAHS.[18] Nonetheless, it is important to compare the effectiveness of MADs with no treatment in milder disease, where the balance of costs and benefits may be more marginal.[19] The considerable heterogeneity in MAD treatment, caused by variation in device design, production processes and specialist involvement, also results in cost variability. There is therefore a need to explore the clinical and cost-effectiveness of a range of MADs. This study aimed to determine whether:

1. MADs are clinically- and cost-effective compared with no treatment in mild to moderate OSAHS.
2. The degree of MAD sophistication influences outcomes, including cost-effectiveness.

METHODS

Open-label, four-period, crossover, randomised controlled trial (RCT) comparing three non-adjustable MADs with no treatment.

Full details of trial methodology are included in an online supplement.

Participants

Patients aged ≥18 years with mild to moderate OSAHS confirmed by respiratory polysomnography (rPSG) (AHI 5–<30/h) and symptomatic daytime sleepiness (Epworth Sleepiness Scale (ESS) score ≥9) were recruited from Papworth Hospital sleep centre. Newly diagnosed patients not requiring or declining CPAP and existing CPAP intolerant patients were eligible. See online supplement for exclusion criteria.

Procedures

The three MADs were: thermoplastic 'boil and bite' device (SleepPro 1 (SP1); Meditas, Winchester, UK); semi-bespoke device produced from a patient-moulded dental impression kit (SleepPro 2 (SP2); Meditas); bespoke MAD (bMAD) device fitted and manufactured by National Health Service (NHS) Maxillofacial Team at Addenbrooke's Hospital, Cambridge, UK.

After dental eligibility was confirmed, patients were randomised via telephone by the hospital's R&D unit using Williams' Latin Squares with allocations generated by computer, using permuted blocks of eight.

Period duration was 6 weeks (4 weeks for no treatment): 2 weeks acclimatisation and 4 weeks treatment. One week's washout followed active treatments. Outcomes were obtained at baseline and at the end of each treatment period.

Outcomes

Primary outcome was AHI, measured by domiciliary rPSG (Embletta, Embla Systems, Kanata, Ontario, Canada) and scored by a polysomnographer blinded to treatment. Other outcomes included rPSG indices, blood pressure, subjective sleepiness (ESS), sleep-related quality of life (Functional Outcomes of Sleep Questionnaire (FOSQ); Calgary Sleep Apnoea Quality of Life Index (SAQLI)) and generic HRQoL (Short Form 36 (SF36); EuroQol (EQ-5D-3L)). Healthcare usage, road traffic accidents and treatment satisfaction were recorded at the end of each treatment period. Treatment compliance was calculated at the end of each treatment period from a patient-completed diary. Device preference was documented at trial exit.

Statistical analysis

Calculations indicated 72 patients for 80% power to detect treatment effects of 1/3 SDs between MADs (two-sided 5% significance). A total of 90 patients were recruited, allowing 20% loss to follow-up. All randomised patients were followed up and available results from periods included in analysis irrespective of treatment compliance ('intention to treat').

Mixed effects Poisson regression was used to estimate effects of treatment, period and treatment by period interactions for AHI. Mixed effects logistic regression was used to analyse response (complete (AHI <5) or partial (AHI ≥5 with ≥50% reduction from baseline)). All other outcomes were analysed using linear mixed models.

Regressions explored the effects on AHI response of baseline AHI, ESS, age, gender, compliance and body mass index (BMI), and BMI changes over time.

Economic analysis

Using an NHS perspective, device costs, fitting time and other healthcare usage within each treatment period were compared against the costs of no treatment. Patient-specific healthcare resource use data were collected and valued using NHS reference costs, standard unit costs and published literature.[20 21] HRQoL was measured and valued using the EQ-5D-3L and UK social tariff and converted to quality-adjusted life years (QALYs).[22] A random effects model estimated differences in costs and QALYs for each MAD against control. Bootstrapping was used to construct cost-effectiveness acceptability curves (CEACs) and cost-effectiveness acceptability frontier. The impact of changes in MAD cost, device lifespan and use of SF6D QALYs on net monetary benefit (NMB) was assessed.

RESULTS

In all, 90 patients were randomised between December 2010 and July 2012 (figure 1). A total of 16 (18%) withdrew from the trial; 7 (8%) of these did not complete any treatment and were excluded from analyses. Two additional patients who withdrew between periods one and two had a failed sleep study but completed secondary outcomes, making 9 (10%) patients who provided no AHI data after baseline. Seven patients withdrew later in the trial. It has been assumed that the data were missing at random for these patients. There were no differences in baseline characteristics between those who completed the trial and those who withdrew, and the pattern of withdrawal did not correspond to any particular MAD (see online supplement). Seven other sleep studies failed, leaving 305 studies (85% of 360) from 81 patients (90%) for AHI analysis. For all other outcomes, 314 (87%) measurements and 83 (92%) patients were

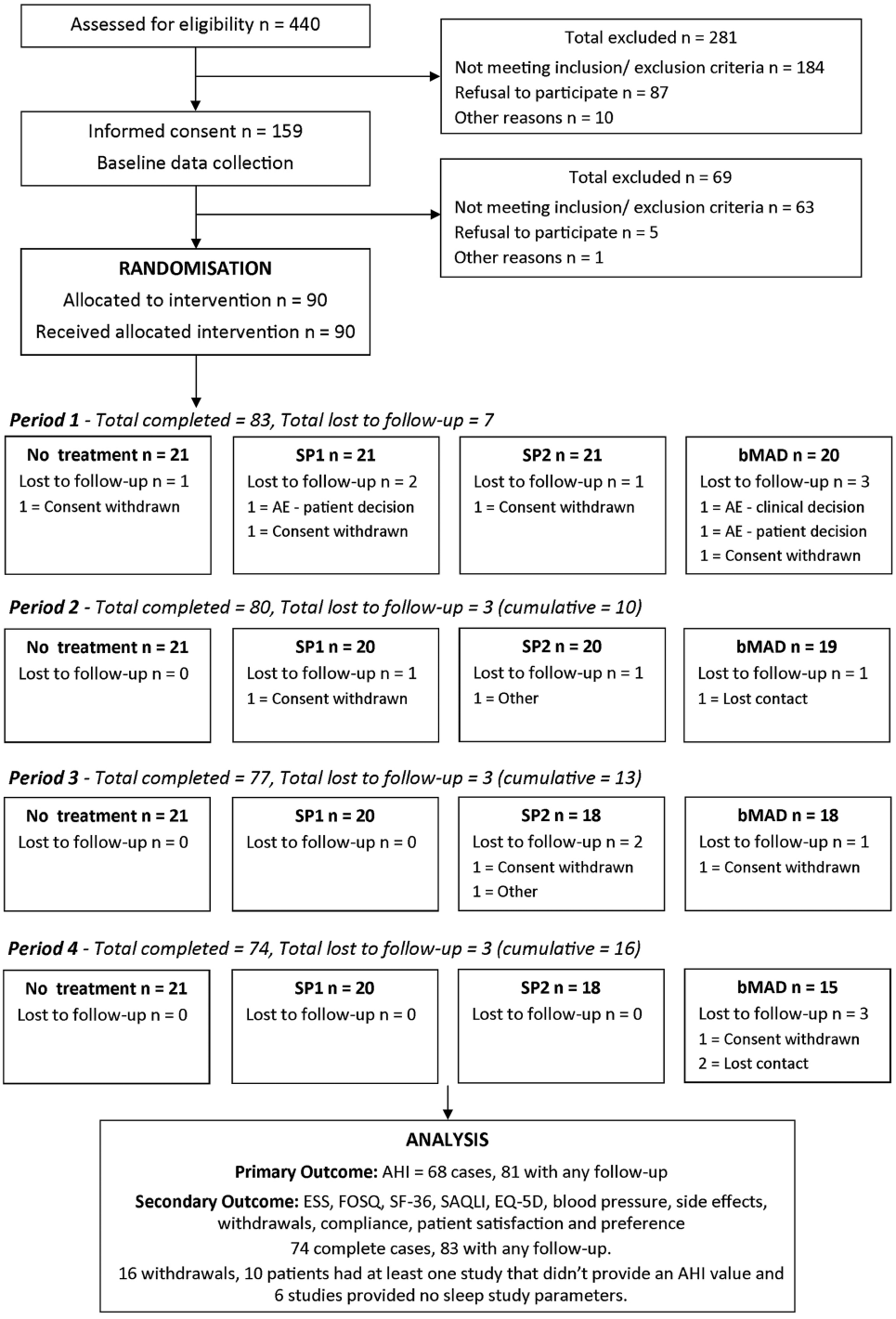


Figure 1 Flow of patients through the trial.

available for analyses. Baseline characteristics for complete cases and patients with missing data were similar.

Baseline characteristics for randomised patients are given in table 1.

Although 12 patients had baseline ESS below 9, they were eligible based on their ESS being 9 or more at screening. One patient was erroneously randomised following a screening ESS of 8 (ESS=10 at baseline). Three patients were randomised on the basis of more sensitive electroencephalographically-guided AHI scoring. When later rescored according to rPSG criteria, the AHI was below 5. There were no significant period-by-treatment interactions and, for AHI and ESS, no period effects. Complete case results were almost identical to main analysis of cases with at least one measurement, so only the latter is reported.

All three MADs significantly decreased the AHI against no treatment by 26% (95% CI 11% to 38%) for the SP1, 33% (95% CI 24% to 41%) for the SP2 and 36% (95% CI 24% to 45%) for the bMAD (table 2, figure 2A).

A similar effect was found for all devices against no treatment for 4% oxygen desaturation index (4% ODI). There were no significant differences between devices for either AHI or 4% ODI. Compared with no treatment, patients spent significantly less time with nocturnal oxygen saturation <90% when using SP2 and bMAD (SP2 vs no treatment p=0.040 and bMAD vs no treatment p<0.001, respectively), and there were some

Table 1 Baseline characteristics

Unit/category	Total (N=90)	Min	Max
Demographic information			
Gender			
Male	72 (80%)		
Female	18 (20%)		
Age at randomisation			
Years	50.9 (11.6)	26.1	79.6
BMI			
Kg/m^2	30.6 (27.9–35.1)	23.9	54.5
Smoking history			
Non-smoker	44 (49%)		
Ex-smoker	39 (43%)		
Smoker	7 (8%)		
Clinical history			
Previous CPAP	4 (4%)		
Asthma	14 (16%)		
Diabetes			
Type I	1 (1%)		
Type II	7 (8%)		
Cardiovascular disease			
Previous stroke	2 (2%)		
Previous TIA	1 (1%)		
Ischaemic heart disease	5 (6%)		
Hypertension	23 (26%)		
Sleep study			
Apnoea-Hypopnoea Index			
Events per hour	13.8 (6.2)	2.9	27.7
Missing*	1		
Oxygen Desaturation Index			
Events per hour	9.8 (5.2)	0.6	22
Minimum SpO_2			
Per cent	83.7 (4.7)	71	91
Missing*	2		
Mean SpO_2			
Per cent	94.2 (1.3)	89.8	97.7
Missing*	1		
Time <90% of nocturnal SpO_2			
Minutes	8.3 (2.9–24.8)	0	315.4
Missing*	1		
Epworth Sleepiness Scale (ESS)			
ESS			
Unit score	11.9 (3.5)	3	20

Categorical variables show frequency (%) and continuous variables show either mean (SD) or median (IQR).
*One sleep study failed and inclusion was based on the Desaturation Index.
BMI, body mass index; TIA, Transient Ischaemic Attack.

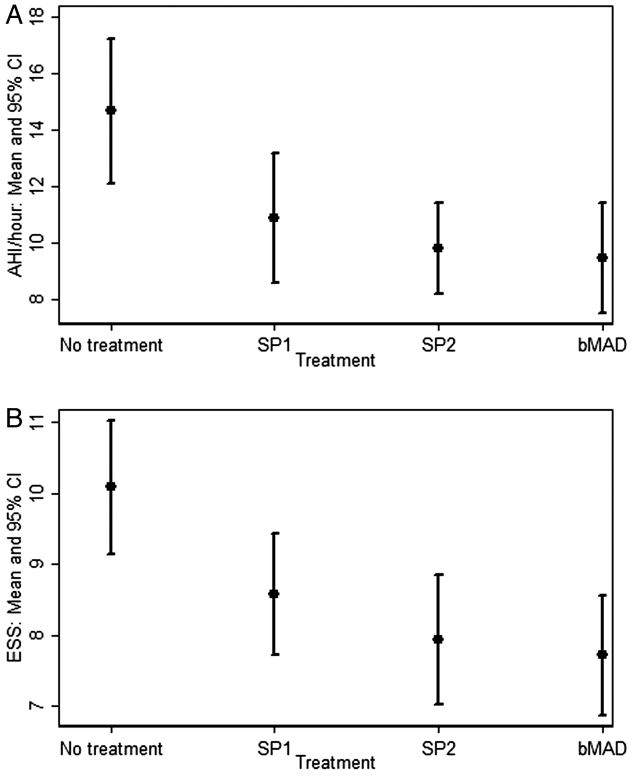


Figure 2 (A) Mean Apnoea-Hypopnoea Index (AHI) and (B) mean Epworth Sleepiness Scale (ESS) score (with 95% CIs for each treatment) from the Poisson mixed effects model.

differences between MADs (bMAD vs SP2, uncorrected for multiple testing p=0.037; bMAD vs SP1, uncorrected multiple testing p=0.006). The bMAD had a significant effect on minimum oxygen saturation compared with no treatment and the other devices (see online supplement).

Complete response (AHI <5) or partial response (≥50% reduction in AHI from baseline but AHI ≥5) to treatment was observed in 17 of 76 (22%) cases after no treatment, and 29/77 (38%), 38/78 (49%) and 33/74 (45%) patients after SP1, SP2 and bMAD, respectively (p=0.0006 for all vs no treatment; see online supplement). Response was significantly associated with baseline BMI (OR 0.89/kg/m^2, 95% CI 0.80 to 0.98, p=0.014) and contemporaneous BMI (OR 0.88/kg/m^2, 95% CI 0.80 to 0·96, p=0·007). There was a weak association with protrusion (OR 1.03 per % maximal protrusion, 95% CI 1.00 to 1.05, p=0.034), although there was no association between % protrusion and AHI as a continuous variable (HR 0.997 (0.991 to 1.001), p=0.206). Baseline AHI, ESS, gender, age and compliance were not associated with treatment response.

Median (quartiles) number of nights (of 28) that SP1, SP2 and bMAD were used was 25 (17, 28), 27 (23, 28) and 26 (23, 28), respectively. Mean (SD) nightly use of each device was 4.4 (2.4), 5.7 (2.0) and 5.7 (2.0) hours, respectively (p<0.001 for both SP2 and bMAD vs SP1). Patients stopped treatment early during 14 of 81 (17%) SP1, four of 78 (5%) SP2 and six of 76 (8%) bMAD periods (p=0.034).

All MADs decreased ESS significantly compared with no treatment, by 1.51 units (95% CI 0.73 to 2.29) for the SP1, 2.15 units (95% CI 1.31 to 2.99) for the SP2 and 2.37 units (95% CI 1.53 to 3.22) for the bMAD (table 2, figure 2B). Based on Bonferroni corrected comparisons, there were no significant differences between devices.

Compared with the no treatment arm, total FOSQ score was significantly higher (better) by 0.50 (95% CI 0.08 to 0.92, p=0.018) for SP1, 1.10 (95% CI 0.65 to 1.55, p<0.001) for SP2 and 1.31 (95% CI 0.84 to 1.78, p<0.001) for bMAD. Total SAQLI score was improved in a similar pattern compared with no treatment: 0.27 (95% CI 0.07 to 0.48, p=0.008) for the SP1, 0.62 (95% CI 0.38 to 0.86, p<0.001) for the SP2 and 0.65 (95% CI 0.41 to 0.90, p<0.001) for the bMAD. The SP2 and bMAD FOSQ and SAQLI scores were significantly higher (better) than the SP1 (uncorrected p≤0.01). While most of the SF36 dimensions did not show differences between MADs, there were important differences in the Vitality Scores for SP2 and bMAD against no treatment (see online supplement).

Table 2 Summary of results from mixed effects models for AHI (n=81) and ESS (n=83)

	Mean (SD)	Coefficient	95% CI	p Value	Global p value
AHI (n=81)					
Constant		14.22	(11.66 to 17.34)	<0.001	
		Relative AHI compared with no treatment			
No treatment	14.6 (10.5)	–	–	–	<0.001
SP1	10.8 (9.5)	0.74	(0.62 to 0.89)	0.001	
SP2	9.7 (8.9)	0.67	(0.59 to 0.76)	<0.001	
bMAD	9.5 (8.4)	0.64	(0.55 to 0.76)	<0.001	
ESS (n=83)					
Constant		10.65	(9.64 to 11.66)	<0.001	
		Difference in ESS compared with no treatment			
No treatment	10.1 (4.3)	–	–	–	<0.001
SP1	8.5 (4.0)	−1.51	(−2.29 to −0.73)	<0.001	
SP2	8.0 (4.1)	−2.15	(−2.99 to −1.31)	<0.001	
bMAD	7.7 (3.8)	−2.37	(−3.22 to −1.53)	<0.001	

AHI, Apnoea-Hypopnoea Index; bMAD, bespoke mandibular advancement device; ESS, Epworth Sleepiness Scale.

Mean (SD) mandibular protrusion for the SP1 was greater than for the other two devices (SP1, 62.6% (22.1); SP2, 51.7% (26.4), uncorrected p<0.001; bMAD, 55.2% (19.7), uncorrected p=0.012). Patients found the SP1 less comfortable and were less satisfied with it than the SP2 and bMAD (see online supplement). The SP1 was more likely to fall out or be removed.

Of 74 trial completers, 30 (41%) ranked the bMAD highest in terms of preference and 23 (31%) ranked it second (see online supplement). SP2 was ranked highest by 22 (30%) and second by 34 (46%) patients, while 10 (14%) favoured no treatment. After the trial, 56 of 74 (76%) completers continued treatment with their preferred MAD and 4 (5%) others chose treatment with the MAD that achieved the best AHI.

There were four serious adverse events during the trial. Minor adverse events were experienced by 86 (96%) patients. Most common were mouth problems/discomfort (83, 92%) and excess salivation (48, 53%) with SP2 performing best for both (see online supplement).

All devices were cost-effective compared with no treatment at a willingness-to-pay (WTP) of £20 000/QALY, based on mean costs and QALYs. SP2 achieved the highest NMB at £33 per 4 weeks (table 3).

On average, the SP1 and SP2 point estimates were associated with more QALYs and lower mean costs compared with no treatment, although QALY differences between devices were small and non-significant. These results are robust to: changes in a device's price and lifespan; increasing the WTP per QALY to £30 000; and using only complete case analysis. When the bMAD price exceeds £525 or average lifespan falls below 14 months, it no longer has a positive NMB. The CEAC (figure 3A) and NMB (figure 3B) show SP2 to be most cost-effective up to a WTP of £39 800/ QALY, at which point bMAD supersedes it (39% likelihood of being cost-effective vs 35% for the SP2). Below a WTP of £5000/QALY, only SP2 is more cost-effective than no treatment. The finding that SP2 was the most cost-effective option was strengthened considerably when using SF6D QALYs.

DISCUSSION

This trial showed that in mild to moderate OSAHS, non-adjustable MADs improve objective and subjective health outcomes over no treatment. Additional improvements diminished with increasing MAD sophistication but the consistent results across outcomes suggest genuine effects. All devices were cost-effective against no treatment based on the point estimates of costs and QALYs. However, differences in EQ-5D-3L results between devices were small and non-significant, although significant using SF6D QALYs. Probabilistic analysis, accounting for uncertainty in costs and QALYs, showed SP2 was the most cost-effective up to a WTP of £39 800/QALY. Above this WTP, bMAD appeared most cost-effective in the short term, although not using SF6D QALYs.

All MADs reduced AHI to between 64% and 75% of no treatment AHI, which is modest. Pneumatic splinting of CPAP efficiently controls multilevel pharyngeal collapse. The mechanisms of action of MADs are more complex, probably involving airway stiffening, splinting and enlargement.[23] These factors and level of obstruction vary between patients, impacting MAD efficacy. Greater AHI effects have been reported[24 25] but methodological and device heterogeneity complicate interpretation of the comparisons. Many studies also include patients with more severe disease at baseline,[24 26] which gives more potential for useful treatments to show an effect. The focus of the trial of oral mandibular advancement devices for obstructive sleep apnoea-hypopnoea (TOMADO) was on milder disease, where a no treatment control was particularly relevant. The RCT most comparable with TOMADO, which used a maximally titrated adjustable MAD in mild to moderate OSAHS, reported a mean AHI of 67% against placebo tablet, which is consistent with our results.[9]

Reduction of AHI can be proportional to mandibular protrusion.[27] Mean (SD) protrusion in this trial (4.8 (2.5) (SP2) to 5.7 (2.1) (SP1) mm (52.5 (27.8)% to 63.4 (22.6)% maximal advancement)) was lower than others have achieved,[18 28] often using adjustable MADs.[9 17 24 26 29] However, greater protrusion in those trials did not always achieve greater AHI reduction than TOMADO.[9 18] The results from this trial did not demonstrate a convincing association between protrusion and AHI. There is also evidence that maximal protrusion may not be necessary in milder OSAHS.[30]

All MADs studied were associated with a statistically significant improvement in ESS. The SP1 improvement was of borderline clinical significance compared with no treatment. The ESS

Table 3 Comparison of costs and QALYs from devices against control

	Intervention							
Cost component (£)	**No treatment n=78**		**SP1 n=81**		**SP2 n=78**		**bMAD n=77**	
Device (fixed)	–		£21		£128		£350	
Measurement for device (fixed)	–		–		–		£110.37	
Fitting of device (fixed)	–		–		–		£92.04	
Additional visit if required (average across all patients)	–		–		–		£5.98	
Subtotal	–		£21		£128		£350	
Device lifespan (months) (fixed)	–		12		12		18	
Cost of intervention subtotal pro rata (4 weeks) (fixed)*	–		£1.62		£9.85		£28.64	
Summary of costs (£)	**Mean**	**SE**	**Mean**	**SE**	**Mean**	**SE**	**Mean**	**SE**
Resource use cost (4 weeks) Mean (SD)†	£78.50	£19.97	£73.02	£10.47	£53.58	£8.05	£76.25	£24.40
Total cost (4 weeks) Mean (SD)†	£78.50	£19.97	£74.64	£10.47	£63.43	£8.05	£104.89	£24.39
Total cost difference intervention vs control	–	–	−£3.87	£21.38	−£15.08	£20.62	£26.39	£27.94
Health-related quality of life measure								
EQ-5D-3L Utility score	0.85	0.2	0.86	0.2	0.86	0.23	0.87	0.19
EQ-5D-3L QALY score (4-week trial period)‡	0.0649	0.0017	0.0658	0.0017	0.0658	0.0019	0.0667	0.0017
QALY score difference intervention vs control	–	–	0.00094	0.00105	0.00088	0.00123	0.00177	0.00147
Cost effectiveness measure								
Incremental Cost-Effectiveness Ratio	–		−£4093		−£17 104		£14 876	
Net monetary benefit (willingness-to-pay=£20 000)	–		£23		£33		£9	

*Device and fitting costs are pro rata to be comparable over the 4-week trial period.
†Resource use and total costs by intervention estimated using a mixed effects model controlling for baseline data. All costs in 2011/2012 (£).
‡QALY scores calculated using area under the curve method to represent the true QALY score for the 4-week intervention period and to be consistent with the costs presented.
bMAD, bespoke mandibular advancement device; QALY, quality-adjusted life year.

effects for the SP2 and bMAD were greater and similar to some of those reported for CPAP in OSAHS.[17] Placebo effects cannot be ignored. A recent study attributed up to 29% of ESS response to expectation of benefit from high CPAP compliance;[31] but the associated objective AHI reductions in this study suggest real effects.

Patient-reported MAD compliance was good, and the reliability of subjective measurement has been objectively demonstrated.[32] Compliance was lower for SP1, which was less well tolerated and often fell out or was removed. Poor retention of a non-bMAD has been related to inferior compliance,[25] while superior MAD compliance is considered key to matching CPAP health outcomes.[17] The SP1's unpopularity at trial exit and poorer in-trial compliance suggest there are significant obstacles to its longer term effectiveness.

This relatively large study used an efficient design to estimate the short-term effectiveness of MADs on the most relevant outcomes in OSAHS: AHI and EDS. The results were clear, unambiguously significant and robust to assumptions for incomplete data. Seven patients failed to complete any treatment and two dropped out after one period. We therefore estimate that 10% of patients eligible for a MAD will prove intolerant. Another 8% may become intolerant later. This highlights the need for longer term data regarding MAD usage. Current ongoing follow-up of trial patients will eventually assess longer term device durability and compliance but until this information becomes available, modelling of long-term outcomes must be relied upon.

This study was conducted at a specialist centre, potentially limiting generalisability. However, participants were recruited from our usual clinic population, mostly referred from primary care. The SP1 and SP2 are available in many countries and similar to other thermoplastic and 'semi-bespoke' MADs. The bMAD was similar in design to other available monobloc devices. It was fitted and manufactured by a hospital maxillofacial laboratory, but using skills, materials and facilities common to dental sleep services.

Although the aim for the bMAD was at least 50% maximal protrusion, this was often lower and similar to that achieved independently by patients with the other devices. This reflects the pragmatic nature of the trial, making its findings more applicable to the wider NHS. Including an adjustable MAD may have achieved greater protrusion and would have extended this effectiveness evaluation. However, the need to adequately cover the range of non-adjustable MADs available to NHS sleep services precluded this. Although they are more costly, adjustable MADs are increasingly recommended.[33 34] Titration is thought to optimise protrusion and tolerance[35] but their superiority remains unproven. A large retrospective review of mild to severe OSAHS patients fitted with a bMAD reported slightly greater AHI effects for an adjustable MAD compared with a non-adjustable device. However, differences were often not statistically significant and clinical significance was doubtful given that ESS reductions were no different.[34] The inconsistent evidence regarding greater protrusion in milder disease has already been discussed, and the high reported compliance with our bespoke devices leaves little room for improvement. Nonetheless, the increasing popularity of more expensive adjustable MADs means rigorous prospective comparisons should be a priority.

In the short term, MADs achieve clinically significant improvements in mild to moderate OSAHS and appear cost-effective at £20 000/QALY compared with no treatment. Inferior tolerance and retention of the SP1 device may limit its effectiveness. Minor differences between the two more sophisticated devices suggest that a semi-bespoke non-adjustable MAD could be a practical and efficient first choice in most patients. A

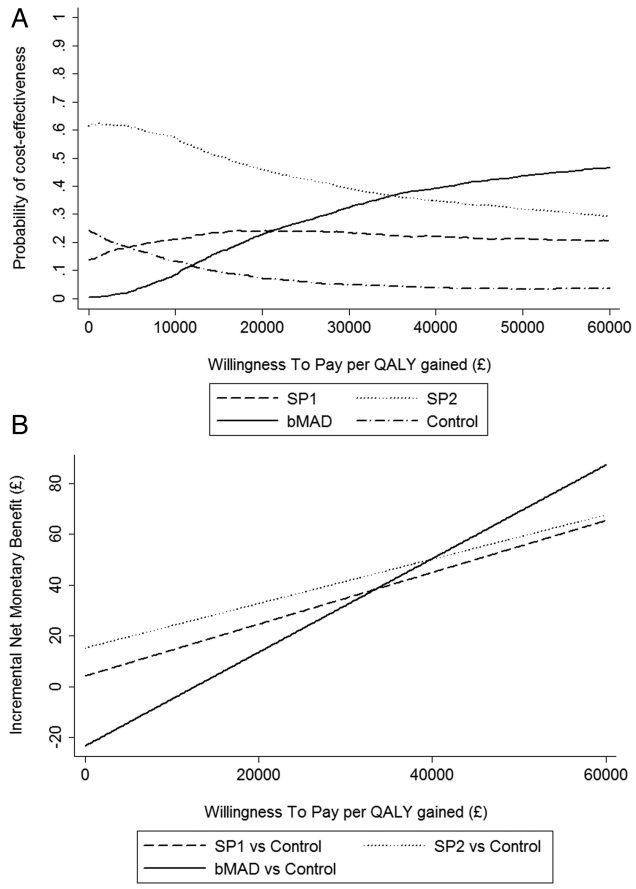


Figure 3 (A) Cost-effectiveness acceptability curves (CEACs) of each device compared with all alternatives and (B) net monetary benefit (NMB) for each device compared with no treatment (control). The CEAC is derived from the joint distribution of incremental costs and incremental quality-adjusted life years (QALYs) and shows the uncertainty associated with both. The curve shows the probability that a specific treatment (SP1, SP2, bespoke mandibular advancement device (bMAD) or no treatment) will be cost-effective compared with all the other alternatives assessed for a specific amount that a service provider is willing to pay per QALY gained. The NMB is the monetary value of the devices incremental health effect (ie, the willingness-to-pay for a QALY gain × total QALYs gained minus the incremental cost). It is measured over the 4-week period of the trial. A higher NMB is associated with a more cost-effective treatment.

bespoke device may be necessary for those patients who require dental help with production and fitting of a MAD, or where there are concerns regarding dental eligibility and oral health, and this approach should be considered for inclusion in clinical guidelines. However, longer term compliance and the potential impact of unknown differences in device durability are being explored in a longer term evaluation of this cohort to help determine whether minor differences between patient-fitted and dentally-produced MADs are important. In addition, the effects on cardiovascular events and road traffic accidents are being studied in a long-term cost-effectiveness model. Finally, whether adjustable devices are cost-effective and offer clinically significant advantages over bespoke non-adjustable devices in the real life setting still need to be explored.

Acknowledgements We are grateful to the TOMADO Trial Steering Committee (Professor John Gibson (Chair), Dr Justin Pepperell and Professor Mary Morrell), the Data Monitoring Committee (Dr Andrew Cummin (Chair), Dr Andrew Hall and Professor John Matthews) and the public patient involvement (PPI) representative, Tanya Wanstall, for their contribution to the trial. We thank the staff of Meditas and the oral and maxillofacial surgery department at Addenbrooke's Hospital for making the devices. Thanks also go to the patients and staff in the Respiratory Support & Sleep Centre at Papworth Hospital who made the trial possible, especially Clare East and Joanna Rayner. The protocol has been published in the Lancet Protocol Reviews (010PRT/4998).

Contributors TGQ conceived the study. TGQ, LDS, RC, MC and MJM designed the study. The systematic review of literature was written by MB, MJG, MAP and LDS. MAP, ALC-J and RC collected the data. MB, JJ, MJG and LDS analysed the data and all authors interpreted the data. TGQ, LDS and JAF-R drafted the manuscript. ALC-J edited the manuscript and all authors critically revised it.

Funding This project was funded by the National Institute for Health Research Health Technology Assessment (NIHR HTA) Programme (project number 08/110/03) and will be published in full in the *Journal of Health Technology Assessment*.

Competing interests MC provided the bespoke mandibular advancement device service at Addenbrooke's Hospital, but has no personal or financial interests in the device. The manufacturer of the SP1 and SP2 supplied them at a discounted price but were not involved in the study design, analysis or reporting of results, nor did they have exclusive access to any part of the potential study data in return for the discount. As such, there are no conflicts of interest. LDS was partly funded by the Medical Research Council programme number U015232027. MB was funded by an NIHR Clinical Trials Methodology fellowship.

Ethics approval The National Research Ethics Service (NRES) Committee East of England — Cambridge Central (formerly known as Cambridgeshire 2 Research Ethics Committee) approved the protocol and subsequent amendments. The trial was registered with the International Standard Randomised Controlled Trial Number Register before the trial began (ISRCTN02309506).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement This work was funded by the NIHR and all data will be published in full in the form of a single whole issue of *Journal of Health Technology Assessment* within the NIHR Journals Library (full draft report submitted in December 2013 for peer review and subsequent publication; no expected publication date yet).

REFERENCES

1 Douglas NJ, Polo O. Pathogenesis of obstructive sleep apnoea/hypopnoea syndrome. *Lancet* 1994;344:653–5.
2 Punjabi NM. The epidemiology of adult obstructive sleep apnea. *Proc Am Thorac Soc* 2008;5:136–43.
3 Stradling JR, Pepperell JC, Davies RJ. Sleep apnoea and hypertension: proof at last? *Thorax* 2001;56(Suppl 2):ii45–9.
4 Dong JY, Zhang YH, Qin LQ. Obstructive sleep apnea and cardiovascular risk: meta-analysis of prospective cohort studies. *Atherosclerosis* 2013;229:489–95.
5 Redline S, Yenokyan G, Gottlieb DJ, *et al*. Obstructive sleep apnea-hypopnea and incident stroke: the sleep heart health study. *Am J Respir Crit Care Med* 2010;182:269–77.
6 Ellen RL, Marshall SC, Palayew M, *et al*. Systematic review of motor vehicle crash risk in persons with sleep apnea. *J Clin Sleep Med* 2006;2:193–200.
7 Moyer CA, Sonnad SS, Garetz SL, *et al*. Quality of life in obstructive sleep apnea: a systematic review of the literature. *Sleep Med* 2001;2:477–91.
8 Tarasiuk A, Greenberg-Dotan S, Simon-Tuval T, *et al*. The effect of obstructive sleep apnea on morbidity and health care utilization of middle-aged and older adults. *J Am Geriatr Soc* 2008;56:247–54.
9 Barnes M, McEvoy RD, Banks S, *et al*. Efficacy of positive airway pressure and oral appliance in mild to moderate obstructive sleep apnea. *Am J Respir Crit Care Med* 2004;170:656–64.
10 Bazzano LA, Khan Z, Reynolds K, *et al*. Effect of nocturnal nasal continuous positive airway pressure on blood pressure in obstructive sleep apnea. *Hypertension* 2007;50:417–23.
11 Bradley TD, Floras JS. Obstructive sleep apnoea and its cardiovascular consequences. *Lancet* 2009;373:82–93.
12 National Institute for Health and Care Excellence. *Continuous positive airway pressure for the treatment of obstructive sleep apnoea/hypopnoea syndrome*. London: National Institute for Health and Care Excellence. TA139. http://www.nice.org.uk/nicemedia/live/11944/40085/40085.pdf
13 Weaver TE, Mancini C, Maislin G, *et al*. Continuous positive airway pressure treatment of sleepy patients with milder obstructive sleep apnea: results of the CPAP Apnea Trial North American Program (CATNAP) randomized clinical trial. *Am J Respir Crit Care Med* 2012;186:677–83.
14 Patel SR, White DP, Malhotra A, *et al*. Continuous positive airway pressure therapy for treating sleepiness in a diverse population with obstructive sleep apnea: results of a meta-analysis. *Arch Intern Med* 2003;163:565–71.
15 Weaver TE, Grunstein RR. Adherence to continuous positive airway pressure therapy: the challenge to effective treatment. *Proc Am Thorac Soc* 2008;5:173–8.

16 Lim J, Lasserson TJ, Fleetham J, *et al*. Oral appliances for obstructive sleep apnoea. *Cochrane Database Syst Rev* 2006:CD004435.
17 Phillips CL, Grunstein RR, Darendeliler MA, *et al*. Health Outcomes of Continuous Positive Airway Pressure versus Oral Appliance Treatment for Obstructive Sleep Apnea. *Am J Respir Crit Care Med* 2013;187:879–87.
18 Petri N, Svanholt P, Solow B, *et al*. Mandibular advancement appliance for obstructive sleep apnoea: results of a randomised placebo controlled trial using parallel group design. *J Sleep Res* 2008;17:221–9.
19 McDaid C, Griffin S, Weatherly H, *et al*. Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea-hypopnoea syndrome: a systematic review and economic analysis. *Health Technol Assess* 2009;13:iii-iv, xi-xiv, 1–119, 43–274.
20 Department of Health. NHS Reference costs (2011/12): Published 8 November 2012. https://www.gov.uk/government/publications/nhs-reference-costs-financial-year-2011-to-12
21 Personal Social Services Research Unit (PSSRU). Unit Costs of Health and Social Care (2012). http://www.pssru.ac.uk/project-pages/unit-costs/2012/index.php
22 Dolan P, Gudex C, Kind P, *et al*. *A social tariff for EuroQol: results from a UK general population survey University of York*. Center for Health Economics, 1995:1–24.
23 Sutherland K, Deane SA, Chan AS, *et al*. Comparative effects of two oral appliances on upper airway structure in obstructive sleep apnea. *Sleep* 2011;34:469–77.
24 Mehta A, Qian J, Petocz P, *et al*. A randomized, controlled study of a mandibular advancement splint for obstructive sleep apnea. *Am J Respir Crit Care Med* 2001;163:1457–61.
25 Vanderveken OM, Devolder A, Marklund M, *et al*. Comparison of a custom-made and a thermoplastic oral appliance for the treatment of mild sleep apnea. *Am J Respir Crit Care Med* 2008;178:197–202.
26 Gotsopoulos H, Chen C, Qian J, *et al*. Oral appliance therapy improves symptoms in obstructive sleep apnea: a randomized, controlled trial. *Am J Respir Crit Care Med* 2002;166:743–8.
27 de Almeida FR, Bittencourt LR, de Almeida CI, *et al*. Effects of mandibular posture on obstructive sleep apnea severity and the temporomandibular joint in patients fitted with an oral appliance. *Sleep* 2002;25:507–13.
28 Andren A, Hedberg P, Walker-Engstrom ML, *et al*. Effects of treatment with oral appliance on 24-h blood pressure in patients with obstructive sleep apnea and hypertension: a randomized clinical trial. *Sleep Breath* 2013;17:705–12.
29 Aarab G, Lobbezoo F, Hamburger HL, *et al*. Oral appliance therapy versus nasal continuous positive airway pressure in obstructive sleep apnea: a randomized, placebo-controlled trial. *Respiration* 2011;81:411–19.
30 Tegelberg A, Walker-Engstrom ML, Vestling O, *et al*. Two different degrees of mandibular advancement with a dental appliance in treatment of patients with mild to moderate obstructive sleep apnea. *Acta Odontol Scand* 2003;61:356–62.
31 Crawford MR, Bartlett DJ, Coughlin SR, *et al*. The effect of continuous positive airway pressure usage on sleepiness in obstructive sleep apnoea: real effects or expectation of benefit? *Thorax* 2012;67:920–4.
32 Vanderveken OM, Dieltjens M, Wouters K, *et al*. Objective measurement of compliance during oral appliance therapy for sleep-disordered breathing. *Thorax* 2013;68:91–6.
33 Marklund M, Verbraecken J, Randerath W. Non-CPAP therapies in obstructive sleep apnoea: mandibular advancement device therapy. *Eur Respir J* 2012;39:1241–7.
34 Lettieri CJ, Paolino N, Eliasson AH, *et al*. Comparison of adjustable and fixed oral appliances for the treatment of obstructive sleep apnea. *J Clin Sleep Med* 2011;7:439–45.
35 Schmidt-Nowara W. Recent Developments in Oral Appliance Therapy of Sleep Disordered Breathing. *Sleep Breath* 1999;3:103–06.

Manuscript ID thoraxjnl-2014-205464 – SUPPLEMENTARY INFORMATION

REVISION – CLEAN COPY

A Crossover Randomised Controlled Trial of Oral Mandibular Advancement Devices for Obstructive Sleep Apnoea-Hypopnoea (TOMADO)

Timothy G. Quinnell, Maxine Bennett, Jake Jordan, Abigail L. Clutterbuck-James, Michael G. Davies, Ian E. Smith, Nicholas Oscroft, Marcus A. Pittman, Malcolm Cameron, Rebecca Chadwick, Mary J. Morrell, Matthew Glover, Julia A. Fox-Rushby, Linda D. Sharples

Online Data Supplement

E1. Inclusion/ Exclusion Criteria

Inclusion Criteria

- Age ≥18 years old.
- Obstructive sleep apnoea-hypopnoea (OSAH) confirmed by respiratory or complete polysomnography (PSG) with an apnoea hypopnoea index (AHI) of 5 - <30/hour.
- Excessive daytime sleepiness: Epworth Sleepiness Scale (ESS) score of ≥9.

Exclusion Criteria

- Central sleep apnoea as predominant form of sleep disordered breathing
- Coexistent sleep disorder, poor sleep hygiene or drug treatment considered likely to have significant impact on symptoms (especially sleepiness) or assessment of mandibular advancement device (MAD) effectiveness.
- Severe and/or unstable cardiovascular disease judged by clinician to warrant immediate continuous positive airway pressure (CPAP).
- Other medical or psychiatric disorder judged likely to adversely interact with MAD or confound interpretation of its effectiveness.
- Significant periodontal disease or tooth decay; partial or complete edentulism; presence of fixed orthodontic devices.
- Temporomandibular joint pain or disease
- Clinical history suggestive of severe bruxism
- Restriction in mouth opening or advancement of mandible.
- Respiratory failure
- Inability to give informed consent or comply with the protocol

- Pregnancy
- Previous exposure to MAD treatment
- Disabling sleepiness leading to significant patient-specific safety concerns

E2. Experimental Interventions

Three types of MAD of varying sophistication, complexity and cost, were made before the patient was randomised.

1 SleepPro 1 (SP1) (*Meditas Ltd., Winchester, UK*): A thermoplastic 'boil and bite' device fitted by the patient following the manufacturer's printed instructions. The patient softened the device in hot water then placed it into their mouth and, having bitten down on it, advanced the mandible to an individually-determined 'comfortable' position. The device was then manually moulded against the teeth and set by immersion in cold water. Rewarming allowed remoulding. (http://www.sleeppro.com).

2 SleepPro 2 (SP2) *(Meditas Ltd., Winchester, UK)*: A semi-bespoke device, formed from a dental impression mould made by the patient. An impression kit was posted to the patient. It consisted of a SP1 with holes to allow the injection of dental putty. The patient was instructed to mould the SP1 (as for the SP1 device) then wear it for two nights to ensure optimum position and fit, remoulding if necessary. The patient then made up the putty and injected it into the SP1, sending the resulting impression back to the manufacturer. The SP2 was produced from this mould. It was designed to grip the entire dentition. Thinner walls than the SP1 were intended to result in a more comfortable fit. Involvement of the patient's dentist in taking the impression was suggested if necessary (http://www.sleeppro.com).

3 Bespoke Device (bMAD) *(Maxillofacial Laboratory, Department of Oral and Maxillofacial Surgery, Cambridge, UK)*: Custom made MAD, professionally fitted by specialists in the NHS Maxillofacial laboratory at Addenbrooke's Hospital, UK. A positional 'wax bite' was taken from the patient and the degree of mandibular advancement (50-70% of the maximal protrusive distance from centric occlusion, i.e. the "normal" bite where the teeth all interdigitate maximally) was determined. Upper and lower full dental impressions were taken in alginate by a suitably qualified dental professional and cast in dental stone. The casts were trimmed and articulated using the positional wax bite. A blow down splint in soft acrylic was created upon each cast and then fused with a further acrylic blow down to ensure the upper and lower dentition are positioned in the predetermined optimal position to hold the mandible forward. The patient returned roughly 2 weeks later for the fitting. The fitting allowed for optimal balance between advancing the mandible sufficiently to bring the tongue base off the posterior pharyngeal wall, and patient comfort. Device adjustment, short of remoulding, was available during treatment.

The SP2 and bMAD were made before randomisation. Protrusion of the SP1 and SP2 devices was determined by the patient, according to manufacturer's instructions. Patients attended the Maxillofacial Laboratory for the impressions of the bMAD approximately 1-2 weeks after their baseline visit with fitting 1-2 weeks later. The bMAD was fitted by dental experts, aiming for maximal comfortable mandibular advancement and at least 50% maximal protrusion. In order to ensure devices were not used outside the designated treatment period the bMAD was sent directly to the

study team at Papworth Hospital, and all devices were held there outside their relevant trial period.

E3. Randomisation

A commonly used randomisation strategy for crossover trials is based on Latin Squares designs in which patients are randomised in blocks of 4, with each treatment being represented in each period. These designs are both efficient and well balanced for period. Williams' Latin Squares are particular types of Latin Squares that are efficient and have attractive properties if some of the patients fail to complete all 4 periods (providing most patients do complete all periods). For this reason the randomisation was based on 2 related Williams' Latin Squares designs, with patients randomised in blocks of 8 to ensure good treatment by period balance. Sequences for each block of 8 patients were as follows.

Sequence	Period 1	Period 2	Period 3	Period 4
1	A	C	D	B
2	B	D	C	A
3	C	B	A	D
4	D	A	B	C
5	A	D	C	B
6	B	C	D	A
7	C	A	B	D
8	D	B	A	C

Although randomisation in blocks of 8 meant that for every eighth patient the sequence was predictable, this was considered to be less important for a crossover trial.

E4. Secondary outcome questionnaires

Subjective daytime sleepiness was measured using the Epworth Sleepiness Score (ESS) questionnaire. The questionnaire asks patients to rate their chances of dozing off (on a scale of 0 to 3, where 0 would never doze, and 3 would have a high chance of dozing) in eight different scenarios over the past 4 weeks. The questionnaire has been shown to significantly distinguish normal subjects from patients in various diagnostic groups including OSAHS [E1], and can be used clinically to demonstrate response to CPAP treatment [E2].

Sleep-related quality of life was assessed by the Functional Outcomes of Sleep Questionnaire (FOSQ) and the Calgary Sleep Apnoea Quality of Life Index (SAQLI). The FOSQ provides a subjective assessment of how excessive sleepiness impacts upon daytime function [E3]. It consists of 30 questions covering five topics: activity level, vigilance, intimacy and sexual relationships, general productivity, and social outcome. It has been used as an outcome measure in therapeutic intervention trials for OSAHS [E4, E5], and FOSQ scores have been shown to significantly improve in patients on treatment for OSAHS compared to before treatment [E6].

The SAQLI contains 35 questions grouped into four dimensions: daily function, social interactions, emotional functioning and symptoms. It was specifically developed to evaluate health related quality of life for sleep apnoea patients in clinical trials [E7], and can detect significant differences in quality of life in response to treatment with either CPAP or MAD [E8].

Generic health related quality of life was assessed by the Medical Outcomes Study 36-item Short Form (SF-36) and the EuroQol 5-dimensions, 3 Levels of severity instrument (EQ-5D-3L). The SF-36 measures eight multi-item dimensions: physical

functioning, role limitations due to physical problems, role limitations due to emotional problems, social functioning, mental health, energy/vitality, pain and general health perception. It has been shown to detect substantial adverse effects on the subjective health of patients with OSAHS as well as detecting improvements in relation to treatment with CPAP [E9, E10]. The SF36 health state responses were converted to the Short Form 6 Dimensions (SF-6D) utility scale using values from a random sample of the general population [E11]. These utility values were used to calculate SF6D quality adjusted life years (QALYs) as part of the sensitivity analysis conducted for the economic evaluation.

The EQ-5D-3L was developed by a group of researchers from five European countries. It consists of five questions relating to mobility, self-care, usual activity, pain/discomfort and anxiety/depression, along with a visual analogue scale known as the 'Euro-thermometer' (EQ-VAS) for participants to indicate their current health state [E12]. The EQ-5D-3L is used to calculate QALYs for health economic evaluation of therapeutic interventions in health care systems. However, its applicability to sleep-related disorders has been queried as it has no questions that specifically address the aspects of life thought to be affected by OSAHS [E9]. The EQ-VAS (a thermometer like visual analogue scale) is considered a useful addition to the questionnaire as it is considered to reflect the disease state of OSAHS patients better than questionnaires alone [E13].

E5. Study outcomes

The primary outcome was AHI, measured by domiciliary rPSG (Embletta, Medcare) and scored manually according to The American Academy of Sleep Medicine (AASM) criteria [E14] by a Polysomnographer blinded to treatment. Subjective daytime sleepiness was measured using the ESS questionnaire. Other outcomes included rPSG indices (4% Oxygen Desaturation Index (ODI), mean, minimum and time <90% of nocturnal oxygen saturation (SpO2)) and blood pressure. Sleep-related quality of life was assessed by the Functional Outcomes of Sleep Questionnaire (FOSQ) and the Calgary Sleep Apnoea Quality of Life Index (SAQLI). Generic HRQoL was assessed by the Short Form 36 (SF-36) and the EuroQol (EQ-5D-3L). Data on healthcare use were collected on a study specific form and included type of device, number and type of visits, admissions (with cause) and length of stay, telephone calls, and diagnostic tests. Driving and road traffic accident questionnaires informed cost-effectiveness modelling. Treatment compliance was recorded in sleep diaries by participants as hours of use per night. Treatment satisfaction was recorded by patients using a visual analogue scale (VAS) and, when applicable, snoring by bed partners. Device preference was recorded at trial exit. Adverse events and withdrawals were reviewed by an independent sleep physician.

E6. Statistical analysis

A mixed effects Poisson regression was used to assess the effects of treatment, period and treatment by period interactions (fixed effects). Gamma (1,α) patient random effects were included. Patients were classified as complete responders (AHI <5), partial responders (≥50% reduction in AHI from baseline but AHI ≥5) or non-responders for each MAD. Mixed effects logistic regression was used to analyse response (complete or partial) with Normal random effects for patients (on the logit scale) and fixed effects for treatment, period and their interaction. All other outcomes were analysed using linear mixed models with patients included as (Normal) random effects and treatment, period and their interaction as fixed effects.

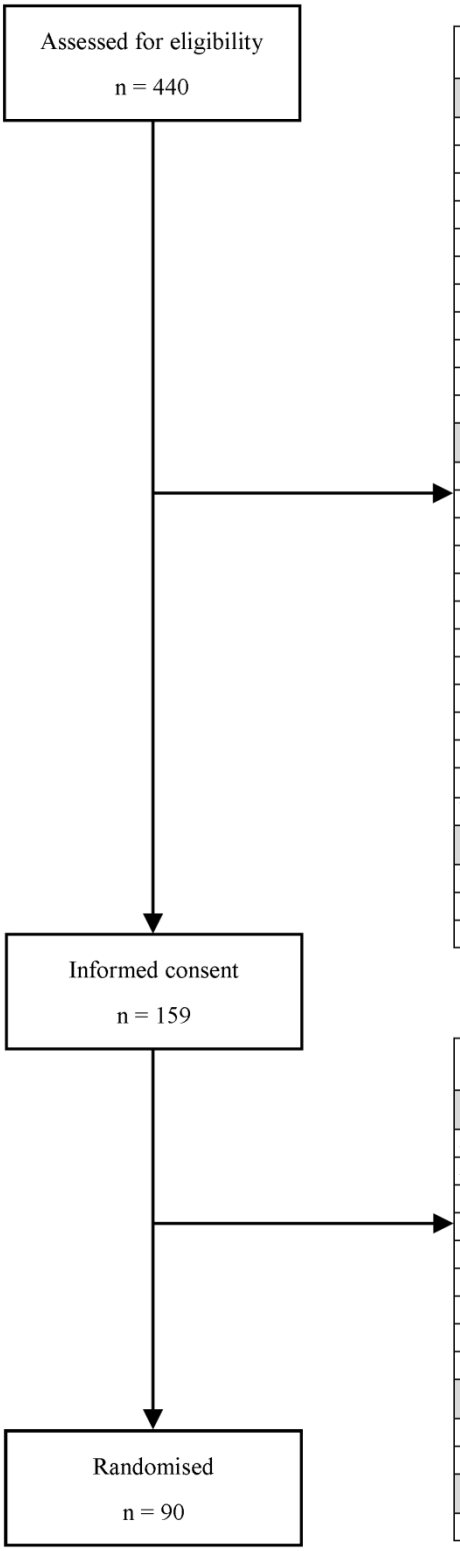
The effects on AHI response of baseline AHI, ESS, age, gender, compliance and BMI, and BMI changes over time were explored using regressions. There were no subgroup analyses.

Our approach to multiple testing was as follows. For each of the general(ised) mixed models performed treatment effects were described as "statistically significant" if the global test comparing the model against one with no treatment effects was less than 0.05. Our protocol states that comparison of each MAD against no treatment was important so that, for models that were "significant" overall, we present the significance level without adjustment. For comparisons between MAD the (conservative) Bonferroni correction should be applied, that is, standard p-values for these comparisons should be multiplied by 3. We have not routinely applied corrections so that readers may make their preferred corrections and have indicated where our results are uncorrected.

All data analyses were made using Stata 13 statistical software (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP).

E7. Reasons for exclusion

Figure E1: Reasons for exclusion at screening and after baseline visit



Reason for exclusion at screening	Number of patients (n = 281)
INELIGIBLE	**184**
Significant periodontal disease	51
Coexistent sleep disorder	39
Disabling sleepiness	35
Previous MAD	26
Medical/ psychiatric disorder	11
Severe cardiovascular disease	8
Inability to consent/ comply	6
Previous/ current CPAP (>6 weeks or 1st review)	5
Respiratory failure	1
Severe bruxism	1
Temporomandibular joint pain	1
REFUSAL	**87**
Work commitments	26
Too time consuming	18
Prefers CPAP	13
Not keen on MAD/ device in mouth	13
Too far to travel	4
Doesn't want to participate	3
Family commitments	3
Other commitments	2
Wants to try weight loss first	1
Wants to watch and wait	1
Other comorbidities	1
Querying OSA diagnosis	1
No reason given	1
OTHER	**10**
Not referred to research team	6
Unable to contact	3
Recruitment completed	1

Reason for exclusion following baseline visit	Number of patients (n = 69)
INELIGIBLE	**63**
AHI > = 30	43
AHI < 5	6
Inability to consent/ comply	5
Significant periodontal disease	3
Coexistent sleep disorder	2
Disabling sleepiness	1
Medical/ psychiatric disorder	1
Severe bruxism	1
Severe cardiovascular disease	1
REFUSAL	**5**
Unable to tolerate SP2 moulding process	4
Family commitments	1
OTHER	**1**
Wants bariatric surgery	1

E8. Summary of Withdrawals

Sixteen patients withdrew from the TOMADO trial. The reasons for withdrawal are described in the table below.

Table E2: Summary of patients who withdrew from the trial

Patient ID	Period	Reason	Explanation	Future care
005	1 (bMAD)	AE - Clinical decision	Bleeding gums due to poor oral hygiene	Discharged, conservative management
012	4 (bMAD)	Consent withdrawn	No time for final visit	Assigned SP1
013	between 1 + 2 (SP1 and bMAD)	Lost to Follow-up	Could not contact patient	Discharged
014	3 (SP2)	Consent withdrawn	Time constraints - did not complete SP2 moulding	Continue with current device
017	between 1 + 2 (SP1 and SP2)	Other	Withdrawn due to unreliability	Continue with current device
024	1 (control)	Consent withdrawn	Could not tolerate SP2 moulding	No further treatment
039	3 (SP2)	Other	Unable to attend visits	Weight Loss
040	1 (bMAD)	Consent withdrawn	Patient unable to complete trial visits due to co-morbidities	Start CPAP
042	1 (bMAD)	AE - Patient decision	Concerns about crowns & bridges moving/ breaking	Start CPAP
043	4 (bMAD)	Lost to Follow-up	Could not contact patient	Recommend use SP2
047	1 (SP1)	Consent withdrawn	Did not like device and did not want to try any others	Weight Loss
049	1 (SP1)	AE - Patient decision	Broke tooth crown whilst wearing device	Start CPAP
050	2 (SP1)	Consent withdrawn	Did not like device and did not want to try any others	Start CPAP
066	3 (bMAD)	Consent withdrawn	Personal issues	Start CPAP
086	4 (bMAD)	Lost to Follow-up	Could not contact patient	Recommend use SP2
089	1 (SP2)	Consent withdrawn	Did not like device and did not think it worked	Start CPAP

Table E3: Summary of baseline characteristics for patients who withdrew from the study

	Unit/Category	Total (n=16)	Min	Max
Demographic Information				
Gender	Male	11 (69%)		
	Female	5 (31%)		
Age at randomisation	Years	52·8 (9·8)	35·3	71·9
Smoking history	Non-smoker	6 (38%)		
	Ex-smoker	10 (63%)		
Clinical Measurements				
BMI	Kg/m^2	30·6 (29·6 - 35·8)	27·1	47·8
Sleep Study				
AHI (1) baseline	Events per hour	14·1 (6·2)	6·3	27·5
	Missing *	1		
ODI (2) baseline	Events per hour	10·5 (5·8)	3·3	20·8
Minimum SpO_2	Percent	83·7 (3·9)	75	89
	Missing*	1		
Mean SpO_2	Percent	93·7 (1·5)	89·8	96·5
	Missing*	1		
Time <90% of nocturnal SpO_2	Minutes	9·7 (2 - 31)	0·8	315·4
	Missing*	1		
Epworth Sleepiness Score				
ESS (3)	Unit score	12·1 (4·9)	3	20

E9. AHI for complete cases

Table E4: Summary of results from mixed effects model for AHI using complete cases (n=68)

	Coefficient	95% CI	P value	Global P value
Constant	14·23	(11·67, 17·36)	<0·001	
No treatment	-	-	-	<0·001
SP1	0·75	(0·63, 0·89)	0·001	
SP2	0·70	(0·61, 0·81)	<0·001	
bMAD	0·67	(0·55, 0·81)	<0·001	
Time period 1	-	-	-	0·688
Time period 2	1·06	(0·92, 1·21)	0·418	
Time period 3	0·97	(0·81, 1·17)	0·758	
Time period 4	1·05	(0·86, 1·28)	0·659	

Footnote: The table (E4) shows evidence that all MADs are effective in reducing AHI compared to no treatment·.

Table E5: Comparison of AHI between MAD for complete cases (n=68)

Comparison	Observed contrast	95% CI	P value
SP2 to SP1	0·94	(0·81, 1·08)	0·379
bMAD to SP1	0·89	(0·72, 1·10)	0·276
bMAD to SP2	0·95	(0·79, 1·13)	0·569

Footnote : Comparisons between different MAD have not been corrected for multiple testing in this table. In order to retain overall type I error of 5% comparisons in this table will be described as “significant” if $p<0.017$.

E10. Predictors of response

Table E6. Summary from mixed effects logistic regression models to determine predictors of complete or partial response

	Odds ratio (bootstrap SE)	**95% CI**	**P value**
Age at randomisation (n=81)	0.97 (0.03)	(0.92, 1.02)	0.287
Sex (n=81)	0.86 (0.64)	(0.20, 3.73)	0.836
Baseline BMI (n=81)	0.89 (0.04)	(0.81, 0.98)	0.014
Baseline AHI (n=81)	1.01 (0.05)	(0.92, 1.11)	0.833
Baseline ESS (n=81)	0.95 (0.07)	(0.82, 1.09)	0.450
Compliance (n=80)	1.00 (0.004)	(0.99, 1.01)	0.783
BMI throughout study (n=81)	0.88 (0.04)	(0.80, 0.96)	0.007
Percentage protrusion (n=79)	1.03 (0.01)	(1.00, 1.05)	0.034

E11. Other sleep study outcomes

Table E7: Complete or partial response to treatment (n=81)

	Odds ratio (S.E)	**95% CI**	**P value**	**Global P value**
Constant	0·12	(0·04, 0·32)	<0·001	
No treatment (A)	-	-		0·0006
SP1 (B)	2·90	(1·16, 7·25)	0·022	
SP2 (C)	5·75	(2·48, 13·33)	<0·001	
bMAD (D)	4·64	(1·79, 12·02)	0·002	
Time period 1	-	-		0·532
Time period 2	1·51	(0·60, 3·80)	0·381	
Time period 3	1·84	(0·81, 4·19)	0·144	
Time period 4	1·23	(0·54, 2·82)	0·621	

Table E8: Summary of results from mixed effects model for 4% oxygen desaturation rate (n=81)

	Coefficient (Bootstrap S.E)	**95% CI**	**P value**	**Global P value**
Constant	11·03	(9·00, 13·52)	<0·001	
No treatment	-	-	-	<0·001
SP1	0·75	(0·60, 0·92)	0·007	
SP2	0·65	(0·55, 0·77)	<0·001	
bMAD	0·60	(0·50, 0·72)	<0·001	
Time period 1	-	-	-	0·951
Time period 2	1·06	(0·83, 1·35)	0·634	
Time period 3	1·00	(0·83, 1·19)	0·984	
Time period 4	1·02	(0·85, 1·23)	0·825	

Table E9: Comparison of 4% oxygen desaturation rate between MADs (n=81)

Comparison	Observed contrast	95% CI	P value
SP2 to SP1	0·87	(0·73, 1·05)	0·144
bMAD to SP1	0·80	(0·64, 1·00)	0·051
bMAD to SP2	0·92	(0·78, 1·09)	0·319

Table E10: Mean (SD) of the minimum oxygen saturation (SpO_2) by treatment

Treatment	N	Mean of the minimum SpO_2 (S.D)	Min	Max
No treatment	77	84·0 (5·4)	69	92
SP1	76	84·3 (5·6)	63·7	91
SP2	76	84·8 (5·4)	64	92
bMAD	75	86·0 (4·5)	68	93

Table E11: Mean (SD) of mean oxygen saturation (SpO_2) by treatment

Treatment	N	Mean of mean SpO_2(S.D)	Min	Max
No treatment	77	94·3 (1·1)	90·7	96·6
SP1	76	94·2 (1·1)	91·5	96·6
SP2	77	94·3 (1·2)	91·0	97·1
bMAD	75	94·3 (1·2)	91·0	97

Table E12: Median Time <90% of nocturnal oxygen saturation (SpO_2) by treatment

Treatment	N	Median Time <90% of nocturnal SpO_2 minutes (25th percentile , 75th percentile)	Min	Max
No treatment	77	8·4 (2·1, 19·9)	0	176·5
SP1	76	5·6 (1·7, 17·7)	0	95·6
SP2	77	5·8 (1·2, 14·2)	0	130·1
bMAD	75	3·1 (0·6, 13)	0	96·7

E12. Epworth Sleepiness Scale (ESS) for complete cases

Table E13: Summary of results from mixed effects model for ESS (n=74)

	Coefficient (S.E)	**P value**	**Global P value**	**95% CI**
Constant	10·58 (0·54)	<0·001		(9·53, 11·63)
No treatment (A)	-	-	<0·001	
SP1 (B)	-1·47 (0·41)	<0·001		(-2·28, -0·66)
SP2 (C)	-2·30 (0·42)	<0·001		(-3·13, -1·46)
bMAD (D)	-2·41 (0·43)	<0·001		(-3·26,-1·56)
Time period 1	-	-	0·172	
Time period 2	-0·65 (0·38)	0·086		(-1·40, 0·09)
Time period 3	-0·91 (0·42)	0·029		(-1·73, -0·09)
Time period 4	-0·49 (0·43)	0·258		(-1·33, 0·36)

Table E14: Comparison of ESS between MADs

Comparison	Observed contrast	95% CI	P value
SP2 to SP1	-0·82	(-1·62, -0·02)	0·044
bMAD to SP1	-0·94	(-1·72, -0·16)	0·019
bMAD to SP2	-0·11	(-0·86, 0·63)	0·762

Note that comparisons between different MAD have not been corrected for multiple testing in this table. In order to retain overall type I error of 5% comparisons in this table will be described as “significant” if $p<0.017$.

E13. Adverse events

Table E15: Serious adverse events

Adverse event	Dates	Treatment receiving at the time	Classification
Sick Sinus Syndrome and Atrial Flutter	25/09/2011 – 28/09/2011	No treatment	Possibly related to OSAHS
Hypoglycaemia	13/10/2011 – 13/10/2011	No treatment	Possibly related to OSAHS
Complete Heart Block	03/11/2011 – 04/11/2011	bMAD	Possibly related to OSAHS or MAD
Non-Specific Chest Pain	11/02/2012 – 17/02/2012	bMAD	Possibly related to OSAHS or MAD

Table E16: Minor Adverse events

Type of adverse event	**No treatment (n=78)**	**SP1 (n=81)**	**SP2 (n=78)**	**bMAD (n=77)**	**Total**
General (1)	32 (24)	38 (24)	35 (25)	34 (26)	139 (47)
Dryness/Bad taste/Numbness (2)	12 (10)	26 (20)	30 (24)	21 (18)	89 (39)
Discomfort/ Mouth problems (3)	18 (13)	135 (60)	124 (52)	148 (74)	425 (83)
Excessive salivation (4)	2 (2)	37 (32)	19 (18)	34 (29)	92 (48)
Cold related (5)	14 (13)	25 (17)	34 (26)	24 (18)	97 (46)
Infection (6)	2 (2)	6 (6)	0 (0)	1 (1)	9 (8)
Total	80 (45)	267 (73)	242 (68)	262 (76)	851 (86)

Table E17: Minor adverse events (classified by an independent sleep physician)

Type of adverse event	**No treatment (n=78)**	**SP1 (n=81)**	**SP2 (n=78)**	**bMAD (n=77)**	**Total**
Possibly related to OSAHS	3 (3)	1 (1)	3 (3)	1 (1)	8 (5)
Probably related to MAD	22 (16)	179 (66)	143 (59)	184 (75)	528 (85)
Possibly related to OSAHS or MAD	29 (18)	50 (34)	55 (35)	40 (27)	174 (54)
Probably unrelated	26 (21)	37 (24)	41 (30)	37 (27)	141 (59)
Total	80 (45)	267 (73)	242 (68)	262 (76)	851 (86)

E14. Driving

Eighty-seven (97%) patients in the TOMADO study reported that they drove at baseline. Eighty-six drove a car, two a motorbike, three a HGV and 16 drove other vehicles (including: fork lift truck, van, minibus and tractor). There were five driving collisions during the trial. No collisions resulted in an injury to anyone else involved. One collision resulted in whiplash injury to the patient, who required treatment and advice from a health care professional. Three incidents were classed as not the fault of the patient and were caused by other drivers colliding with them. The other two incidents involved patients driving into the back of the car in front due to distraction rather than sleepiness.

Table E18: patient reported sleepiness associated with driving

		Treatment			
		No treatment (n=75)	**SP1 (n=78)**	**SP2 (n=75)**	**bMAD (n=74)**
Sleepy whilst driving	Never	43 (59%)	54 (72%)	55 (75%)	56 (78%)
	Rarely	16 (22%)	11 (15%)	13 (18%)	5 (7%)
	Occasionally	11 (15%)	9 (12%)	5 (7%)	10 (14%)
	Frequently	3 (4%)	1 (1%)	0	1 (1%)
	Always	0	0	0	0
	Missing*	2	3	2	2

Nodded off driving	Yes	1 (1%)	1 (1%)	1 (1%)	0
	No	72 (99%)	74 (99%)	72 (99%)	72 (100%)
	Missing	2	3	2	2
Pulled off road	Yes	11 (15%)	4 (5%)	7 (10%)	4 (6%)
	No	62 (85%)	71 (95%)	66 (90%)	68 (94%)
	Missing	2	3	2	2
Collisions	Yes	1 (1%)	0	2 (3%)	2 (3%)
	No	72 (99%)	75 (100%)	71 (97%)	70 (97%)
	Missing	2	3	2	2

*All missing data were because the patient had not driven in the past 4 weeks

E15. Device measurements

Table E19: Device Measurements

	Unit/Category	Total (n=90)	Min	Max
Malocclusion type	Class I (Neutrocclusion)	51 (60%)		
	Class II (Distocclusion)	28 (33%)		
	Class III (Mesiocclusion)	6 (7%)		
	Missing	5		
Overjet	mm	3·49 (2·50)	-4·5	10
	Missing	1		
Maximum protrusion	mm	9·06 (2·06)	5	14
	Missing	2		
Measured protrusion for bMAD device	mm	4·99 (1·89)	1	10
	Missing	9		
Percentage protrusion for bMAD device	%	55·18 (19·72)	9·09	100
	Missing	11		

Measured protrusion for SP2 device	mm	4·75 (2·50)	-2	11·5
	Missing	11		
Percentage protrusion for SP2 device	%	51·66 (26·42)	-25	100
	Missing	12		
Measured protrusion for SP1 device	mm	5·65 (2·12)	1	11
	Missing	14		
Percentage protrusion for SP1 device	%	62·63 (22·08)	10	100
	Missing	15		

E16. Treatment Evaluation

Treatment comfort was measured on a 0-100 scale with 0 being very uncomfortable and 100 being very comfortable.

Table E20: Median treatment comfort

Treatment	n	Median treatment comfort (IQR)	Min	Max
No treatment	78	50 (50 – 97)	1	100
SP1	81	34 (16 – 50)	0	91
SP2	78	52 (36 – 82)	0	100
bMAD	77	50 (25 – 76)	0	97

Overall treatment satisfaction was also measured on a 0-100 scale with 0 being very dissatisfied and 100 being very satisfied.

Table E21: Median treatment satisfaction

Treatment	n	Median treatment satisfaction (IQR)	Min	Max
No treatment	78	50 (25 – 50)	0	100
SP1	81	43 (14 – 65)	0	99
SP2	78	67 (41 – 87)	0	100
bMAD	77	71 (38 – 87)	0	100

Table E22: Patient-estimated device retention during treatment period

Treatment	On average, how often did the device fall out?	Frequency (%)
SP1 (n=81)	Never fell out	27 (33%)
	Fell out occasionally, but not every night	35 (43%)
	Fell out 1-2 times every night	11 (14%)
	Fell out more than 2 times every night	8 (10%)
SP2 (n=78)	Never fell out	43 (56%)
	Fell out occasionally, but not every night	26 (34%)
	Fell out 1-2 times every night	5 (6%)
	Fell out more than 2 times every night	3 (4%)
bMAD (n=77)	Never fell out	51 (66%)
	Fell out occasionally, but not every night	22 (29%)
	Fell out 1-2 times every night	4 (5%)
	Fell out more than 2 times every night	0 (0%)

*1 missing value for the SP2 as the patient didn't wear the device for longer than 1 minute

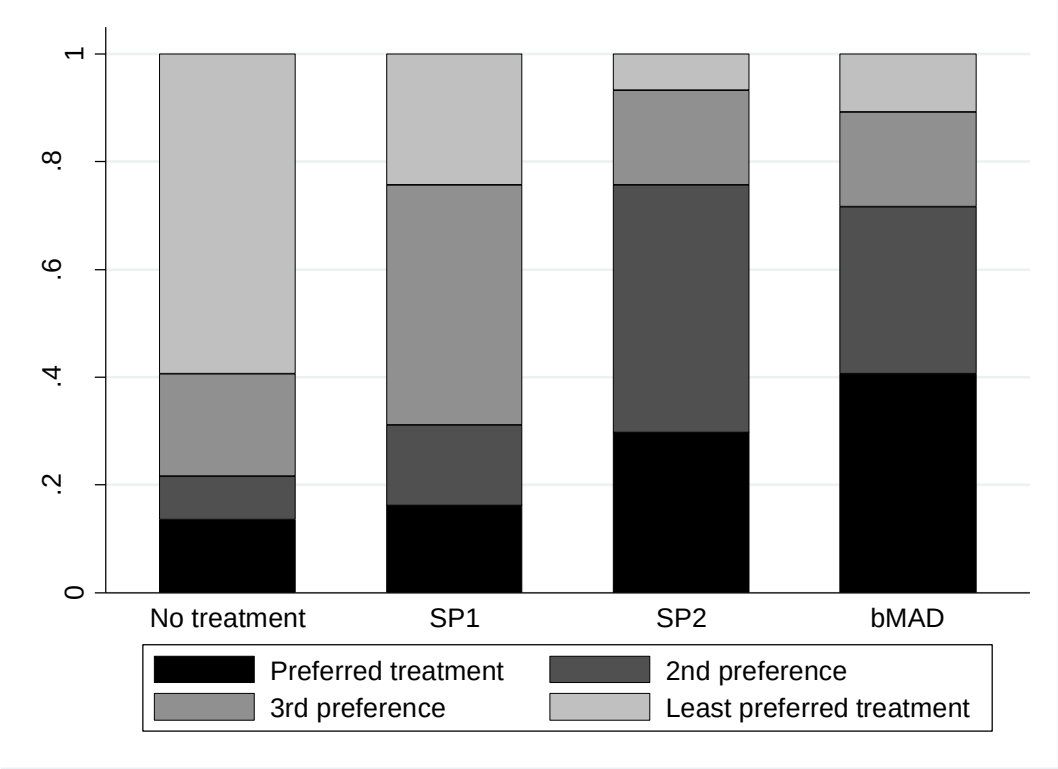
Table E23: Patient-estimated device removal during treatment period

Treatment	On average, how often was the device removed?	Frequency (%)
SP1 (n=81)	Never removed	25 (31%)
	Removed 1-3 nights/week	33 (41%)
	Removed 4-6 nights/week	10 (12%)
	Removed every night	13 (16%)
SP2 (n=78)	Never removed	40 (52%)
	Removed 1-3 nights/week	23 (30%)
	Removed 4-6 nights/week	9 (12%)
	Removed every night	5 (6%)
bMAD (n=77)	Never removed	34 (44%)
	Removed 1-3 nights/week	28 (36%)
	Removed 4-6 nights/week	8 (10%)
	Removed every night	7 (9%)

*1 missing value for the SP2 as the patient didn't wear the device for longer than 1 minute

E18. Patient treatment preferences

Figure E24: Bar chart showing patient preference for n=74 patients who completed the trial



E18. Quality of life indices

Table E25: Quality of life summaries

Mean (SD)	No treatment (n=78)	SP1 (n=81)	SP2 (n=78)	bMAD (n=77)
FOSQ				
General Productivity	3·48 (0·45)	3·57 (0·44)*	3·66 (0·40)*	3·73 (0·36)*
Social Outcome	3·53 (0·58) ‡	3·61 (0·58) †	3·71 (0·53)* ‡	3·74 (0·49)* §
Activity Level	3·11 (0·68)	3·25 (0·59)*	3·37 (0·53)*	3·40 (0·48)*
Vigilance	3·25 (0·57)	3·33 (0·54)	3·48 (0·47)*	3·53 (0·42)*
Intimate Relationships	3·20 (0·87) ‖	3·34 (0·80) ‖	3·45 (0·73)* ¥	3·49 (0·68)* ††
Total Score	16·62 (2·55)	17·13 (2·42)*	17·70 (2·14)*	17·90 (1·92)*
SAQLI				
Daily Activities	4·83 (1·49)	5·16 (1·38)*	5·56 (1·23)*	5·47 (1·33)*
Social Interactions	5·31 (1·25)	5·49 (1·34)	5·85 (1·16)*	5·89 (1·12)*
Emotions	5·40 (1·25)	5·46 (1·25)	5·70 (1·25)*	5·79 (1·09)*
Symptoms	4·47 (1·72)	4·82 (1·59)*	5·23 (1·52)*	5·37 (1·47)*
Total Score	5·01 (1·24)	5·25 (1·20)*	5·60 (1·12)*	5·64 (1·06)*
SF-36				
Physical	82·37 (23·56)	81·42 (23·27)	83·91 (22·53)*	81·36 (23·49)

Function				
Role physical	66·99 (39·58)	68·83 (36·13)	75·00 (37·15)	72·40 (38·39)
Bodily pain	70·04 (24·26)	70·09 (25·21)	75·56 (24·05)*	70·65 (28·93)
General health	61·45 (23·13)	60·38 (22·57)	62·21 (22·85)	62·27 (22·04)
Vitality	42·95 (23·86)	45·80 (21·94)	51·67 (22·11)*	54·03 (21·35)*
Social function	69·90 (21·42)	69·54 (20·30)	71·31 (21·69)	70·62 (22·07)
Role emotional	79·91 (34·96)	77·08 (31·19) †	80·13 (34·30)	80·09 (32·56)
Mental health	71·33 (18·25)	73·98 (17·00)	72·97 (18·43)	75·06 (18·02)
Physical component	43·06 (12·86)	42·73 (12·22) †	45·11 (12·33)*	43·12 (13·81)
Mental component	46·20 (10·78)	46·87 (9·63) †	47·34 (11·24)	48·81 (9·00)*
EQ-5D-3L				
Utility score	0·85 (0·20)	0·86 (0·20)	0·86 (0·23)	0·87 (0·19)
EQ-VAS score	74.32 (17.04)	73.77 (17.38)	77.00 (14.67)*	77.29 (15.06)

* indicates P<0·05 compared with no treatment

† indicates n=80 for this component due to missing data

‡ indicates n=77 for this component due to missing data

§ indicates n=76 for this component due to missing data

‖ indicates n=72 for this component due to missing data

¥ indicates n=71 for this component due to missing data

†† indicates n=70 for this component due to missing data

E19. Summary of economic analysis

For each individual and each intervention, unit costs (see Table E26) were multiplied by resource use (recorded on case report forms) to give total costs for each item, accounting for device use of 4 weeks over an expected life span (52 weeks for SP1 and SP2 and 78 weeks for bMAD). These were summed by intervention and divided by the number of participants in each intervention group for a raw (i.e. unadjusted for differences at baseline) average cost per participant by intervention (see Table E27). As the treatment period was a fixed 4 week duration for each intervention and generic quality of life measures (EQ-5D-3L and SF36) were only collected at one time point for each, the 4-week QALYs are calculated as a 4 week proportion of the 52 week year using the 'area under the curve' method. QALYs were not annualised given the short time period.

Two mixed effects models were used to estimate within-patient differences in total costs and within-patient differences in total QALYs. The mean costs of each treatment and the 'no treatment' control group were compared. Baseline EQ-5D-3L scores, patient weight and the time-period were included as covariates. As all coefficients on time period were insignificant, a period effect could be excluded from the model. Results in Table E28 show that differences in EQ-5D-3L QALYs for each MAD against no treatment were not statistically significant ($p<0{\cdot}05$). However, the SP2 showed the greatest change, using SF6DQALYs, compared with control and was also the only intervention with a statistically significant difference ($p = 0.013$). As residuals were not serially correlated, coefficients should not be biased and are therefore usable as a measure of uncertainty in sensitivity analysis.

Figure E29 shows that, comparing each intervention with control group over the 4 week intervention period, the SP1 was £4 less (SE 21), and the SP2 was £15 less on average (SE 21) but the mean cost of bMAD was £26 greater (SE 28) (see Table E29). Differences were not statistically significant. Figure E30 shows box plots of total costs for each group.

The incremental cost-effectiveness ratio (ICER) was estimated for each MAD against no treatment as the mean of within-patient difference in total 4-week costs, divided by the within-patient difference in 4-week QALYs. Table E31 shows that the ICERs were negative for SP1 and SP2 compared with control i.e. costs were lower and outcomes better for the two interventions compared with no treatment. Of these two, SP2 is more beneficial as costs were lower than SP1. Table E28 also shows that bMADs have the greatest impact on QALY gain, but at a cost of £14,900 per additional QALY gained and would therefore be considered a cost-effective buy compared with control. However, compared with SP2, bMAD costs an additional £46,000 per QALY (£105-£64)/(0.0667-0.0658 QALYs). These results are mirrored by the net monetary benefit, which shows that SP2 achieved the highest INMB, compared with no treatment, at £33 per 4 weeks assuming a willingness-to-pay (WTP) of £20,000 per QALY (Table E31). For comparisons between each treatment, the incremental net monetary benefit (INMB), over 4 weeks, was estimated across a range of values of decision-makers' WTP per QALY (see main text, Figure 3).

Probabilistic sensitivity analysis was conducted to incorporate the uncertainty in estimates cost and effects. Samples (with replacement) of patients were generated and for each sample the mixed effect model was rerun and unit costs were resampled from the estimated Gamma distributions. Two thousand bootstrap samples produced a set of possible costs and effects for each intervention, each of which were used to estimate an incremental cost (difference in total cost) and incremental effect (difference in QALYs) (see E32-E34) and show the uncertainty of both. These were used to construct a series of cost-effectiveness acceptability curves (CEACs) which plot the probability that each MAD is cost-effective against the maximum WTP for one QALY. In addition a cost-effectiveness acceptability frontier (CEAF) was constructed, which shows the most cost-effective option by WTP per QALY and both whether and at which WTP per QALY the most cost-effective option switches to another option. The CEAF plots the uncertainty associated with the optimal option at different values of the cost-effectiveness threshold (see main text, Figure 3).

Deterministic sensitivity analysis assessed the impact of varying the purchase price and expected lifespan (from 6 to 60 months) of each device, the use of complete case analysis and SF-6D QALYs on net monetary benefit (NMB), which is the benefit of an intervention in monetary terms (i.e. the WTP for a QALY gain x total QALYs gained) minus the monetary cost. The results are robust to using only complete case analysis as well as changes in a device's price and lifespan (see E35-E44). When the bMAD price exceeds £525 or its average lifespan falls <14 months, it no longer has a positive INMB. When the price of bMAD falls to below £60, or its length of life extends to beyond 3 years (with no change in SP1) the bMAD becomes more

cost-effective than the SP1. However even when assuming the same price for the bMAD of £60 or that its lifetime is at least 5 years the bMAD remains less cost-effective than the SP2. The move from using EQ5D to SF6D QALYs (see figures E40 and E43) considerably strengthened the base case conclusion that SP2 is the most cost-effective of options.

Table E26: Unit Costs (2011-2012, £ sterling)

	Mean (2011/12 £)	SD	Source	Notes
MAD Device				
SP1	£1·62			Pro-rata 4 weeks
SP2	£9·85			Pro-rata 4 weeks
bMAD	£17·95			Pro-rata 4 weeks
Measurement Consult (Maxillofacial Surgeon)	£5·66	£7·42	NHS Ref 144: First attendance	Pro-rata 4 weeks
Fitting Consult (Maxillofacial Surgeon)	£4·72	£7·43	NHS Ref 144: Follow-up	Pro-rata 4 weeks
Dentist visit, SP2 moulding	£11·52	£13·77	NHS Ref CZ38Y	Pro-rata 4 weeks
Additional visit to Addenbrooke's (bMAD)	£4·72	£7·43	NHS Ref 144: Follow-up	Pro-rata 4 weeks
Visits				
General Practitioner (GP) Visits	£43·40	£8·68	PSSRU 10·8b	Assumes 14 minute appointment
GP Home Visits	£28·23	£5·65	PSSRU 10·8b	Assumes 14 minute appointment
Nurse (GP Practice) Visits	£9·10	£1·82	PSSRU 10·6	Assumes 14 minute appointment
Nurse (Specialist Community) Home Visits	£11·67	£2·33	PSSRU 10·4	Assumes 14 minute appointment
Dentist (Normal visit)	£105·04	£43·96	NHS Ref: 450	
Accident & Emergency Visit	£64·09	£15·00	NHS Ref: VB11Z	
Outpatient Clinical Visit	£105·89	£47·08	NHS Ref: Average of all Outpatient procedures	
Other Hospital visit	£105·89	£47·08	NHS Ref: Average of all Outpatient procedures	
Telephone Calls				
GP telephone calls	£22·00	£4·40	PSSRU 10·8b	Assumes 7·1 minute call
NHS Direct calls	£22·00	£4·40	PSSRU 10·8b	Assumes 7·1 minute call
Contacted trial helpline	£22·00	£4·40	PSSRU 10·8b	Assumes 7·1 minute call
Hospital Admissions				
Heart Attack				
EI:	£2,251·13	£1,073·39	NHS Ref EB10Z	

XS bed days	£312·29	£111·89	NHS Ref EB10Z	
NEI	£1,966·78	£674·38	NHS Ref EB10Z	
XS bed days	£242·46	£67·30	NHS Ref EB10Z	
Road Traffic Accident (RTA)	£64·09	£15·00	NHS Ref VB11Z	
Stroke				
EI:	£3,302·62	£2,855·17	NHS Ref AA22A/B	
XS bed days	£283·34	£82·35	NHS Ref AA22A/B	
NEI	£3,082·45	£900·66	NHS Ref AA22A/B	
XS bed days	£236·16	£71·92	NHS Ref AA22A/B	
Diagnostic Tests				
MRI	£157·24	£51·20	NHS Ref: Avg of all MRI Codes	
CT Scan	£136·62	£48·84	NHS Ref Avg of all CT scan codes	
X-Ray	£32·21	£6·44	P. Auguste et al. HTA (2011)	
Angiogram	-	-		
Other Service Use				
Ambulance Call out	£214·02	£53·96	NHS Ref: ASS01/02	
Hospital Overnight Stay	-	-		
Hospital Overnight Stay (Emergency case)	-	-		
Other' classified resource use				
Acupuncture	£80·33	£118·33	NHS Ref HB63Z used as proxy	
Counsellor session	£60·00		PSSRU 2·7	
Echocardiogram	£84·01	£17·34	NHS Ref: RA60A	
Pre-op assessment	£120·71	£35·00	NHS Ref: 100	
Blood test	£2·95	£1·77	NHS Ref: DAP839	
Occupational health session	£60·68	£29·16	NHS Ref: 651	
Ophthalmologist session	£85·12	£19·23	NHS Ref: 130	
Osteopath appointment	£40·70	£13·20	NHS Ref: 650 as proxy	
Physiotherapist appointment	£40·70	£13·20	NHS Ref: 650	
Health trainer session	£40·70	£13·20	NHS Ref: 650 as proxy	
Nasal polyp removal	£132·34	£51·16	NHS Ref: CZ12Y	
Podiatrist session	£41·17	£18·96	NHS Ref: 651	
Minor surgery	£132·34	£51·16	NHS Ref: CZ12Y used as proxy	
Contacted dentist over the phone	£105·04	£43·96	NHS Ref: 450	
Complete heart block, pacemaker fitted overnight stay	£1,708·17	£901·74	NHS Ref: EA39Z	

Atrial flutter 3 days hospital stay	£1,360·98	£802·18	NHS Ref: EB07I	
Tonsillitis Overnight Hospital stay	£338·62	£159·30	NHS Ref: CZ01Y	
Chest pain, hypertension day case	£446·00	£156·91	NHS Ref EB04I	
NB: PSSRU: Personal Social Service Research Unit.2011.				

Table E27: Summary of resource use costs

	Mean Unit cost	Baseline n=83 Mean cost/ participant	SD	No Treatment n=78 Mean cost/ participant	SD	SP1 n=81 Mean cost/ participant	SD	SP2 n=78 Mean cost/ participant	SD	bMAD n=77 Mean cost/ participant	SD
General Practitioner (GP) Visits[1]	£43·4	£11·0	£23·3	£12·2	£26·1	£15·5	£30·2	£16·7	£29·9	£14·7	£39·0
GP Home Visits[1]	£28·2	£0·3	£3·1	£0·4	£3·2	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0
Nurse (GP Practice) Visits[1]	£9·1	£0·0	£0·0	£0·4	£2·3	£0·7	£2·8	£0·2	£1·4	£0·4	£1·8
Nurse (Specialist Community) Home Visits[1]	£11·7	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·2	£1·3
GP telephone calls[1]	£22·0	£0·0	£0·0	£0·3	£2·5	£0·0	£0·0	£0·3	£2·5	£0·3	£2·5
NHS Direct calls[1]	£22·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·3	£2·5
Contacted trial helpline[1]	£22·0	£0·0	£0·0	£0·0	£0·0	£0·5	£4·9	£0·3	£2·5	£1·4	£7·4
Ambulance Call out[2]	£214·0	£0·0	£0·0	£2·7	£24·2	£2·6	£23·8	£0·0	£0·0	£0·0	£0·0
Accident & Emergency Visit[2]	£64·1	£0·0	£0·0	£0·8	£7·3	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0
Outpatient Clinical Visit[2]	£105·9	£51·0	£74·6	£14·9	£44·3	£19·6	£44·6	£10·9	£32·3	£12·4	£45·4
Dentist (Normal visit)[2]	£105·0	£8·9	£29·4	£16·2	£38·1	£22·0	£54·4	£16·2	£41·7	£12·3	£34·0
Other											
Accupuncture[2]	£80·3	£0·0	£0·0	£1·0	£9·1	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0
Minor surgery[3]	£132·3	£0·0	£0·0	£0·0	£0·0	£1·6	£14·7	£0·0	£0·0	£0·0	£0·0
Blood test[2]	£3·0	£0·0	£0·0	£0·3	£2·7	£0·1	£1·3	£0·0	£0·0	£0·0	£0·0
Counsellor session[2]	£60·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·8	£6·8	£0·0	£0·0
Echocardiogram[2]	£84·0	£0·0	£0·0	£0·0	£0·0	£1·0	£9·3	£0·0	£0·0	£0·0	£0·0
Pre op assessment[2]	£120·7	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£4·6	£23·4	£3·1	£19·3
Health trainer session[2]	£40·7	£0·0	£0·0	£1·0	£9·2	£0·0	£0·0	£1·0	£9·2	£0·0	£0·0
Occupational health session[2]	£60·7	£0·0	£0·0	£0·0	£0·0	£0·7	£6·7	£0·0	£0·0	£0·0	£0·0
Ophthalmologist session[2]	£85·1	£0·0	£0·0	£0·0	£0·0	£1·1	£9·5	£0·0	£0·0	£0·0	£0·0

Optician[4]	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0
Osteopath appointment[2]	£40·7	£0·5	£4·5	£2·1	£18·4	£0·0	£0·0	£0·5	£4·6	£0·0	£0·0
Physiotherapist appointment[2]	£40·7	£1·0	£6·3	£2·1	£14·5	£0·5	£4·5	£0·0	£0·0	£1·1	£6·5
Nasal polyp removal[2]	£132·3	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£1·7	£15·1
Podiatrist session[2]	£41·2	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·5	£4·7
Contacted dentist over the phone[2]	£105·0	£0·0	£0·0	£1·3	£11·9	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0
Hospital overnight, length of staty[7]											
Heart attack_ [7]											
Road traffic accident (RTA) requiring medical treatment[5]	£64·1	£0·0	£0·0	£0·8	£7·3	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0
Stroke[2]											
Other											
Tonsillitis Overnight Hospital stay[2]	£338·6	£0·0	£0·0	£0·0	£0·0	£4·2	£37·6	£0·0	£0·0	£0·0	£0·0
Atrial flutter 3 days hospital stay[2]	£1,361·0	£0·0	£0·0	£17·4	£154·1	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0
Complete heart block, pacemaker fitted overnight stay[2]	£1,708·2	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£22·2	£194·7
Chest pain, hypertension day case[2]	£446·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0	£5·8	£50·8
MRI[2]	£157·2	£0·0	£0·0	£2·0	£17·8	£0·0	£0·0	£0·0	£0·0	£0·0	£0·0
CT scan[2]	£136·6										
X-ray[6]	£32·2	£0·0	£0·0	£0·0	£0·0	£0·4	£3·6	£0·0	£0·0	£0·4	£3·7
Angioplasty[7]											
Angiogram[7]											
Total		£72·7	£80·8	£76·1	£175·1	£70·7	£91·5	£51·5	£67·5	£76·7	£214·8

1=PSSRU costs 2011; 2=NHS Ref cost 2011/12; 3=NHS Ref applied cost of Nasal poly removal as proxy; 4=No NHS cost for optician visit; 5=RTA 1 event, A&E visit only, no treatment NHS ref A&E cost applied; 6=taken from literature "P. Auguste et al. HTA (2011)"; 7=no events, no unit cost sourced

Table E28: Differences in QALYs compared to no treatment

		Coefficient (SE)	P value	Global P value
EQ-5D-3L QALYs	Constant	0.0649 (0.002)	<0.001	
	Baseline	0.0005 (0.001)	0.69	0.76
	SP1	0.0009 (0.001)	0.37	
	SP2	0.0009 (0.001)	0.47	
	bMAD	0.0018 (0.002)	0.23	
SF-6D QALYs	Constant	0.0527 (0.001)	<0.001	
	Baseline	-0.0011 (0.001)	0.1	0.00
	SP1	0.00039 (0.001)	0.63	
	SP2	0.0019 (0.001)	0.01	
	bMAD	0.0009 (0.001)	0.31	

Table E 29: Comparison of costs incurred over 4 weeks

	No treatment n=78	**SP1 n=81**	**SP2 n=78**	**bMAD n=77**
Device costs (Fixed)	·	21	128	350
Measurement for device	-	-	-	110·37
Fitting of device	-	-	-	92·04
Additional fitting visit if required (average across all patients)	-	-	-	5·98[1]
Sub total	**-**	**21**	**128**	**558.39**
Device Lifespan (months) (Fixed)	-	12	12	18
Fixed cost of intervention -pro rata (4 weeks) sub total	**-**	**1·62**	**9·85**	**28·64**
	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
Variable resource use cost (4 weeks)	78·50 (19·97)	73·02 (10·47)	53·58 (8·05)	76·25 (24·40)
Total 4 week costs	**78·50 (19·97)**	**74·64 (10·47)**	**63·43 (8·05)**	**104·89 (24·39)**

1 Five participants required additional fitment visit for bMAD, cost @ £92.04. Average across all participants = £5.98

Table E30: Box plots of total cost during each 4-week treatment period

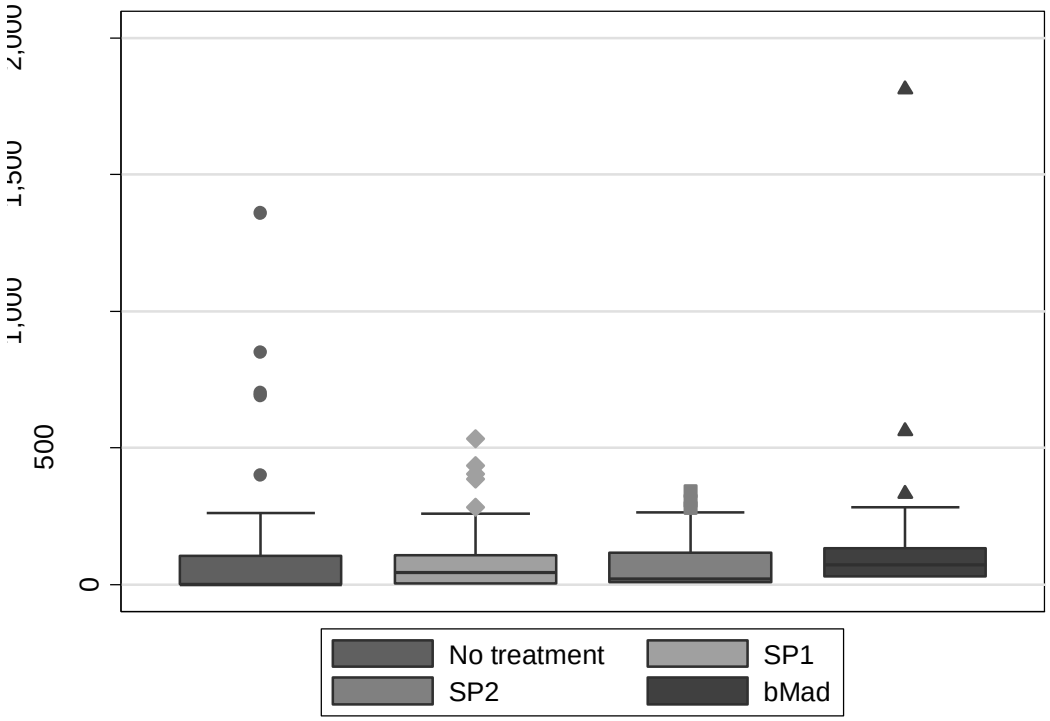


Table E31: Comparison of costs and QALYs of devices against control

	No treatment n=78	**SP1 n=81**	**SP2 n=78**	**bMAD n=77**
	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
Total costs over 4 weeks				
Total cost [1]	78·50 (19·97)	74·64 (10·47)	63·43 (8·05)	104·89 (24·39)
Incremental cost (MAD – no treatment)	-	-3·87 (21·38)	-15·08 (20·62)	26·39 (27·94)
Total utility over 4 weeks				
QALY[2]	0·0649 (0·0017)	0·0658 (0·0017)	0·0658 (0·0019)	0·0667 (0·0017)
Incremental QALY (MAD – no treatment)	-	0·00094 (0·00105)	0·00088 (0·00123)	0·00177 (0·00147)
Cost-effectiveness measure (UK£, 2011)				
Incremental Cost-Effectiveness Ratio (ICER)	-	-£4,093	-£17,104	£14,876
Incremental Net Monetary Benefit	-	£23	£33	£9

(Willingness to pay = £20,000) vs. no treatment				
1. Resource use & Total costs by intervention, estimated using a Mixed effects model controlling for baseline data. All costs in 2011/12 (£) 2. QALY scores calculated using the Area Under the Curve method to represent the true QALY score for the 4 week intervention period to be consistent with the costs presented. Based on EQ-5D-3L responses.				

Figure E32: Scatter plot of estimated joint density incremental costs and incremental effects of SP1 vs No Treatment obtained from bootstrap sampling

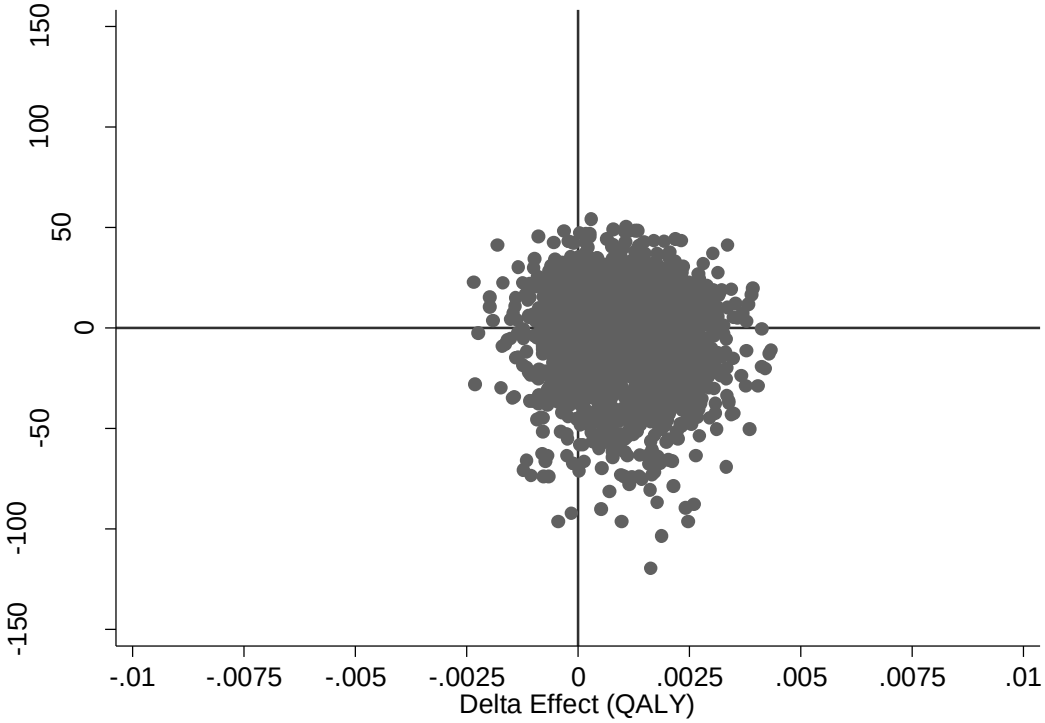


Figure E33: Scatter plot of estimated joint density incremental costs and incremental effects of SP2 vs No Treatment obtained from bootstrap sampling

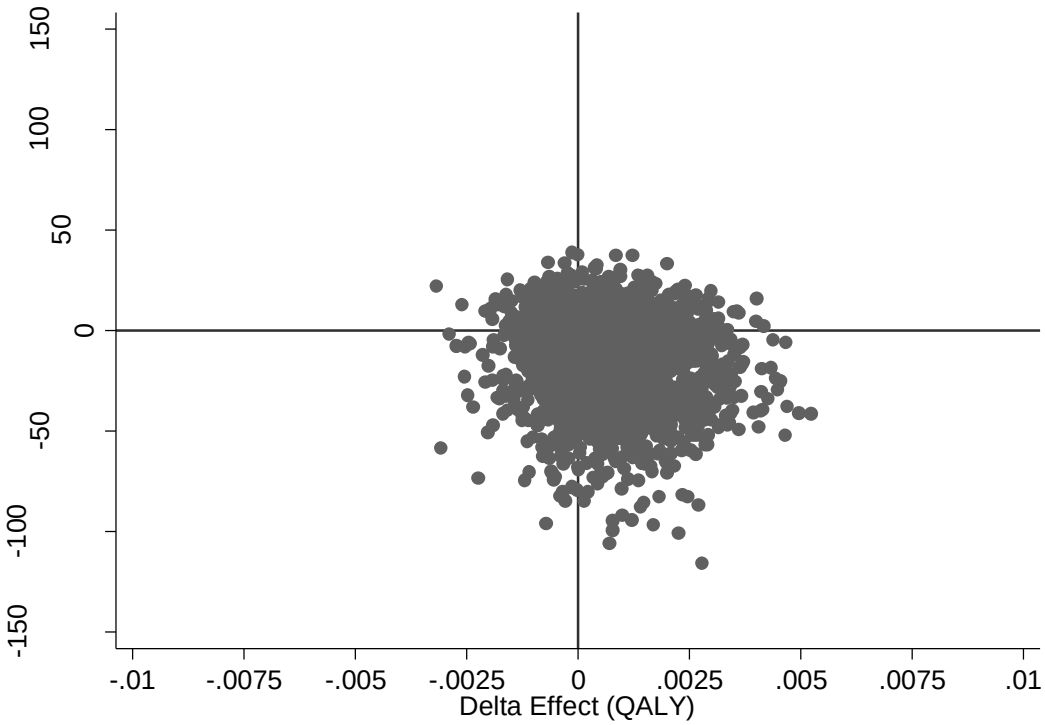


Figure E34: Scatter plot of estimated joint density incremental costs and incremental effects of bMAD vs No Treatment obtained from bootstrap sampling

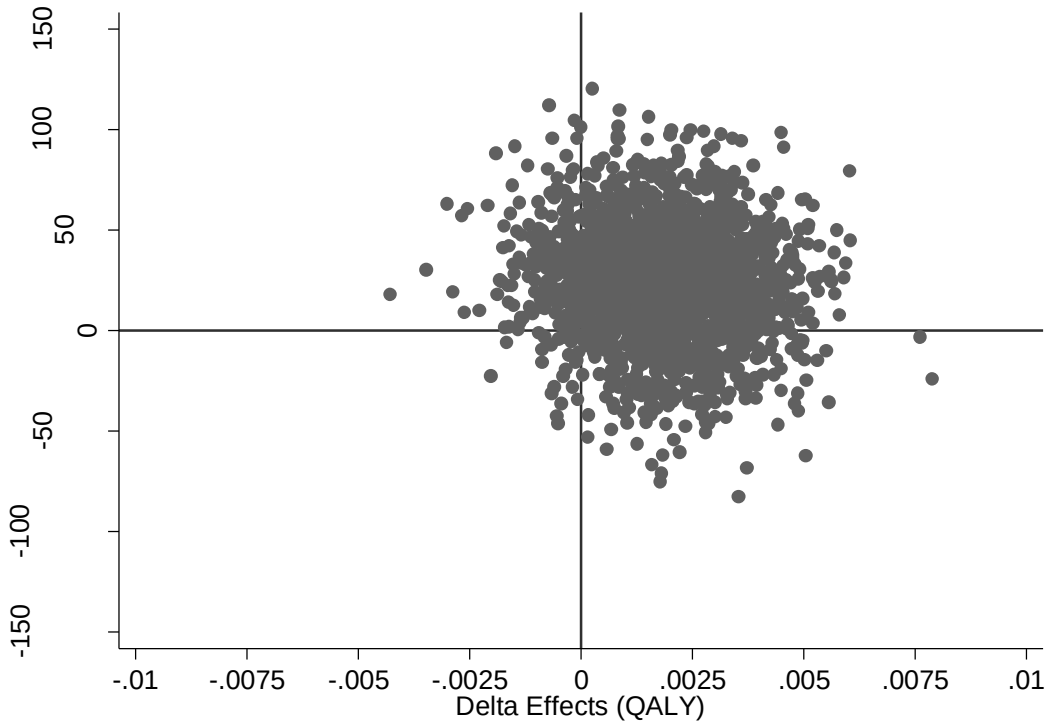


Figure E35: Sensitivity analysis: Lifespan of devices

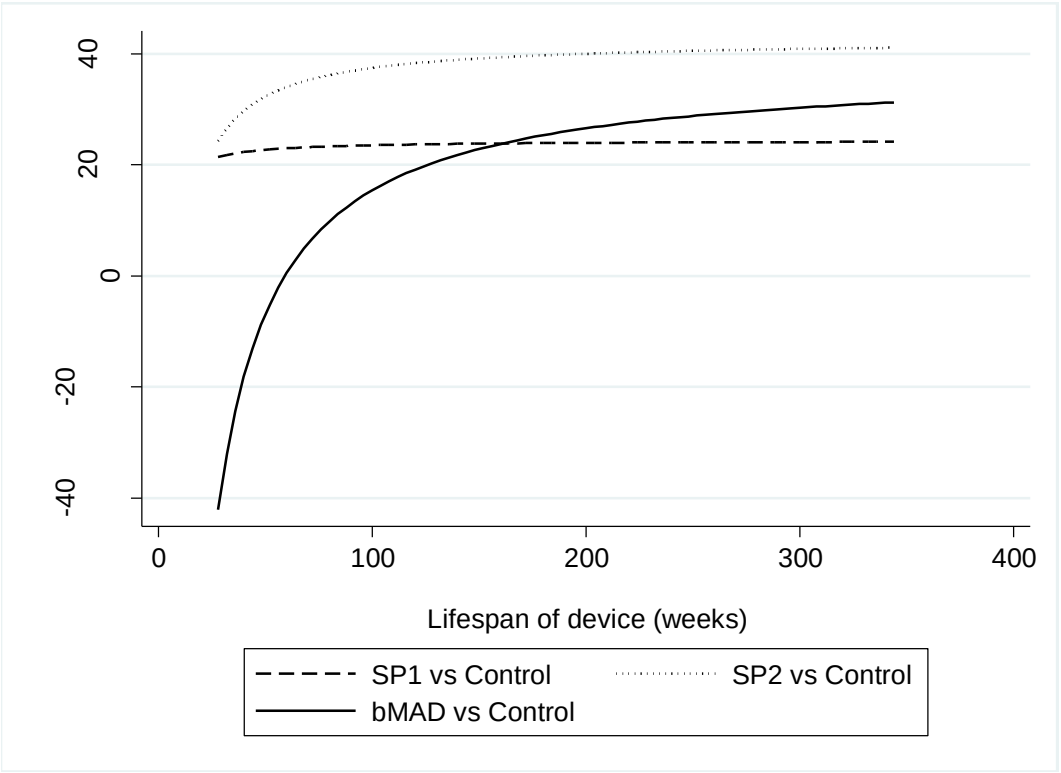


Figure E36: Sensitivity analysis: Varying cost of SP1

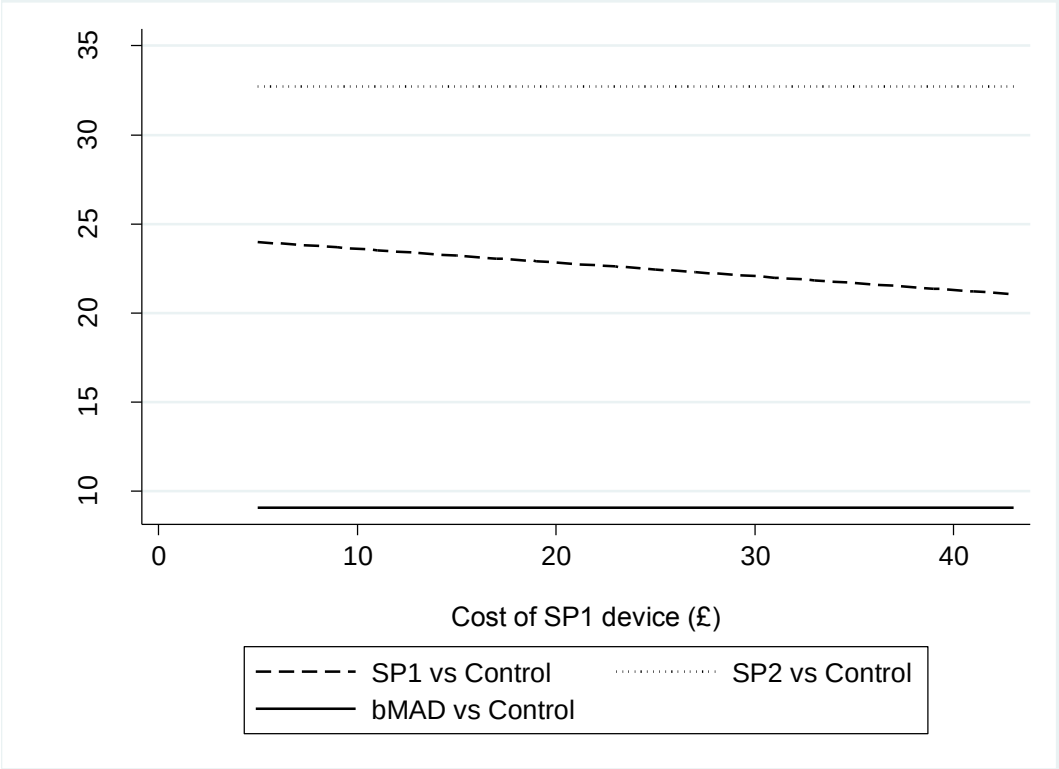


Figure E37: Sensitivity analysis: Varying cost of SP2

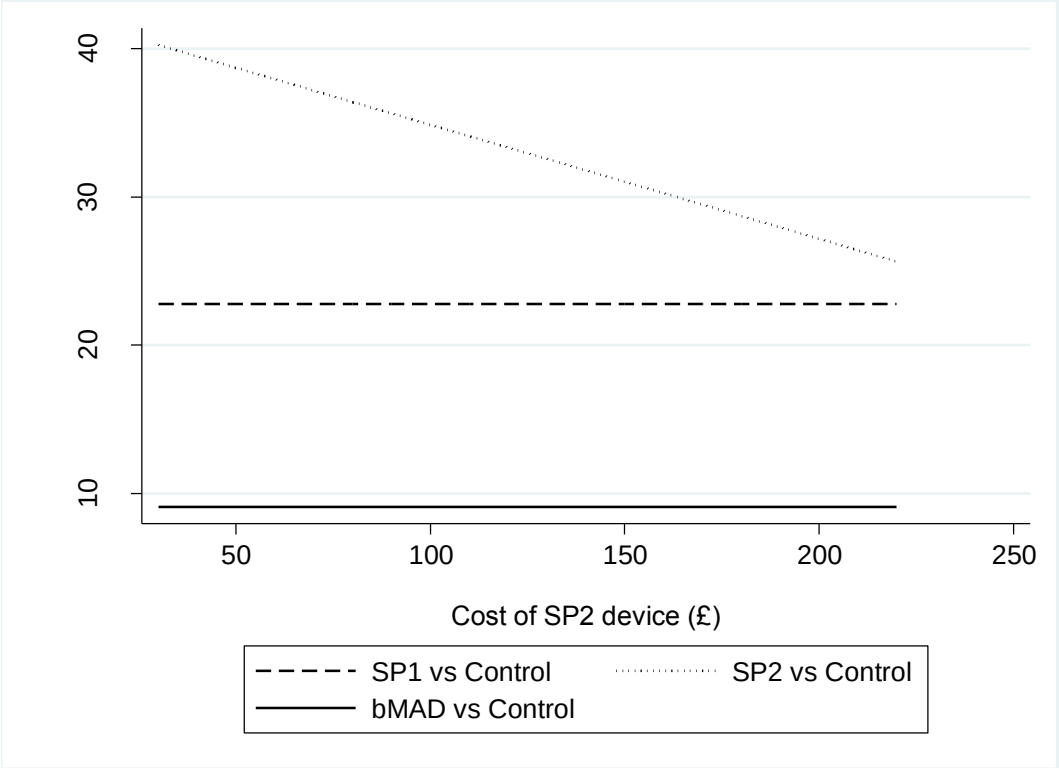


Figure E38: Sensitivity analysis: Varying cost of bMAD

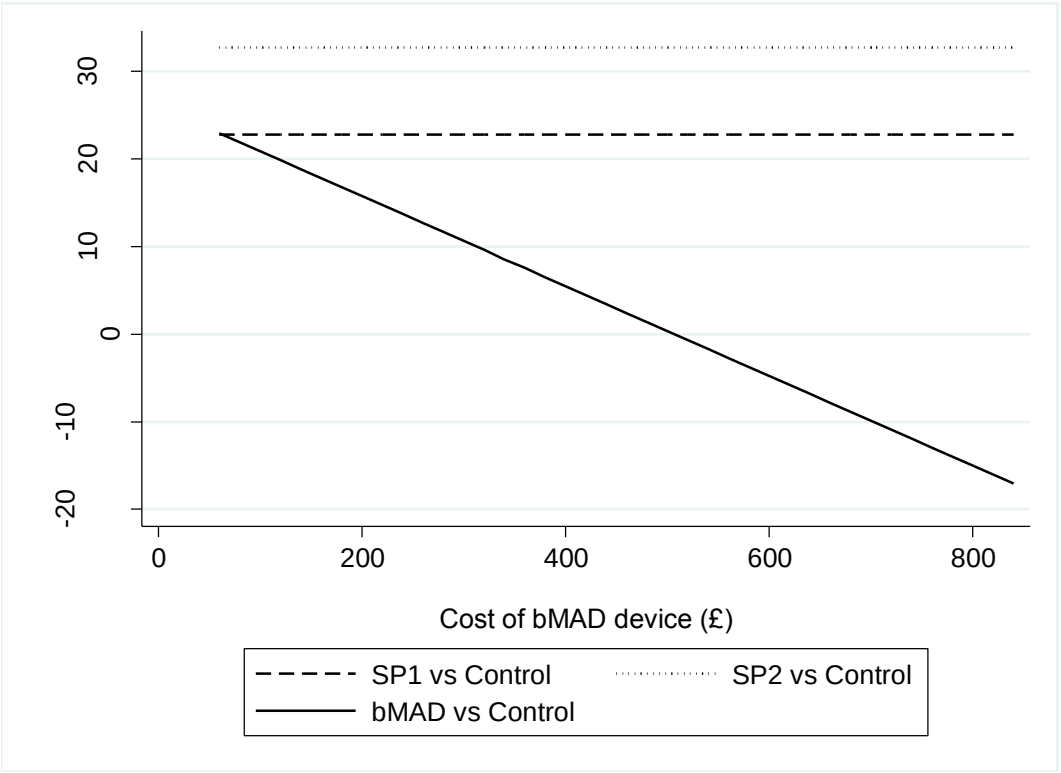


Figure E39: Sensitivity analysis: NMB devices vs. control (EQ-5D-3L)

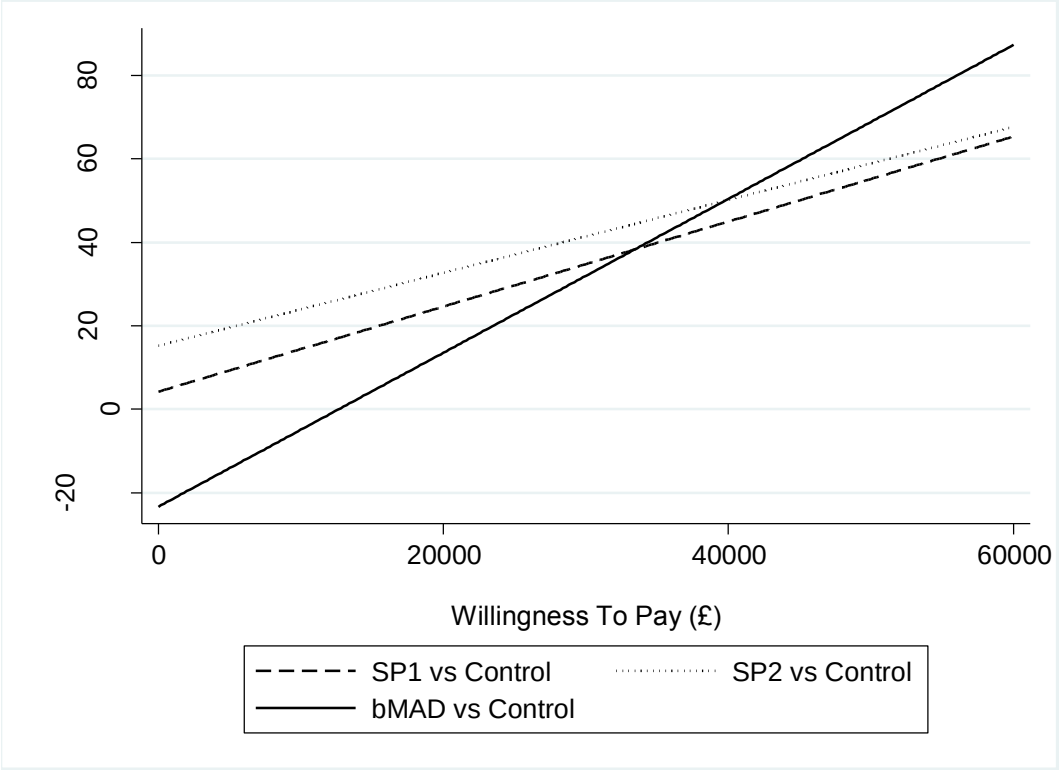


Figure E40: Sensitivity analysis: CEAC between all devices (EQ-5D-3L)

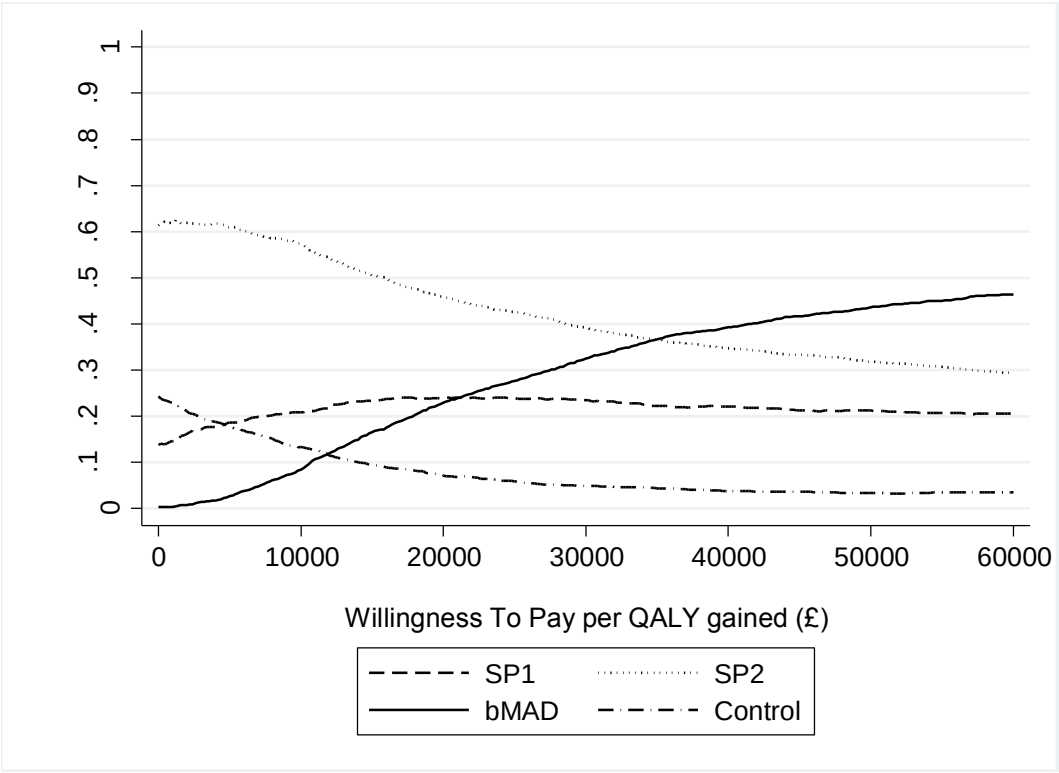


Figure E41: Sensitivity analysis: EVPI (EQ-5D-3L)

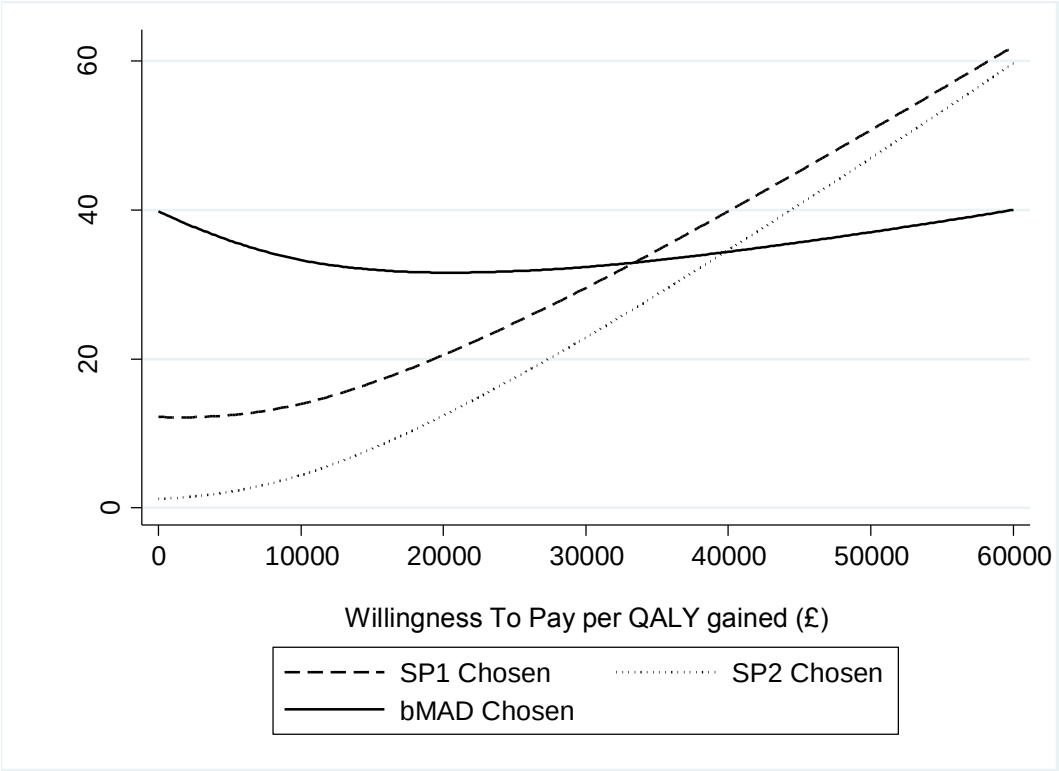


Figure E42: Sensitivity analysis: NMB vs. control (SF6D)

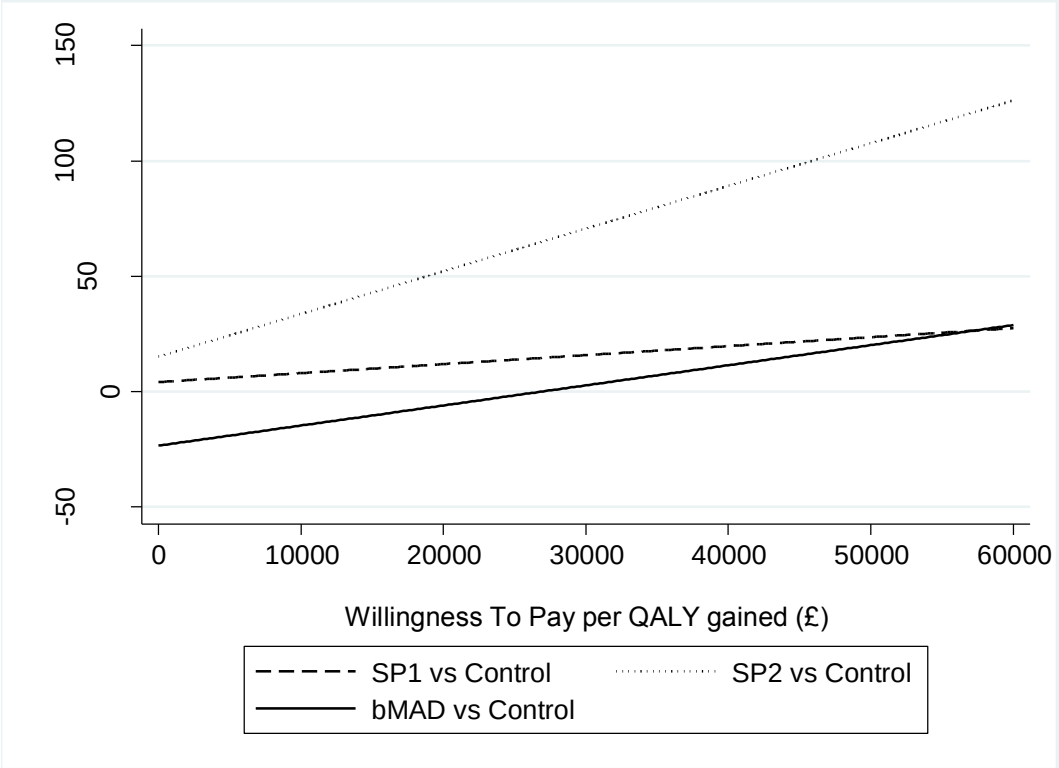


Figure E43: Sensitivity analysis: CEAC between all devices (SF6D)

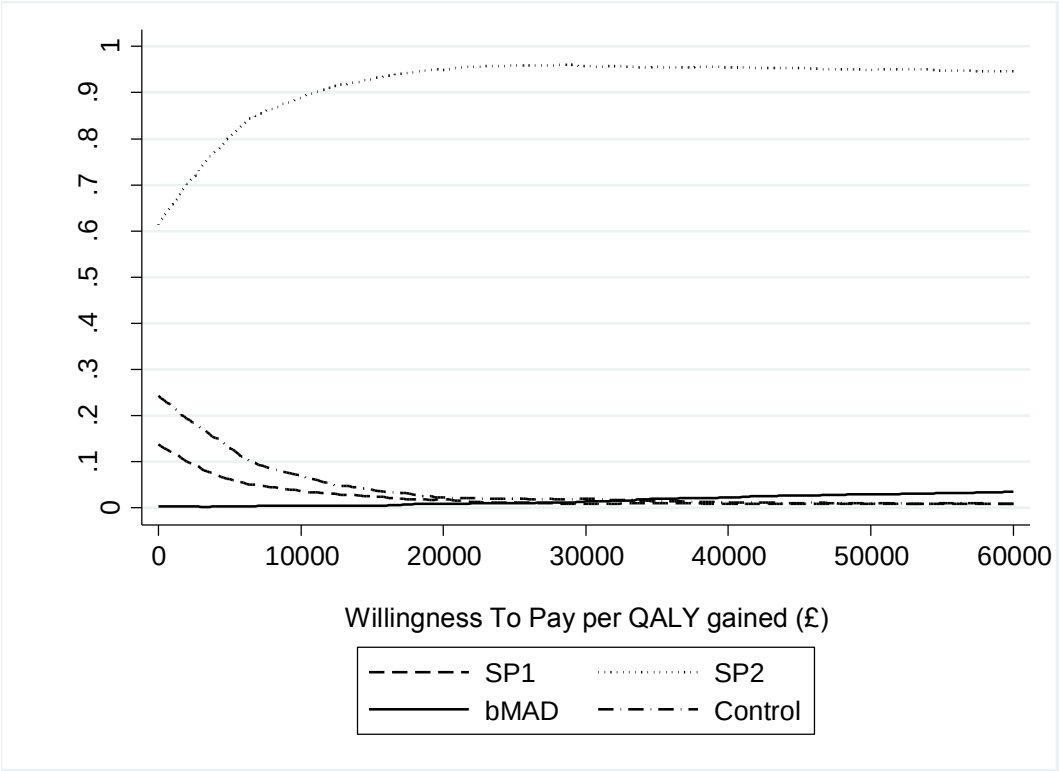
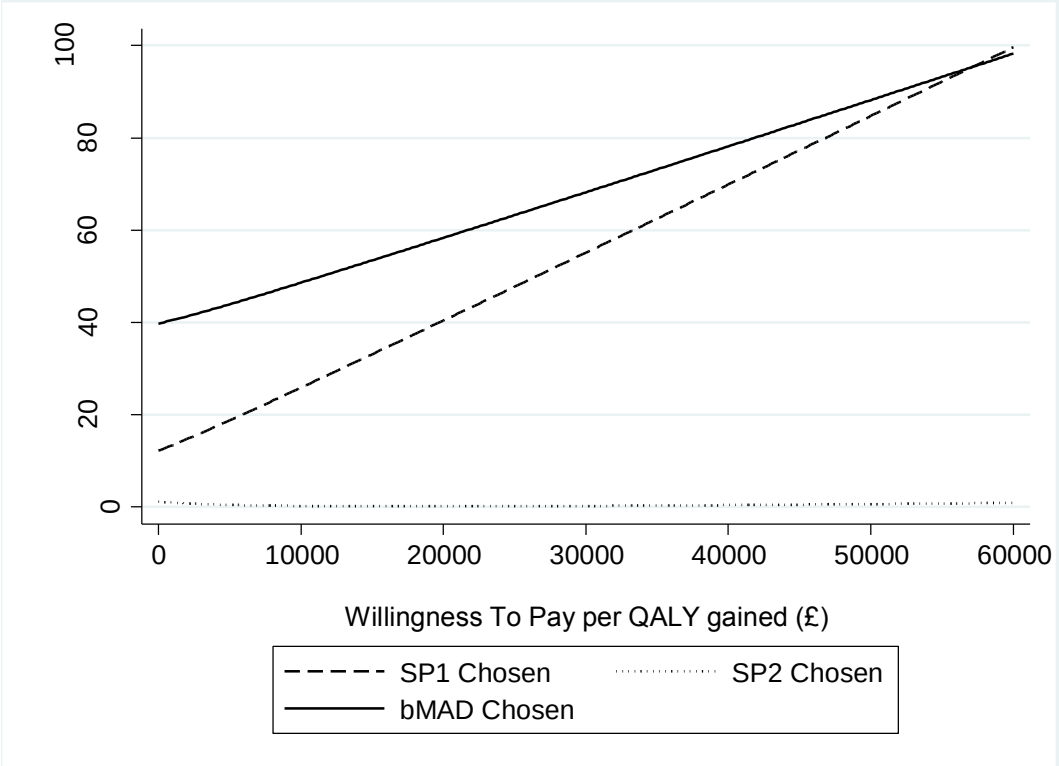


Figure E44: Sensitivity analysis: EVPI (SF6D)



Supplementary Information References

E1. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991;14:540-5.

E2. Hardinge FM, Pitson DJ, Stradling JR. Use of the Epworth Sleepiness Scale to demonstrate response to treatment with nasal continuous positive airways pressure in patients with obstructive sleep apnoea. *Respir Med* 1995;89:617-20.

E3. Weaver TE, Laizner AM, Evans LK*, et al.* An instrument to measure functional status outcomes for disorders of excessive sleepiness. *Sleep* 1997;20:835-43.

E4. Montserrat JM, Ferrer M, Hernandez L*, et al.* Effectiveness of CPAP treatment in daytime function in sleep apnea syndrome: a randomized controlled study with an optimized placebo. *Am J Respir Crit Care Med* 2001;164:608-13.

E5. Blanco J, Zamarron C, Abeleira Pazos MT*, et al.* Prospective evaluation of an oral appliance in the treatment of obstructive sleep apnea syndrome. *Sleep Breath* 2005;9:20-5.

E6. Gauthier L, Laberge L, Beaudry M*, et al.* Efficacy of two mandibular advancement appliances in the management of snoring and mild-moderate sleep apnea: a cross-over randomized study. *Sleep Med* 2009;10:329-36.

E7. Flemons WW, Reimer MA. Development of a disease-specific health-related quality of life questionnaire for sleep apnea. *Am J Respir Crit Care Med* 1998;158:494-503.

E8. Lam B, Sam K, Mok WY*, et al.* Randomised study of three non-surgical treatments in mild to moderate obstructive sleep apnoea. *Thorax* 2007;62:354-9.

E9. Jenkinson C, Stradling J, Petersen S. Comparison of three measures of quality of life outcome in the evaluation of continuous positive airways pressure therapy for sleep apnoea. *J Sleep Res* 1997;6:199-204.

E10. Smith IE, Shneerson JM. Is the SF 36 sensitive to sleep disruption? A study in subjects with sleep apnoea. *J Sleep Res* 1995;4:183-88.

E11. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;21:271-92.

E12. The EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy* The EuroQol Group,1990;16:199-208.

E13. Schmidlin M, Fritsch K, Matthews F*, et al.* Utility indices in patients with the obstructive sleep apnea syndrome. *Respiration* 2010;79:200-8.

E14. Iber C, Ancoli-Isreal S, Chesson A*, et al.* The American Academy of Sleep Medicine Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, 1st ed.: Westchester, Illinois: American Academy of Sleep Medicine. 2007.