The efficacy of baking soda dentifrice in controlling plaque and gingivitis: A systematic review

Valkenburg, C.; Kashmour, Y.; Dao, A.; van der Weijden, G.A.; Slot, D.E.

DOI
10.1111/idh.12390

Publication date
2019

Document Version
Other version

Published in
International Journal of Dental Hygiene

License
CC BY

Citation for published version (APA):
Valkenburg, C., Kashmour, Y., Dao, A., van der Weijden, G. A., & Slot, D. E. (2019). The efficacy of baking soda dentifrice in controlling plaque and gingivitis: A systematic review. International Journal of Dental Hygiene, 17(2), 99-116. https://doi.org/10.1111/idh.12390
Appendix S1.
Post hoc changes to the protocol.

- Several subgroup and sensitivity analyses were planned but could not be performed due to limited data.
- As sub-analysis, network meta-analyses were performed post hoc and are appropriately noted. The Cochrane Handbook (6) expresses a preference for direct evidence, highlighting the observational nature of indirect evidence, and encourages the use of indirect evidence as a supplemental analysis (78).

A network meta-analysis (NMA) is a method to build on findings from a conventional meta-analysis (22). When multiple interventions have been used and compared for the same disease and outcomes, NMA offers a set of methods to visualize and interpret the wider picture of the evidence and to understand the relative merits of these multiple interventions (19, 75, 76). NMA enables investigators to compare the effects of multiple health care interventions including treatments that were not previously compared in head-to-head trials. Additionally, combining indirect and direct evidences can sometimes provide more precise estimates of treatment effects to support decision-making (77). NMA has been recommended as the highest level of evidence for treatment guidelines (21, 22). Zooming out toward larger scales of evidence may help us to understand the strengths and limitations of the data guiding the medical care of individual patients (19).

The comparisons between different treatments constitute a network (18). For a valid analysis of the network, consistent information from different pathways is assumed (29).

Important in understanding a NMA is assessing the network geometry, finding out which interventions are included in the network and figuring out which interventions have been compared against each other in randomized trials (18, 20). The geometry of the network may show that not all treatments represented equally in the network (18), that some (head to head) comparisons have been ignored or that some comparators were preferred (19, 20).

One appealing feature of a NMA is the rank ordering of interventions (20). However, an intervention may be ranked highly even though it was assessed in only a few trials and with a few patients, which resulted in a misleadingly strong endorsement for the intervention despite large uncertainty (20).

Clinical researchers and physicians should always be able to examine the uncertainty in the ranking of interventions, because the difference between interventions might be small and not clinically relevant (20).

Direct and indirect treatment estimates are calculated in netmeta. This function combines and prints these estimates. A comparison of direct and indirect treatment estimates can serve as a check for consistency of NMA (28). This function provides a back-calculation method to derive indirect estimates from direct pairwise comparisons and network estimates. The direct evidence proportion as described in König et al. (2013) (27) is used to calculate the indirect evidence (24).

Another specific issue in NMA is the importance of the transitivity assumption to validate NMA. The distribution of effect modifiers should be similar (11, 30). This was accomplished by an analysis on ingredients. An exception to this is the group Non-BS, the exact composition of which is not clear.
### ONLINE Appendix S2. Overview of the studies processed for data extraction.

| Authors (year) | Study design, duration | # Participants baseline (end), gender, age (mean/range), oral prophylaxis (OP) | Groups Brands | Regimen: use & instructions | Conclusions of the original authors |
|----------------|------------------------|--------------------------------------------------------------------------------|---------------|-----------------------------|------------------------------------|
| Akwagyiram et al. (2018) (37) | RCT, Parallel (6m) Full mouth 12 hours of no oral hygiene procedures | 247(228) Population Silverstone Research Group, Las Vegas, NV, USA ♀: 141♦(57.1%) ♂: 106♦(42.9%) Mean(sd) age: 36.0(13.63) Age range: 18-79 Smoking status: reported OP: yes (after baseline) | TB+ BS-67% DF TB+ NaF-DF BS-67% DF: 67% sodium bicarbonate, sodium fluoride; experimental RDA: NR Fluoride: 1150 ppm F NaF DF: sodium fluoride; Crest Cavity Protection; Procter & Gamble Co, USA RDA: NR Fluoride: 1100 ppm F TB: Oral-B® 35 Soft Standard Toothbrush; Procter & Gamble Co, USA | Participants had to have at least 20 bleeding points at baseline. Self-brushing, own technique, twice a day for 1 minute. Refraining from other forms of oral hygiene except for flossing if this was part of the subject’s usual oral hygiene routine. Prior to each visit, participants were instructed not to consume anything for at least four hours. | Twice-daily brushing with sodium bicarbonate containing dentifrices significantly improved indices of gingival bleeding and plaque compared with a non-BS dentifrice in subjects with moderate gingivitis. |
| Bosma et al. (2018) (38) | 56(56) Population Salus Research, Inc., Fort Wayne, IN, USA | TB+ BS-67% DF | Participants had to be a whole mouth PI ≥ 2.00 based on the TQ&H. |
|-------------------------|-----------------------------------------------------------|----------------|-----------------------------------------------------------------|
| RCT                     | Coke Crossover (w.o. 3 days)                              | TB+ BS-67% DF  | Plaque removal was significantly greater with sodium bicarbonate containing dentifrices compared with a non-BS dentifrice after a single timed brushing. There was no effect of herbal tinctures. |
| Single-brushing exercise | Single-brushing exercise under supervision with 1.5 g of the assigned dentifrice for 60 seconds at the study site. | TB+ BS-62% DF  | Prior to each visit, participants were instructed not to consume anything for at least four hours. |
| Full mouth (PI results divided in six sites) | Participants had to be a whole mouth PI ≥ 2.00 based on the TQ&H. | TB+ NaF-DF     | Plaque removal was significantly greater with sodium bicarbonate containing dentifrices compared with a non-BS dentifrice after a single timed brushing. There was no effect of herbal tinctures. |
| 24 hours of no oral hygiene procedures | Participants had to be a whole mouth PI ≥ 2.00 based on the TQ&H. | A) BS-67% DF: herbal mixture, 67% sodium bicarbonate, sodium fluoride; experimental | Plaque removal was significantly greater with sodium bicarbonate containing dentifrices compared with a non-BS dentifrice after a single timed brushing. There was no effect of herbal tinctures. |
|                          | A) BS-67% DF: herbal mixture, 67% sodium bicarbonate, sodium fluoride; experimental | B) BS-67% DF: 67% sodium bicarbonate, sodium fluoride; experimental | Single-brushing exercise under supervision with 1.5 g of the assigned dentifrice for 60 seconds at the study site. |
|                          | RDA: NR                                                   | RDA: NR       | Plaque removal was significantly greater with sodium bicarbonate containing dentifrices compared with a non-BS dentifrice after a single timed brushing. There was no effect of herbal tinctures. |
|                          | Fluoride: 923 ppm F                                      | Fluoride: 923 ppm F | Plaque removal was significantly greater with sodium bicarbonate containing dentifrices compared with a non-BS dentifrice after a single timed brushing. There was no effect of herbal tinctures. |
|                          | C) BS-62% DF: herbal mixture, 67% sodium bicarbonate, sodium fluoride; experimental | D) NaF DF: sodium fluoride; experimental | Plaque removal was significantly greater with sodium bicarbonate containing dentifrices compared with a non-BS dentifrice after a single timed brushing. There was no effect of herbal tinctures. |
|                          | RDA: NR                                                   | RDA: NR       | Plaque removal was significantly greater with sodium bicarbonate containing dentifrices compared with a non-BS dentifrice after a single timed brushing. There was no effect of herbal tinctures. |
|                          | Fluoride: 923 ppm F                                      | Fluoride: 923 ppm F | Plaque removal was significantly greater with sodium bicarbonate containing dentifrices compared with a non-BS dentifrice after a single timed brushing. There was no effect of herbal tinctures. |
|                          | TB: Oral-B 40 Soft Compact Toothbrush; Procter & Gamble Co. Cincinnati, OH, USA |              | Plaque removal was significantly greater with sodium bicarbonate containing dentifrices compared with a non-BS dentifrice after a single timed brushing. There was no effect of herbal tinctures. |

Population: Salus Research, Inc., Fort Wayne, IN, USA

♀: 36 (64.3%)  ♂: 20 (35.7%)
Mean(sd) age: 44.1(11.87)  Age range: 19-66
Smoking: no exclusion criteria  OP: NR
| Jose et al. (2018) (54) | 246 (216♀) | TB+ BS-67% DF  
TB+ NaF-DF  
TB: Oral-B® 35 Soft Standard Toothbrush; Procter & Gamble Co., Cincinnati, OH, USA  
BS-67% DF: 67% sodium bicarbonate, sodium fluoride; experimental; GSK Consumer Healthcare, Weybridge, Surrey, UK  
RDA: NR  
Fluoride: 1,150 ppm F  
NaF-DF: sodium fluoride; Colgate® Triple Action; Colgate-Palmolive Co, New York, NY, USA  
RDA: NR  
Fluoride: 1,150 ppm F  
Participants had to be a whole mouth BI ≥ 20 bleeding sites based on the modified BI (68).  
Self-brushing, own technique, twice a day for 1 minute.  
Gingival bleeding, gingivitis, and plaque indices were significantly improved at 24 weeks with twice-daily brushing with a 67% sodium bicarbonate dentifrice in participants with moderate gingivitis. |
|-------------------------|------------|--------------------------------------------------|
| RCT                     | Parallel (6m) | Full mouth  |
| 24 hours of no oral hygiene procedures | Population Salus Research, Inc., Fort Wayne, IN, USA |
| ♂: 152 (61.8%)  
♀: 94 (38.2%)  
Mean(sd) age: 38.58(13.5)  
Age range: 18-?  
Smoking: no exclusion criteria  
OP: optional | TB: Oral-B® 35 Soft Standard Toothbrush; Procter & Gamble Co., Cincinnati, OH, USA  
BS-67% DF: 67% sodium bicarbonate, sodium fluoride; experimental; GSK Consumer Healthcare, Weybridge, Surrey, UK  
RDA: NR  
Fluoride: 1,150 ppm F  
NaF-DF: sodium fluoride; Colgate® Triple Action; Colgate-Palmolive Co, New York, NY, USA  
RDA: NR  
Fluoride: 1,150 ppm F  
Participants had to be a whole mouth BI ≥ 20 bleeding sites based on the modified BI (68).  
Self-brushing, own technique, twice a day for 1 minute.  
Gingival bleeding, gingivitis, and plaque indices were significantly improved at 24 weeks with twice-daily brushing with a 67% sodium bicarbonate dentifrice in participants with moderate gingivitis. |
| Hosadurga et al. (2017) (36) | Parallel (1m) | Full mouth | Hours of no oral hygiene procedure: NR |
|----------------------------|--------------|------------|---------------------------------------|

Patients Yenepoya Dental College, Mangalore, India
♀: 30◊ (66.7◊)  
♂: 15◊ (33.3◊)  
Mean(sd) age: 24.1◊(8.3◊)  
Age range: 18-?

Smoking: exclusion criteria  
OP: no

**TB+ BS-DF**  
TB+ MFP-DF

**TB:** NR  

**BS-DF:** herbal mixture, sodium bicarbonate, sodium fluoride, Parodontax® (GlaxoSmithKline, Middlesex, United Kingdom)  
RDA: NR  
Fluoride: 1,400 ppm F

**MFP-DF:** herbal mixture, calcium carbonate, Colgate® Herbal (Colgate-Palmolive India Limited, Mumbai, India)  
RDA: NR  
Fluoride: NR

Participants had to be a whole mouth PI ≥ 1.5 based on the TQ&H and a GI ≥ 1.0 based on the Löe-Silness Gingival Index.

Self-brushing, own technique, twice a day for a period of 5-10 minutes.

There was no significant difference between the groups in GI and PI levels.
| Mason et al. (2017) (53) | For all sub-studies: Population: NR  Location: Research Center, Cheshire, UK  Smoking status: NR  OP: no | For all sub-studies:  TB: (Oral-B® 35 Indicator, Compact, Soft; Procter & Gamble Co., Cincinnati, OH, USA)?  Participants had to be a whole mouth PI ≥ 2.00 based on the TQ&H.  Refraining from other forms of oral hygiene.  Single-brushing exercise under supervision with 1.5 g of the assigned dentifrice for 60 seconds at the study site.  Prior to each visit, participants were instructed not to consume anything for at least four hours.  The commercially available dentifrices containing 67% or 45% sodium bicarbonate exerted a significantly greater effect on plaque removal than commercially available dentifrices without sodium bicarbonate. |
|------------------------|--------------------------------------------------------------------------------|-----------------------------------------------------------------|
| RCT(2X)                | 24 hours of no oral hygiene procedures                                      | Sub-study 1:  TB+ BS-67% DF  TB+ BS-45% DF  TB+ NaF-DF |
| Crossover (w.o. 3 days)|                                                                                   | A) BS-67% DF: 67% sodium bicarbonate, sodium fluoride with herbs; Parodontax®-F *  RDA: 68  Fluoride: 1,425 ppm F |
| Single-center          |                                                                                   | B) BS-45% DF: 45% sodium bicarbonate, sodium fluoride with herbs and silica; Parodontax® Gentle Whitening *  RDA: 107  Fluoride: 1,425 ppm F |
| Full mouth             |                                                                                   | C) NaF-DF: sodium fluoride; Crest® Decay prevention **  RDA: 129  Fluoride: 1,425 ppm F |
| Single-brushing exercise |                                                                                   |                                                                                     |

**Note:**
- NR: Not reported.
- PI: Plaque Index.
- TQ&H: Thomas Quantitative & Hygiene.
- BS: Brushing.
- DF: Dentifrice.
- RDA: Recommended Daily Allowance.
- ppm F: Parts per million of Fluoride.
- *: Indicates product availability.
- **: Indicates product availability and its purpose.
Sub-study 2:
55 (54♀)
♀: 41♀(74.5%)  
♂: 14♂(25.5%)  
Mean(sd) age: 37.3(11.41)  
Age range: 18-?

Sub-study 2:
TB+ BS+67% DF  
TB+ SnF-DF  

BS-67% DF: 67% sodium bicarbonate, sodium fluoride with herbs; Parodontax ®-F  
RDA: 68  
Fluoride: 1,425 ppm F  

SnF-DF: stannous fluoride, amine fluoride with silica; Meridol® ♦  
RDA: 55  
Fluoride: 1,400 ppm F
| Lomax et al. (2016) (52) | 148 (135◊) | TB+ BS DF |
|-------------------------|------------|----------|
| RCT                     | Non-dental: ? | TB+ NaF-DF |
| Parallel (6w)           | New Delhi (India)? | |
| Single-center           | ♂: 84◊ (56.8%◊) | |
| Full mouth              | ♀: 64◊ (43.2%◊) | |
| 24 hours of no oral     | Mean(sd) age: | |
| hygiene procedures      | 28.2◊(9.14◊) | |
|                         | Age range: 18-? | |
| Smoking status: reported| Smoking status: reported | |
| OP: yes (after baseline)| OP: yes (after baseline) | |

TB: Medium toothbrush
(provided)

BS-DF: 67% sodium bicarbonate,
sodium fluoride; Parodontax® Daily
Toothpaste (GlaxoSmithKline Asia
Pvt. Ltd, Punjab, India)
RDA: NR
Fluoride: 923 ppm F

NaF-DF: Experimental non-sodium
bicarbonate, silica sodium fluoride;
DF, not commercially available.
RDA: NR
Fluoride: 923 ppm F

Participants with mild-to-
moderate gingivitis, a
positive response to
bleeding on brushing and
≥ 20 bleeding sites.

Self-brushing,
own technique, twice a
day.

Refraining from other
forms of oral hygiene

11 p.m. on the evening
prior to each visit,
participants were
instructed not to consume
anything.

A 67% BS dentifrice
provided statistically
significant
improvements in
gingival health and
bleeding after 6 weeks
of use compared to the
NaF dentifrice.
| Study                        | Sample Size | Intervention | Characteristics | 6-month outcomes |
|------------------------------|-------------|--------------|-----------------|------------------|
| Triratana et al. (2015)      | 135 (135♀) | TB+ BS-DF    | Non-dental: ?   | The triclosan/copolymer/flouride dentifrice was superior to the BS dentifrice. |
| (51)                         |             | TB+ MFP+NaF-DF| Bangkok (Thailand) | The BS dentifrice was equivalent to the fluoride dentifrice in its effect on plaque and gingivitis over 6 months of product use. |
| RCT                          |             | TB+ Tcs-DF   | Mean(sd) age: 38.5 (9.15♀) |                  |
| Parallel (6m)                |             |              | Age range: 18-70 |                  |
| Single-center                |             |              | Smoking status: NR |                  |
| Full mouth                   |             |              | OP: NR           |                  |
| Hours of no oral hygiene procedure: NR |           |              |                  |                  |
|                              |             |              | TB: Soft-bristle toothbrush (provided) |                  |
|                              |             |              | BS-DF: herbal mixture, bicarbonate, sodium fluoride; Parodontax® |                  |
|                              |             |              | RDA: NR |                  |
|                              |             |              | Fluoride: 1,400 ppm F |                  |
|                              |             |              | A) MFP+NaF-DF: sodium fluoride and monofluorophosphate; Colgate Cavity Protection® |                  |
|                              |             |              | RDA: NR |                  |
|                              |             |              | Fluoride: 450 ppm F (sodium fluoride) and 1,000ppm F (sodium monofluorophosphate). |                  |
|                              |             |              | B) Tcs-DF: 0.3% triclosan, 2.0% copolymer, silica sodium fluoride; Colgate Total®; |                  |
|                              |             |              | RDA: NR |                  |
|                              |             |              | Fluoride: 1,450 ppm F |                  |
|                              |             |              | Participants had to be a whole mouth PI ≥ 1.5 based on the TQ&H and a GI ≥ 1.0 based on the Loe-Silness Gingival Index. |                  |
|                              |             |              | Self-brushing, own technique. |                  |
|                              |             |              | Refraining from other forms of oral hygiene |                  |
|                              |             |              | Brush one minute twice-daily (morning and evening) with the assigned TB and dentifrice. |                  |
| Study | N | Population |干预 | Treatment | Outcomes |
|-------|---|------------|-----|----------|----------|
| Al-Kholani (2011) (35) | 48 (NR) | Patients with chronic marginal gingivitis from a private clinic in Sana’a, Yemen | TB+ BS-DF | Participants with chronic marginal gingivitis. | The BS dentifrice was more effective for the reduction of dental plaque than with the (non-BS) herbal dentifrice; however, differences were non-significant. |
| | | | TB+ NaF-DF | Self-brushing, given oral hygiene instruction. |  |
| | | | TB+ Non-BS Herbal DF | Toothbrushing for two minutes, three times/day |  |
| | | | TB: NR | Refraining from other forms of oral hygiene: NR | Signs of gingival inflammation reduced significantly in both herbal test groups compared to the conventional group. |
| | | | BS-DF: sodium bicarbonate, sodium fluoride, herbal ingredients (chamomile, echinacea, sage rhatany, myrrh, peppermint oil); Parodontax®; (GlaxoSmithKline, Middlesex, United Kingdom) |  |
| | | | RDA: NR |  |
| | | | Fluoride: 1400ppm F- |  |
| | | | A) Non-BS Herbal-DF: herbal ingredients (chamomile, calendula, sage); Silca® Herbal◊ Toothpaste |  |
| | | | B) NaF-DF: 0.32% sodium fluoride, hydrated silica, other ingredients (sorbitol, water, SLS, Peg-32, flavor, cellulose gum, sodium saccharin, trisodium phosphate, vitamin-E-acetate) |  |
| | | | RDA: NR |  |
| | | | Fluoride: 1450 F◊ |  |
| Ghassemi et al. (2008) | 218 (207) Non-dental: ? University Park Research Center, Fort Wayne, IN, (USA)? Sex of participants who finished the study: ♀: 144(66.1%)♂: 63(28.9%) Mean age: 39 Age range: 18-71 Smoking status: NR OP: NR |
|----------------------|---------------------------------------------------------------|
| RCT Parallel (4w) & Single-brushing exercise: A) Baseline single-brushing B) End single-brushing | TB+ BS-DF TB+ Tcs-DF TB: (Oral-B 35 Compact Soft) BS-DF: BS; Arm & Hammer® Advance White® Brilliant Sparkle; (Church & Dwight Co., Inc., Princeton, NJ, USA) RDA: NR Fluoride: ? ppm F- |
| Single-center Full mouth 24 hours of no oral hygiene procedures | Tcs-DF: 0.3% triclosan/ 2.0% copolymer; Colgate® Total® Clean Mint Dentifrice; (Colgate-Palmolive Company, New York, NY, USA) RDA: NR Fluoride: ? ppm F- |
| | Participants had to be a whole mouth PI ≥ 2 based on the TQ&H. Self-brushing at home, own technique; 60 seconds twice-daily. Refraining from other forms of oral hygiene. At baseline visit and 4w visit (supervised): Single-brushing with 1.5 g of the assigned dentifrice for 60 seconds at the study site. The BS dentifrice was significantly more effective in reducing and controlling plaque accumulation compared to a clinically proven antimicrobial (triclosan/copolymer) dentifrice. |
### Putt et al. (2008) (49)

RCT (5X)
Crossover (w.o. 3 to 5 days)
Single-brushing exercise
Multi-center
Full mouth

Participants refrained from oral hygiene for 24 hours, and did not eat, drink, or smoke the morning of their examination visit.

| Sub-study 1: | Sub-study 2: |
|-------------|-------------|
| Population: 71 (66) | Population: 68 (66) |
| Gender: ♂ 15 (23.1%) | Gender: ♂ 17 (25.8%) |
| Gender: ♂ 50 (76.9%) | Gender: ♂ 49 (74.2%) |
| Mean(sd) age: 42 (13) | Mean(sd): 40 (13) |
| Age range: 18-68 | Age range: 22-81 |

For all sub-studies:
- Population: NR
- Location: NR
- Smoking status: NR
- OP: NR

**TB:** Oral-B® 35 Compact Toothbrush; ***
**RDA:** NR
**Fluoride:** NR ppm F-

The sodium bicarbonate dentifrices enhanced plaque removal effectiveness of toothbrushing to a significantly greater extent than the non-BS dentifrice products.

For all sub-studies:
- Population: NR
- Location: NR
- Smoking status: NR
- OP: NR

### Sub-study 1:
- **A) BS-65% DF:** BS 65%; Arm & Hammer® Dental Care®;
- **B) BS-20% DF:** BS 20%; Arm & Hammer® Advance White™ Brilliant Sparkle;
- **C) Tcs-DF:** 0.3% triclosan, copolymer; Colgate® Total® Clean Mint;

### Sub-study 2:
- **A) BS-48% DF:** BS 48%; Arm & Hammer® Advance White™ BS and Peroxide;
- **B) BS-20% DF:** BS 20%; Arm & Hammer Advance White Brilliant Sparkle;

Participants had to be a whole mouth PI ≥ 1.5 based on the TQ&H and had no periodontitis affecting more than two teeth.

Self-brushing, own technique.
Each visit involved a baseline (pre-brushing) plaque assessment, a one-minute supervised brushing with the assigned test dentifrice, and a post brushing plaque assessment.

Refraining from other oral hygiene products: NR.

During each w.o. (3 to 5 days): follow normal oral hygiene procedures using BS-free fluoride DF and toothbrush provided.
Sub-study 3:
69 (64)
♀: 46 (71.9%)  
♂: 18 (28.1%)
Mean(sd) age: 41 (11)
Age range: 22-68

Sub-study 3:
TB+ BS-48% DF
TB+ BS-27% DF
TB+ Tcs-DF
TB+ NaF-DF

| Treatment       | Description                                                                 |
|-----------------|-----------------------------------------------------------------------------|
| A) BS-48% DF    | BS 48%; Arm & Hammer® Advance White™ BS and Peroxide; *                     |
| B) BS-27% DF    | BS 27%; Arm & Hammer® Advance White™ Liquid Gel; *                          |
| C) Tcs-DF       | Colgate® Total® Clean Mint; 0.3% triclosan, copolymer; **                   |
| D) NaF-DF       | sodium fluoride, silica; Crest® Regular Cavity Protection; ***             |

C) Tcs-DF: 0.3% triclosan, copolymer; Colgate® Total® Clean Mint; **
D) SnF-DF: stannous fluoride and sodium hexametaphosphate; Crest® Pro-Health™; ***
| Sub-study 4:  | Sub-study 4:  |
|------------|-------------|
| 36 (36)    | 36 (36)     |
| ♀: 27(75%) | ♀: 27(75%)  |
| ♂: 9(25%)  | ♂: 9(25%)   |
| Mean(sd): 38(11) | Mean(sd): 38(11) |
| Age range: 19-64 | Age range: 19-64 |
| TB+ BS-65% DF | TB+ BS-65% DF |
| TB+ NaF-DF | TB+ NaF-DF |
| BS-65% DF: BS 65% | BS-65% DF: BS 65% |
| Arm & Hammer® Dental Care®; * | Arm & Hammer® Dental Care®; * |
| NaF-DF: sodium fluoride, silica Aim® Cavity Protection Gel; * | NaF-DF: sodium fluoride, silica Aim® Cavity Protection Gel; * |

| Sub-study 5:  | Sub-study 5:  |
|------------|-------------|
| 45 (41)    | 45 (41)     |
| ♀: 33(80.5%) | ♀: 33(80.5%) |
| ♂: 8(19.5%) | ♂: 8(19.5%) |
| Mean(sd): 42(12) | Mean(sd): 42(12) |
| Age range: 20-64 | Age range: 20-64 |
| TB+ BS-20% DF | TB+ BS-20% DF |
| TB+ Tcs-DF | TB+ Tcs-DF |
| BS-20% DF: BS 20% | BS-20% DF: BS 20% |
| Arm & Hammer Advance White Brilliant Sparkle; * | Arm & Hammer Advance White Brilliant Sparkle; * |
| Tcs-DF: 0.3% triclosan, copolymer; Colgate® Total® Clean Mint; ** | Tcs-DF: 0.3% triclosan, copolymer; Colgate® Total® Clean Mint; ** |
| *Church & Dwight Co., Inc., Princeton, NJ, USA | *Church & Dwight Co., Inc., Princeton, NJ, USA |
| **Colgate-Palmolive Company, New York, NY, USA | **Colgate-Palmolive Company, New York, NY, USA |
| ***Procter & Gamble, Cincinnati, OH, USA | ***Procter & Gamble, Cincinnati, OH, USA |

Ozaki et al. (2006) (48) 48 (42) TB+ BS-DF Participants had to be a Both dentifrices were
RCT
Parallel (4w)
Single-center
Full mouth (GI and PI results divided in lingual+ buccal and mesial+ distal)

Hours of no oral hygiene procedures: *NR*

Valkenburg et al.
Dental Clinic
São Caetano (Brazil)
♀: 28 (58.3%◊)
♂: 20 (41.7%◊)
Mean(sd) age: 33.19 (13.57)
Age range: 18-69
Smoking status: *NR*
OP: *NR*

TB+ Tcs-DF
TB: *NR*

**BS-DF:** sodium bicarbonate, sodium fluoride, herbal ingredients (chamomile, echinacea, sage, rhatany, myrrh and peppermint oil); Parodontax® (GlaxoSmithKline, Middlesex, United Kingdom)
RDA: *NR*
Fluoride: 1,400 ppm F-

**Tcs-DF:** containing 0.3% triclosan, 2% copolymer and 0.243% sodium fluoride; Colgate Total ® (Colgate-Palmolive Company, New York, United States of America); Colour and taste modified similar to Parodontax ®
RDA: *NR*
Fluoride: 1,400 ppm F-

whole mouth PI ≥ 1.5 based on the TQ&H, a GI ≥ 1.0 based on the Löe-Silness Gingival Index and presence of established gingivitis.

Self-brushing, own technique

The participants were instructed to brush with the assigned dentifrice three times a day. Rinsing with water after toothbrushing was not allowed, as well as using any other dentifrice or mouthrinse during the experimental period.

All participants used dental floss during the study.

Mankodi et al. (1998)
34 (NR)
TB+ BS-DF

Self-brushing, own technique under

BS is an effective dentifrice abrasive
### Valkenburg et al.

RCT Crossover (w.o. 1w)
Single-brushing exercise
partial mouth
48 hours no oral hygiene procedures

| Population: NR | Location: NR |
| --- | --- |
| ♂: NR |
| ♀: NR |
| Mean age: NR |
| Age range: 18-65 |
| Smoking status: NR |
| OP: NR |

TB+ NaF-DF
TB+ MFP+NaF-DF
TB: NR

**BS-DF:** 65% sodium bicarbonate; Arm & Hammer Dental Care® (Church & Dwight Co., Inc., Princeton, NJ)
RDA: 30-40
Fluoride: NR

**NaF-DF:** hydrated silica, sodium fluoride ◊; Crest® Regular DF (Procter & Gamble, Cincinnati, OH)
RDA: 100-110
Fluoride: NR ppm F-

**MFP+NaF-DF:** dicalcium phosphate, (monofluorophosphate & sodium fluoride?); Colgate® Regular Toothpaste (Colgate-Palmolive Co., New York, NY)
RDA: 70-80
Fluoride: NR ppm F-

Refraining from other forms of oral hygiene

### Beiswanger et al. (1997)

| 570 (554) |

TB+ BS-DF
TB+ NaF-DF

Participants did have at least 5 gingival bleeding

The sodium bicarbonate and NaF/peroxide

system for removing accumulated plaque and maintaining good oral health. superior to both other dentifrices tested.
| Study                        | Design          | Geographical Area          | Participants | Intervention Comparison | Brushing Frequency | Additional Instructions                                                                 |
|------------------------------|-----------------|-----------------------------|--------------|--------------------------|--------------------|----------------------------------------------------------------------------------------|
| Valkenburg et al. (1990) (6) | CCT Parallel (6m) | Indianapolis metropolitan area (Indiana, USA) | ♀: 413 (72.5%)  
♂: 157 (27.5%)  
Mean (sd) age: 36.3 (10.5)  
Age range: NR  
Smoking status: NR  
OP: yes (after baseline) | TB+ SnF-DF  
TB: NR  
BS-DF: BS, peroxide, 0.243% sodium fluoride; Mentadent® Fluoride dentifrice; (Chesebrough-Ponds, Greenwich, CT).  
RDA: NR  
Fluoride: 1,000 ppm F-  
A) NaF-DF: 0.243% sodium fluoride, silica; Crest® Regular  
RDA: NR  
Fluoride: 1,100 ppm F-  
B) SnF-DF: 0.454% stabilized stannous fluoride, silica; Crest® Plus Gum Care; (Procter & Gamble, Cincinnati, OH, USA)  
RDA: NR  
Fluoride: 1,000 ppm F- | sites based on the Löe-Silness Index.  
New toothbrush provided at baseline and three-month examination.  
Self-brushing at home, own technique  
Brushing twice-daily  
Refraining from other dentifrices, mouthrinses or oral irrigation products.  
dentifrice did not provide reductions in gingivitis, plaque or gingival bleeding as compared with the conventional NaF dentifrices. |

| Mullally et al. (1995) (5)   | RCT Parallel (6w) | Northern-Ireland? Belfast (Northern-Ireland) | 70 (70) | TB+ BS-DF  
TB+ MFP+NaF-DF  
TB: Oral B 35 (Oral B laboratories,  
Participants had to be a whole mouth PI > 2 based on the TQ&H.  
The herbal based (sodium bicarbonate) dentifrice is as effective as the conventionally formulated dentifrice in


| Partial mouth (index teeth) | Hours of no oral hygiene procedures: NR |
|-----------------------------|----------------------------------------|
| ♂: 36 (51.4%) ♀: 34 (48.6%)  |
| Mean age: NR Age range: 18-65|
| Smoking status: NR OP: yes (after baseline and at the end of the study) |
| Aylesbury, Bucks, England |

**BS-DF:** sodium bicarbonate, herbal ingredients (camomile, echinacea, sage, myrrh, rhatany, peppermint oil); Parodontax® (Madaus, Cologne, Germany)

**RDA:** NR

Fluoride: **NR ppm F-**

**MFP+NaF-DF:** monofluorophosphate, sodium fluoride; Colgate (Colgate-Palmolive, London, England)

**RDA:** NR

Fluoride: **NR ppm F-**

Self-brushing, own technique, twice a day. Refraining from other oral hygiene products.

Participants were asked to keep a diary to record compliance with the brushing instructions

the control of plaque and gingivitis, however, it has no significant clinical advantage over a traditionally formulated dentifrice.

---

**Saxer et al. (1995) (45)**

**RCT**

Parallel (2m) Single Centre Half mouth, 60 (46%)

Factory workers Köln (Germany)?

♂: NR

TB+ BS-DF TB+ Non-BS Herbal-DF TB: Oral-B35

Participants with clinical signs of gingivitis and no pockets exceeding 5 mm.

Self-brushing own

The amount of plaque did not appear to be influenced by the dentifrices used.
| Valkenburg et al. | Saxer et al. (1994) (44) | 22 (22) Zürich, Switzerland? | Participants with clinical signs of gingivitis and no pockets exceeding 5 mm. | There was a significant decrease found for the sodium bicarbonate dentifrice in BOP compared to the control group and baseline |
|-----------------|--------------------------|-------------------------------|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| contralateral   | BS-P: NR                 | TB+ BS-DF                    | Self-brushing at home, own technique                                      |
| Hours of no oral hygiene procedures: NR | BS-DF: sodium bicarbonate, herbal ingredients; Parodontax®; (Madaus AG, Cologne, Germany) | TB+ Non-BS F-DF | Brushing twice a day. Refraining from mouthwashes or chewing gum |
| Mean age: NR    | RDA: NR                  | Fluoride: NR ppm F-         | The gingivitis and bleeding indices decreased significantly by 40% in both groups compared to the baseline examination. |
| Age range: NR  | Fluoride: NR ppm F-      |                              |                                                                         |
| Smoking status: NR | TB: Parodontax® toothbrush |                              |                                                                         |
| OP: NR          |                          |                              |                                                                         |
| Non-BS Herbal: calcium hydrogen phosphate, herbal ingredients, non-marketed new DF |                          |                              |                                                                         |
| RDA: NR         |                          |                              |                                                                         |
| Fluoride: NR ppm F- |                          |                              |                                                                         |
| Study                  | Design | Duration | Location | Population | Mean age | Age range | Smoking status | Oral hygiene | Product | Examiner | Hours of no oral hygiene procedures | Notes |
|------------------------|--------|----------|----------|------------|----------|-----------|----------------|--------------|---------|----------|-----------------------------------|--------|
| Valkenburg et al.      | RCT    | 42 (40◊) | NR       | 17 (42.5%◊) | 31.0     | NR        | NR             | NR           | NR      | NR       | NR                  | Refraining from other forms of oral hygiene: NR | There was a significant decrease found for the sodium bicarbonate dentifrice in bleeding sites compared to control group and baseline |
| Taller (1993) (43)     | RCT    | Parallel (5w) | Full mouth (examined 6 teeth) | 42 (40◊) | 46       | 21-71     | NR             | TB+ BS-DF    | NR      | NR       | NR                  | In this study, the sodium bicarbonate dentifrice did not show a significant drop in bleeding points compared to the baseline. Additionally, there was no difference compared to the control groups. |
| Yankell et al. (1993) (42) | RCT    | 128 (123◊) | NR       | NR         | NR       | NR        | NR             | TB+ BS-DF    | NR      | NR       | NR                  | The sodium bicarbonate dentifrice was found to be significantly more effective in reducing plaque gingivitis and bleeding on probing than the placebo. |
| Study | Duration | Population | Oral Hygiene | Ingredients | Notes |
|-------|----------|------------|--------------|-------------|-------|
| Valkenburg et al. | 10-12 hours of no oral hygiene procedures | Mean age: NR Age range: 18-65 (as inclusion) Smoking status: NR OP: NR | Non-BS Herbal-DF: calcium carbonate, herbal ingredients; non-marketed modified (similar consistency, colour, taste as Parodontax®); (Madaus AG, Cologne, Germany) | commercial mouthrinse or other oral hygiene solutions with the exception of floss. | |
| Emling and Yankell (1988) (41) | 12 (NR) Population: ?, Philadelphia, Pennsylvania USA? | TB+ BS-DF TB+ NaF-DF TB: NR BS+DF: Parodontax® RDA: NR Fluoride: NR ppm F- | GI: Statistical significant difference favoring the BS dentifrice. BOP: the BS dentifrice was significantly more effective than both placebo and sodium fluoride dentifrice. | |
| Yankell and Emling (1988) (40) | ±60° (NR) Northern-American University employees and members of the public at large. University of Pennsylvania School of Dental Medicine (USA)? | TB+ BS-DF TB+ Non-BS DF TB+ NaF-DF TB: Own toothbrush used BS-P: Parodontax® | There was no significant difference found between brushing with the sodium bicarbonate dentifrice and sodium fluoride dentifrice in plaque scores. | |
| Valkenburg et al. | Winer et al. (1986) (39) |
|----------------|------------------------|
| **10-12 hours of no oral hygiene procedures** | **RCT** |
| ♀: NR | 50 (39%) |
| ♂: NR | Essex county jail, Salem, MA |
| Mean age: NR | ♀: 0 |
| ♂: 50 | ♂: 50 |
| Smoking status: NR | 50 (39%) |
| Age range: NR | MA |
| OP: NR | Mean age: NR |
| **RDA:** NR | **TB+ BS-DF** |
| **Fluoride:** NR ppm F- | **TB+ Non-BS DF** |
| **A) Non-BS DF:** modified dentifrice with laboratory abrasivity indices similar to Parodontax® with no known active ingredients. | **TB:** Butler no. 411 |
| **RDA:** NR | **BS-DF:** silica-based, essential oils, sodium bicarbonate, sodium chloride, zinc chloride, and sodium fluoride; |
| **Fluoride:** NR ppm F- | Participants with varying degrees of gingival inflammation and plaque accumulation. |
| **B) Na-DF:** sodium fluoride◊; Crest Tartar Control® | Self-brushing, own technique |
| **RDA:** NR | For both the BS and non-BS dentifrice groups the decrease in the PI was significant. However, there was no significant difference between the two dentifrices. |
| Age range: 18-? | Kolynos Periodontal Dental Cream® |
|----------------|----------------------------------|
| Median age: 24 | RDA: NR                           |
| Smoking status: NR | Fluoride: NR ppm F-             |
| OP: NR          |                                   |

**Non-BS DF:**
hydrated calcium phosphate-base, essential oils.
RDA: NR
Fluoride: NR ppm F-

| Brush teeth in usual manner twice a day. |
| Refrained from all other oral hygiene products. |
| On average three quarter of a 6-oz. tube was used during the test period. |

Ø: calculated by the authors of this review based on the data presented in the selected paper

?: unclear

BOP: Bleeding on Probing

CHX: Chlorhexidine

d: day

DF: Dentifrice

F: Fluoride

GI: Gingival Index

m: month

MFP: Monofluorophosphate

NaF: Sodium fluoride

NR: not reported

OP: Oral Prophylaxis; at the initial appointment, all teeth were thoroughly scaled and polished.

Pi: Plaque Index

RDA: Relative (or Radioactive) Dentine Abrasivity (or Abrasion)

SnF: Stannous fluoride

TB: Toothbrush

Tcs: Triclosan

w: week

w.o.: washout period (crossover design)
### Appendix S3: Methodological quality and potential risk of bias scores of the individual included studies.

| Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Akwagyiram et al. (2018) (37) | Bosina et al. (2018) (38) | Jose et al. (2017) (54) | Hosada et al. (2017) (36) | Mason et al. (2017) (52) | Lomax et al. (2016) (52) | Triratana et al. (2015) (51) | Al-Kholani et al. (2011) (35) | Ghassemi et al. (2009) (49, 50) | Putt et al. (2008) (49) | Ozaki et al. (2006) (48) | Mantouk et al. (1998) (47) | Beiswenger et al. (1997) (46) | Saxer et al. (1995) (45) | Saxer et al. (1994) (44) | Taller (1993) (43) | Yankell et al. (1993) (42) |
| Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      | Internal validity      |
| Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          | Study design          |
| Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    | Random allocation*    |
| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment| Allocation concealment|
| Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   | Blinded to product*   |
| Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  | Blinded to examiner*  |
| Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis | Blinding during statistical analysis |
| Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* | Balanced experimental groups* |
| Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* | Reported loss to follow-up* |
| # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    | # (% of drop-outs)    |
| 19(7.7%)              | 30(12.2%)              | 5(10%)                | 11(5.0%)               | 0(0%)                 | NR                    | 11(5.0%)              | 0(0%)                 | NR                    | 16(2.8%)              | 0(0%)                 | 7(11.7%)              | 2(4.8%)               | NR                    | NR                    | NR                    | 11(22%)               |
| 30(12.2%)             | 5(10%)                 | 12(8.8%)              | 0(0%)                  | NR                    | NR                    | 13(8.8%)              | 0(0%)                 | NR                    | 13(12.5%)             | 0(0%)                 | 7(11.7%)              | 2(4.8%)               | NR                    | NR                    | NR                    | 11(22%)               |
|                                                  | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
|-------------------------------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Treatment identical, except for intervention*   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Representative population group                  | + | + | + | + | + | + | + | + | NR | + | + | + | + | + | NR | + |
| Eligibility criteria defined*                    | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | - |
| Sample size calculation and power                | + | + | + | + | + | + | + | NR | NR | NR | NR | NR | NR | NR | NR | NR |
| Point estimates presented for the primary outcome| + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Measures of variability presented for the primary outcome | + | + | + | + | + | + | + | + | - | + | + | + | + | + | + | + |
| Unit of analysis                                 | participant |
| Included a per protocol analysis                 | + | + | + | + | + | + | + | NR | + | + | + | + | + | + | + | + |
| Included an intention-to-treat analysis          | + | + | - | - | + | - | NR | - | - | - | - | - | NA | ? | ? | - |
| Validated measurement                            | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Calibration examiner                             | + | + | + | + | NR | NR | NR | NR | NR | NR | + | NR | + | NR | + | + |
| Reproducibility data shown                       | + | + | + | - | + | - | - | - | - | - | - | - | - | - | - | - |
| Authors' estimated risk of bias                  | Low | Low | Low | Low | Low | Moderate | Low | High | Moderate | High | Low | Moderate | High | Low | Moderate | Low | Moderate | High | High | High |

As stated by Van der Weijden et al. (2009) (9), a low risk of bias was defined as presence of random allocation, blinding of both participants and examiners, balanced experimental groups, reporting loss to follow-up, identical treatment between groups (apart from the intervention) and defined eligibility criteria. Absence of one criterion was considered a possible moderate risk of bias. Studies meeting five or less criteria were given the high-risk estimate.
Appendix S4: Mean (SD) scores for the different intervention groups with reported significance of changes within groups, including various indices and their modifications extracted from the included papers.

Appendix S4a (PI): Turesky et al. (1970) modification of the Quigley & Hein Plaque Index (1962).

| # | DF GROUP | N       | Baseline | End        | Mean Difference | Significant within groups |
|---|----------|---------|----------|------------|----------------|---------------------------|
|   |          |         |          |            | Mean         | %                        |
| 1 | Akwagiyaram et al. (2018) (37) | BS67% NaF | (N=118) | 3.04 (0.206) | 2.52 (0.348) | -0.52 (0.05) | 17.1% (0.05) | NR |
|   |          |         | (N=122) | 3.00 (0.210) | 2.95 (0.221) | -0.05 (0.05) | 17.1% (0.05) | NR |
|   | Bosma et al. (2018) (38) | Single-brushing | A) BS67%herbal | (N=56) | 2.90 (0.308) | 1.76 (0.514) | -1.14 (0.430) | 39.0% (0.05) | NR |
|   |          |         | B) BS67% | (N=56) | 2.92 (0.312) | 1.75 (0.492) | -1.14 (0.426) | 39.0% (0.05) | NR |
|   |          |         | C) BS62%herbal | (N=56) | 2.89 (0.304) | 1.77 (0.469) | -1.11 (0.419) | 39.4% (0.05) | NR |
|   |          |         | D) NaF | (N=56) | 2.86 (0.309) | 1.96 (0.470) | -0.90 (0.409) | 31.5% (0.05) | NR |
|   | Jose et al. (2018) (54) | BS67% NaF | (N=118) | 3.05 (0.403) | 2.47 (0.437) | -0.58 (0.13) | 19.0% (0.05) | NR |
|   |          |         | (N=117) | 3.05 (0.353) | 2.92 (0.399) | -0.13 (0.13) | 4.3% (0.05) | NR |
|   | Hosadurga et al. (2017) (36) | BS MFP | (N=25) | 1.99 (0.44) | 1.57 (0.356) | -0.42 (0.34) | 21.0% (0.05) | YES |
|   |          |         | (N=20) | 2.07 (0.44) | 1.41 (0.316) | -0.66 (0.34) | 31.8% (0.05) | YES |
|   | Mason et al. (2017) (53) | Single-brushing | A) BS67% | (N=60) | 2.54 (0.50) | 0.83 (0.50) | -1.71 (0.36) | 61.0% (0.05) | NR |
|   |          |         | B) BS45% | (N=60) | 2.48 (0.45) | 0.80 (0.51) | -1.68 (0.36) | 61.8% (0.05) | NR |
|   |          |         | C) NaF | (N=60) | 2.54 (0.49) | 0.99 (0.57) | -1.55 (0.36) | 61.0% (0.05) | NR |
|   | Mason et al. (2017) (53) | Single-brushing | B) BS67% SnF | (N=54) | 2.59 (0.47) | 0.80 (0.51) | -1.80 (0.43) | 68.4% (0.05) | NR |
|   |          |         | (N=54) | 2.62 (0.42) | 1.10 (0.51) | -1.52 (0.43) | 58.0% (0.05) | NR |
|   | Triratana et al. (2015) (51) | BS MFP+NaF | (N=45) | 3.56 (0.36) | 3.23 (0.35) | -0.33 (0.46) | 9.0% (12.52%) | YES |
|   |          |         | (N=45) | 3.53 (0.37) | 3.40 (0.39) | -0.13 (0.57) | 4.2% (16.48%) | YES |
|   |          |         | (N=45) | 3.55 (0.36) | 1.65 (0.45) | -1.90 (0.56) | 53.5% (13.58%) | YES |
|   | Ghassemi et al. (2008) (50) | daily brushing | BS Tcs | (N=105) | 2.90 (0.40) | 2.55 (0.43) | -0.34 (0.32) | 11.8% (?| YES |
|   |          |         | (N=102) | 2.90 (0.39) | 2.74 (0.40) | -0.15 (0.24) | 5.3% (?| YES |
|   | Ghassemi et al. (2008) 1 | Single-brushing (baseline visit) | BS Tcs | (N=105) | 2.90 (0.40) | 2.36 (0.46) | -0.54 (0.26) | 18.5% (?| YES |
|   |          |         | (N=102) | 2.90 (0.39) | 2.61 (0.40) | -0.28 (0.18) | 9.8% (?| YES |
|   | Ghassemi et al. (2008) 2 | BS | (N=105) | 2.55 (0.43) | 2.09 (0.45) | -0.47 (0.21) | 18.3% (?| YES |
| **Single-brushing (final visit)** |   | Tcs   | (N=102) | 2.74(0.40) | 2.52(0.39) | -0.23(0.15) | -8.2%(?) | YES |
|----------------------------------|---|-------|---------|------------|------------|-------------|---------|-----|
| **Putt et al. (2008) (49)**     | Sub-study 1 | Single-brushing | A) BS65% | (N=65) | 3.16(0.35) | 2.43(0.42) | -0.73(0.26) | -23.1%(?) | YES |
|                                  | B) BS20%    |         | (N=65)  | 3.18(0.37) | 2.55(0.48) | -0.64(0.26) | -20.1%(?) | YES |
|                                  | C) Tcs      |         | (N=65)  | 3.16(0.37) | 2.64(0.42) | -0.52(0.20) | -16.5%(?) | YES |
| **Putt et al. (2008)**           | Sub-study 2 | Single-brushing | A) BS48% | (N=66) | 3.05(0.35) | 2.33(0.42) | -0.72(0.28) | -23.6%(?) | YES |
|                                  | B) BS20%    |         | (N=66)  | 3.02(0.36) | 2.35(0.39) | -0.68(0.27) | -22.5%(?) | YES |
|                                  | C) Tcs      |         | (N=66)  | 3.03(0.33) | 2.45(0.35) | -0.58(0.25) | -19.1%(?) | YES |
|                                  | D) SnF      |         | (N=66)  | 3.04(0.40) | 2.45(0.38) | -0.59(0.26) | -19.4%(?) | YES |
| **Putt et al. (2008)**           | Sub-study 3 | Single-brushing | BS48%   | (N=66) | 3.14(0.31) | 2.36(0.35) | -0.78(0.31) | -24.8%(?) | YES |
|                                  | BS20%       |         | (N=66)  | 3.15(0.34) | 2.42(0.36) | -0.73(0.32) | -23.2%(?) | YES |
|                                  | C) Tcs      |         | (N=66)  | 3.12(0.40) | 2.51(0.35) | -0.62(0.27) | -19.9%(?) | YES |
|                                  | D) NaF      |         | (N=66)  | 3.13(0.34) | 2.54(0.31) | -0.58(0.24) | -18.5%(?) | YES |
| **Putt et al. (2008)**           | Sub-study 4 | Single-brushing | BS65%   | (N=36) | 2.94(0.35) | 2.14(0.35) | -0.80(0.31) | -27.2%(?) | YES |
|                                  | NaF         |         | (N=36)  | 2.97(0.34) | 2.50(0.36) | -0.47(0.23) | -15.8%(?) | YES |
| **Putt et al. (2008)**           | Sub-study 5 | Single-brushing | BS65%   | (N=41) | 2.28(0.36) | 1.21(0.50) | -1.07(0.38) | -46.9%(?) | YES |
|                                  | NaF         |         | (N=41)  | 2.25(0.33) | 1.36(0.47) | -0.87(0.37) | -38.7%(?) | YES |
| **Ozaki et al. (2006) * (48)**   | **whole mouth** | BS   | (N=20)  | 2.28(0.68) | 1.82(0.58) | -0.46(0.19) | -20.6%(8.14) | YES* |
|                                  | Tcs         |         | (N=22)  | 2.26(0.60) | 1.89(0.57) | -0.38(0.33) | -16.5%(14.22) | YES* |
| **Mullally et al. (1995) (5)**   | BS           | MFP+NaF | (N=35)  | 2.4(0.4) | 1.6(0.3) | -0.8(0.3) | -33.3%(?) | YES |
|                                  | Tcs         |         | (N=35)  | 2.6(0.5) | 1.8(0.4) | -0.8(0.3) | -30.8%(?) | YES |
| **Saxer et al. (1995) (45)**     | BS           | Non-BS  | (N=23)  | 2.22(0.44) | 2.19(0.40) | -0.03(0.3) | -1%(?) | NO |
|                                  | Tcs         |         | (N=23)  | 2.21(0.50) | 2.15(0.44) | -0.06(0.2) | -3%(?) | NO |
| **Yankell et al. (1993) (42)**   | BS           | Non-BS  | (N=60)  | 3.08(0.65) | 2.45(0.69) | -0.63(0.3) | -20.5%(?) | YES |
|                                  | Tcs         |         | (N=63)  | 3.18(0.53) | 2.75(0.58) | -0.43(0.2) | -13.5%(?) | YES |
| **Emling and Yankell (1988) (41)** | Single-brushing | BS-P DF | (N=60)  | 2.71(0.83) | 2.13(1.15) | -0.58(0.7) | -21.4%(?) | YES |
|                                  | NaF-DF      |         | (N=60)  | 2.71(0.85) | 1.82(1.27) | -0.89(0.7) | -32.8%(?) | YES |
| **Yankell and Emling (1988) (40)** | **Non-BS** | BS     | (N=20)  | 1.06(0.52) | 0.98(0.69) | -0.08(0.7) | -7.5%(?) | NO |
|                                  | A) Non-BS   |         | (N=20)  | 1.15(0.71) | 1.10(0.44) | -0.05(0.3) | -4.3%(?) | NO |
|                                  | B) NaF      |         | (N=20)  | 1.16(0.54) | 0.95(0.72) | -0.21(0.2) | -18.1%(?) | NO |

* based on pooled data ♦ additional data provided by the original authors ♣ additional data from ClinicalTrials.gov
Appendix S4b (PI): Distal Mesial Plaque Index Fischman et al. (1987).

| GROUP | Baseline | End | Difference | Significant within groups |
|-------|----------|-----|------------|--------------------------|
| Mankodi et al. (1998) (47) single-brushing | BS-DF(N=34◊) NaF-DF(N=34◊) MFP+NaF-DF(N=34◊) | NR | NR | NR | -66.93%(?) | -58.39%(?) | -58.65%(?) | NR |

Appendix S4c (PI): Löe (1967) modification of the Silness & Löe (Silness & Löe 1964) Plaque Index.

| GROUP | Baseline | End | Difference | Significant within groups |
|-------|----------|-----|------------|--------------------------|
| Beiswanger et al. (1997) (46) | BS-DF(N=147) A) NaF-DF(N=140) B) SnF-DF(N=267) | 0.73(0.36◊) 0.67(0.35◊) 0.73(0.33◊) | 0.58(0.24◊) 0.54(0.24◊) 0.55(0.33◊) | -0.15◊(?) -0.13◊(?) -0.18◊(?) | -20.5%◊(?) -19.4%◊(?) -24.7%◊(?) | NO | NO | NO |
| Winer et al. (1986) (39) | BS-DF(N=18) Non-BS DF(N=21) | 2.20◊(1.07◊) 2.04◊(0.93◊) | 1.37◊(0.96◊) 1.21◊(0.76◊) | -0.83◊(?) -0.83◊(?) | -37.6%(?) -40.5%(?) | YES | YES |
Appendix S4d (PI): Podshadley & Haley (1968) Patient Hygiene Performance Index.

| #          | GROUP                          | Mean (SD)       | Difference          | Significant within groups |
|------------|--------------------------------|-----------------|---------------------|---------------------------|
|            |                                | Baseline        | End                 | Mean %                    |                           |
|            |                                |                 |                     |                           |                           |
|            | Al-Kholani. (2011) (35)        |                 |                     |                           |                           |
|            | BS-DF(N=16)                    | 2.56(0.06)      | 0.96(0.07)          | -1.60(?) -61.2%(?)        | YES                       |
|            | A) Non-BS Herbal DF(N=16)      | 2.57(0.09)      | 1.07(0.06)          | -1.50(?) -57.2%(?)        | YES                       |
|            | B) NaF-DF(N=16)                | 2.71(0.12)      | 1.77(0.06)          | -0.94(?) -33.3%(?)        | NO                        |
### Appendix S4e (PI): Lange et al. (1977) Approximal Plaque Index.

| # | GROUP | Baseline | End | Difference | Significant within groups |
|---|-------|----------|-----|------------|--------------------------|
|   | Al-Kholani. (2011) (35) | BS-DF(N=16) |   |   |   |
|   |       | A) Non-BS Herbal DF(N=16) | 65.3 (2.1) | 26.1 (1.1) | -39.2%(?) | YES |
|   |       | B) NaF-DF(N=16) | 63.5 (1.7) | 28.3 (1.2) | -35.2%(?) | YES |
|   |       |           | 61.2 (2.2) | 45.6 (1.5) | -15.6%(?) | NO |
## Appendix S4f (GI): Löe & Silness (1963) and Löe (1967) Gingival Index.

| #   | GROUP                                      | Mean (SD)     | Significant within groups |
|-----|-------------------------------------------|---------------|--------------------------|
|     |                                           | Baseline      | End          | Difference               | Mean    | %       |
|     |                                           |               |              |                          |         |         |
|     | **Hosadurga et al. (2017)**' (36)         |               |              |                          |         |         |
|     | BS-DF(N=25)                               | 1.468(0.329)  | 1.088(0.289) | -0.38★(0.35★)            | -25.92%?|         |
|     | MFP-DF(N=20)                              | 1.493(0.327)  | 1.207(0.114) | -0.29★(0.34★)            | -19.14%?|         |
|     | **Triratana et al. (2015)**' (51)         |               |              |                          |         |         |
|     | BS-P DF(N=45)                             | 1.77(0.20)    | 1.47(0.30)   | -0.30★(0.34★)            | -17.4%(18.60%) | YES   |
|     | A) MFP+NaF-DF(N=45)                       | 1.77(0.15)    | 1.57(0.28)   | -0.20★(0.35★)            | -11.8%(19.06%) | YES   |
|     | B) Tcs-DF(N=45)                           | 1.78(0.20)    | 0.95(0.26)   | -0.83★(0.22★)            | -46.6%(12.01%) | YES   |
|     | **Al-Kholani. (2011)**' (35)              |               |              |                          |         |         |
|     | BS-DF(N=16)                               | 1.86 (0.11)   | 0.57 (0.04)  | -1.29?(?)                | -68.0%(?) | YES   |
|     | A) Non-BS-Herbal DF(N=16)                 | 1.75 (0.08)   | 0.49 (0.03)  | -1.26?(?)                | -70.6%(?) | YES   |
|     | B) NaF DF (N=16)                          | 1.87 (0.11)   | 1.09 (0.9)   | -0.78?(?)                | -40.5%(?) | YES   |
|     | **Ozaki et al. (2006)** *'(48)            |               |              |                          |         |         |
|     | BS-DF(N=20)                               | 1.10★(0.25★)  | 0.82★(0.35★) | -0.28★(0.21★)            | -27.2%(20.83%) | YES*  |
|     | Tcs-DF(N=22)                              | 1.11★(0.27★)  | 0.73★(0.27★) | -0.38★(0.13★)            | -35.7%(9.11%) | YES*  |
|     | **Beiswanger et al. (1997)** *'(46)       |               |              |                          |         |         |
|     | BS-DF(N=147)                              | 0.89(0.24◊)   | 0.74(0.24◊)  | -0.15?(?)                | -16.7%(?) | NO    |
|     | A) NaF-DF(N=140)                          | 0.84(0.24◊)   | 0.78(0.24◊)  | -0.06?(?)                | -7.1%(?) | NO    |
|     | B) SnF-DF(N=267)                          | 0.86(0.33◊)   | 0.64(0.16◊)  | -0.22?(?)                | -25.6%(?) | NO    |
|     | **Mullally et al. (1995)**' (5)           |               |              |                          |         |         |
|     | BS-DF(N=35◊)                              | 0.86(0.09)    | 0.41(0.19)   | -0.45(0.17)              | -52.3%(?) | YES   |
|     | MFP+NaF-DF(N=35◊)                         | 0.86(0.08)    | 0.47(0.17)   | -0.39(0.17)              | -45.3%(?) | YES   |
|     | **Saxer et al. (1995)** (45)              |               |              |                          |         |         |
|     | BS-DF (N=23◊)                             | 1.58(0.62)    | 0.89(0.30)   | -0.69?(?)                | -44%(?) | YES   |
|     | Non-BS Herbal-DF(N=23◊)                   | 1.46(0.58)    | 0.87(0.35)   | -0.59?(?)                | -40%(?) | YES   |
|     | **Yankell et al. (1993)** *'(42)          |               |              |                          |         |         |
|     | BS-DF(N=60)                               | 2.10(0.15◊)   | 0.84(0.64◊)  | -1.26(0?)                | -60.0%(0%) | YES*  |
|     | Non-BS Herbal-DF(N=63)                    | 2.11(0.17◊)   | 0.84(0.48◊)  | -1.27(0?)                | -60.2%(0%) | YES*  |
|     | **Yankell and Emling (1988)** *'(40)      |               |              |                          |         |         |
|     | BS-DF(N=20◊)                              | 0.56(0.25)    | 0.51(0.27)   | -0.05(0?)                | -8.9%(0?) | NO    |
|     | A) Non-BS DF (N=20◊)                      | 0.53(0.25)    | 0.56(0.31)   | +0.03(0?)                | +5.7%(0?) | NO    |
|     | B) NaF-DF(N=20◊)                          | 0.51(0.26)    | 0.55(0.25)   | +0.04(0?)                | +7.8%(0?) | YES   |
|     | **Winer et al. (1986)** *'z (39)          |               |              |                          |         |         |
|     | BS-DF(N=18 )                             | 2.95(0.49)    | 2.56(0.59)   | -0.40(0?)                | -13.5%(?) | YES   |
|     | Non-BS DF(N=21)                           | 2.87 (0.64)   | 2.61 (0.73)  | -0.26(0?)                | -9.1%(?) | YES   |

* based on pooled data  
' Löe & Silness Gingival Index (1963)  
'  Löe Gingival Index (1967)
Appendix S4g (GI): Modified Gingival Index (MGI). Gingival Index of Löe & Silness (1963) modified by Lobene et al. (1986).

| # | GROUP | Baseline | End | Difference | Significant within groups |
|---|--------|----------|-----|------------|--------------------------|
| | | Mean (SD) | Mean (SD) | Mean | % |
| | | Mean | % |
| | | | | | |
| Akwagyiram et al. (2018) (37) | BS67%(N=118) NaF(N=122) | 2.53(0.185) 2.54(0.199) | 2.23(0.228) 2.43(0.228) | -0.30(?) -0.11(?) | NR NR |
| Jose et al. (2018) (54) | BS67%(N=118) NaF(N=117) | 2.32(0.345) 2.35(0.327) | 1.26(0.505) 1.93(0.408) | -1.06(?) -0.42(?) | NR NR |
| Lomax et al. (2016) (52) | BS-DF(N=66) NaF-DF(N=69) | 1.26(0.28) 1.26(0.23) | 0.64(0.24) 0.91(0.27) | -0.61(0.14) -0.34(0.16) | NR NR |
### Appendix S4h-1 (BI): Ainamo & Bay (1975) Gingival Bleeding Index.

| #   | GROUP                  | Mean (SD)          | Significant within groups |
|-----|------------------------|--------------------|---------------------------|
|     | Baseline               | End                | Difference                |
|     |                        | Mean (SD)          | %                         |
| 1   | Mullally et al. (1995) | BS-DF(N=35◊)       | NR                        | NR                       | -32.3(14.6)      | -29.1(13.4)      | YES          |
|     | MFP+NaF-DF(N=35◊)      |                    |                           |                          |                       |              |
| 2   | Taller (1993)          | BS-DF(N=14)        | 0.57◊(0.16◊)              | 0.51◊(0.18◊)             | -0.06(0.20)         | -11%(?)       | NO           |
|     | Non-BS DF(N=13)        | 0.57◊(0.07◊)       | 0.52◊(0.09◊)              | -0.05(0.13)              | -9%(?)              |              |

1. Bleeding points on four areas of 12 teeth (17, 15, 13, 21, 24, 26, 37, 35, 33, 42, 44, 46); % of sites bleeding calculated for each participant.
2. Bleeding (Ainamo & Bay Gingival Bleeding Index (1975) points scores (yes/no) on four areas of six teeth (16, 21, 24, 36, 41, 44) totaled per participant.

### Appendix S4h-2 (BI): Abrams, Caton and Polson (1984) Bleeding on Probing Index.

| #   | GROUP                  | Mean (SD)          | Significant within groups |
|-----|------------------------|--------------------|---------------------------|
|     | Baseline               | End                | Difference                |
|     |                        | Mean (SD)          | %                         |
| 1   | Yankell et al. (1993)  | BS-DF(N=60)        | 0.26(0.25)                | 0.24(0.27)              | -0.02 ◊(?)         | -7.7%◊(?)      | NO           |
|     | Non-BS Herbal-DF(N=63)| 0.31(0.23)         | 0.36(0.32)                | +0.05 ◊(?)              | +16.1%◊(?)         |              |
| 2   | Yankell and Emling (1988) | BS-DF (N=200)     | 0.50(0.22)                | 0.26(0.20)              | -0.24◊(?)          | -48.0%◊(?)     | YES          |
|     | A) Non-BS DF (N=200)  | 0.44(0.26)         | 0.44(0.28)                | 0.0%◊(?)                | 0.0%◊(?)           |              |
|     | B) NaF-DF (N=200)     | 0.31(0.23)         | 0.34(0.33)                | +0.03◊(?)               | +9.7%◊(?)          |              |
Appendix S4i-1 (BI): Ainamo & Bay Gingival Bleeding Index (1975) and the modification by Saxer et al. (1977).

| GROUP | Baseline | End | Difference Mean | Difference % |
|-------|----------|-----|----------------|--------------|
| Saxer et al. (1995) | 1.19(0.46) | 0.73(0.32) | -0.46(?) | -39%(?) | YES |
| Saxer et al. (1994) | 1.06(0.28) | 0.83(0.31) | -0.56(?) | -22.0%(?) | YES |

1 Bleeding points on four areas of 12 teeth (17, 15, 13, 21, 24, 26, 37, 35, 33, 42, 44, 46); % of sites bleeding calculated for each participant.

Appendix S4i-2 (BI): Gingival Bleeding Index by Saxton & van der Ouderaa (1989).

| GROUP | Baseline | End | Difference Mean | Difference % |
|-------|----------|-----|----------------|--------------|
| Akwagyiram et al. (2018) | 0.45(0.152) | 0.19(0.130) | -0.26(?) | -57.8%(?) | NR |
| Jose et al. (2018) | 0.38(0.167) | 0.17(0.116) | -0.21(?) | -55.3%(?) | NR |
| Lomax et al. (2016) | 1.00(0.27) | 0.36(0.16) | -0.64(0.18) | -64.0%(?) | NR |
### Appendix S4j (BI): Mühlemann & Son (1971) Sulcus Bleeding Index.

| # | GROUP | Mean (SD) | Difference | Significant within groups |
|---|---|---|---|---|
| | | Baseline | End | Mean | % |
| | | | | | |
| | Al-Kholani (2011) (35) | BS-DF(N=16) | | | |
| | A) Non-BS Herbal DF(N=16) | 45.2 (3.2) | 9.1 (1.2) | -36.1(?) | -79.5%(?) |
| | B) NaF-DF (N=16) | 46.4 (2.5) | 8.3 (1.0) | -38.1(?) | -81.2%(?) |
| | | 41.1 (3.4) | 22.8 (2.1) | -18.3(?) | -42.1%(?) |

### Appendix S4k (BI): Number of Bleeding Sites.

| # | GROUP | Mean (SD) | Difference | Significant within groups |
|---|---|---|---|---|
| | | Baseline | End | Mean | % |
| | | | | | |
| | Akwagyiram et al. (2018) (37) | BS67% | 45.13(1.393) | 18.67(1.142) | -26.46(?) | -58.6%(?) |
| | NaF | 45.84(1.448) | 36.39(1.510) | -9.45(?) | 20.6%(?) |
| | Jose et al. (2018) (54) | BS67%(N=118) | 29.76(8.238) | 13.24(7.663) | -16.52(?) | -55.5%(?) |
| | NaF(N=117) | 29.05(8.390) | 24.90(9.044) | -4.15(?) | -14.3%(?) |
| | Lomax et al. (2016) (52) | BS-DF(N=66) | 76.0(13.06) | 33.2(12.21) | -42.7(10.15) | -56.2%(?) |
| | NaF-DF(N=69) | 73.5(12.25) | 42.5(12.85) | -31.0(8.71) | -42.2%(?) |
| | Beiswanger et al. (1997) (46) | BS-DF(N=147) | 27.42(21.95) | 21.82(10.67) | -5.6(?) | -20.4%(?) |
| | A) NaF-DF(N=140) | 23.36(16.21) | 22.25(10.65) | -1.10(?) | -4.8%(?) |
| | B) SnF2-DF (N=267) | 24.87(17.16) | 16.13(10.62) | -8.74(?) | -35.1%(?) |
| | Saxer et al. (1994) (44) | BS-DF(N=110) | 21.30 (3.77) | 17.10 (5.63) | -4.20(?) | -19.7%(?) |
| | Non-BS F-DF (N=110) | 20.33 (3.73) | 20.42 (4.08) | +0.08 (?) | +0.4%(?) |
Appendix S5: Overview of all reported indices included in this systematic review.

| Plaque indices                                      | Gingival indices                                      | Bleeding indices                                      |
|-----------------------------------------------------|-------------------------------------------------------|-------------------------------------------------------|
| The Turesky et al. (1970) modification of the Quigley and Hein Plaque Index (1962) (TQ&H) | The Löe & Silness Gingival Index (1963)                | The Ainamo & Bay Gingival Bleeding Index (1975)         |
| The Fischman et al. (1987) Distal Mesial Plaque Index | The Löe & Silness Gingival Index (1967)                | The Saxer et al. (1977) Papillary bleeding Index modification of the Ainamo & Bay Bleeding Index (1975) |
| The Löe. (1967) modification of the Silness & Löe Plaque Index (1964) | The Lobene et al. (1986) modification of the Gingival Index (MGI) | The Saxton & Van der Ouderaa (1989) Gingival Bleeding Index |
| The Podshadley & Haley Patient Hygiene Performance Index (1968) |                                                                 | The Abrams, Caton and Polson Bleeding on Probing Index (1984) |
| The Lange et al. Approximal Plaque Index (1977)       |                                                                 | The Mühlemann & Son Sulcus Bleeding Index (1971).      |
Appendix S6: PI
Forest plots of BASELINE, END and DIFFERENCE PLAQUE scores for the comparisons of dentifrices containing BS with negative and positive controls without BS.

Appendix S6-A) (PI): Turesky et al. (1970) modification of the Quigley & Hein (1962) Plaque Index analysing the single-brushing experiments.

Appendix S6-A1) Forest plot of the BASELINE PLAQUE scores of the SINGLE-BRUSHING exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; no significant difference was observed between groups.

| Study                  | BS DF       | Control DF    | Mean Difference | Mean Difference |
|------------------------|-------------|---------------|-----------------|-----------------|
|                        | BS DF       | Control DF    | Weight          | IV, Random, 95% CI |
|                        | Mean | SD | Total | Mean | SD | Total |                |                  |
| Putt 4 2008            | 2.94 | 0.35 | 36 | 2.97 | 0.34 | 36 | 14.2% | −0.03 [−0.19; 0.13] |
| Putt 3 A vs D 2008     | 3.14 | 0.31 | 32 | 3.13 | 0.34 | 32 | 14.2% | 0.01 [−0.15; 0.17] |
| Bosma C 2018           | 2.89 | 0.30 | 56 | 2.86 | 0.31 | 19 | 14.1% | 0.03 [−0.13; 0.19] |
| Bosma A 2018           | 2.90 | 0.31 | 56 | 2.86 | 0.31 | 19 | 14.0% | 0.04 [−0.12; 0.20] |
| Bosma B 2018           | 2.92 | 0.31 | 56 | 2.86 | 0.31 | 19 | 13.9% | 0.06 [−0.10; 0.22] |
| Putt 3 B vs D 2008     | 3.15 | 0.34 | 32 | 3.13 | 0.34 | 32 | 13.0% | 0.02 [−0.15; 0.19] |
| Mason 1 B 2017         | 2.48 | 0.45 | 60 | 2.54 | 0.49 | 30 | 8.4% | −0.06 [−0.27; 0.15] |
| Mason 1 A 2017         | 2.54 | 0.50 | 60 | 2.54 | 0.49 | 30 | 7.8% | 0.00 [−0.22; 0.22] |
| Emling & Yankell 1988  | 2.71 | 0.83 | 6  | 2.71 | 0.85 | 6  | 0.4% | 0.00 [−0.95; 0.95] |
| **Total (random effects, 95% CI)** | **394** | **223** | 100.0% | **0.01 [−0.05; 0.07]** |

**Prediction interval**

Heterogeneity: Tau² = 0; Chisq = 1.28, df = 8 (P = 1.00); I² = 0% [0%; 0%]

Test for overall effect: Z = 0.42 (P = 0.67)

Favours sodium bicarbonate  Favours negative control dentifrice
PI TQH BASELINE NEGATIVE SINGLE BRUSHING
Appendix S6-A2) Forest plot of the BASELINE PLAQUE scores of the SINGLE-BRUSHING exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S6-A3) Forest plot of the END PLAQUE scores of the SINGLE-BRUSHING exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; a significant difference was observed in favour of the use of BS.

| Study               | BS DF  | Control DF | Mean Difference | Mean Difference |
|---------------------|--------|------------|-----------------|-----------------|
|                     | Mean   | SD         | Total           | Mean           | SD         | Total           | Weight | IV, Random, 95% CI | IV, Random, 95% CI |
| Putt 3 A vs D 2008  | 2.36   | 0.35       | 32              | 2.54           | 0.31       | 32              | 20.2%   | -0.18 [-0.34; -0.02] |
| Putt 3 B vs D 2008  | 2.42   | 0.36       | 32              | 2.54           | 0.31       | 32              | 19.5%   | -0.12 [-0.28; 0.04]  |
| Putt 4 2008         | 2.14   | 0.43       | 36              | 2.50           | 0.36       | 36              | 15.8%   | -0.36 [-0.54; -0.18] |
| Mason 1 A 2017      | 0.83   | 0.50       | 60              | 0.99           | 0.57       | 30              | 9.2%    | -0.16 [-0.40; 0.08]  |
| Mason 1 B 2017      | 0.80   | 0.51       | 60              | 0.99           | 0.57       | 30              | 9.1%    | -0.19 [-0.43; 0.05]  |
| Bosma C 2018        | 1.77   | 0.47       | 56              | 1.96           | 0.47       | 19              | 8.9%    | -0.19 [-0.43; 0.05]  |
| Bosma B 2018        | 1.78   | 0.49       | 56              | 1.96           | 0.47       | 19              | 8.6%    | -0.18 [-0.43; 0.07]  |
| Bosma A 2018        | 1.76   | 0.51       | 56              | 1.96           | 0.47       | 19              | 8.4%    | -0.20 [-0.45; 0.05]  |
| Emling & Yankell 1988 | 2.13 | 1.15       | 6               | 1.82           | 1.27       | 6               | 0.3%    | 0.31 [-1.06; 1.68]   |

**Total (random effects, 95% CI)**

| Study               | BS DF  | Control DF | Mean Difference | Mean Difference |
|---------------------|--------|------------|-----------------|-----------------|
|                      | 394    | 223        | 100.0%          | -0.20 [-0.27; -0.12] |

**Prediction interval**

-1 -0.5 0 0.5 1

Heterogeneity: Tau² = 0; Chi² = 4.57, df = 8 (P = 0.80); I² = 0% [0%; 38%]

Test for overall effect: Z = -5.31 (P < 0.01)

Favours sodium bicarbonate  Favours negative control dentifrice
PI TQH SINGLE BRUSHING END NEGATIVE
Appendix S6-A4) Forest plot of the END PLAQUE scores of the SINGLE-BRUSHING exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS; a significant difference was observed in favour of the use of BS.
Appendix S6-A5) Forest plot of the DIFFERENCE PLAQUE scores of the SINGLE-BRUSHING exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; a significant difference was observed between groups.

| Study            | BS DF   | Control DF | Mean Difference | Mean Difference |
|------------------|---------|------------|-----------------|-----------------|
|                  | Mean    | SD         | Total           | Weight          | IV, Random, 95% CI |
| Putt 4 2008      | -0.80   | 0.31       | 36              | 19.9%           | -0.33 [-0.46; -0.20] |
| Putt 3 A vs D 2008 | -0.78  | 0.31       | 32              | 17.1%           | -0.20 [-0.34; -0.06] |
| Putt 3 B vs D 2008 | -0.73  | 0.32       | 32              | 16.5%           | -0.15 [-0.29; -0.01] |
| Mason 1 A 2017   | -1.71   | 0.36       | 60              | 13.0%           | -0.16 [-0.32; 0.00]  |
| Mason 1 B 2017   | -1.70   | 0.36       | 60              | 13.0%           | -0.15 [-0.31; 0.01]  |
| Bosma C 2018     | -1.11   | 0.42       | 56              | 6.9%            | -0.21 [-0.42; 0.00]  |
| Bosma B 2018     | -1.14   | 0.43       | 56              | 6.8%            | -0.24 [-0.46; -0.02] |
| Bosma A 2018     | -1.13   | 0.43       | 56              | 6.8%            | -0.23 [-0.45; -0.01] |

Total (random effects, 95% CI) 388 217 100.0% -0.21 [-0.27; -0.16]

Prediction interval [-0.28; -0.14]

Heterogeneity: Tau² = 0; Chi² = 5.29, df = 7 (P = 0.62); I² = 0% [0%; 57%]
Test for overall effect: Z = -7.37 (P < 0.01)
Appendix S6-A6) Forest plot of the DIFFERENCE PLAQUE scores of the SINGLE-BRUSHING exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS; a significant difference was observed between groups.
Appendix S6-B) (PI): Turesky et al. (1970) modification of the Quigley & Hein (1962) Plaque Index analysing the FOLLOW-UP brushing experiments.

Appendix S6-B1) Forest plot of the BASELINE PLAQUE scores of the FOLLOW-UP brushing exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S6-B2) Forest plot of the BASELINE PLAQUE scores of the FOLLOW-UP brushing exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S6-B3) Forest plot of the meta-subanalysis on ingredients showing END PLAQUE scores of the FOLLOW-UP brushing exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; a significant difference was observed between groups.

| Study or Subgroup | BS DF | Control DF | Mean Difference | Mean Difference |
|-------------------|-------|------------|-----------------|-----------------|
|                   | Mean  | SD         | Total           | Mean           | SD         | Total | Weight| IV, Random, 95% CI | IV, Random, 95% CI |
| **Non-BS**        |       |            |                 |                |            |       |       |                  |                  |
| Yankell 1993      | 2.45  | 0.69       | 60              | 2.75           | 0.58       | 63     | 11.4% | -0.30 [-0.53; -0.07] |
| Saxer 1995        | 2.19  | 0.40       | 23              | 2.15           | 0.44       | 23     | 10.9% | 0.04 [-0.20; 0.28] |
| Yankell & Emling A 1988 | 0.98  | 0.69       | 10              | 1.10           | 0.44       | 20     | 6.2%  | -0.12 [-0.59; 0.35] |
| **Total (random effects, 95% CI)** | 93    | 106        | 28.5%           | -0.13 [-0.37; 0.11] |        |       |       |                  |                  |
| Heterogeneity: Tau² = 0.0219; Chi² = 4.04, df = 2 (P = 0.13); I² = 51% [0%; 86%] |
| Test for overall effect: Z = -1.09 (P = 0.28) |

| NaF                |       |            |                 |                |            |       |       |                  |                  |
|-------------------|-------|------------|                 |                |            |       |       |                  |                  |
| Akwagiyaram 2018  | 2.52  | 0.35       | 118             | 2.95           | 0.22       | 122    | 14.7% | -0.43 [-0.50; -0.36] |
| Jose 2018         | 2.47  | 0.44       | 109             | 2.92           | 0.40       | 113    | 14.1% | -0.45 [-0.56; -0.34] |
| Yankell & Emling B 1988 | 0.98  | 0.69       | 10              | 0.95           | 0.72       | 20     | 5.3%  | 0.03 [-0.50; 0.56]  |
| **Total (random effects, 95% CI)** | 237   | 255        | 34.1%           | -0.43 [-0.51; -0.34] |        |       |       |                  |                  |
| Heterogeneity: Tau² = 0.0022; Chi² = 3, df = 2 (P = 0.22); I² = 33% [0%; 78%] |
| Test for overall effect: Z = -9.27 (P < 0.01) |

| MFP                |       |            |                 |                |            |       |       |                  |                  |
|-------------------|-------|------------|                 |                |            |       |       |                  |                  |
| Mulally 1995      | 1.60  | 0.30       | 35              | 1.80           | 0.40       | 35     | 12.9% | -0.20 [-0.37; -0.03] |
| Triratana A 2015  | 3.23  | 0.35       | 23              | 3.40           | 0.39       | 45     | 12.5% | -0.17 [-0.35; 0.01] |
| Hosadurga 2017    | 1.57  | 0.36       | 25              | 1.41           | 0.32       | 20     | 12.1% | 0.16 [-0.04; 0.36]  |
| **Total (random effects, 95% CI)** | 83    | 100        | 37.4%           | -0.07 [-0.29; 0.14] |        |       |       |                  |                  |
| Heterogeneity: Tau² = 0.0284; Chi² = 8.61, df = 2 (P = 0.01); I² = 77% [24%; 93%] |
| Test for overall effect: Z = -0.67 (P = 0.50) |

| Total (random effects, 95% CI) | 413  | 461        | 100.0%          | -0.19 [-0.34; -0.04] |     |       |       |                  |                  |
| Prediction interval          | [-0.69; 0.31] |                          | |
| Heterogeneity: Tau² = 0.0391; Chi² = 51.95, df = 8 (P < 0.01); I² = 85% [72%; 91%] |
| Test for overall effect: Z = -2.46 (P = 0.01) |
Appendix S6-B4) Forest plot of the END PLAQUE scores of the FOLLOW-UP brushing exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S6-B5) Forest plot of the DIFFERENCE PLAQUE scores of the FOLLOW-UP brushing exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S6-B6) Forest plot of the DIFFERENCE PLAQUE scores of the FOLLOW-UP brushing exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S6-C) (PI): The modification by Loe (1967) of the Silness & Loe (Silness & Loe 1964) Plaque Index.

Appendix S6-C1) Forest plot of the BASELINE PLAQUE scores of the FOLLOW-UP brushing exercises analysing the modification by Loe (1967) of the Silness & Loe (Silness & Loe 1964) Plaque Index for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S6-C2) Forest plot of the END PLAQUE scores of the FOLLOW-UP brushing exercises analysing the modification by Löe (1967) of the Silness & Löe (Silness & Löe 1964) Plaque Index for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S7: GI
Forest plots of BASELINE, END and DIFFERENCE gingival index scores for the comparisons of dentifrices containing BS with negative and positive controls without BS.
(GI): The Löe & Silness (1963) Gingival Index and the modification of this index by Löe (1967).

Appendix S7-1) Forest plot of the BASELINE GINGIVAL index scores of the FOLLOW-UP brushing exercises analysing the Löe & Silness (1963) Gingival Index and the modification of this index by Löe (1967) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S7-2) Forest plot of the BASELINE GINGIVAL index scores of the FOLLOW-UP brushing exercises analysing the Löe & Silness (1963) Gingival Index and the modification of this index by Löe (1967) for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S7-3) Forest plot of the meta-subanalysis on ingredients showing END GINGIVAL index scores of the FOLLOW-UP brushing exercises analysing the Löe & Silness (1963) Gingival Index and the modification of this Index by Löe (1967) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S7-4) Forest plot of the meta-subanalysis on ingredients showing END GINGIVAL index scores of the FOLLOW-UP brushing exercises analysing the Loe & Silness (1963) Gingival Index and the modification of this index by Löe (1967) for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS; no significant difference was observed between groups.

| Subgroup     | BS DF | Control DF | Mean Difference |
|--------------|-------|------------|-----------------|
|              | Mean | SD | Total | Mean | SD | Total | Weight | IV, Random, 95% CI |
| Tcs          |      |     |       |      |    |       |        |                   |
| Tiritana B 2015 | 1.47 | 0.30 | 23    | 0.95 | 0.26 | 45    | 33.1%  | 0.52 [0.38; 0.66]  |
| Ozaki 2006   | 0.82 | 0.35 | 20    | 0.73 | 0.27 | 22    | 31.0%  | 0.09 [-0.10; 0.28] |
| Total (random effects, 95% CI) | 43   | 67  | 64.1% | 0.31 [-0.11; 0.73] |

Heterogeneity: Tau² = 0.0850; Chi² = 12.45, df = 1 (P < 0.01); I² = 92%

| SnF          |      |     |       |      |    |       |        |                   |
| Beiswanger B 1997 | 0.74 | 0.24 | 74    | 0.64 | 0.16 | 267   | 35.9%  | 0.10 [0.04; 0.16]  |
| Total (random effects, 95% CI) | 74   | 267 | 35.9% | 0.10 [0.04; 0.16] |

Heterogeneity: not applicable

Total (random effects, 95% CI) | 117 | 334 | 100.0% | 0.24 [-0.04; 0.51] |

Heterogeneity: Tau² = 0.0540; Chi² = 28.51, df = 2 (P < 0.01); I² = 93% [83%; 97%]

Test for overall effect: Z = 1.68 (P = 0.09)
Appendix S7-5) Forest plot of the DIFFERENCE GINGIVAL index scores of the FOLLOW-UP brushing exercises analysing the Loe & Silness (1963) Gingival Index and the modification of this index by Loe (1967) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; a significant difference was observed between groups.
Appendix S7-6) Forest plot of the DIFFERENCE GINGIVAL index scores of the FOLLOW-UP brushing exercises analysing the Lëe & Silness (1963) Gingival Index and the modification of this index by Lëe (1967) for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S7-7A) Forest plot of the BASELINE GINGIVAL index scores of the FOLLOW-UP brushing exercises analysing the Lobene et al. (1986) modification of the Gingival Index (MGI) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; a significant difference was observed between groups.

| Study          | BS DF       | Control DF    | Mean Difference | Mean Difference |
|----------------|-------------|---------------|-----------------|-----------------|
|                | Mean       | SD | Total | Mean | SD | Total | Weight | IV, Random, 95% CI | IV, Random, 95% CI |
| Akwagyiram 2018| 2.53 ± 0.18 | 118|       | 2.54 ± 0.20 | 122|       | 61.2%  | −0.01 [−0.06; 0.04] |                       |
| Jose 2018      | 2.32 ± 0.34 | 118|       | 2.35 ± 0.33 | 117|       | 19.6%  | −0.03 [−0.12; 0.06] |                       |
| Lomax 2016     | 1.26 ± 0.28 | 66 |       | 1.26 ± 0.23 | 69 |       | 19.2%  | 0.00 [−0.09; 0.09]  |                       |

Total (random effects, 95% CI) 302 308 100.0% −0.01 [−0.05; 0.03] [−0.26; 0.23]  

Heterogeneity: $\tau^2 = 0$; $\chi^2 = 0.25$, df = 2 ($P = 0.88$); $I^2 = 0\%$ [0%; 16%]  
Test for overall effect: $Z = -0.62$ ($P = 0.54$)  
Favours sodium bicarbonate  
Favours negative control dentifrice  
Modified Gingival Index Baseline
Appendix S7-7B) Forest plot of the END GINGIVAL index scores of the FOLLOW-UP brushing exercises analysing the Lobene et al. (1986) modification of the Gingival Index (MGI) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; a significant difference was observed between groups.

| Study          | BS DF Mean | SD Total | Control DF Mean | SD Total | Weight IV, Random, 95% CI | Mean Difference IV, Random, 95% CI |
|----------------|------------|----------|-----------------|----------|---------------------------|-----------------------------------|
| Akwagiyiram 2018 | 2.23       | 0.23     | 118             | 2.43     | 0.23                      | 34.3% -0.20 [-0.26; -0.14]         |
| Lomax 2016      | 0.64       | 0.24     | 66              | 0.91     | 0.27                      | 33.5% -0.27 [-0.36; -0.18]         |
| Jose 2018       | 1.26       | 0.50     | 109             | 1.93     | 0.41                      | 32.2% -0.67 [-0.79; -0.55]         |
| Total (random effects, 95% CI) | 293 | 100.0% | 304 | 100.0% | -0.37 [-0.61; -0.14] [-3.41; 2.66] |

Prediction interval
Heterogeneity: Tau² = 0.0424; Chi² = 47.18, df = 2 (P < 0.01); I² = 96% [91%; 98%]
Test for overall effect: Z = -3.07 (P < 0.01)
Appendix S8: BI
Forest plots of BASELINE and END BLEEDING scores for the comparisons of dentifrices containing BS with negative controls without BS.

Appendix S8-A) (BI): The Ainamo & Bay (1975) Gingival Bleeding Index and the Abrams, Caton and Polson (1984) Bleeding on Probing Index.

Appendix S8-A1) Forest plot of the BASELINE BLEEDING scores of the FOLLOW-UP brushing exercises analysing The Ainamo & Bay (1975) Gingival Bleeding Index (See Appendix S4h-1) and the Abrams, Caton and Polson (1984) Bleeding on Probing Index (See Appendix S4h-2) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; no significant difference was observed between groups.
Appendix S8-A2) Forest plot of the END BLEEDING scores of the FOLLOW-UP brushing exercises analysing The Ainamo & Bay (1975) Gingival Bleeding Index (See Appendix S4h-1) and the Abrams, Caton and Polson (1984) Bleeding on Probing Index (65) (See Appendix S4h-2) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; a significant difference was observed between groups.
Appendix S8-B) (BI): The modification by Saxer et al. (1977) of the Ainamo & Bay (1975) Gingival Bleeding Index and the Gingival Bleeding Index by Saxton & van der Ouderaa (1989).

Appendix S8-B1) Forest plot of the BASELINE BLEEDING scores of the FOLLOW-UP brushing exercises analysing the modification by Saxer et al. (1977) of the Ainamo & Bay (1975) Gingival Bleeding Index (See Appendix S4i-1) and the Gingival Bleeding Index by Saxton & van der Ouderaa (1989) (See Appendix S4i-2) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; no significant difference was observed between groups.

| Study            | BS DF Mean | SD  | Total | Control DF Mean | SD  | Total | Weight | IV, Random, 95% CI | Mean Difference IV, Random, 95% CI |
|------------------|------------|-----|-------|-----------------|-----|-------|--------|-------------------|----------------------------------|
| Akwagyiram 2018  | 0.45       | 0.15| 118   | 0.47           | 0.16| 122   | 47.4%  | −0.02            | [−0.06; 0.02]                     |
| Jose 2018        | 0.38       | 0.17| 118   | 0.37           | 0.17| 117   | 39.0%  | 0.01             | [−0.03; 0.05]                     |
| Lomax 2016       | 1.00       | 0.27| 66    | 0.94           | 0.22| 69    | 10.9%  | 0.06             | [−0.02; 0.14]                     |
| Saxer 1994       | 1.06       | 0.28| 11    | 1.03           | 0.26| 11    | 1.5%   | 0.03             | [−0.20; 0.26]                     |
| Saxer 1995       | 1.19       | 0.46| 23    | 1.07           | 0.43| 23    | 1.2%   | 0.12             | [−0.14; 0.38]                     |
| Total (random effects, 95% CI) | 336 |     | 342 | 100.0% | 0.00 | [−0.02; 0.03] | [−0.05; 0.05] |

Prediction interval

Heterogeneity: Tau^2 < 0.0001; Chi^2 = 4.09, df = 4 (P = 0.39); I^2 = 2% [0%; 80%]

Test for overall effect: Z = 0.20 (P = 0.84)
Appendix S8-B2) Forest plot of the END BLEEDING scores of the FOLLOW-UP brushing exercises analysing the modification by Saxer et al. (1977) of the Ainamo & Bay (1975) Gingival Bleeding Index (See Appendix S4i-1) and the Gingival Bleeding Index by Saxton & van der Ouderaa (1989) (See Appendix S4i-2) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS; a significant difference was observed between groups.

| Study            | BS DF     | Control DF | Mean Difference | Mean Difference |
|------------------|-----------|------------|-----------------|-----------------|
|                  | BS DF     | Control DF | Mean Difference | Mean Difference |
| Akwagyiram 2018  | 0.19 0.13 | 0.37 0.16  | 118 31.6%       | –0.18 [-0.22; –0.14] |
| Jose 2018        | 0.17 0.12 | 0.31 0.16  | 109 31.5%       | –0.14 [-0.18; –0.10] |
| Lomax 2016       | 0.36 0.16 | 0.47 0.16  | 66 26.6%        | –0.11 [-0.16; –0.06] |
| Saxer 1995       | 0.73 0.32 | 0.64 0.27  | 23 7.4%         | 0.09 [-0.08; 0.26]  |
| Saxer 1994       | 0.83 0.31 | 1.08 0.38  | 11 3.0%         | –0.25 [-0.54; 0.04] |
| Total (random effects, 95% CI) | 327 | 338 | 100.0%       | –0.13 [-0.18; –0.08] |
|                  |           |            |                 | [–0.29; 0.03]   |

Prediction interval

Heterogeneity: $\text{Tau}^2 = 0.0019$; $\text{Chi}^2 = 13.00$, df = 4 ($P = 0.01$); $I^2 = 69\%$ [21\%; 88\%]

Test for overall effect: $Z = –4.93$ ($P < 0.01$)

Favours sodium bicarbonate
Favours control dentifrice

BLEEDING INDEX END SAXER & SAXTON
Appendix S9.

S9-A) 3. Contour-enhanced funnel plot with trimfill of the meta-analysis showing BASELINE PLAQUE scores of the SINGLE-BRUSHING exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS. Egger’s test shows a non-significant p-value (p=0.85).*.

*A test for funnel plot asymmetry (sometimes referred to as a test for small study effects) examines whether the association between estimated intervention effects and a measure of study size is greater than might be expected to occur by chance. These tests typically have low power, so even when a test does not provide evidence of asymmetry, bias cannot be excluded (71).

The Counter-Enhanced Funnel plot is an enhancement to the usual funnel plot proposed to allow considering the statistical significance of study estimates (70). The above funnel is centered at the model estimate. Color of line represent random effects estimate. The filled circles represent estimated treatment effect (mean difference) and its precision (standard error) for each individual study. Contour lines representing well established levels of statistical significance are added to the funnel plot to indicate regions where a test of treatment effect is significant. Contour lines indicating conventional milestones in levels of statistical significance (e.g., <0.01, <0.05, <0.1) are added (73). P-values correspond to a trial’s treatment effect (72). The unshaded (i.e., white) region in the middle
corresponds to non-significant results, the medium gray region to significant results at the 5% level and the dark gray region to the 1% level.

The open circles are missing studies filled in by a Trim-and-Filled Method. The basic idea of the trim-and-fill method is to add studies to the funnel plot until it becomes symmetric (73). The counter-funnel plot could be used naturally in conjunction with the trim-and-fill method because the latter informs the likely location of missing studies (70).

Visually, shows the funnel plot asymmetry. However, Egger’s test for asymmetry is not significant ($p=0.85$), indicating a publication bias mechanism is not a major cause for concern here (71).
Appendix S9-B)

S9-B) Contour-enhanced funnel plot with trimfill of the meta-analysis showing END PLaque scores of the SINGLE-BRUSHING exercises analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and POSITIVE control dentifrices without BS. Egger’s test shows a non-significant p-value (p = 0.08).

The Counter-Enhanced Funnel plot is an enhancement to the usual funnel plot proposed to allow considering the statistical significance of study estimates (70). The above funnel is centered at the model estimate. Color of line represent random effects estimate. The filled circles represent estimated treatment effect (mean difference) and its precision (standard error) for each individual study. Contour lines representing well established levels of statistical significance are added to the funnel plot to indicate regions where a test of treatment effect is significant. Contour lines indicating conventional milestones in levels of statistical significance (e.g., <0.01, <0.05, <0.1) are added. The unshaded (i.e., white) region in the middle corresponds to p-values greater than .10, the light gray-shaded region corresponds to p-values between .10 and .05, the gray-shaded region corresponds to p-values between .05 and .01, and the region outside of the funnel corresponds to p-values below.

The open circles are missing comparisons filled in by a Trim-and-Filled Method. The basic idea of the trim-and-fill method is to add studies to the funnel plot until it becomes symmetric (73). The counter-funnel plot could be used naturally in conjunction with the trim-and-fill method because the latter informs the likely location of missing studies (70).
Visually, shows the funnel plot minor asymmetry. However, Egger’s test for asymmetry is not significant ($p=0.08$), indicating a publication bias mechanism is not a major cause for concern here (71). Also, the contour-enhanced funnel plot shows that the “missing” studies are expected to lie in areas of high statistical significance (whereas the majority of available studies are nonsignificant), indicating that the observed asymmetry may not be due to publication bias based on statistical significance (70).
Appendix S10.

S10-A) A Contour-enhanced funnel plot with trimfill of the meta-analysis showing BASELINE gingival index scores of the FOLLOW-UP brushing exercises analysing the Løe & Silness (1963) Gingival Index and the modification of this index by Løe (1967) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS. Egger's test shows a non-significant $p$-value ($p = 0.31$).

Visually, shows the funnel plot minor asymmetry. However, Egger's test for asymmetry is not significant ($p=0.31$), indicating a publication bias mechanism is not a major cause for concern here (71). Also, the contour-enhanced funnel plot shows that the “missing” studies are expected to lie in areas of high statistical significance (whereas the majority of available studies are nonsignificant), indicating that the observed asymmetry may not be due to publication bias based on statistical significance (70).
S10-B) A Contour-enhanced funnel plot with trimfill of the meta-analysis showing END gingival index scores of the FOLLOW-UP brushing exercises analysing the Löe & Silness (1963) Gingival Index and the modification of this index by Löe (1967) for the comparison of dentifrices with BS and NEGATIVE control dentifrices without BS. Egger's test shows a significant $p$-value ($p = 0.009$).

Visually, the plot shows asymmetry. Missing studies probably lie within the white non-significant white region and in the significant colored and shaded region (71). Egger's test shows a significant $p$-value ($p = 0.009$) leading to rejection of the null hypothesis of symmetry in the funnel plot (73).
Appendix S11. Network meta-analysis.

S11-A) 1. Network meta-analysis graph for the Single-brushing Exercises end scores analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and control dentifrices without BS.

The treatments are equally spaced on the perimeter of the circle. Any two treatments are connected by a line when there is at least one study comparing the two treatments. The thickness of the line is proportional to the inverse standard error of the direct treatment comparison. Each comparison may have data from several studies that may be combined in a traditional meta-analysis. The shading indicates the studies with more than two arms. The set of nodes corresponds to the considered treatments and the edges display the treatment comparisons which are directly compared of all included trials.

Quantifying heterogeneity / inconsistency:
\( \tau^2 = 0; \ I^2 = 0\%

Tests of heterogeneity (within designs) and inconsistency (between designs):
|               | Q d.f. | p-value |
|---------------|--------|---------|
| Total         | 7.81   | 0.7306  |
| Within designs| 4.03   | 0.1336  |
| Between designs| 3.78  | 0.9253  |
S11-A) 2. Ranking for the Single-brushing TQ&H End plaque scores

| Treatment ranking | # Experiments | # Participants | P-score ranking * | MD      | 95%-CI     |
|-------------------|---------------|----------------|-------------------|---------|------------|
| BS unknown %      | 3             | 112            | 0.99              | 0.00    |            |
| BS 62%-67% **     | 5             | 271            | 0.83              | 0.11    | (-0.03; 0.27) |
| BS 45%-48%        | 3             | 190            | 0.66              | 0.19    | (0.04; 0.33)  |
| BS 20%-27%        | 4             | 236            | 0.52              | 0.23    | (0.10; 0.37)  |
| Tcs               | 6             | 338            | 0.22              | 0.34    | (0.22; 0.45)  |
| SnF               | 2             | 120            | 0.17              | 0.35    | (0.19; 0.50)  |
| NaF               | 3             | 222            | 0.11              | 0.37    | (0.22; 0.51)  |

Comparison: Other Treatments vs “BS unknown %” Prediction Interval: (0.05; 0.49)

* The frequentist P-scores allow ranking treatments on a continuous 0-1 scale. The P-score of treatment is defined as the mean extent of certainty that treatment is better than another treatment; scale: 0 (worst) to 1 (best). At least under normal assumption the order depends largely on the point estimates ((23).

** data from Mason et al. (2017) and Bosma et al. (2018) are pooled

Forest plot of the Ranking for the Single-brushing TQ&H End plaque scores sorted by ranking.

Comparison: other vs 'BS %'

| Treatment (Random Effects Model) | MD   | 95%-CI | P-Score |
|---------------------------------|------|--------|---------|
| BS ?%                           | 0.00 | 0.99   |         |
| BS 62%-67%                      | 0.12 | [-0.03; 0.27] | 0.83 |
| BS 45%-48%                      | 0.19 | [0.04; 0.33]  | 0.66 |
| BS 20%-27%                      | 0.23 | [0.10; 0.37]  | 0.52 |
| Tcs                             | 0.34 | [0.22; 0.45]  | 0.22 |
| SnF                             | 0.35 | [0.19; 0.50]  | 0.17 |
| NaF                             | 0.36 | [0.22; 0.51]  | 0.11 |

Prediction interval [ 0.05; 0.49]
3. Heterogeneity/ inconsistency between designs.

The $Q_{\text{total}}$ statistic (of the "whole network") can be decomposed into a $Q$ statistic for assessing the heterogeneity between studies with the same design ("within designs") and a $Q$ statistic for assessing the design inconsistency ("between designs"). Designs are defined by the subset of treatments compared in a study (73).

In the 11 studies included in the network meta-analysis, 9 different designs are used. Since there are only 2 designs for which we have more than two studies, the remaining design specific $Q$ statistics are equal to zero and have no degrees of freedom (31).

| design | $Q$ | df | pval |
|--------|-----|----|------|
| BS ?%:NaF | 0.0 | 0 | NA |
| BS ?%:Tcs | 2.3 | 1 | 0.13 |
| BS 20%-27%:Tcs | 0.0 | 0 | NA |
| BS 62%-67%:NaF | 1.7 | 1 | 0.19 |
| BS 62%-67%:SnF | 0.0 | 0 | NA |
| BS 20%-27%:BS 45%-48%:NaF:Tcs | 0.0 | 0 | NA |
| BS 20%-27%:BS 45%-48%:SnF:Tcs | 0.0 | 0 | NA |
| BS 20%-27%:BS 62%-67%:Tcs | 0.0 | 0 | NA |
| BS 45%-48%:BS 62%-67%:NaF | 0.0 | 0 | NA |
Inconsistency-detecting heat map function `netheat` from the `netmeta` package applied to the Single-brushing TQ&H End scores data set. “Hotspots” of inconsistency are absent in the `netheat` plot.

The net heat plot shows a marked reduction in inconsistency. For a valid analysis of the network, consistent information from different pathways is assumed. Consistency can be checked by contrasting effect estimates from direct comparisons with the evidence of the remaining network. If corresponding treatment effect estimates of various connections, differ between two treatments, there is inconsistency (29).

Treatment comparisons for which there is only one source of evidence are omitted (31). The area of the gray squares displays the contribution of the direct estimate in design (shown in the column) to the network estimate in design (shown in the row). A clustering procedure is applied to the heat matrix in order to find warm colored hot spots of inconsistency. Blue colors indicate low level of inconsistency, while red colors indicate “hot spots” of high inconsistency (29).
S11-B) 1. Network meta-analysis graph for the Follow-up End scores analysing the Turesky et al. (1970) modification of the Quigley and Hein (1962) Plaque Index for the comparison of dentifrices with BS and control dentifrices without BS.

Studies: 10; treatments: 6; pairwise comparisons: 14.

Treatments are represented by nodes and head-to-head studies between treatments are represented by edges.

Quantifying heterogeneity / inconsistency:
\[\tau^2 = 0.3311; I^2 = 97.9\%

Tests of heterogeneity (within designs) and inconsistency (between designs):

|                | Q  | d.f. | p-value  |
|----------------|----|------|----------|
| Total          | 327.61 | 7    | < 0.0001 |
| Within designs | 4.53  | 3    | 0.002    |
| Between designs| 323.07 | 4    | < 0.0001 |
### S11-B) 2. Ranking for the Follow-up TQ&H End plaque scores

| Treatment | # Experiments | # Participants | P-score ranking * | MD | 95%-CI        |
|-----------|---------------|----------------|------------------|----|---------------|
| Tcs       | 3             | 169            | 0.90             | -0.50 | (-1.15; 0.15) |
| MFP       | 1             | 20             | 0.64             | -0.16 | (-1.30; 0.98) |
| BS        | 10            | 560            | 0.57             | 0.00  |               |
| Non-BS    | 3             | 106            | 0.35             | 0.19  | (-0.46; 0.83) |
| NaF       | 4             | 255            | 0.32             | 0.28  | (-0.37; 0.93) |
| MFP+NaF   | 2             | 80             | 0.18             | 0.49  | (-0.28; 1.26) |

**Comparison: Other Treatments vs “BS”**  
Prediction Interval: [-0.71; 0.85]

* The frequentist P-scores allow ranking treatments on a continuous 0-1 scale. The P-score of treatment is defined as the mean extent of certainty that treatment is better than another treatment; scale: 0 (worst) to 1 (best). At least under normal assumption the order depends largely on the point estimates (23).

---

### Forest plot of for the Follow-up TQ&H End plaque scores sorted by ranking *.

#### Comparison: other vs 'BS'

| Treatment | MD  | 95%–CI       | P–Score |
|-----------|-----|--------------|---------|
| Tcs       | -0.50 | [-1.15; 0.15] | 0.90    |
| MFP       | -0.16 | [-1.30; 0.98] | 0.64    |
| BS        | 0.00  |              | 0.57    |
| Non-BS    | 0.19  | [-0.46; 0.83] | 0.39    |
| NaF       | 0.28  | [-0.37; 0.93] | 0.32    |
| MFP+NaF   | 0.49  | [-0.28; 1.26] | 0.18    |

Prediction interval: [-0.71; 0.85]

* The frequentist P-scores allow ranking treatments on a continuous 0-1 scale. The P-score of treatment is defined as the mean extent of certainty that treatment is better than another treatment; scale: 0 (worst) to 1 (best). At least under normal assumption the order depends largely on the point estimates (23).
S11-B) 3. Heterogeneity/ inconsistency between designs.

The $Q_{\text{total}}$ statistic (of the "whole network") can be decomposed into a $Q$ statistic for assessing the heterogeneity between studies with the same design ("within designs") and a $Q$ statistic for assessing the design inconsistency ("between designs"). Designs are defined by the subset of treatments compared in a study (73).

In the 10 studies included in the network meta-analysis, 8 different designs are used. Since there are only 3 designs for which we have more than two studies, the remaining design specific $Q$ statistics are equal to zero and have no degrees of freedom. Except for design BS: Tcs ($p=0.521$) and BS: NaF ($p=0.768$), for the other design BS: Non-BS ($p=0.045$) there is more heterogeneity between the contributing studies than we would expect by chance (31).

The total within-design heterogeneity further decomposed into the contribution from each design:

| design          | $Q$  | df | pval |
|-----------------|------|----|------|
| 1 BS: MFP       | 0.000| 0  | NA   |
| 2 BS: MFP+NaF   | 0.000| 0  | NA   |
| 3 BS: NaF       | 0.087| 1  | 0.768|
| 4 BS: Non-BS    | 4.035| 1  | 0.045|
| 5 BS: Tcs       | 0.412| 1  | 0.521|
| 6 BS: MFP+NaF:Tcs| 0.000| 0  | NA   |
| 8 BS: NaF:Non-BS| 0.000| 0  | NA   |
S11-B) 4. Net heat plot for detecting inconsistency between designs of the Follow-up TQ&H End plaque scores from a random effects model.

A heat map plot for the detection of inconsistency (73). Several “hotspots” of inconsistency can be seen between the effects in designs.
S11-C) 1. Network meta-analysis graph for the **Follow-up End** scores analysing the Løe & Silness (1963) and Løe (1967) **Gingival** Index for the comparison of dentifrices with BS and control dentifrices without BS.

Studies: 10; treatments: 7; pairwise comparisons: 18

Treatments are represented by nodes and head-to-head studies between treatments are represented by edges.

**Quantifying heterogeneity / inconsistency:**

\[ \tau^2 = 0.0110; \ I^2 = 67.8\% \]

**Tests of heterogeneity (within designs) and inconsistency (between designs):**

|                  | Q   | d.f. | p-value  |
|------------------|-----|------|----------|
| Total            | 24.82 | 8    | 0.0017   |
| Within designs   | 7.21  | 4    | 0.1251   |
| Between designs  | 17.61 | 4    | 0.0015   |
S11-C) 2. Ranking for the Follow-up End Scores according the Loe & Silness Gingival Index.

| Treatment   | # Experiments | # Participants | P-score ranking * | MD   | 95%-CI          |
|-------------|---------------|----------------|-------------------|------|-----------------|
| Tcs         | 2             | 67             | 1.00              | -0.36| (-0.54; -0.19)  |
| SnF         | 1             | 267            | 0.72              | 0.11 | (-0.28; 0.11)   |
| Non-BS      | 5             | 143            | 0.60              | 0.19 | (-0.15; 0.10)   |
| BS          | 10            | 409            | 0.53              | 0.00 |                 |
| NaF         | 3             | 176            | 0.28              | 0.07 | (-0.08; 0.22)   |
| MFP         | 1             | 20             | 0.22              | 0.35 | (-0.12; 0.36)   |
| MFP+Naf     | 2             | 80             | 0.16              | 0.37 | (-0.03; 0.28)   |

Comparison: Other Treatments vs “BS”

Prediction Interval: (-0.50; 0.44)

Forest plot of for the Follow-up End scores according the Loe & Silness Gingival Index.

Comparison: other vs 'BS'

(Random Effects Model) MD 95%-CI P-Score

| Treatment   | MD   | 95%-CI | P-Score |
|-------------|------|--------|---------|
| Tcs         | -0.36| [-0.54; -0.19] | 1.00    |
| SnF         | -0.08| [-0.28; 0.11]  | 0.72    |
| Non-BS      | -0.02| [-0.15; 0.10]  | 0.60    |
| BS          | 0.00 |        | 0.53    |
| NaF         | 0.07 | [-0.08; 0.22]  | 0.28    |
| MFP         | 0.12 | [-0.12; 0.36]  | 0.22    |
| MFP+Naf     | 0.13 | [-0.03; 0.28]  | 0.16    |

Prediction interval [-0.50; 0.44]

* The frequentist P-scores allow ranking treatments on a continuous 0-1 scale. The P-score of treatment is defined as the mean extent of certainty that treatment is better than another treatment; scale: 0 (worst) to 1 (best). At least under normal assumption the order depends largely on the point estimates (23).
S11-C) 3. Heterogeneity/ inconsistency between designs.
In the 10 studies included in the network meta-analysis, 9 different designs are used. Since there are only two designs for which we have more than two studies, the remaining design specific Q statistics are equal to zero and have no degrees of freedom. Except for design BS: Non-BS (p=0.954), for the other design BS: NaF: Non-BS (p=0.028) there is more heterogeneity between the contributing studies than we would expect by chance (31).

The total within-design heterogeneity further decomposed into the contribution from each design:

```
> round(decomp.design(net81)$Q.decomp, 3)
  Q  df  pval
Total  24.819  8  0.002
Within designs  7.212  4  0.125
Between designs 17.607  4  0.001
> print(decomp.design(net81)$Q.het.design, digits=2)
       design       Q  df  pval
      1       BS:MFP 0.000  0   NA
      2 BG:MF + NaF 0.000  0   NA
      3       BS:Non-BS 0.094  2 0.954
      4       BS:Tcs 0.000  0   NA
      5 BS:MF + NaF:Tcs 0.000  0   NA
      6 BS:NaF:Non-BS 7.117  2 0.028
      7       BS:NaF:SnF 0.000  0   NA
```
S11-C) 4. Net heat plot for detecting inconsistency between designs of the Follow-up Loe & Silness End gingival index scores from a random effects model.

A heat map plot for the detection of inconsistency. Warm orange colors indicate inconsistency of evidence.