CRYING, UNSETTLED, DISTRESSED INFANTS: EFFECTIVENESS STUDY

Statistical report

Chief Investigator: Dawn Carnes

Senior Trial Statistician: Paul Vaucher

Data Manager and Statistician: Kevin Brownhill

Version and Date V0.4 23.09.22

Brief overview of data collection

Around 84 osteopaths were trained to participate in the trial, with 22 osteopaths (18 female and 4 males were engaged as trial practitioners) going on to enrol infants in the trial. The range of practitioner years' of experience was 3–31 years (at start of trial training), with a mean of 13.9 years. All had post-graduate training in paediatric care (courses varied from CPD to diplomas). There were 56 infants recruited in the UK, six in Australia and four in Switzerland. Eight osteopaths recruited one infant each, five recruited two, three recruited three, two recruited five, one recruited six, another recruited seven, and two recruited eight infants.

Around one in four parents who were approached to participate in the trial were interested in being part of the study with 66 enrolled and randomised.

The study recruited 66 participants who reported 975 days of crying time from September 2019 to July 2022. Thirty-two infants were recruited in private practices in the UK, six in Australia and four in Switzerland. Sixteen were enrolled by the European School of Osteopathy and eight by the University College of Osteopathy. All participants received the treatment they were allocated to. Eleven crying diaries were not returned, two were summarised as daily crying time by parents and three reported crying at specific hours without providing any details on the duration within that hour. Baseline questionnaire was incomplete for two participants and ten did not complete the follow-up questionnaire. Protocol deviations could have impacted results for three participants: two received external osteopathic care on the last day of follow-up, and a second was recruited even if they were 12 weeks of age. Missing data was evenly balanced between both groups (p=0.333). Blinding was successful with parents reporting beliefs on allocation in similar proportion for both groups (p=0.730). Details on follow-up is provided in **Figure 1.**

Description of groups

Age at entry was similar between both groups with a mean age of 40 days in the test group and of 43 days in the control group (p=0.441). Most participants were unsure about expectations on treatment results in both groups (53.1% vs. 50.0%, p=0.527). Baseline crying time was also similar in both groups (252 min vs. 235 min, p=0.523). Parenting confidence scores [scores of 0–45] revealed lack of confidence in both groups with an average score of 32.2 (SD 3.3) in the test group and 32.5 (SD 3.3) in the control group (p=0.653). Randomisation achieved group balance as no significance difference was observed between groups (**Table 1**).

Effects on daily crying time

A mixed-effects linear regression model was used to predict daily crying times with data entered at a daily level. Baseline crying time, age at entry, expectation for outcome and days within trial were used as fixed effects, and babies were treated as random effects (random intercepts model). Suspected cofactors (fixed effects or predictors) were linearly associated to daily crying time except for expectations. Plotting the fitted predicted crying times against the true observed crying time (Figure A) and residuals over observed crying time (Figure B) showed that the random effect linear model had difficulties fitting appropriately for longer crying times (r=0.675, p<0.001). Residuals were independent of duration within trial (Figure C), revealed similar heteroscedasticity within each group (Figure D), and were normally distributed (Figure E).

Changes over time during the follow-up period were assessed and explored to make sure linear regression could model differences between groups appropriately. Reduction in crying time could be

considered as linear and was similar between groups. Interaction terms between days within the trial and group allocation showed no differences in trend for crying times between groups.

Primary outcome – Daily crying time

There was an overall reduction of crying time in both groups with an important change between the day before the treatment and on the day the treatment took place in both groups (TTR= - 62 minutes, GTR = -72 minutes). Following this first drop, crying time seemed to decrease by approximately 4 minutes per day over the next two weeks in both groups. Crying was not randomly distributed across the hours of the day with more frequent observed cries between 4 PM and 10 PM. Cries became less frequent in both groups over time with a slight cluster of increased crying time (< 6 minutes) for the test group concentrated at the end of the trial in the 4-10PM time slot when compared to the control group (**Figure 2**). The average daily crying time was however very similar between groups throughout the follow-up period (**Figure 3**).

```
Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log restricted-likelihood = -3763.1735
Iteration 1:   log restricted-likelihood = -3763.1735

Computing standard errors:

Mixed-effects REML regression                   Number of obs     =         687
Group variable: id                              Number of groups  =          55

                                                Obs per group:
                                                              min =           7
                                                              avg =        12.5
                                                              max =          13

                                                Wald chi2(5)      =      114.81
Log restricted-likelihood = -3763.1735          Prob > chi2       =      0.0000

------------------------------------------------------------------------------
 crying_time |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |   2.191086   11.54348     0.19   0.849    -20.43371    24.81589
    baseline |   .3691646   .0587057     6.29   0.000     .2541037    .4842256
expectation2 |  -4.343393   6.100005    -0.71   0.476    -16.29918    7.612398
     age_cat |   -17.0678   5.892208    -2.90   0.004    -28.61632   -5.519285
         day |  -4.351613   .5612858    -7.75   0.000    -5.451713   -3.251513
       _cons |   117.7239   21.88063     5.38   0.000      74.8387    160.6092
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                  var(_cons) |   1546.397   359.4884      980.4906    2438.925
-----------------------------+------------------------------------------------
               var(Residual) |   2961.684   166.8095      2652.143    3307.353
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 159.86        Prob >= chibar2 = 0.0000
```

The group who received the test treatment (TTR) cried in average very slightly more than the control group (GTR) with an average non-significant increase in daily crying time of 2 minutes (CI95% -20 to 25, p=0.849). These results were consistent throughout the 13 days of follow-up (**Table 2**).

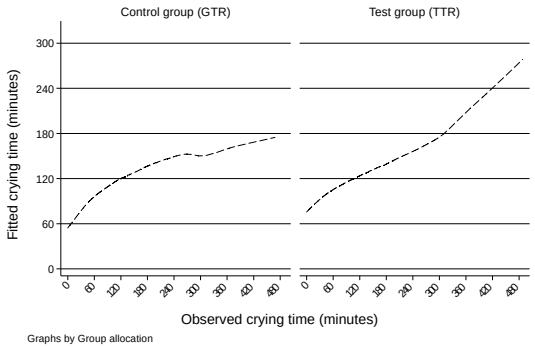

***Figure A** : Predicted fitted value over observed crying time (with lowess weighted regression line)*

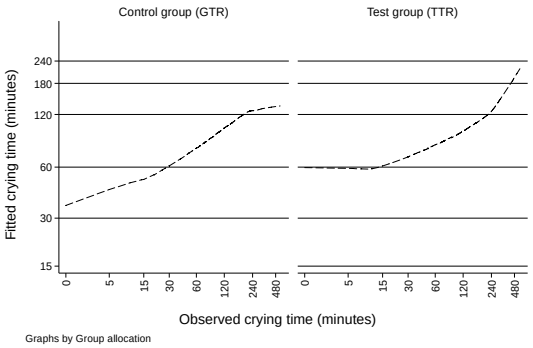

***Figure Ab** : Logged fitted value over observed crying time (with lowess weighted regression line)*

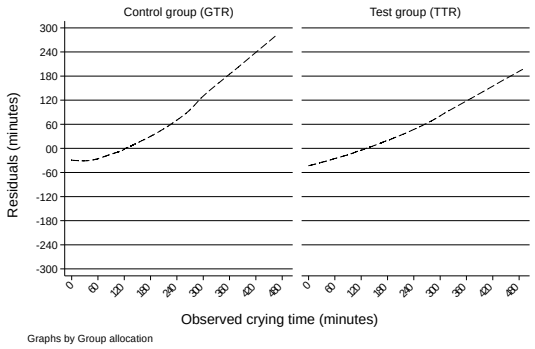

***Figure B** : Residuals over observed crying time (with lowess weighted regression line)*

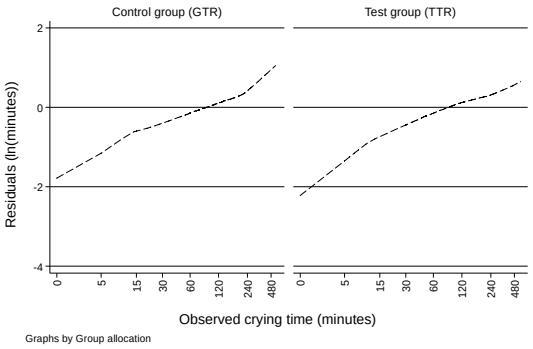

***Figure Bb** : Residuals over logged observed crying time (with lowess weighted regression line)*

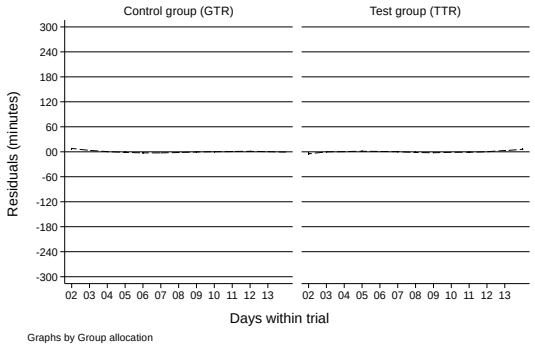

***Figure C** : Residuals over trial days (with lowess weighted regression line)*

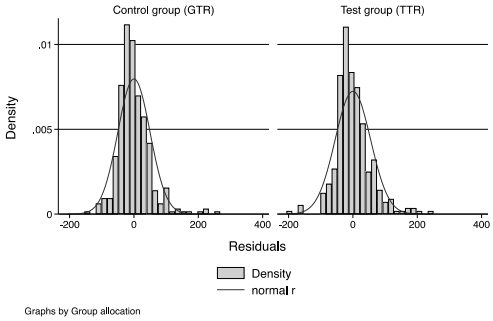

***Figure E**: Distribution of residual*

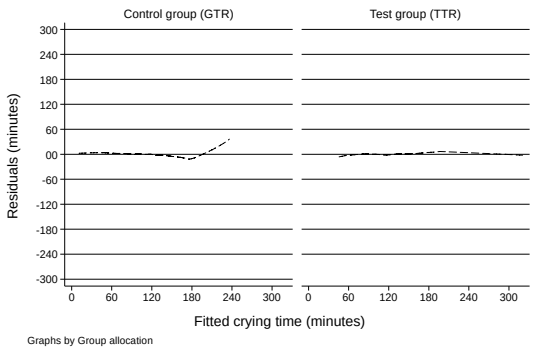

***Figure D**: Residuals over predicted fitted values (with lowess weighted regression line)*

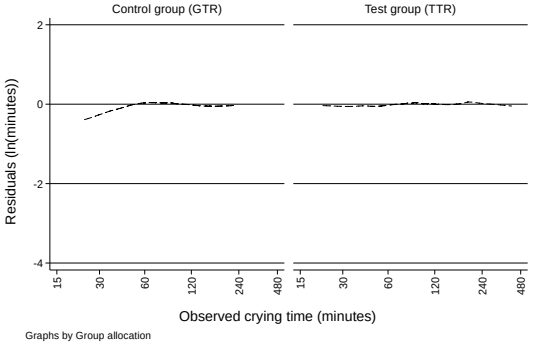

***Figure Db**: Residuals over logged predicted fitted values (with lowess weighted regression line)*

Secondary outcomes

Results from regression analysis for all other outcomes were consistent with no significant treatment effects over control groups for parenting confidence scores, perceived changes in symptoms, satisfaction with received care or patient experience of care (**Table 3**).

Unexpected reactions were equally distributed in both groups but excess was always observed in disfavour of the test group (**Table 3**). Two serious adverse events, one in each group, were observed and were deemed unrelated to patient care and management.

Patient management

The number of sessions were not equally distributed in both groups (exact Fisher p=0.023) with a higher number of visits in the treatment group (TTR 2.5 vs GTR 2.2; p=0.217). Advice was frequent with approximately 90% of parents receiving reassurance in both groups (**Table 4**). Medication and remedies were similar in both groups. Proportions of infants having visited their paediatrician for additional care were more frequent in the test group (**Table 4**).

Sensitivity analysis

To account for missing data, worst and best-case scenarios were run in which the 25th and 75th percentiles values of average crying time were imputed respectively for missing values in each group. In the worse-case scenario, the test group would have an added 22 minutes of daily crying time over the control group (CI95% -6 to 49, p=0.122); in the best-case scenario, a reduction of 7 minutes (-35 to 21; p=0.618).

Log transformed data (ln(crying time +e)) showed that between group difference was also negligeable (0.012; CI95% -0.257 to 0.281 min; p=0.930). Back transformation was not done given it is difficult to interpret correctly. Per protocol analysis, in which only eligible patients who received the allocated treatment were maintained in the analysis (n=51), revealed similar results than the intention-to-treat approach (4.4 min; CI95% -20 to 28; p=0.718). Results were also similar when excluding the three participants whose hourly crying duration was imputed (3.6 min; CI95% -20 to 27; p=0.766).

To account for eventual performance bias (i.e. systematic differences in the way groups received care other than the tested intervention), an analysis was run adjusting for number of treatment sessions, use of hypoallergic milk supplements, and having visited the paediatrician. Mixed-effect logistic regression modelling showed these factors not to alter results (8 min; CI95% -20 to 36; p=0.597).

Finally, we verified that between institution differences did not affect results. Using multilevel mixed-effects linear modelling, we observed that most variance was explained at the participant level and that results differed little when accounting for eventual institutional level differences (2.7 min; CI95% -18 to 24; p=0.804).

9. Results analysis

Provisional tables

Table 1 Baseline characteristics for each group. GTR=Generic Tension Release, TTR=Targeted Tension Release, SD=standard deviation

Characteristics	Test group (TTR) N=32	Control group (GTR) N= 34	Group imbalance (p-values)
Sex; n (%)			p=0.900
Female	17 (53.1%)	19 (55.9%)	
Male	14 (43.8%)	15 (44.1%)	
Unknown	1 (3.1%)	0 (0%)	
Age of infant; n (%)			p=0.720
1–14 days (1–2 weeks)	0 (0%)	1 (2.9%)	
15–28 days (3–4 weeks)	5 (15.6%)	5 (14.7%)	
29–42 days (5–6 weeks)	15 (46.9%)	11 (32.3%)	
43–56 days (7–8 weeks)	8 (25.0%)	12 (35.3%)	
57–70 days (9–10 weeks)	4 (12.5%)	4 (11.8%)	
71–84 days (11–12 weeks)	0 (0%)	1 (2.9%)	
Infant weight in kg; mean (SD)			
At birth	3.4 (0.5)	3.4 (0.4)	p=0.697
At baseline	4.3 (0.7) †	4.5 (0.8) ‡	p=0.272
Age of mother; n (%)			p=0.959
21-25 years	1 (3.2%)	2 (5.9%)	
26-30 years	7 (21.9%)	9 (26.5%)	
31-35 years	14 (43.7%)	15 (44.1%)	
36-40 years	7 (21.9%)	8 (23.5%)	
41-45 years	1 (3.1%)	0 (0%)	
46-50 years	1 (3.1%)	0 (0%)	
Missing	1 (3.1%)	0 (0%)	
Siblings; n (%)			p=0.445
None	14 (43.8%)	21 (61.8%)	
1	14 (43.8%)	9 (26.5%)	
2	2 (6.2%)	2 (5.9%)	
3 or more	1 (3.1%)	2 (5.9%)	
Missing	1 (3.1%)	0 (0%)	
Type of parenting; n (%)			p=0.485
Co-parenting	31 (96.9%)	34 (100%)	
Missing	1 (3.1%)	0 (0%)	
Expected response to osteopathic care; n (%)			p=0.827
Very well	2 (6.2%)	1 (2.9%)	
Well	7 (21.9%)	10 (29.4%)	
Unsure	17 (53.1%)	17 (50.0%)	
Not very well	2 (6.2%)	4 (11.8%)	
Not well	3 (9.4%)	1 (2.9%)	
Missing	1 (3.1%)	1 (2.9%)	
Excessive crying, distress or unsettlement; mean (SD)			
Reported for day prior to treatment (minutes)	252 (119)	235 (94)	p=0.523
Parenting Confidence Score; n (%)			p=0.819
Non-clinical range (40 – 45)	0 (0%)	0 (0%)	
Mild clinical range (36 – 39)	4 (12.5%)	6 (17.6%)	
Moderate clinical range (32 – 35)	15 (46.9%)	17 (50%)	
Severe clinical range (31 or less)	12 (37.5%)	11 (32.4%)	
Missing	1 (3.1%)	0 (0%)	

* Significant group imbalance at baseline (Student T-Test | Fischer exact test); † n=29, ‡ n=33

Table 2 Effects of treatment on crying time in minutes

Outcomes	Test group (TTR) N=26	Control group (GTR) N= 29	Adjusted* between group difference	Unadjusted between group difference
	Mean (Std.Err)	**Mean (Std.Err)**	**Mean (CI 95%; P-value)**	**Mean (CI 95%; P-value)**
Mean daily crying time (minutes)				
From day 1 to day 13	**124 (14)**	**115 (9)**	**2.2 (-20 to 25; p=0.849)**	**8.8 (-23 to 40; p=0.583)**
From day 1 to day 6	139 (13)	129 (11)	3.0 (-21 to 27; p=0.810)	10 (-24 to 44; p=0.557)
From day 7 to day 13	110 (15)	104 (9)	1.3 (-26 to 29; p=0.926)	7.5 (-26 to 41; p=0.663)
Daily crying times (minutes)				
Day 1 after treatment	147 (13)†	148 (17)	-10 (-47 to 27; p=0.595)	-0.7 (-43 to 42; p=0.973)
Day 2 after treatment	137 (18)‡	141 (15)	-10 (-48 to 28; p=0.604)	-4.1 (-48 to 40; p=0.856)
Day 3 after treatment	164 (24)‡	134 (12)	19 (-15 to 53; p=0.274)	30 (-20 to 840; p=0.237)
Day 4 after treatment	145 (17)	117 (14)§	16 (-20 to 53; p=0.382)	27 (-15 to 70; p=0.203)
Day 5 after treatment	116 (13)	128 (14)¶	-18 (-51 to 14; p=0.271)	-11 (-49 to 27; p=0.554)
Day 6 after treatment	132 (18)†	111 (13)	16 (-18 to 51; p=0.343)	21 (-20 to 63; p=0.316)
Day 7 after treatment	140 (18) †	116 (11)	17 (-15 to 49; p=0.305)	24 (-15 to 63; p=0.231)
Day 8 after treatment	111 (18)	105 (10)§	2 (-33 to 36; p=0.927)	5.7 (-33 to 45; p=0.776)
Day 9 after treatment	110 (17)‡	115 (14)§	-13 (-49 to 23; p=0.483)	-5 (-47 to 36; p=0.794)
Day 10 after treatment	108 (16)	104 (9)	-3 (-33 to 26; p=0.826)	4.3 (-30 to 38; p=0.800)
Day 11 after treatment	111 (18)‡	97 (15)¶	4.6 (-34 to 43; p=0.814)	14 (-31 to 59; p=0.540)
Day 12 after treatment	114 (13)†	93 (13)¶	11 (-23 to 46; p=0.524)	20 (-60 to 31; p=0.531)
Day 13 after treatment	97 (18)††	87 (13)§	10 (-28 to 49; p=0.598)	10 (-31 to 52; p=0.634)

Positive values are in favour of control group. * Adjusted for baseline crying time, infant age, prior expectations for osteopathic care, and days within trial; † n=25, ‡ n=24, § n=28, ¶ n=27, †† n=21

Table 3 Effects of treatment on other outcomes

Outcomes	Test group (TTR)	Control group (GTR)	Between group difference*
	N=32	N= 34	(CI 95%; P-value)
Parenting Confidence Score [0–45]; mean (SD)	35.9 (2.8) †	36.2 (2.6) ‡	-0.4 (-1.1 to 1.8; p=0.627)
Perceived changes in symptoms; n(%)			p=0.896
Completely recovered	0 (0%)	0 (0%)	
Much improved	16 (50%)	14 (41.2%)	
Slightly improved	5 (15.6%)	9 (26.5%)	
No change	6 (18.7%)	5 (14.7%)	
Slightly worse	1 (3.1%)	0 (0%)	
Much worse	1 (3.1%)	0 (0%)	
Vastly worse	0 (0%)	0 (0%)	
Did not respond	3 (9.4%)	7 (17.6%)	
Satisfaction with received care; n (%)			p=0.906
Very satisfied	24 (75%)	24 (70.6%)	
Fairly satisfied	2 (6.2%)	4 (11.8%)	
Neither satisfied nor dissatisfied	3 (9.4%)	1 (2.9%)	
Fairly dissatisfied	0 (0%)	0 (0%)	
Very dissatisfied	0 (0%)	0 (0%)	
Did not respond	3 (9.4%)	5 (14.7%)	
Parent's experience of care; n(%)			p=0.863
Very good	22 (68.8%)	22 (61.8%)	
Fairly good	4 (12.5%)	6 (17.6%)	
Neither good nor bad	2 (6.2%)	2 (5.9%)	
Fairly poor	1 (3.1%)	0 (0%)	
Very poor	0 (0%)	0 (0%)	
Did not respond	3 (9.4%)	5 (14.7%)	
Unexpected reactions; n(%)			
More distress	2 (6.2%)	1 (2.9%)	p=0.608
Crying more	4 (12.5%)	3 (8.8%)	p=0.705
More unsettled	7 (21.9%)	2 (5.9%)	p=0.079
Vomiting more	0 (0%)	1 (2.9%)	p=1.000
Increased feeding difficulties	1 (3.1%)	1 (2.9%)	p=1.000
Increased difficulties sleeping	1 (3.1%)	1 (2.9%)	p=1.000
Other	1 (3.1%)	2 (5.9%)	p=1.000

* P-values measured using likelihood ratio test in linear regression, Mann-Whitney U test for ordinal outcomes, Fisher's exact test for categorical; † n=29, ‡ n=28

Table 4 Number of sessions, advice given, additional treatment and effectiveness of blinding

	Test group (TTR) N=32	**Control group (GTR)** N= 34	**Between group difference** (CI 95%; P-value)
Number of sessions; mean (SD)	2.6 (1.0)	2.1 (0.9) †	0.4 (-0.0 to 0.9; p=0.062)
Advices; n(%)			
Managing the baby's sleep pattern	9 (28.1%)	9 (26.5%)	p=1.000
Managing the baby's feeding pattern	14 (43.7%)	11 (32.3%)	p=0.447
Mother's diet	15 (46.9%)	10 (29.4%)	p=0.205
Handling the baby	17 (53.1%)	17 (50%)	p=0.811
Managing the baby's environment	8 (25.0%)	10 (29.4%)	p=0.785
Parenting behaviour	9 (28.1%)	12 (35.3%)	p=0.603
Reassure parent/guardian	29 (90.6%)	30 (88.2%)	p=1.000
Other	14 (43.7%)	11 (32.3%)	p=0.447
Medication and remedies; n(%)			
Prescribed medication	5 (15.6%)	3 (8.8%)	p=0.469
Anti-gas drops	5 (15.6%)	11 (32.3%)	p=0.154
Herbal supplements	0 (0%)	0 (0%)	p=1.000
Probiotics	6 (18.7%)	6 (17.6%)	p=1.000
Non-cow or anti-allergenic formula milk	6 (18.7%)	1 (2.9%)	p=0.051
Omeprazol	2 (6.2%)	0 (0%)	p=0.231
Additional care; n(%)			
General practitioner	8 (25.0%)	6 (17.6%)	p=0.554
Accident and Emergency	0 (0%)	0 (0%)	p=1.000
Unplanned hospital admission	0 (0%)	1 (2.9%)	p=1.000
Another osteopath	1 (3.1%)	1 (2.9%)	p=1.000
Paediatrician	5 (15.6%)	0 (0%)	p=0.023
Pharmacist	0 (0%)	1 (2.9%)	p=1.000
Lactation consultation or midwife	1 (3.1%)	1 (2.9%)	p=1.000
Blinding; n(%)			p=0.730
Thinks the baby was in test group (TTR)	7 (21.9%)	4 (11.8%)	
Thinks the baby was in the control group (GTR)	5 (15.6%)	6 (17.6%)	
Doesn't know / unsure	16 (53.1%)	19 (55.9%)	
Did not respond	3 (9.4%)	5 (14.7%)	

* P-values measured using likelihood ratio test in linear regression, or Fischer exact test for nominal variables; † n=32

Figures

Figure 1 Flow chart of participant inclusion and follow-up

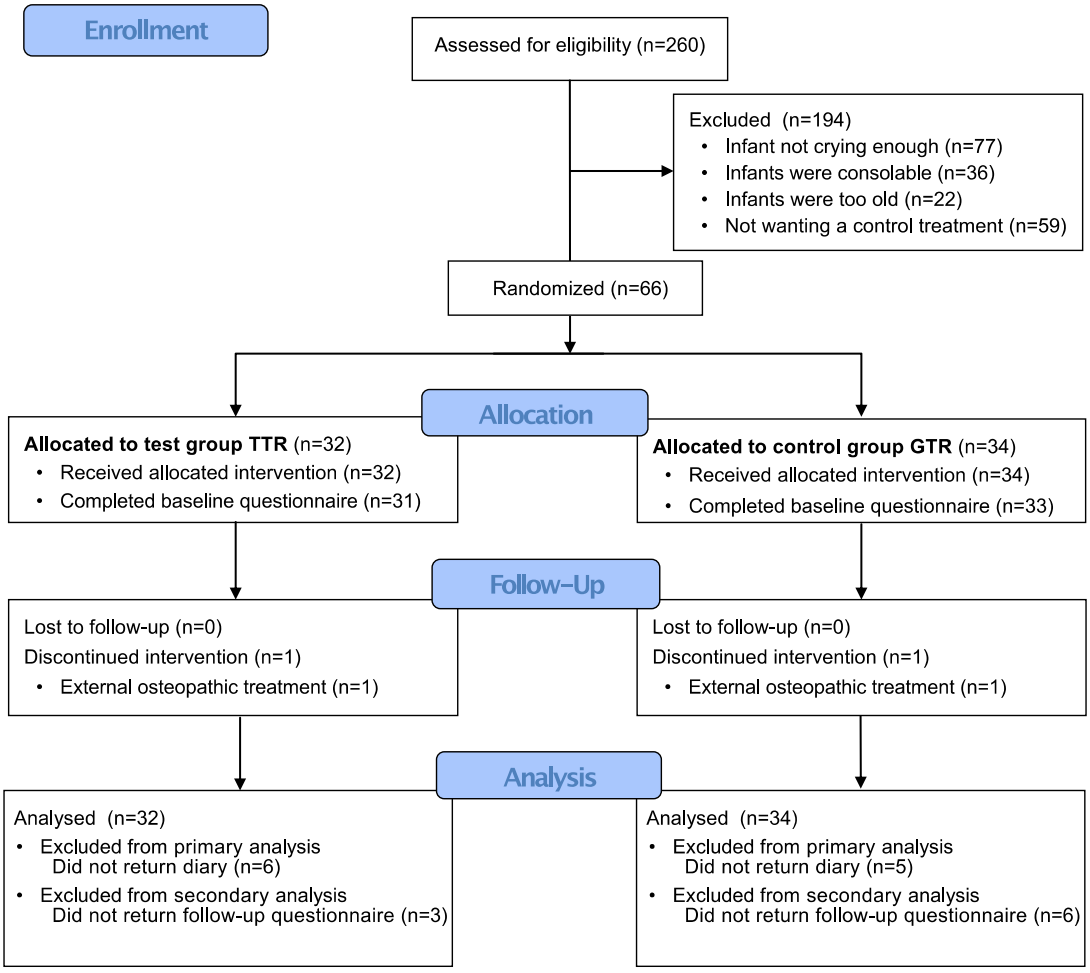

Figure 2 Crying duration spatial distribution over time

Average crying times were measured for each time slots for each group. Values were then smoothed by averaging the value with the adjacent slots. The top figure represents crying times in the test group (TTR). In the middle figure, values for the control group (GTR). The lower figure provides a visual representation of the effects of the test treatment over the control treatment.

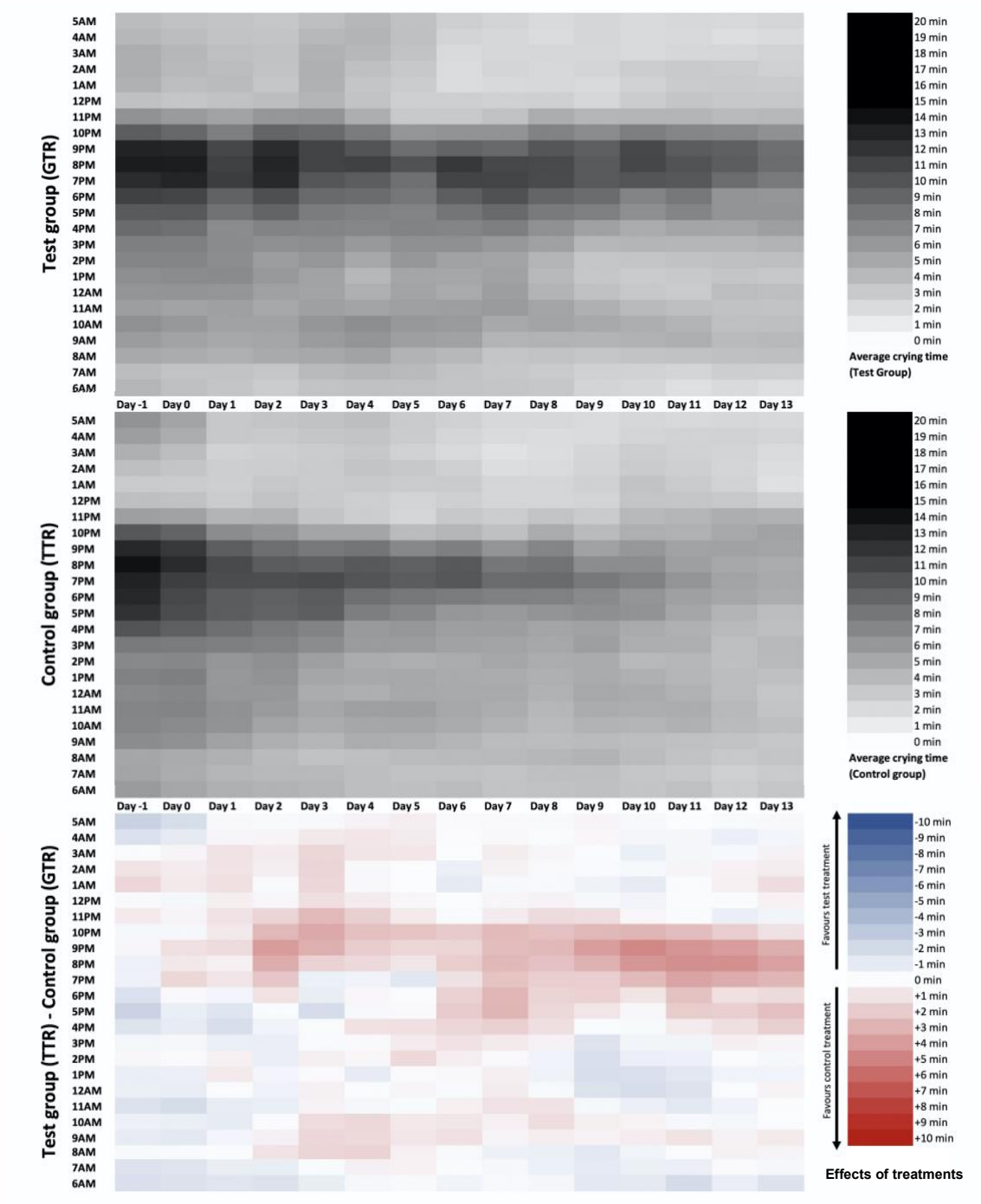

Figure 3 Predicted daily crying time during follow-up for each treatment group

Between group differences were measured using random effect linear regression adjusting for lack of independence between measures from the same participant and an interaction term between group allocation and day of follow-up.

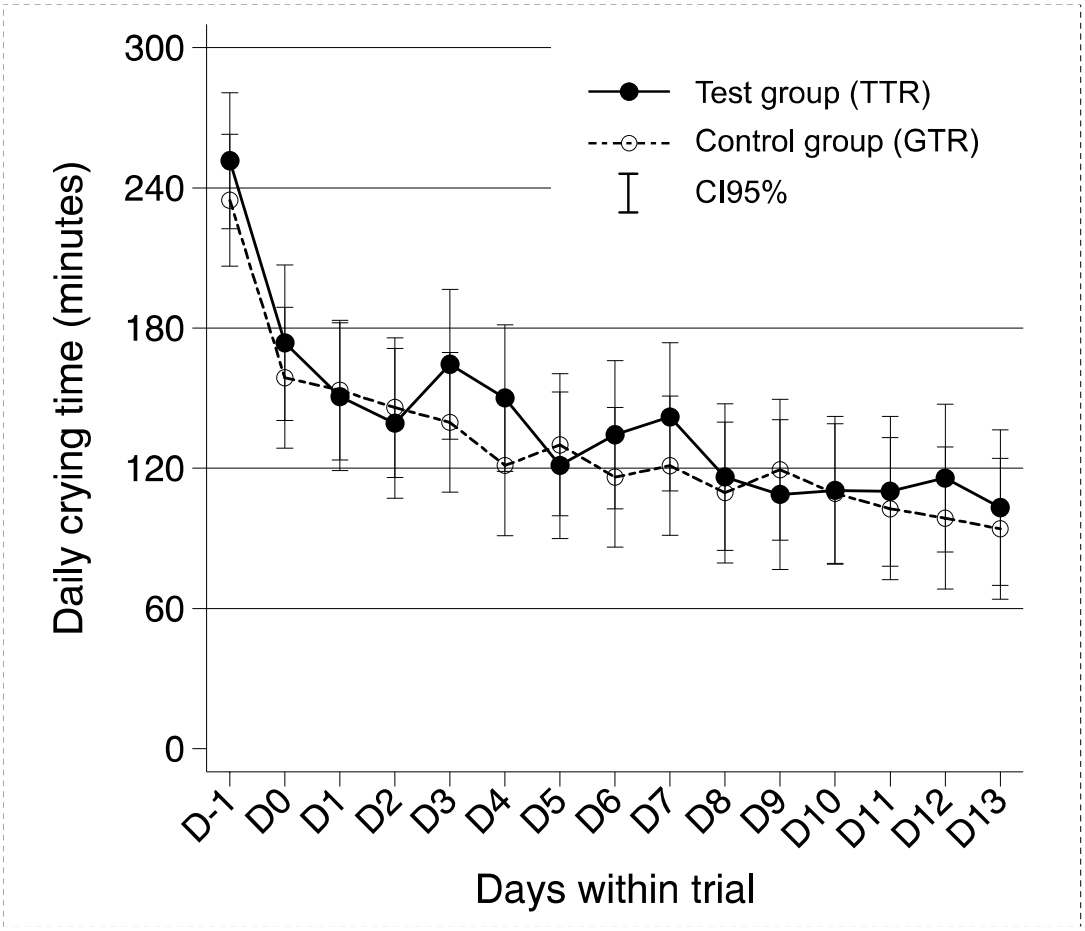

Figure 4 Types of osteopathic manipulative treatment delivered (n=32)