APPENDIX: SENSITIVITY ANALYSES

Several sensitivity analyses have been performed, random-effects pooled estimates calculated according to Williamson et al[9]:

1. Study quality: excluding studies with "Unclear" or "High" risk of bias in the domain *Flow and Timing* since too long time between measurements was assumed to be a source of bias. Twenty-eight comparisons with this setting resulted in random-effects pooled mean temperature difference -0.16°C (95% LoA -1.18 to 0.85°C).

2. Some studies included replicated data collected in pairs. Provided appropriate methods of analysis have been applied,[11] such studies can be included in the meta-analysis, along with the studies with one measurement per test. For example, studies that used replicated data in pairs using differences for each pair of measurements and accounted for within-person correlation of observations, were eligible for the meta-analysis. Excluding studies with no information on how they dealt with replicated nature of the data,[2 24 33 35 50] 35 comparisons, random-effects pooled mean temperature difference -0.18°C (95% LoA -1.14 to 0.78°C).

3. Several studies had repeated measurements by each of the two methods on the same subjects and calculated the mean for each method on each subject and used these pairs of means to evaluate the agreement between the two methods.[31 47 51] This does not affect the estimates of bias, but makes the estimates of the SD of the differences too small, since some of the effect of repeated measurement error is removed. The estimates of SD of difference can be corrected as proposed by Bland and Altman,[11 12] however, it is not clear whether the correction was applied in some studies. Excluding these studies, 40 comparisons, random-effects pooled mean temperature difference -0.19°C (95% LoA -1.15 to 0.76°C).

4. Including only studies that included Bland-Altman plots and did not show evidence for deviation from random scatter: 22 comparisons, random-effects pooled mean temperature difference -0.09°C (95% LoA -1.01 to 0.82°C).

5. Excluding one study as an outlier:[50] 39 comparisons, random-effects pooled mean temperature difference -0.10°C (95% LoA -1.04 to 0.83°C).

Sensitivity analysis 2 and 3 combined, 32 comparisons: random-effects pooled mean temperature difference -0.17°C (95% LoA -1.12 to 0.77°C).

Sensitivity analysis 1 to 5, 9 comparisons: random-effects pooled mean temperature difference 0.02°C (95% LoA -0.87 to 0.88°C).