**Supplemental Material**

## Additional information for methods

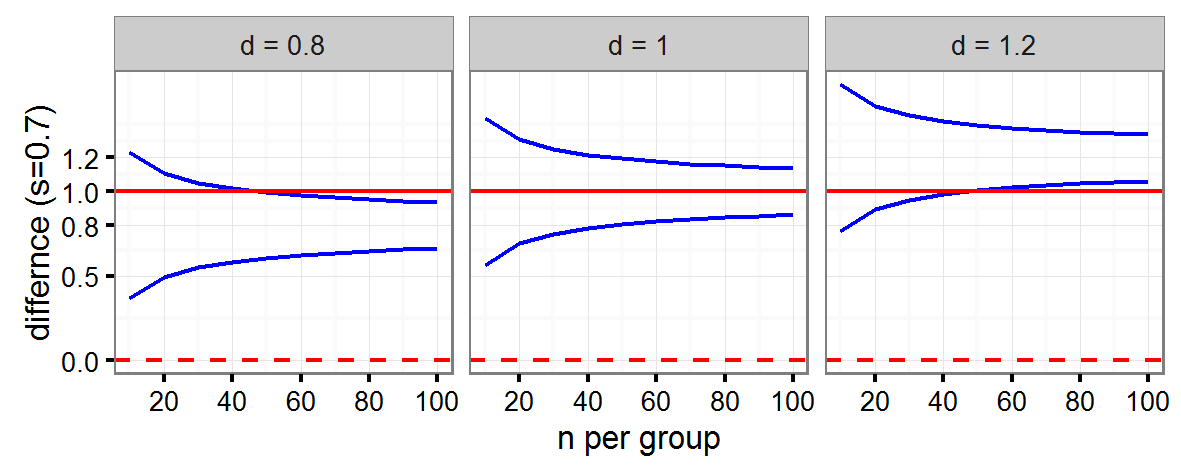
## Determination of Sample Size

Both the primary and secondary endpoints was analyzed using Bland Altman analysis as outlined below, i.e. by calculating measures of bias and precision. The width of the 95% confidence intervals (95% CI) of these measures depended on the standard deviation of the mean difference (*s*) and the sample size *n*. From previous similar studies we expected a maximal standard deviation of *s* = ±0.7 kPa [1, 2]. Considering the nature of the primary endpoint (evaluation of bias and precision of measurements), no hypothesis testing could be performed, and hence, no formal power calculation based on hypothesis testing could be provided [3]. However, we aimed to determine bias and precision of measurements with clinically satisfying precision (95% CI) as discussed in the following. As detailed below, we aimed to include 60 patients for the transcutaneous-arterial comparison (included in this group if arterial line was in place) and 60 patients for the transcutaneous-capillary comparison (included into this group if no arterial line was in place). Assuming normal distribution of the calculated measurement differences, a samples size of at least 60 patients per group (in each the arterial and capillary blood sample group) would allow to:

From S1 Figure/ S 1 Table A/B:

* Show for a calculated bias (mean difference *d*) of 0.8 kPa that this is significantly smaller than the maximal clinically acceptable bias of 1.0 kPa by the upper limit of the 95% CI.
* Show for a calculated bias (mean difference *d*) of 1.2 kPa that this is significantly larger than the maximal clinically acceptable bias of 1.0 kPa by the lower limit of the 95% CI.
* Show for a calculated bias (mean difference *d*) of 1.0 kPa, the clinically maximal acceptable bias, that the “true” bias is not larger than 1.13 kPa by the upper limit of the 95% CI.

In the last two cases (bias equal or larger than 1 kPa), substraction (for *d* ≥ 1 kPa) or addition (for *d* ≤ 1 kPa) of this systematic deviation may allow correcting transcutaneous measurements, assuming that precision is considered clinically acceptable.



**S1 Figure**: 95% confidence intervals of mean difference (*blue lines*) over sample size *N*, for an assumed maximal standard deviation *s* of 0.7 kPa. Confidence intervals will be tighter for lower values of *s*. *Red solid lines*: limits of target accuracy (1 kPa). *Red dashed line*: reference line indicating no bias ( = 0).

|  |
| --- |
| **S1 Table A**: **Single measurement per subject.** 95% confidence intervals (95% CI) of calculated mean difference (*d*) for different sample sizes *N* and a maximal expected standard deviation of the mean difference of *s=0.7 kPa* (“worst case scenario”). |

|  |  |  |  |
| --- | --- | --- | --- |
| **N** | **95% CI for *d* = 0.8 kPa** | **95% CI for *d* = 1.0 kPa** | **95% CI for *d* = 1.2 kPa** |
| **50** | 0.61-0.99 | 0.86-1.14 | 1.01-1.39 |
| **60** | 0.62-0.98 | 0.87-1.13 | 1.02-1.38 |
| **70** | 0.64-0.96 | 0.88-1.12 | 1.04-1.36 |
| **80** | 0.65-0.95 | 0.89-1.11 | 1.05-1.35 |

Assuming additionally a within-patient variance (sw2) of 0.12 (i.e. <<than the variance of the method differences *s2*), and applying repeated-measures ANOVA, the same sample size of patients would be required (with m=2 measurements per subjects) to calculate the confidence interval of the mean difference with the same precision:

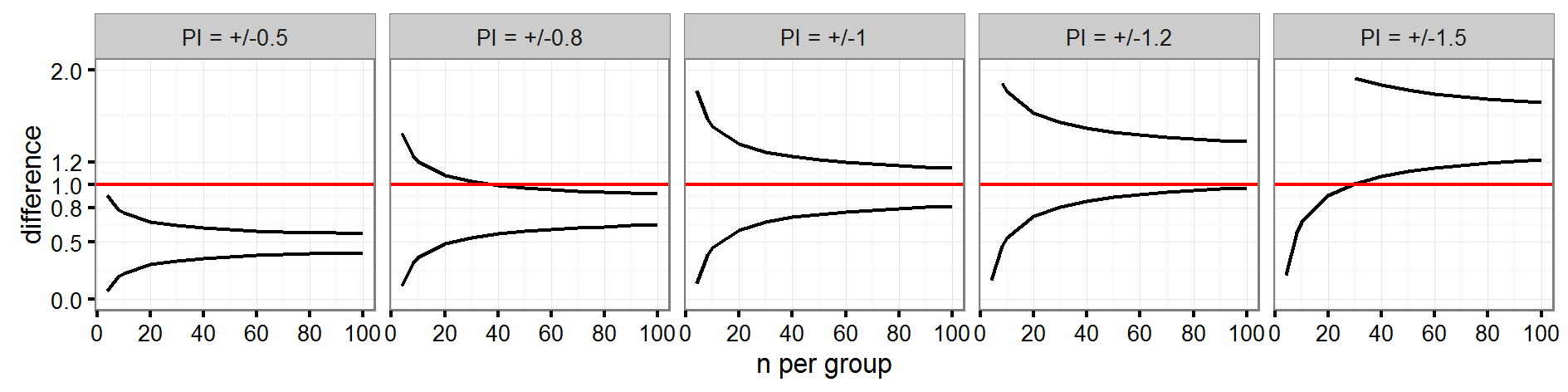
|  |
| --- |
| **S1 Table B**: **Repeated versus single measurement per subject (m=2 measurements per subject).** 95% confidence intervals (95% CI) of calculated mean difference (*d*) for different sample sizes *N* and a maximal expected standard deviation of the mean difference of *s=0.7 kPa* (“worst case scenario”) from repeated measures analysis, assuming a within-patient variance sw2=0.12. |

|  |  |  |  |
| --- | --- | --- | --- |
| **N (m=2)** | **95% CI for *d* = 0.8 kPa** | **N (m=1)** | **95% CI for *d* = 0.8 kPa** |
| **20** | **0.5-0.11** | **40** | 0.58-1.02 |
| **25** | 0.53-1.07 | **50** | 0.61-0.99 |
| **30** | 0.56-1.06 | **60** | 0.62-0.98 |
| **35** | 0.57-1.03 | **70** | 0.64-0.96 |
| **40** | 0.58-1.01 | **80** | 0.65-0.95 |
| **50** | 0.60-0.99 |  |  |
| **60** | 0.63-0.98 |  |  |

This sample size of 60 patients per group would also allow determining the limits of agreement (precision intervals *PI*) with good precision, assuming the effect of within-patient variance on the width of 95% confidence intervals to be negligible:

From S2 Figure/ S2 Table, i.e. for the comparison of monitor values to arterial blood values, this would allow to

* Show for a calculated precision interval *PI* of ±0.8 kPa that this is significantly *smaller* than the maximal clinically acceptable precision interval of ±1.0 kPa by the upper limit of the 95% CI.
* Show for a calculated precision interval *PI* of ±1.5 kPa that this is significantly *larger* than the maximal clinically acceptable precision interval of ±1.0 kPa by the lower limit of the 95% CI.
* Show for a calculated precision interval *PI* of ±1.0 kPa, the clinically maximal acceptable precision interval, that the “true” precision interval is not larger than ±1.20 kPa by the upper limit of the 95% CI.



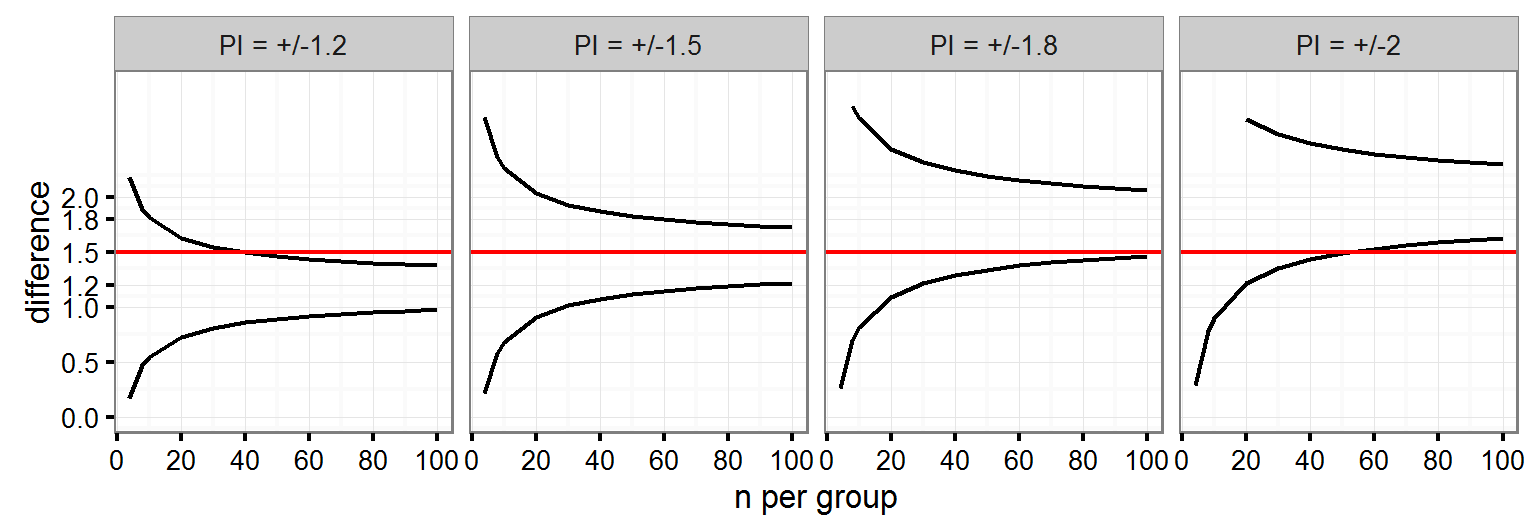
**S2 Figure**: 95% confidence intervals of precision intervals (*black lines,* only the upper limit of the acceptance interval is illustrated) over sample size n for different standard deviations of the mean difference ( = 0.3 to 0.7 kPa) resulting in precision limits of ±0.6 to 1.4 kPa. *Red line*: maximal acceptable precision interval for the arterial measurement group (± 1 kPa).

|  |
| --- |
| **S2 Table**: 95% confidence intervals (95% CI) of calculated precision intervals (*PI*) for different sample sizes *n* and standard deviation of mean difference *s*. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **N** | **95% CI for**  ***PI* ± 0.5 kPa**  (*s* = 0.25 kPa) | **95% CI for**  ***PI* = ±0.6 kPa**  (*s* = 0.3 kPa) | **95% CI for**  ***PI* = ±1.0 kPa**  (*s* = 0.5 kPa) | **95% CI for**  ***PI* = ±1.2 kPa**  (*s* = 0.6 kPa) | **95% CI for**  ***PI* = ±1.5 kPa**  (*s* = 0.75 kPa) |
| **50** | ±0.37-0.61 | ±0.59-0.98 | ±0.74-1.22 | ±0.89-1.46 | ±1.11-1.83 |
| **60** | ±0.38-0.60 | ±0.61-0.96 | ±0.76-1.20 | ±0.91-1.44 | ±1.15-1.79 |
| **70** | ±0.39-0.59 | ±0.62-0.95 | ±0.78-1.18 | ±0.93-1.42 | ±1.17-1.77 |
| **80** | ±0.40-0.58 | ±0.63-0.94 | ±0.79-1.17 | ±0.95-1.40 | ±1.19-1.75 |

From S3 Figure/ S3 Table, i.e. for the comparison of monitor values to capillary blood values, this would allow to:

* Show for a calculated precision interval *PI* of ±1.2 kPa that this is significantly *smaller* than the maximal clinically acceptable precision interval of ±1.5 kPa by the upper limit of the 95% CI.
* Show for a calculated precision interval *PI* of ±2.0 kPa that this is significantly *larger* than the maximal clinically acceptable precision interval of ±1.5 kPa by the lower limit of the 95% CI.
* Show for a calculated precision interval *PI* of ±1.5 kPa, the clinically maximal acceptable precision interval, that the “true” precision interval is not larger than ±1.79 kPa by the upper limit of the 95% CI.



**S3 Figure**: 95% confidence intervals of precision intervals (*black lines,* only the upper limit of the acceptance interval is illustrated) over sample size n for different standard deviations of the mean difference ( = 0.6 to 1.0 kPa) resulting in precision limits of ±1.2 to 2.0 kPa. *Red line*: maximal acceptable precision interval for the capillary measurement group (± 1.5 kPa).

|  |  |  |  |
| --- | --- | --- | --- |
| **S3 Table:** 95% confidence intervals (95% CI) of calculated precision intervals (*PI*) for different sample sizes *n* and standard deviation of mean difference *s*. | | | |
|  | **95% CI for *PI* = ±1.2 kPa**  (*s* = 0.6 kPa) | | **95% CI for *PI* = ±1.5 kPa**  (*s* = 0.75 kPa) | **95% CI for *PI* = ±1.8 kPa**  (*s* = 0.9 kPa) | **95% CI for *PI* = ±2.0 kPa**  (*s* = 1.0 kPa) | |
| **50** | ±0.89-1.46 | | ±1.11-1.83 | ±1.34-2.19 | ±1.49-2.43 | |
| **60** | ±0.91-1.44 | | ±1.15-1.79 | ±1.37-2-15 | ±1.53-2.39 | |
| **70** | ±0.93-1.42 | | ±1.17-1.77 | ±1.40-2.12 | ±1.56-2.36 | |
| **80** | ±0.95-1.40 | | ±1.19-1.75 | ±1.43-2.10 | ±1.59-2.33 | |

In summary, considering the repeated measurement design with (at least) 2 measurements per patient, assuming variance of mean differences *s2* to be ≤0.72, and the within-subject variance (*sw2*) to be much smaller than *s2* (*sw2* ≤ 0.12), at least 60 patients per group would provide reliable estimates of bias and precision (primary outcome), i.e. each 60 patients for comparison of transcutaneous and arterial, and transcutaneous and capillary measurements, respectively (120 patients in total).

## Study variables

The following variables were obtained. For baseline assessment: postconceptional age; postnatal age; birth weight; and actual weight. For primary endpoint at the time of blood sampling: transcutaneous partial carbon dioxide (PtcCO2) and oxygen pressures (PtcO2) as provided by the Sentec OxiVenTTM system; arterial partial carbon dioxide (PaCO2) and oxygen pressures (PaO2); capillary partial carbon dioxide (PcapCO2); and oxygen pressures (pcapO2) as provided by blood gas analyses (BGA; ABL Flex 800, Radiometer, Copenhagen, Denmark). For secondary endpoint: sensor operational temperature and application time; presence of skin and soft tissue edema, determined by the presence of relevant edema on anterior-posterior chest radiographs, i.e. doubled horizontal distance between skin and rip surface on level of the sixth thoracic vertebra in the time from admission to time of blood sampling; vasoactive drugs, quantified by an inotropic score for the amount of vasopressors established by Wernovsky et al. and modified by Choong et al.: one point for every µg/kg/min dopamine or dobutamine, 10 points for every µg/kg/min milrinone and 100 points for every µg/kg/min adrenaline or noradrenaline [4, 5]; blood levels of indirect bilirubin; capillary refill time in the same anatomical region where the sensor was applied; presence of intra- or extracardiac right to left shunt present as confirmed by echocardiography by pediatric cardiologist; skin color type as assessed by the Fitzpatrick-Scale [6]. For safety outcome: signs of burns: redness; blisters; and/or necrosis at the transcutaneous measurement sites while sensor was in place (48 hours in total) and within 4 hours after sensor removal (total surveillance period: 52 h).

**Study Flow Diagram**

## Enrollment

## Analysis

Excluded (n=385)

  Not meeting inclusion criteria (n=369)

  Declined to participate (n=11)

  Other reasons (n=5)

Included (n=113)

Assessed for eligibility (n=498)

## Additional information for results

**Skin Colour Type**

The distribution of skin color types was skewed towards lightish-skinned neonates: 88.2% of patients had Fitz-Patrick skin colour type I or II (both light skin) and very few patients (11.8%) had darker skin (type III-IV), thus skin colour was not analysed as covariate.

**Sensitivity analyses for outlying data.**

After visual inspection, outlying data points of three patients were deleted from database to evaluate the influence on bias and precision and the analyses were rerun. For pCO2 this did not change significantly bias (overlapping 95% confidence intervals), precision intervals were 10% smaller (±1.61 instead of ±1.79 kPa). For pO2, this did not significantly change bias neither, precision intervals were 13% smaller (±4.25 instead of ±4.88 kPa).

**S4 Table**: Outlying data points

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlier #** | **BGA type** | **Sample** | **BGA (kPa)** | **Sensor (kPa)** |
| 1 | arterial | pCO2 | 6.7 | 13.3 |
| 2 | arterial | pO2 | 24.4 | 8.0 |
| 3 | arterial | pO2 | 19.7 | 4.2 |

**Outlying case presentation:**

Outlier #1 was a term birth neonate (37 5/7 weeks of GA, birth weight 2750g) with prenatally unknown d-transposition of the great arteries and microdeletion syndrome 22q11.2 including severe T-cell immunodeficiency. The arterial switch operation was performed on the 5th day of life with consecutive renal failure requiring peritoneal dialysis and with a postoperative III° atrioventricular block requiring permanent pacemaker implantation on 19th day of life. Further, the child developed postoperatively several thromboses between 7th and 18th days of life (right ateria iliaca externa, femoralis communis, and superficialis; right vena jugularis interna, and subclavia; intrahepatic vena cava inferior; left vena iliaca externa) and was anticoagulated, and a pneumothorax requiring revision of the perioperatively inserted thoracic drainage on 12th day of life. On the 11th day of life the child developed decompensated septic shock caused by Pseudomonas aeruginosa found in blood culture, tracheal aspirate and pleural fluid requiring escalation of vasoactives and antibiotic treatment. It was extubated after pacemaker insertion on 20th day of life and discharged from the intensive care unit to the intermediate care neonatology ward on 34th day of life. It was discharged home on day of life 52.

The child was included into the study on 14th day of life into this study, still being mechanically ventilated with pressure support ventilation, dialyzed, without vasoactives, and with significant soft tissue edema. The child was in stable cardiocirculatory conditions with a body temperature of 37.2°C. Following the above mentioned pCO2 monitor readings showing a large discrepancy to the arterial values (S4 Table), the membrane and the sensor of the study monitor were visually inspected and no technical failure could be found. However, the membrane was changed and further readings were clinically plausible. Therefore, besides being in a postoperative state requiring intensive care support, no evident reason for the divergent monitor readings could be found.

Outlier #2 was a term birth neonate (40 0/7 weeks of GA, birth weight 3450g) also with prenatally unknown d-transposition of the great arteries. On 6th day of life the arterial switch operation was performed and the child was included in this study immediately after arrival from operational theatre. After 6 h the child developed supraventricular tachycardia that resolved spontaneously. Further, it was ventilated via pressure support ventilation with an inspiratory fraction on inhaled oxygen of 0.5. The above mentioned blood gas analysis (S4 Table) was taken at 8 h postoperatively under vasoactive support (adrenaline 0.03 µg/kg/min; noradrenaline 0.07 µg/kg/min, milrinone 0.5 µg/kg/min) and showed high paO2, still under FiO2 0.5 which could be reduced afterwards. The child had a targeted body temperature of 35.5-36.5°C, body temperature at time of blood sampling was actually 35.4°C. It needed some volume resuscitation (20 ml/kg cristalloids) and was extubated on 7th day of life. It was discharged to the intermediate care neonatology ward on 9th day of life and discharged home on 21st day of life.

Taken together, at the time of blood sampling, the child was in poor cardiac output state 8 h after arterial switch operation, therefore, compromised peripheral perfusion is the reason most likely being responsible for the highly divergent monitor readings.

Outlier #3 was a term birth neonate (39 3/7 weeks of GA, birth weight 3580g) again with a prenatally unknown d-transposition of the great arteries requiring balloon atrioseptostomy (Rashkind) on 4th day of life for low oxygen saturation readings. The arterial switch operation was performed on 7th day of life and the child was included into the study immediately after arrival on the intensive care ward from operational theatre. Postoperatively it was mechanically ventilated with pressure support ventilation with FiO2 0.35, was on vasoactive support with adrenaline, noradrenaline, milrinone, and developed several episodes of supraventricular tachycardias (sixteen episodes), therefore amiodarone was started and the body temperature was targeted to be 35.5-36.5°C. The above mentioned blood gas analysis was sampled 12 h postoperatively after having received volume resuscitation 100 ml/kg cristalloids and being on vasoactive support with adrenaline (0.12 µg/kg/min), noradrenaline (0.18 µg/kg/min), milrinone (0.5 µg/kg/min), and (amiodarone 10 µg/kg/min). Body temperature at time of blood sampling was 36.0°C. It could be extubated on 9th day of life, discharged to the intermediate care neonatology ward on 11th day of life, and discharged home on 22nd day of life.

In summary, the child was in a postoperative low cardiac output state at the time of blood sampling, aggravated by arrhythmias. Compromised peripheral perfusion was most likely the cause for the paO2 monitor readings highly diverging from blood gas analysis.

**S5 Table**: Numerical summary of bias and precision stratified by sensor temperature.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **BGA type** | **Sensor Temperature**  **(°C)** | **Samples**  **(n)** | **Patients**  **(n)** | **Bias in kPA**  **(95%CI)** | **Precision interval in kPa**  **(95% limits of agreement)** |
| pCO2  (kPa) | arterial | 42°C | 157 | 60 | +0.74  (0.54; 0.94) | ±1.89  (-1.14; 2.63) |
| 43°C | 77 | 40 | +0.34  (0.16; 0.52) | ±1.48  (-1.14; 1.82) |
| capillary | 42°C | 50 | 34 | -0.18  (-0.53; 0.16) | ± 2.42  (-2.06; 2.24) |
| 43°C | 28 | 25 | -0.31  (-0.51; -0.10) | ±1.04  (-1.35; 0.74) |
| pO2  (kPa) | arterial | 42°C | 151 | 59 | -2.96  (-3.50; -2.43) | ±5.32  (-8.28; 2.36) |
| 43°C | 77 | 40 | -1.58  (-0.65; 0.49) | ±3.16  (-4.73; 1.58) |
| capillary | 42°C | 46 | 32 | -0.08  (-0.37; 0.66) | ±3.33  (-3.41; 3.25) |
| 43°C | 24 | 21 | +0.63  (0.001; 1.26) | ±2.89  (-2.26; 3.52) |

*BGA:* blood gas analysis. *Bias*: mean difference Δ(sensor-BGA). *Precision interval*: Calculated as ±1.96 x total standard deviation (s), with total s = √(s2) = √(sinter-individual2+ sresidual2) estimated from repeated measurements ANOVA (random intercept model). *Limits of agreement*: calculated as bias ± precision interval.

**References**

1. Bernet-Buettiker V, Ugarte MJ, Frey B, Hug MI, Baenziger O, Weiss M. Evaluation of a New Combined Transcutaneous Measurement of Pco2/Pulse Oximetry Oxygen Saturation Ear Sensor in Newborn Patients. Pediatrics. 2005 January 1, 2005;115(1):e64-e68.

2. Bernet V, Doell C, Cannizzaro V, Ersch J, Frey B, Weiss M. Longtime performance and reliability of two different PtcCO2 and SpO2 sensors in neonates. Pediatric Anesthesia. 2008;18(9):872-77.

3. Bland M. How can I decide the sample size for a repeatability study? 2010.

4. Wernovsky G, Wypij D, Jonas RA, Mayer JE, Hanley FL, Hickey PR, et al. Postoperative Course and Hemodynamic Profile After the Arterial Switch Operation in Neonates and Infants: A Comparison of Low-Flow Cardiopulmonary Bypass and Circulatory Arrest. Circulation. 1995 October 15, 1995;92(8):2226-35.

5. Choong K, Bohn D, Fraser DD, Gaboury I, Hutchison JS, Joffe AR, et al. Vasopressin in pediatric vasodilatory shock: a multicenter randomized controlled trial. Am J Respir Crit Care Med. 2009 Oct 1;180(7):632-9.

6. Fitzpatrick TB. The validity and practicality of sun-reactive skin types i through vi. Arch Dermatol. 1988;124(6):869-71.