Outcomes after automated oxygen control for preterm infants

Salverda, H.H.

Chapter 5

Comparing descriptive statistics for retrospective studies from one-per-minute and one-per-second data

H.H. Salverda, J. Dekker, R.S.G.M. Witlox, P.A. Dargaville, S. Pauws, A.B. te Pas

Front Pediatr. 2022 May 12;10:845378. doi: 10.3389/fped.2022.845378. eCollection 2022.

Abstract

Background Large amounts of data are collected in neonatal intensive care units which could be used for research. It is unclear whether this data, usually sampled at a lower frequency, is sufficient for retrospective studies. We investigated what to expect when using one-per-minute data for descriptive statistics.

Methods One-per-second inspiratory oxygen and saturation was processed to one-per-minute data and compared on average, standard deviation, target range time, hypoxia, days of supplemental oxygen, and missing signal.

Results Outcomes calculated from data recordings (one-per-minute=92, one-per-second=92) showed very little to no difference. Sub analyses of recordings under 100 and 200 hours showed no difference.

Conclusion In our study descriptive statistics of one-per-minute data were comparable to one-per-second and could be used for retrospective analyses. Comparable routinely collected once per minute data could be used to develop algorithms or find associations retrospectively.

Keywords: neonatology, technology, data, methodology, retrospective.

Introduction

The wealth of routinely collected data in Neonatal Intensive Care Units (NICUs) has great potential. Morbidities such as bronchopulmonary dysplasia, retinopathy of prematurity and sepsis can possibly be predicted when coupling analyses of routinely measured vital signs or derivatives with outcomes. One example is the HeRO symphony system predicting sepsis from variability of heart rate.[1] In real time, algorithms can summarize relevant data, detect anomalies and notify bedside staff of risk factors for certain diseases. Routinely collected data could be used to develop algorithms or find associations retrospectively.

However, it is unclear at what frequency data should be sampled. In our NICU, data is often sampled at least once per second (one-per-second data, f.e. 1 heart rate value per second) for prospective studies, but routinely collected vital parameters are only sampled once per minute (one-per-minute data). This keeps up performance of the clinical patient data management system, and prevents high costs associated with storage of data. Other NICUs may have similar infrastructure in place with data already collected and available. Although the data could be collected at a higher frequency, it is unclear whether lower frequency data is already enough.

We hypothesized that lower frequency data could in some cases be sufficient to run retrospective studies. In this short report we investigated what to expect when using one-per-minute data abstracted from one-per-second data and investigated under what conditions one-per-minute data could be used.

Materials and methods

Routinely collected data from a previous study was used, the ethical review committee of Leiden Den Haag Delft provided a statement of no objection for obtaining and publishing the anonymized data (G19.075).[2] Data recordings were included from infants born under 30 weeks of gestation in our tertiary-level perinatal center between November 1st 2018 and March 15th 2020. Recordings were excluded if they contained no data on peripherally measured oxygen saturation (SpO_2).

Data collection and outcome measures

Parameters collected were 2-4s averaged SpO_2 measured by a weight-appropriate pulse-oximeter probe (LNCS Neo Masimo SET; Masimo Irvine, California, USA), and measured inspiratory fraction of oxygen (FiO_2). These data were sent from a

SLE6000 respirator (SLE Limited, South Croydon, UK) with OxyGenie automated oxygen titration to a MP70 bedside monitor (Philips, Eindhoven, the Netherlands) or, if no respiratory support was given, SpO_2 was measured by a Masimo module on the Philips monitor.

From the bedside monitor data is sent to two databases: a Philips Datawarehouse Connect feed to a database in which numerical data is stored once per second for 1 year; and a once per minute feed (HL7 data transfer protocol) which sends the exact value at the set interval time, which may be between 5-60 seconds (in our situation 1-per-minute). The HL7 message is picked up by our patient data management system (PDMS Metavision; IMDsoft, Tel Aviv, Israel). These data are stored for at least 15 years. No filtering, anti-aliasing, averaging or other processing is done on data prior to entry in the database.

To prevent synchronization issues caused by systems running on different timeclocks we chose to process the one-per-second data into one-per-minute data: one value per minute was extracted from one-per-second data by taking the value at the change of the minute (i.e. at 0 seconds). For both the one-per-second and one-per-minute data for SpO_2 we calculated the average, standard deviation, proportion of time within target range or hypoxia (<80). Within target range was defined as SpO_2 between 91%-95% irrespective of FiO_2, or 96%-100% when room air was being inspired. For the FiO_2 average and oxygen days were calculated. An oxygen day was defined as at least half of the data FiO_2 values of that day above 21%. Please note that this may not represent the true oxygen exposure, as the oxygen sensor can have a deviation of 1%. Finally, the number of data points and the difference between the first and last timepoint in each dataset were noted.

Data are presented as mean (SD) and median [IQR] with standard tests for normality. Data processing and analyses were done by custom written software in MATLAB (Matlab R2020b; The MathWorks Inc., Natick, Massachusetts, USA). No statistical hypothesis testing was done as we were not testing for a differences between treatments, but examining for comparability.

Results

There was data available from 92 patients, with a median of 1151774 [577843 - 2586608] one-per-second data points per patient. An excerpt from a data recording is shown in Figure 1. When processed to one-per-minute data, there were 19462 [9129 – 43162] data points left. The time difference between the first and last entry in the

data recording was 375 hours, 24 minutes and 46 seconds [157:59:33 – 762:23:11] for the one-per-second data, and 367 hours, 58 minutes and 30 seconds [155:11:45 - 757:12:00] for the one-per-minute data.

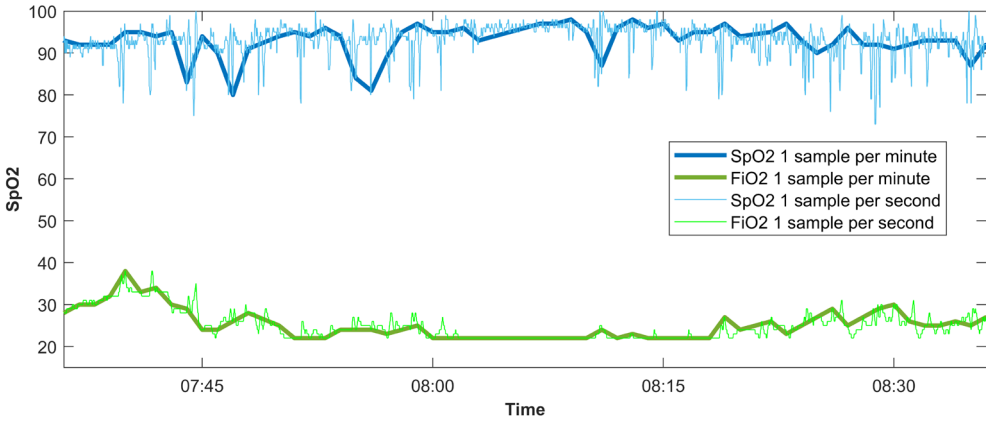


Figure 1. *An example of a data recording displaying the effect of taking one sample every sixty seconds from a 1-per-second data recording. SpO_2, oxygen saturation measured by pulse-oximetry; FiO_2, fraction of inspired oxygen.*

In the one-per-second data, the mean SpO_2 was 94.96 (1.88) vs. 94.96 (1.87) in the one-per-minute data (Table 1) and the standard deviation of SpO_2 was 3.14 (0.92) vs. 3.15 (0.91) respectively. SpO_2 was found to be within the target range in 70.96 [57.16 - 91.50] % of the time in the one-per-second data, and in 71.06 [57.00 - 91.53] % of the time in one-per-minute data. Hypoxic values under 80% were found in 0.36 [0.09 - 0.85] % of SpO_2 values in the one-per-second dataset vs. 0.35 [0.10 - 0.85] % of SpO_2 values for the one-per-minute dataset. Missing values were also similar (2.06 [1.59 - 2.91] % vs. 2.06 [1.52 - 2.87] %). Bland-Altman plots can be found in the supplemental figures S1-S6 at the end of this chapter.

Table 1. Analysis of recordings

n=92		one-per-second	one-per-minute
Average SpO_2	mean (SD)	94.96 (1.88)	94.96 (1.87)
Standard deviation SpO_2	mean (SD)	3.14 (0.92)	3.15 (0.91)
Percentage SpO_2 in target range†	median [IQR]	70.96 [57.16 - 91.50]	71.06 [57.00 - 91.53]
Percentage SpO_2 <80%	median [IQR]	0.36 [0.09 - 0.85]	0.35 [0.10 - 0.85]
Average FiO_2	median [IQR]	22.65 [21.67 - 24.56]	22.70 [21.69 - 24.59]
Days of supplemental oxygen	median [IQR]	0 [0-2]	0 [0-2]
Percentage missing SpO_2	median [IQR]	2.06 [1.59 - 2.91]	2.06 [1.52 - 2.87]
Number of data points	median [IQR]	1151774 [577843 - 2586608]	19462 [9129 - 43162]
Duration hours:min:sec	median [IQR]	375:24:46 [157:59:33 - 762:23:11]	367:58:30 [155:11:45 - 757:12:00]

FiO_2, fraction of inspired oxygen; SpO_2, peripheral oxygen saturation. † 91% <= SpO_2 <= 95% or SpO_2 >= 96% while FiO_2 = 0.21

The per-patient average inspired FiO_2 was 22.65 [21.67 - 24.56] vs. 22.70 [21.69 - 24.59], and there was no difference in oxygen days (0 [0 – 2] in both datasets).

Sub analyses of groups with only less than 100 hours (table 2), 200 hours and half of the total dataset showed similar results with little difference between one-per-second and one-per-minute data.

Table 2. Sub analysis of shorter recordings

Recordings <100 hours (n=11)		**one-per-second**	**one-per-minute**
Average SpO_2	mean (SD)	96.74 (1.74)	96.74 (1.67)
Standard deviation SpO_2	mean (SD)	3.07 (1.04)	3.17 (1.08)
Percentage SpO_2 in target range†	median [IQR]	93.58 [52.76 - 98.80]	93.87 [55.43 - 98.85]
Percentage SpO_2 <80%	median [IQR]	0.60 [0.08 - 0.81]	0.75 [0.14 - 0.96]
Average FiO_2	median [IQR]	21.05 [21.00 - 22.03]	21.05 [21.00 - 22.13]
Days of supplemental oxygen	median [IQR]	0 [0-2]	0 [0-2]
Percentage missing SpO_2	median [IQR]	2.85 [1.43 - 5.16]	3.04 [1.41 - 5.14]
Number of data points	median [IQR]	161742 [82864 - 188130]	2699 [1381 - 3364]
Duration hours:min:sec	median [IQR]	51:18:47 [32:40:20 - 51:17:00]	51:17:00 [24:13:00 - 62:04:00]
Recordings <200 hours (n=28)		**one-per-second**	**one-per-minute**
Average SpO_2	mean (SD)	95.43 (2.06)	95.43 (2.04)
Standard deviation SpO_2	mean (SD)	3.00 (1.01)	3.03 (1.04)
Percentage SpO_2 in target range†	median [IQR]	90.21 [60.06 - 96.47]	90.49 [59.62 - 94.69]
Percentage SpO_2 <80%	median [IQR]	0.32 [0.07 - 0.92]	0.30 [0.07 - 1.01]
Average FiO_2	median [IQR]	21.99 [21.01 - 25.40]	21.95 [21.01 - 25.44]
Days of supplemental oxygen	median [IQR]	0 [0-3]	0 [0-3]
Percentage missing SpO_2	median [IQR]	2.58 [1.65 - 3.45]	2.58 [1.48 - 3.28]
Number of data points	median [IQR]	507425 [175763 - 542365]	8458 [3027 - 9041]
Duration hours:min:sec	median [IQR]	149:49:41 [55:26:45 - 154:20:26]	149:48:30 [55:25:45 - 154:20:00]

FiO_2, fraction of inspired oxygen; SpO_2, peripheral oxygen saturation. † 91% <= SpO_2 <= 95% or SpO_2 >= 96% while FiO_2 = 0.21

Discussion

In this study we found little to no difference when comparing descriptive statistics of one-per-minute data and one-per-second data from the same source. This included clinically relevant outcomes as proportion of time within oxygen saturation target range, hypoxia and days of supplemental oxygen. Sub analyses of recording under 100 or 200 hours showed no difference. The results suggest that routinely collected data recordings of comparable length or longer could be used for retrospective studies.

Although using routinely collected vital parameters for big data analysis and machine learning is increasingly popular, to our knowledge there is no literature available describing the minimum data sampling frequency for our purpose. From the field of data signal processing the Nyquist-Shannon sampling theorem[3] provides us with a guideline for a sufficient sample-rate, but this is aimed at reproducing the original signal, and not the summarizing statistic we often require for our retrospective studies.

One could argue that taking a sample every minute from continuous vital signs monitoring is somewhat analogous to research in general. It is uneconomical to study an entire population, thus we take a representative sample. When the change of being sampled is related to the outcome there is a chance of biased results. Although our sample is not at random, the value is always extracted in the first second of the minute. It is unlikely that a vital parameter like heart rate is systematically lower in the first second of the minute or in other words related to the outcome. There may be a detectable circadian trend in the average heart rate, but the instantaneous heart rate should not be related to a certain second within a minute.

Limitations of our study are that 1-per-minute data cannot be used to calculate the length of vital sign episodes, for example the duration of a hypoxic episode, or other more elaborate outcomes from complex signal processing techniques. We have only investigated descriptive statistics of SpO_2 and FiO_2 and most of the data recordings had a minimum duration of 100 hours. It should also be noted that intra-recording differences were present but these are averaged out over the entire set. Finally, to prevent synchronization issues we did not compare our PDMS data directly with our higher frequency data, but down-sampled the latter. However, because in neither case filtering, anti-aliasing or other processing was done they are comparable.

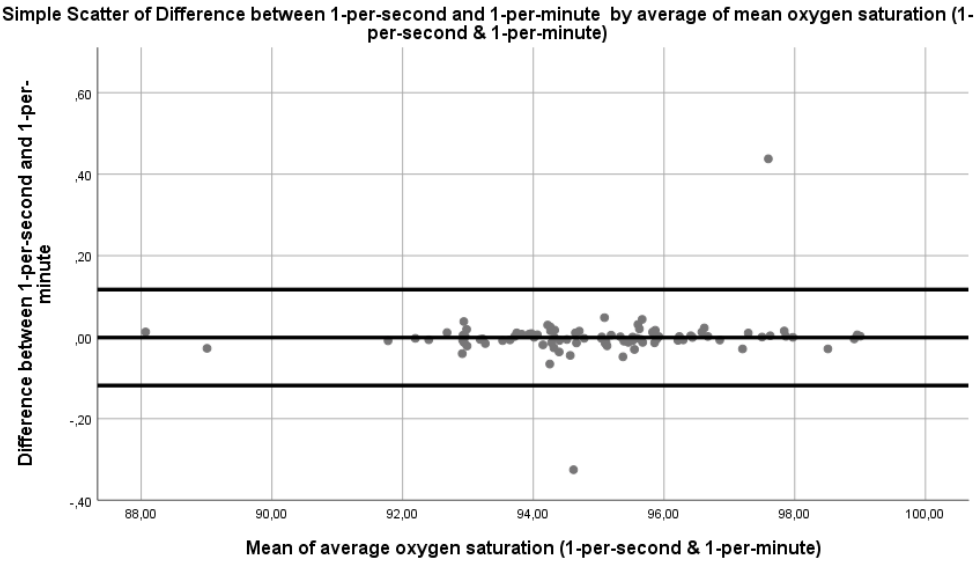
Conclusion

In our study descriptive statistics of lower frequency data were comparable to high frequency data and could be used for retrospective analyses. Comparable routinely collected once per minute data could be used to develop algorithms or find associations retrospectively.

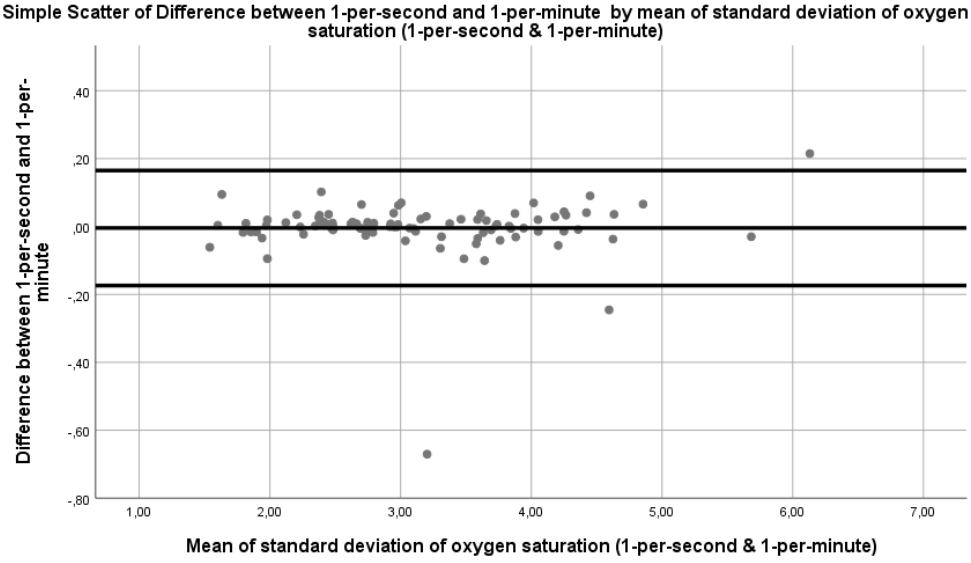
References

1. Moorman JR, Carlo WA, Kattwinkel J, et al. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. *The Journal of pediatrics* 2011;159(6):900-6.e1.
2. Salverda HH, Oldenburger NJ, Rijken M, et al. The effect of automated oxygen control on clinical outcomes in preterm infants: a pre- and post-implementation cohort study. *Eur J Pediatr* 2021;180(7):2107-13.
3. Miller FP, Vandome AF, McBrewster J. Nyquist-Shannon Sampling Theorem: Aliasing, Sine Wave, Signal Processing, Nyquist Rate, Nyquist Frequency, Sampling Rate, Shannon-Hartley Theorem, Whittaker-Shannon Interpolation Formula, Reconstruction from Zero Crossings: Alphascript Publishing 2010.

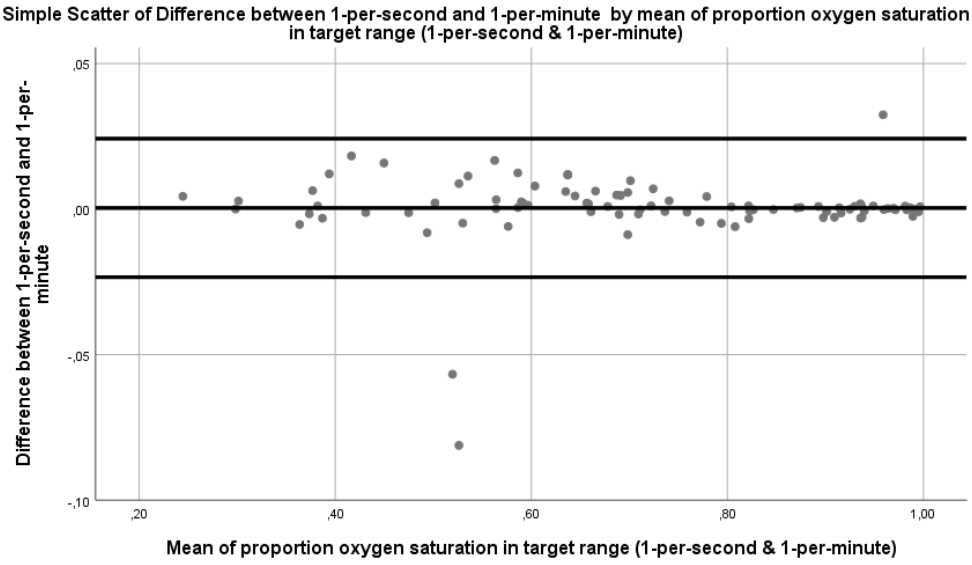
Supplemental material



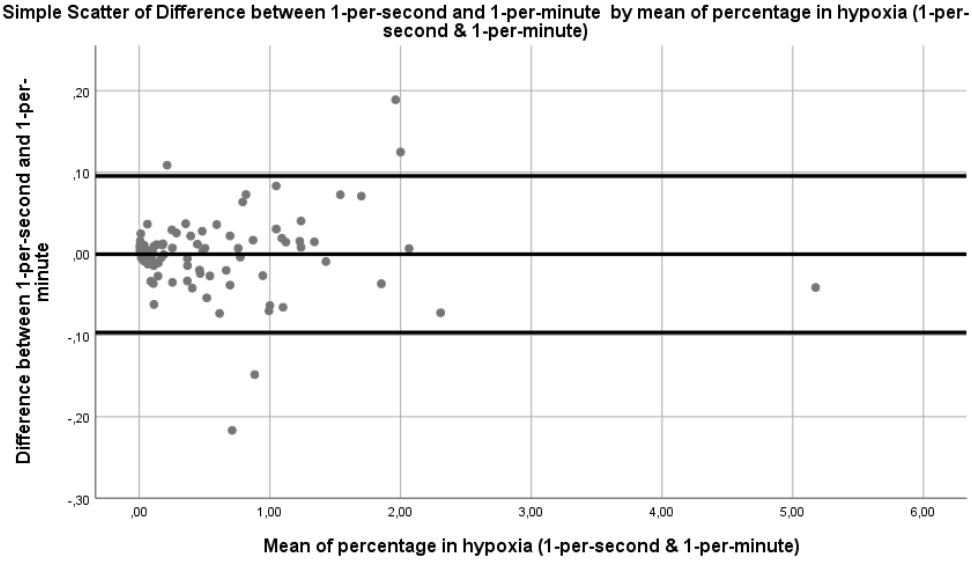
Supplementary Figure 1. *Bland-Altman plot of average oxygen saturation 1-per-second data & 1-per-minute data*



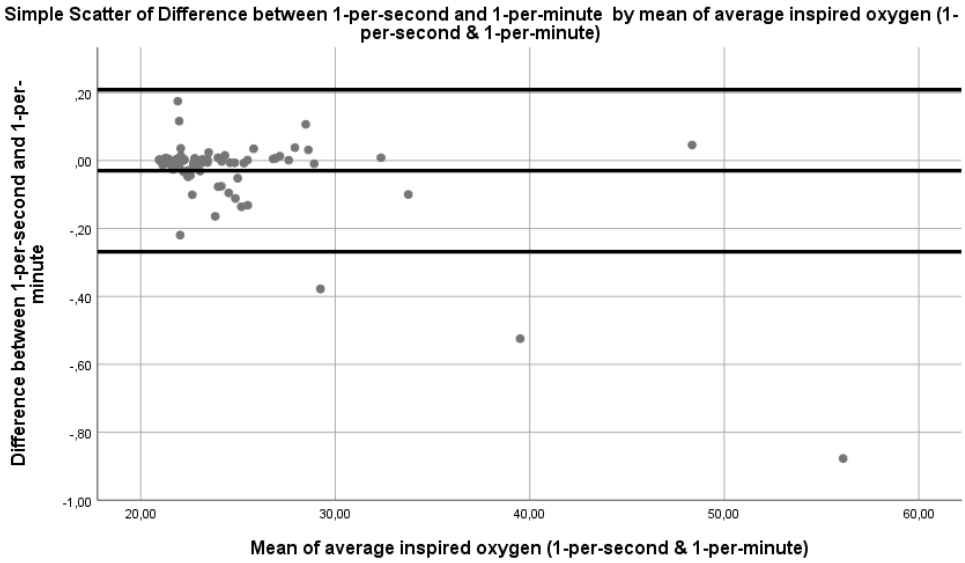
Supplementary Figure 2. *Bland-Altman plot of standard deviation oxygen saturation 1-per-second data & 1-per-minute data*



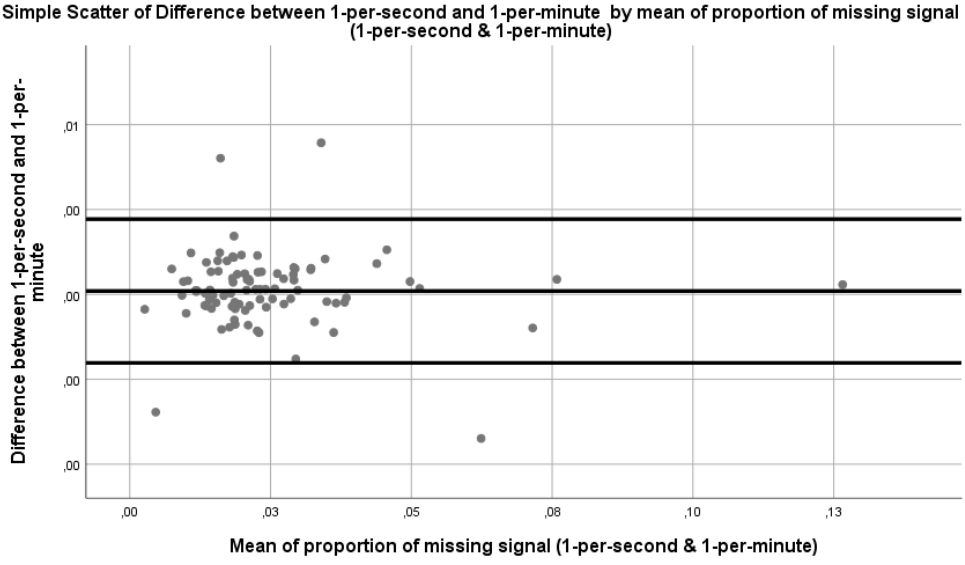
Supplementary Figure 3. *Bland-Altman plot of oxygen saturation in target range 1-per-second data & 1-per-minute data*



Supplementary Figure 4. *Bland-Altman plot of hypoxia 1-per-second data & 1-per-minute data*



Supplementary Figure 5. *Bland-Altman plot of average inspired oxygen 1-per-second data & 1-per-minute data*



Supplementary Figure 6. *Bland-Altman plot of missing signal 1-per-second data & 1-per-minute data*