

Exploring the impact of pulse oximeter selection within the COVID-19 home-use pulse oximetry pathways

David Stell,¹ Jonathan James Noble,² Rebecca Hazell Kay,^{1,3} Man Ting Kwong,¹ Michael John Russell Jeffryes ,^{3,4} Liam Johnston,^{3,4} Guy Glover,⁵ Emmanuel Akinluyi ^{1,3}

To cite: Stell D, Noble JJ, Kay RH, *et al*. Exploring the impact of pulse oximeter selection within the COVID-19 home-use pulse oximetry pathways. *BMJ Open Res Res* 2022;**9**:e001159. doi:10.1136/bmjresp-2021-001159

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjresp-2021-001159>).

Received 22 November 2021
Accepted 16 January 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Medical Physics, Guy's and St Thomas' Hospitals NHS Trust, London, UK

²One Small Step Gait Laboratory, Guy's and St Thomas' Hospitals NHS Trust, London, UK

³School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

⁴Medical Engineering and Physics, King's College Hospital NHS Foundation Trust, London, UK

⁵Department of Critical Care, Guy's and St Thomas' NHS Foundation Trust, London, UK

Correspondence to

Dr Emmanuel Akinluyi;
Emmanuel.Akinluyi@kcl.ac.uk

ABSTRACT

Background During the COVID-19 pandemic, portable pulse oximeters were issued to some patients to permit home monitoring and alleviate pressure on inpatient wards. Concerns were raised about the accuracy of these devices in some patient groups. This study was conducted in response to these concerns.

Objectives To evaluate the performance characteristics of five portable pulse oximeters and their suitability for deployment on home-use pulse oximetry pathways created during the COVID-19 pandemic. This study considered the effects of different device models and patient characteristics on pulse oximeter accuracy, false negative and false positive rate.

Methods A total of 915 oxygen saturation (s_pO_2) measurements, paired with measurements from a hospital-standard pulse oximeter, were taken from 50 patients recruited from respiratory wards and the intensive care unit at an acute hospital in London. The effects of device model and several patient characteristics on bias, false negative and false positive likelihood were evaluated using multiple regression analyses.

Results and conclusions All five portable pulse oximeters appeared to outperform the standard to which they were manufactured. Device model, patient s_pO_2 and patient skin colour were significant predictors of measurement bias, false positive and false negative rate, with some variation between models. The false positive and false negative rates were 11.2% and 24.5%, respectively, with substantial variation between models.

INTRODUCTION

The COVID-19 pandemic created unprecedented demand for bed space within National Health Service (NHS) hospitals. Home-use pulse oximeters were introduced to allow home monitoring of some patients who would otherwise occupy a hospital bed. This was recommended by WHO guidance.¹ There is evidence of uptake of this recommendation in various health systems.^{2,3} Use of similar pathways is likely to increase,⁴ including for non-COVID-19 patients.⁵

There are two virtual wards created by the NHS in England to manage the surge

Key messages

What is already known on this topic?

- Pulse oximeter performance varies with factors like subject skin colour and oxygen saturation.
- The pulse oximeters investigated for this study are known to meet the requirements for “CE marking”, which guarantees a minimum level of safety and performance.

What this study adds

- This study considered the performance, and factors which affect performance, of five devices issued for home-use during the COVID-19 pandemic, for which no similar analysis has previously been published.
- The study also considered the implications of any inaccuracy with respect to the escalation thresholds within the COVID-19 home-use pathways.

How this study might affect research, practice or policy

- This study should improve understanding of the performance and limitations of these devices and the implications when they are deployed for home-use within the COVID-19 home oximetry pathways.
- It is likely that more pathways will be developed, which involve the home-use of diagnostic medical devices. The methods used and the findings reported here, may inform further and larger-scale evaluations into oximeters and other devices.

of COVID-19 patients: ‘COVID-19 Virtual Wards’ and ‘COVID-19 Oximetry @Home’. Both virtual wards were supported by portable pulse oximeters (as opposed to hospital grade pulse oximeters) due to their wider availability. For the rest of the article portable pulse oximeters will be referred to as pulse oximeters for simplicity.

COVID-19 Virtual Wards were operated by secondary care providers, and it contains an ascending and a descending pathways. The descending pathway included patients in the recovery phase, deemed fit for transfer to a virtual ward for home-monitoring. The



ascending pathway included patients with a COVID-19 diagnosis deemed appropriate for home monitoring. A pulse oximeter was issued to patients on both pathways.⁶ Patients were asked to record their pulse oximetry derived oxygen saturation (s_pO_2) three times daily, and they were proactively contacted by phone daily, and asked to contact the hospital if their s_pO_2 fell below 92%² or if other symptoms worsened. Asymptomatic patients were considered for discharge by fourteen days. The other virtual ward, COVID-19 Oximetry @Home, was led by primary care providers,⁷ where pulse oximeters were issued to COVID-19 positive patients. Patients were supported to self-escalate if their s_pO_2 fell to 92% or below, and to call a low-acuity hotline or their general practitioner if it fell to 93% or 94% or if other symptoms worsened. Various media articles^{8,9} suggested that the public purchase pulse oximeters and monitor themselves, hence forming a self-referral pathway. The recommended decision boundaries were the same as for the two virtual wards.

Pulse oximeter performance

The ISO 80601-2-61:2019 standard gives performance requirements for pulse oximeters.¹⁰ Pulse oximeter manufacturers are not mandated to comply with this standard, but compliance is generally expected for NHS devices. All of the virtual ward devices claimed compliance with this standard.⁷

The standard requires that manufacturers perform a study to quantify their pulse oximeter's accuracy. This can be a desaturation study on healthy volunteers, using either a CO-oximeter or another pulse oximeter as reference. Patient studies are also permitted provided that a CO-oximeter is used as reference. Root-mean-square error may not exceed 4% for s_aO_2 in the range 70%–100%. The standard recommends that study subjects 'should vary in their physical characteristics to the greatest extent possible' to permit broad application to different patient groups. It does not prescribe sample size or analysis methodology. Publication and peer-review of these studies is not required.

Assurance of the accuracy of the pulse oximeters deployed during the pandemic was sought, since they had little history of NHS use and the results of previous investigations of pulse oximeters performance have been variable.

A desaturation study on six non-US Food and Drug Administration (FDA)-approved low-cost pulse oximeters¹¹ found that 4/6 did not meet ISO 80601-2-61 requirements. The study included two device models that are earlier versions of devices used for the virtual wards: the Contec CMS50DL (which did meet the requirements) and the Beijing Choice MD300C23 (which did not). Two further clinical studies found that several portable pulse oximeters (including the Contec CMS50D, CMS50DL and ChoiceMMed MD300C52) performed acceptably

in patients with relatively high s_pO_2 ,^{12,13} but offered no definitive comment for patients with lower s_pO_2 .

Other studies which have considered the accuracy of portable pulse oximeters similar to those investigated for this study, have found that their measurements agree closely with gold standard measurements from CO-oximeters. These include studies which consider these devices' use at high altitude (a common use case for these devices, due to their easy portability).^{14,15}

Concerns had also been raised about pulse oximeter performance for patients with darker skin. A California based research group has undertaken desaturation studies to evaluate the relationship between pulse oximeter performance and skin tone,^{16,17} finding that skin tone, gender and s_aO_2 range were consistent predictors of bias for various pulse oximeters. They found that s_pO_2 was overestimated in dark-skinned, hypoxic individuals ($s_aO_2 < 80\%$), and concluded that this effect could be clinically significant. Similar findings were presented in a letter to the *New England Journal of Medicine*,¹⁸ and an early study of ear pulse oximetry also found that performance was poorer for darker skinned patients.¹⁹ Other studies have found no relationship between pulse oximeter performance and skin colour,^{20,21} including one study on the Contec CMS50D.¹²

Factors other than skin tone may affect pulse oximeter performance. These devices rely on arterial flow pulsatility to measure s_pO_2 , so conditions like peripheral arterial disease (PAD), which reduce pulsatility, may affect function.^{22–24} The presence of carboxyhaemoglobin in the blood (for instance, due to smoking) also causes pulse oximeters to overestimate s_pO_2 .²²

While questions remain about the performance characteristics of home-use pulse oximeters, there is evidence that they reduce COVID-19 mortality in some populations. A South African retrospective cohort study of high risk COVID-19 positive individuals found that mortality was 48% lower in individuals issued with a pulse oximeter than in those without one.²⁵

Previous studies have not investigated the performance of pulse oximeters used within the context of these virtual wards, considering the 92% decision threshold. For this study 'false positive' was defined as a s_pO_2 reading less than 92% for a patient whose 'true' s_pO_2 is greater than or equal to 92%. False positive rate (FPR) was the proportion of negative results recorded as positive by the pulse oximeter. A 'false negative' was defined as a s_pO_2 reading greater than or equal to 92% for a patient whose true s_pO_2 is less than 92%. False negative rate (FNR) was defined as the proportion positive results recorded as negative by the pulse oximeter. The objective of this work is to estimate the impact of performance characteristics of these devices on the home-use pathways in terms of FPR and FNR. This work also investigated the effects of factors such as skin tone, smoking status and PAD status on FPR and FNR.

METHODS

The accuracies of the five pulse oximeters were evaluated by comparison with a hospital-use pulse oximeter, an M1191BL digital probe (Philips, Eindhoven, the Netherlands) connected to an IntelliVue X3 Patient Monitor (Philips, Eindhoven, Netherlands). The test pulse oximeters were: (1) Oxywatch MD300C19 (ChoiceMMed, Beijing, China), (2) Oxywatch MD300C29 (ChoiceMMed, Beijing, China), (3) PC-60F (Creative Medical, Shenzhen, China), (4) Contec CMS50D (Contec Medical Systems, Hebei, China) and 5) AM801 (Shenzhen Med-link Electronics Tech, Shenzhen, China).

Use and maintenance of the M1191BL digital probe and IntelliVue X3 Patient Monitor was in accordance with the manufacturer's instructions and standard hospital practice. Measurements made using this device therefore represent 'standard' measurements that would be performed in a hospital setting in the absence of any home-oximetry pathway.

This evaluation was authorised by the Guy's and St Thomas' NHS Foundation Trust Quality Improvement and Patient Safety team. Informed consent was taken from all conscious participants, and the responsible nursing team advised on the inclusion of sedated patients. Participants' clinical care was unaffected by inclusion in the study.

Patient and public involvement

Patients and members of the public were involved in the determination of this study's value, and in the design of the information sheets issues to participants.

Data collection

Patients with $s_pO_2 < 85\%$ (according to inpatient monitoring equipment) were excluded, together with unstable patients likely to experience acute changes to s_pO_2 (eg, due to respiratory support or positional changes). Each participant underwent up to three sets of measurements, at least 90 min apart.

Patients, or their nursing teams, who met the inclusion criteria and were inpatients on the intensive care units, recovery and respiratory wards, were approached for inclusion in the study. Fifty patients were recruited in total.

Measurements were collected in pairs from a test device and reference device, placed simultaneously on the ring and index finger. The measurement was taken 30s after placement of the pulse oximeter, in-line with the greatest stabilisation period recommended by any of the device manufacturers. This was repeated three times for each test device. Test order and finger selection were randomly determined for each session.

Skin pigmentation was quantified using the Fitzpatrick Skin Pigmentation (FSP) scale and recorded alongside s_pO_2 measurements. Patients' ages, genders, PAD and smoking statuses were extracted from their medical notes.

Table 1 A summary of study population characteristics is shown under the "Subjects" heading. A summary of the data pairs (a data pair consists of a measurement from a test pulse oximeter and a simultaneous measurement from a reference pulse oximeter) is shown under the " s_pO_2 data pairs" heading.

Subjects

No subjects	50
No subjects in each FSP group	1: 6 2: 21 3: 9 4: 5 5: 7 6: 2
No current smokers	9
No with known PAD	10
Age range	19 to 88
s_pO_2 data pairs	
No data pairs	915
No with ref. $s_pO_2 < 92\%$	159
No with ref. $s_pO_2 \geq 92\%$	756

FSP, Fitzpatrick Skin Pigmentation; PAD, peripheral arterial disease; spO_2 , oxygen saturation.

Data analysis

The effects of device model, gender, age, smoking status, PAD status, FSP score and s_pO_2 (as recorded using the reference device) on bias, and on the likelihoods of a false positive or false negative result were evaluated using multiple effects analyses.

The effect on bias was modelled using linear multiple regression analysis, with significance calculated using t-tests. The effects on false positive and false negative likelihoods were modelled using binary logistic regression, with significance calculated using the Wald test. Results were considered significant where $p < 0.05$.

In all cases, an initial analysis was conducted on the data set as a whole with subsequent analyses performed for each test device model individually. All analyses were performed using the IBM SPSS software, release 27.0.1.0.

RESULTS

Subject characteristics are shown in [table 1](#).

Multiple regression analyses

Aggregated data set

Table 2 shows the results of the multiple effects analyses with the aggregated data from all five test pulse oximeters. This allows investigation of the differences between the reference device and the general performance of test pulse oximeters.

The relationships between bias and its significant predictors are shown in [figure 1](#).

Bias is plotted against device model (top left), FSP (top right), subject s_pO_2 as measured by the reference device

Table 2 Significance and effect size of factors on bias, false positive (FP) likelihood and false negative (FN) likelihood

Factors	Bias		FP likelihood		FN likelihood	
	P value	Coefficient	P value	OR	P value	OR
Gender	0.735	-0.076 (-0.514 to 0.363)	0.579	0.863 (0.512 to 1.453)	0.938	1.055 (0.272 to 4.098)
Age	0.017	-0.016 (-0.029 to -0.003)	0.261	1.010 (0.992 to 1.029)	0.062	1.038 (0.998 to 1.080)
Smoking status	0.089	0.528 (-0.081 to 1.136)	0.933	1.032 (0.496 to 2.146)	0.007	0.121 (0.026 to 0.560)
PAD status	0.672	-0.121 (-0.680 to 0.439)	0.736	1.115 (0.593 to 2.098)	0.080	7.872 (0.779 to 79.58)
FSP	<0.001	-0.455 (-0.639 to -0.279)	0.071	1.263 (0.980 to 1.628)	0.006	0.344 (0.161 to 0.735)
Device model	<0.001	0.309 (0.167 to 0.451)	0.013	0.801 (0.672 to 0.953)	0.014	1.511 (1.087 to 2.100)
Ref. s_pO_2	<0.001	-0.155 (-0.209 to -0.100)	<0.001	0.694 (0.613 to 0.785)	<0.001	2.415 (1.604 to 3.636)

Bias results are from multiple regression analysis. FP and FN likelihood results are from binary logistic regression analysis. Significant p values are shown in bold. 95% CIs are shown in parentheses after coefficient and OR values. FSP, Fitzpatrick Skin Pigmentation; PAD, peripheral arterial disease; spO_2 , oxygen saturation.

(bottom left) and subject age (bottom right). Zero bias is shown for each plot with a red dashed line. Error bars show the bias SD associated with each data group.

On average, measurements from all five test pulse oximeters were lower than those by the reference device. The mean biases were -1.1% (AM801), -2.1% (CMS50D), -0.6% (MD300C29), -1.3% (MD300C29) and -0.1% (PC-60F).

The associated SD were 2.2%, 3.3%, 2.8%, 3.7% and 2.6%, respectively. These SD overestimate devices' RMS error, as they also include reference device error (quoted by the manufacturer as 2.5% RMS).

Subject s_pO_2 (as measured using the reference device) was a significant predictor of bias, false negative likelihood and false positive likelihood. Bias became more

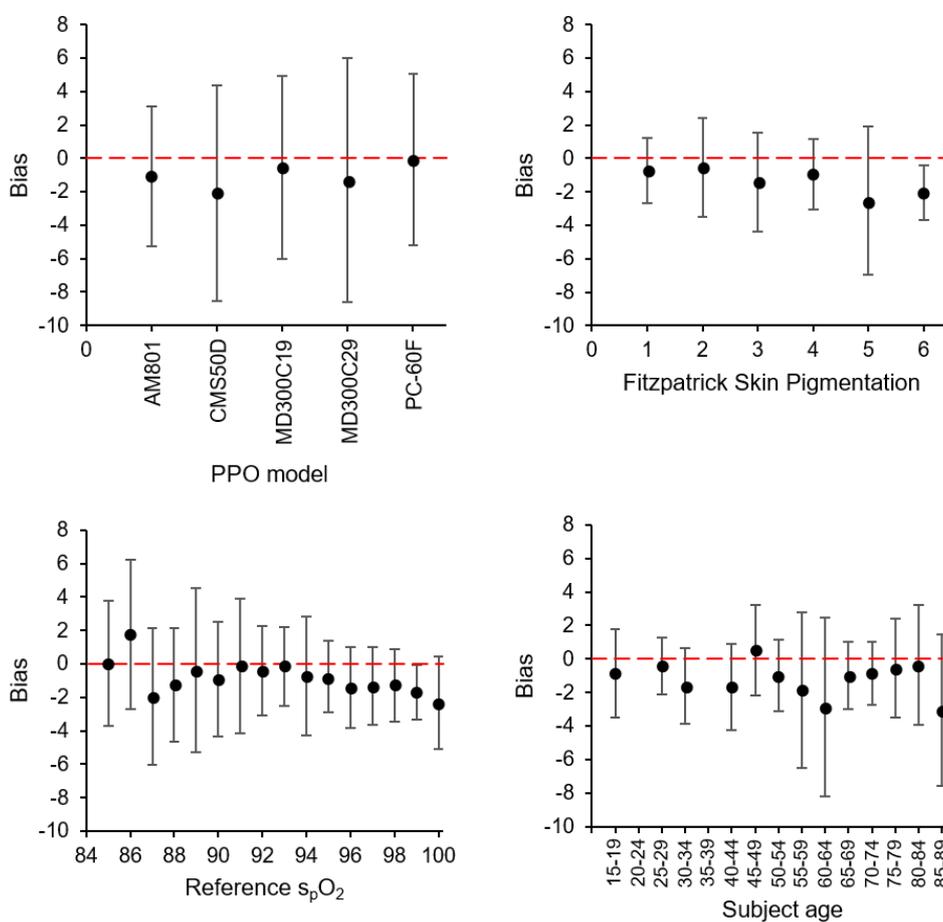


Figure 1 Plots showing the relationship between measurement bias (difference between test oximeter and reference oximeter measurements) vs (1) test pulse oximeter models (top left), (2) Fitzpatrick skin pigmentation scale (top right), (3) subject s_pO_2 as measured by the reference oximeter (bottom left), and (4) subject age (bottom right). Zero bias is shown for each plot with a red dashed line. Error bars show the bias SD associated with each data group. s_pO_2 , oxygen saturation.

Table 3 Statistically significant effects from multiple regression and binary logistic regression analyses

Dependent variable	Independent variable	Device model	P value	Effect size
Bias	Ref. s_pO_2	AM801	0.014	-0.111 (-0.198 to -0.023)
		MD300C19	<0.001	-0.255 (-0.355 to -0.156)
		PC-60F	<0.001	-0.210 (-0.571 to 0.152)
FP likelihood	Ref. s_pO_2	AM801	<0.001	0.369 (0.210 to 0.650)
		CMS50D	<0.001	0.616 (0.485 to 0.783)
		PC-60F	0.043	0.666 (0.449 to 0.987)
FN likelihood	Ref. s_pO_2	AM801	0.045	22.05 (1.068 to 455.1)
Bias	FSP score	MD300C19	0.020	-0.428 (-0.788 to -0.068)
		MD300C29	0.001	-0.886 (-1.412 to -0.360)
Bias	PAD status	MD300C29	0.011	-2.010 (-3.456 to -0.474)
		PC-60F	0.028	1.143 (0.126 to 2.160)
Bias	Smoking status	MD300C19	0.018	1.451 (0.255 to 2.646)

The p value and coefficient is reported for each significant effect. 95% CIs are shown in parentheses after the effect size values. Where the dependent variable is bias effect size is given by the multiple regression coefficient, where it is FN or FP likelihood effect size is given by the OR.

FN, false negative; FP, false positive; FSP, Fitzpatrick Skin Pigmentation; PAD, peripheral arterial disease; spO_2 , oxygen saturation.

negative with increasing s_pO_2 . The directions of the effects on false positive and FNR indicate that false results are more likely for patients whose s_pO_2 is close to the threshold value of 92%, as would be expected. Device model was also a significant predictor of bias, false negative likelihood and false positive likelihood.

Skin tone was a significant predictor of bias and of false negative likelihood. Bias was more negative for subjects with darker skin. False negative likelihood was lower for subjects with darker skin, this is consistent with a stronger tendency for test pulse oximeters to under-read for these subjects.

A significant effect existed between smoking status and FNR, with current smokers less likely to receive a false negative result than non-smokers.

The only other significant effect was between subject age and bias. The test pulse oximeters' tendency to under-read was greater for older subjects. However, the effect size was too small to be clinically significant.

Results for individual test pulse oximeters

Factors (such as age, gender, FSP score smoking status, PAD and reference s_pO_2 range) that may affect measurement bias, FPR and FNR were considered, and [table 3](#) summarises the statistically significant results from the multiple effects analyses for individual test pulse oximeters. A summary of all the results (including non-significant results) is given in the online supplemental material.

Subject s_pO_2 (as measured using the reference device) had a significant effect on bias for all PPOs except the CMS50D and MD300C29. The effect direction was the same in all cases, with bias becoming more negative for greater s_pO_2 . The effect of s_pO_2 on false positive likelihood was also fairly consistent, being significant for all devices other than the MD300C19 and MD300C29. The

direction of this effect was also the same in all cases, indicating that false positive likelihood increases for subjects whose s_pO_2 is close to the 92% threshold value. The equivalent effect on false negative likelihood (with likelihood increasing for s_pO_2 close to the threshold) was significant for the AM801 only.

Skin tone had a significant effect on bias for the MD300C19 and MD300C29, with bias becoming more negative for subjects with darker skin.

PAD status had a significant effect on bias for the MD300C29 and PC-60F. The effects were in opposite directions, with the MD300C29 tending to underestimate saturation level compared with the reference device for subjects with a PAD diagnosis, and the PC-60F tending to overestimate saturation level compared with the reference device.

The only other significant effect was between smoking status and bias for the MD300C19. This PPO was more likely to over-read in current smokers.

Observed error rates

The observed FNRs and FPRs in the study population are shown in [table 4](#). Any attempt to generalise the results shown in [table 4](#) should account for variation in population characteristics, particularly s_pO_2 distribution. The s_pO_2 distribution of the study population is therefore presented in online supplemental figure 2.

False negative and false positive likelihood vary depending on subject s_pO_2 , being greatest for subjects with s_pO_2 close to the threshold value of 92%. The PPOs' tendency to under-read might be expected to produce a higher FNR than FPR, however the opposite was true in this case. This was due to the high proportion of positive cases with s_pO_2 close to the threshold value (46.5% within

**Table 4** Summary of the performance of each test pulse oximeter

	AM801	CMS50D	MD300C19	MD300C29	PC-60F	Overall
+ve cases	38	40	28	25	28	159
-ve cases	148	152	157	150	149	756
Total	186	192	185	175	177	915
FPR	10.8%	21.1%	7.6%	10.7%	6.0%	11.2%
FNR	15.8%	17.5%	25.0%	28%	42.9%	24.5%
PPV	66.7%	50.8%	63.6%	52.9%	64.0%	58.5%
NPV	95.7%	94.5%	95.4%	95.0%	92.1%	94.5%

The rows titled '+ve cases' and '-ve cases' show the total number of data pairs with a respective positive or negative reading from the reference device. The FPR, FNR, PPV and NPV show, respectively, the observed FPR, FNR, PPV and NPV for each device. FNR, false negative rate; FPR, false positive rate; NPV, negative predictive value; PPV, positive predictive value.

2% of this value), so small biases were more likely to cause false negative results. The proportion of negative cases close to the threshold was smaller (20.1% within 2% of this value) so larger biases were generally required to produce false positive results.

The results indicate that, for the study population, both the FPR and FNR were appreciable, with approximately one in four positive cases incorrectly identified as negative, and one in nine negative cases incorrectly identified as positive by the portable pulse oximeters.

DISCUSSION

Level of agreement

All five portable pulse oximeters tended to under-read s_pO_2 relative to the reference device. The bias SD was between 2% and 4% in all cases. This would be expected to include the error associated with the portable pulse oximeters, as well as that associated with the reference device, suggesting that all devices meet the 4% RMS error requirement in the ISO 80601-2-61 standard. However, this error would be enough to affect clinical decisions, particularly for patients with s_pO_2 close to a decision threshold (as a small bias may be sufficient to push these patients' measurements across the threshold).

Pulse oximeter performance

This study evaluated the individual performance of five portable pulse oximeters, and also their aggregated performance, to identify performance attributes they shared in common. More significant effects were identified for the aggregated data set than for the individual devices. This is likely partly due to the smaller sample size, and consequently lower analytical power, for the individual models. Power was particularly limited for the analyses of false negative likelihood, due to the small number of subjects with s_pO_2 readings below 92%.

The aggregate analysis showed that pulse oximeter error (whether expressed in terms of bias, false positive likelihood or false negative likelihood) varied with patient characteristics and, independently, with device model. A personalised approach to s_pO_2 -based decision-making

could account for both effects by modelling these dependencies.

Dependency on s_pO_2

The variation in pulse oximeter performance with s_aO_2 (as measured by CO-oximetry) is well documented, and is permitted by the ISO 80601-2-61 standard.^{10 11 16 17} Pulse oximeters are known to perform more poorly at lower s_aO_2 .^{10 11 16 17} This study suggests that a separate dependency exists between the performance of the portable pulse oximeters, and the 'hospital standard' device used as reference. The direction of this dependency was the same for all five pulse oximeters, with all five under-reading by a greater amount at higher s_pO_2 . However, this effect only reached statistical significance for three of the five devices, when considered individually.

Within the home-use oximetry pathways, the effect of this dependency is likely to be modest. Irrespective of this effect, the patients most likely to be wrongly classified remain those whose s_pO_2 is closest to the decision threshold. The relationship between bias and s_pO_2 is not strong enough to much increase the likelihood of false positive results among patients with high s_pO_2 (and consequently higher bias).

Dependency on skin colour

Previous studies have reported that pulse oximeters tend to over-read s_pO_2 , relative to CO-oximeters, for dark-skinned subjects whose s_aO_2 is very low.^{16 17} This study did not include subjects with s_pO_2 low enough to be influenced by this effect. However, the data do suggest that a separate effect exists for the portable pulse oximeters. These devices appear to under-read s_pO_2 more for subjects with darker skin than those with lighter skin. This effect would be expected to increase the likelihood of false positive results, and reduce the risk of false negative results in this patient group. There was evidence of this effect in the data, although it did not achieve significance in all cases.

In the context of the home oximetry pathways, this effect could mean that failure to appropriately escalate is

less likely for darker-skinned patients, and that inappropriate escalation is more likely.

Other dependencies

Other dependencies were found to exist between age and bias, between smoking status and FNR, between PAD status and bias for specific pulse oximeter models, and between smoking status and bias for one model. These effects were not consistent for different devices or between the aggregate data set and the device specific data sets. It is difficult to draw firm conclusions from these results, however, they would merit further investigation if widespread deployment of the affected devices were planned.

Observed and forecasted error rates

The observed FNR and FPR for all five PPOs are given in [table 4](#). These parameters are strongly dependent on population characteristics, particularly s_pO_2 distribution. Caution should be exercised before generalising these results to other populations. This should be informed by a comparison of respective population characteristics (see [table 1](#) and the online supplemental material). These results provide a first estimate of the likely practical implications of deployment of these devices.

FNR implications

Within a home-oximetry pathway, a high FNR may result in failure to detect clinical deteriorations, representing a patient safety risk. This study found the FNR to be appreciable in the study population ([table 4](#)). The associated risks may be reduced by emphasising the importance of other information available to the responsible clinical team, including regular communications and the symptom diary—in line with elaborated advice.²

FPR implications

The home oximetry pathways were intended to reduce hospital bed occupancy by allowing some COVID-19 patients, who would otherwise occupy a hospital bed, to instead be monitored at home. Use of pulse oximeters with high FPR would tend to increase (re)admission and reduce this benefit. The results of this study suggest that this effect would be modest in size for these pulse oximeters.

Limitations

Choice of reference measurement device

Due to practical constraints on the protocol, a widely used hospital pulse oximeter was used as the reference device in this study, rather than a CO-oximeter. This study, therefore, gives an indication of how portable pulse oximeters compare to hospital-based oximetry. It is not possible to definitively determine the devices' accuracies from these results.

s_pO_2 range

The relevant FDA and ISO standards require that pulse oximeters are validated across an s_pO_2 range of 70%–100%. This study only included patients with $s_pO_2 \geq 85\%$. It is therefore not possible to comment on these devices' performance for $s_pO_2 < 85\%$. In practice, errors large enough for a patient with $s_pO_2 < 85\%$ to receive a measurement $\geq 92\%$ are unlikely; this limitation is therefore unlikely to affect interpretation with respect to the home oximetry pathways.

Practical challenges

This study considered a small subset of the home-use oximeters available on the market. A comprehensive analysis of all available models would be logistically challenging and might, for example, require aggregation of data from multiple centres over longer time periods.

FURTHER WORK

Further research, with a larger sample size and use of CO-oximeter measurements as reference, is required to precisely determine these devices' bias. A simple approach to accounting for this bias would be to apply the offset back to observed data, or to modify the threshold value. More sophisticated techniques, which account for the test's receiver-operator characteristics, are also available.^{26 27} Another alternative would be to use a longitudinal deterioration model as described by Prower *et al.*²⁸

CONCLUSIONS AND FINAL REMARKS

All the portable pulse oximeters investigated outperformed the standard to which they were manufactured. When used within a home-use pathway, pulse oximetry measurements must be interpreted in isolation, without the benefit of the other information available to a clinician reviewing a patient who is physically present (such as observable breathlessness and work of breathing). This creates an increased risk of error. Notwithstanding methodological limitations discussed, the variation between the pulse oximeters upholds this consideration.

This indicates the use of trend analysis for more robust indications that might mitigate inter-oximeter variation. The practice of observing baselines and trends goes beyond the initial SOP documents but is clear in elaborated guidance.² Other risk mitigations include use of conservative escalation thresholds, and inclusion of other clinical information from patients' symptom diaries and from telephone contact with clinicians. Advanced approaches might include use of a-priori data and analyses to tailor these approaches for individual patients and/or devices.

There was also evidence of performance being dependent on subject skin tone. This would tend to increase the likelihood of inappropriate escalations of patients with darker skin, and reduce the risk of these patients not being escalated when appropriate.



Acknowledgements The authors would like to thank the team of training Clinical Scientists who came together to carry out data collection: namely Charlotte Jones, Ruphinder Kaur, Amy Morris and Alexander Mitton. They were supported by Clinical Scientists Adam Shortland, Nicola Fry and Shima Maqsood. The authors would also like to thank Marlies Ostermann and Ruth Thomsen (NHSE&I Medical Directorate) for their support and facilitation of this work. Finally, we would like to thank the members of the public who helped us to verify the importance of this study, and who helped us to redraft our information sheets.

Contributors DS carried out data postprocessing; modelling and analysis; and led in the drafting of manuscript. JJN was the principal investigator, had input to protocol design and advised on ethics and governance. RHK had input into protocol design—particularly around data collection. MTK also supported on drafting, modelling and analysis. MJRJ and LJ, coordinated data collection. GG provided specialist input to the protocol design and to the interpretation of results. EA was the chief investigator, had input to protocol design, modelling and analysis, and the drafting of the manuscript. EA was responsible overall for the content as guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This evaluation (12094) was authorised by the Guy's and St Thomas' NHS Foundation Trust Quality Improvement and Patient Safety team. Non-sedated participants gave informed consent to participate in the study before taking part. Standard ethics processes were followed for consent of sedated patients.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Michael John Russell Jeffries <http://orcid.org/0000-0001-9707-8936>

Emmanuel Akinluyi <http://orcid.org/0000-0002-4215-5559>

REFERENCES

- 1 WHO. Clinical management clinical management living guidance COVID-19. *World Heal Organ* 2021;16–44.
- 2 Greenhalgh T, Knight M, Inda-Kim M, *et al*. Remote management of covid-19 using home pulse oximetry and virtual ward support. *BMJ* 2021;372:n677.
- 3 Jouffroy R, Jost D, Prunet B. Prehospital pulse oximetry: a red flag for early detection of silent hypoxemia in COVID-19 patients. *Crit Care* 2020;24:4–5.
- 4 Jumreornvong O, Yang E, Race J, *et al*. Telemedicine and medical education in the age of COVID-19. *Acad Med* 2020;95:1838–43.
- 5 Mukerji A, Shafey A, Jain A, *et al*. Pulse oximetry screening for critical congenital heart defects in Ontario, Canada: a cost-effectiveness analysis. *Can J Public Health* 2020;111:804–11.
- 6 NHS England & NHS Improvement. Novel coronavirus (COVID-19) standard operating procedure: COVID virtual ward, 2021. Available: <https://www.england.nhs.uk/coronavirus/publication/covid-virtual-ward/>
- 7 NHS England & NHS Improvement. Novel coronavirus (COVID-19) standard operating procedure: COVID Oximetry @home, 2020. Available: <https://www.england.nhs.uk/coronavirus/publication/novel-coronavirus-covid-19-standard-operating-procedure-covid-oximetry-home/>
- 8 Parsons J. £20 pulse oximeter gadget could save lives from Covid-19. *METRO*, 2021. Available: <https://metro.co.uk/2021/01/22/20-pulse-oximeter-gadget-could-save-lives-from-covid-19-13948612/>
- 9 Gallagher J. Covid: how a £20 gadget could save lives, 2021. Available: <https://www.bbc.co.uk/news/health-55733527>
- 10 International Organization for Standardization. ISO 80601-2-61:2019 Medical electrical equipment - Particular requirements for basic safety and essential performance of pulse oximeter equipment 2019.
- 11 Lipnick MS, Feiner JR, Au P, *et al*. The accuracy of 6 inexpensive pulse oximeters not cleared by the food and drug administration: the possible global public health implications. *Anesth Analg* 2016;123:338–45.
- 12 Smith RN, Hofmeyr R. Perioperative comparison of the agreement between a portable fingertip pulse oximeter V. a conventional bedside pulse oximeter in adult patients (COMFORT trial). *S Afr Med J* 2019;109:154–8.
- 13 Hudson AJ, Benjamin J, Jardeleza T, *et al*. Clinical interpretation of peripheral pulse oximeters labeled “not for medical use”. *Ann Fam Med* 2018;16:552–4.
- 14 Ross EM, Matteucci MJ, Shepherd M, *et al*. Measuring arterial oxygenation in a high altitude field environment: comparing portable pulse oximetry with blood gas analysis. *Wilderness Environ Med* 2013;24:112–7.
- 15 Torre-Bouscoulet L, Chávez-Plascencia E, Vázquez-García JC. Precision and accuracy of a ‘pocket’ pulse oximeter in Mexico City. *Rev Investig Clin* 2006;58:28–33.
- 16 Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology* 2005;102:715–9.
- 17 Feiner JR, Severinghaus JW, Bickler PE. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. *Anesth Analg* 2007;105:S18–23.
- 18 Sjoding MW, Dickson RP, Iwashyna TJ. Correspondence: racial bias in pulse oximetry measurement. *N Engl J Med* 2020;383:2477–8.
- 19 Ries AL, Prewitt LM, Johnson JJ. Skin color and ear oximetry. *Chest* 1989;96:287–90.
- 20 Adler JN, Hughes LA, Vtvilecchia R, *et al*. Effect of skin pigmentation on pulse oximetry accuracy in the emergency department. *Acad Emerg Med* 1998;5:965–70.
- 21 Foglia EE, Whyte RK, Chaudhary A, *et al*. The effect of skin pigmentation on the accuracy of pulse oximetry in infants with hypoxemia. *J Pediatr* 2017;182:375–7.
- 22 Hutton P, Clutton-Brock T. The benefits and pitfalls of pulse oximetry. *BMJ* 1993;307:457–8.
- 23 Vegfors M, Lindberg LG CL. The influence of changes in blood flow on the accuracy of pulse oximetry in humans. *Anaesthesiol Scand* 1992.
- 24 Wilson BJ, Cowan HJ, Lord JA, *et al*. The accuracy of pulse oximetry in emergency department patients with severe sepsis and septic shock: a retrospective cohort study. *BMC Emerg Med* 2010;10:2–7.
- 25 Nematswerani N, Collie S, Chen T, *et al*. The impact of routine pulse oximetry use on outcomes in COVID-19-infected patients at increased risk of severe disease: a retrospective cohort analysis. *SSRN Journal* 2021.
- 26 Unal I. Defining an optimal cut-point value in ROC analysis: an alternative approach. *Comput Math Methods Med* 2017;2017:1–14.
- 27 Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561–77.
- 28 Prower E, Grant D, Bisquera A, *et al*. The ROX index has greater predictive validity than NEWS2 for deterioration in Covid-19. *EclinicalMedicine* 2021;35:100828.

SUPPLEMENTAL MATERIAL

Population s_pO_2 distribution

Any attempt to generalise the results shown in Table 4 should account for variation in population characteristics, particularly s_pO_2 distribution. The s_pO_2 distribution of the study population is therefore presented in Figure 2.

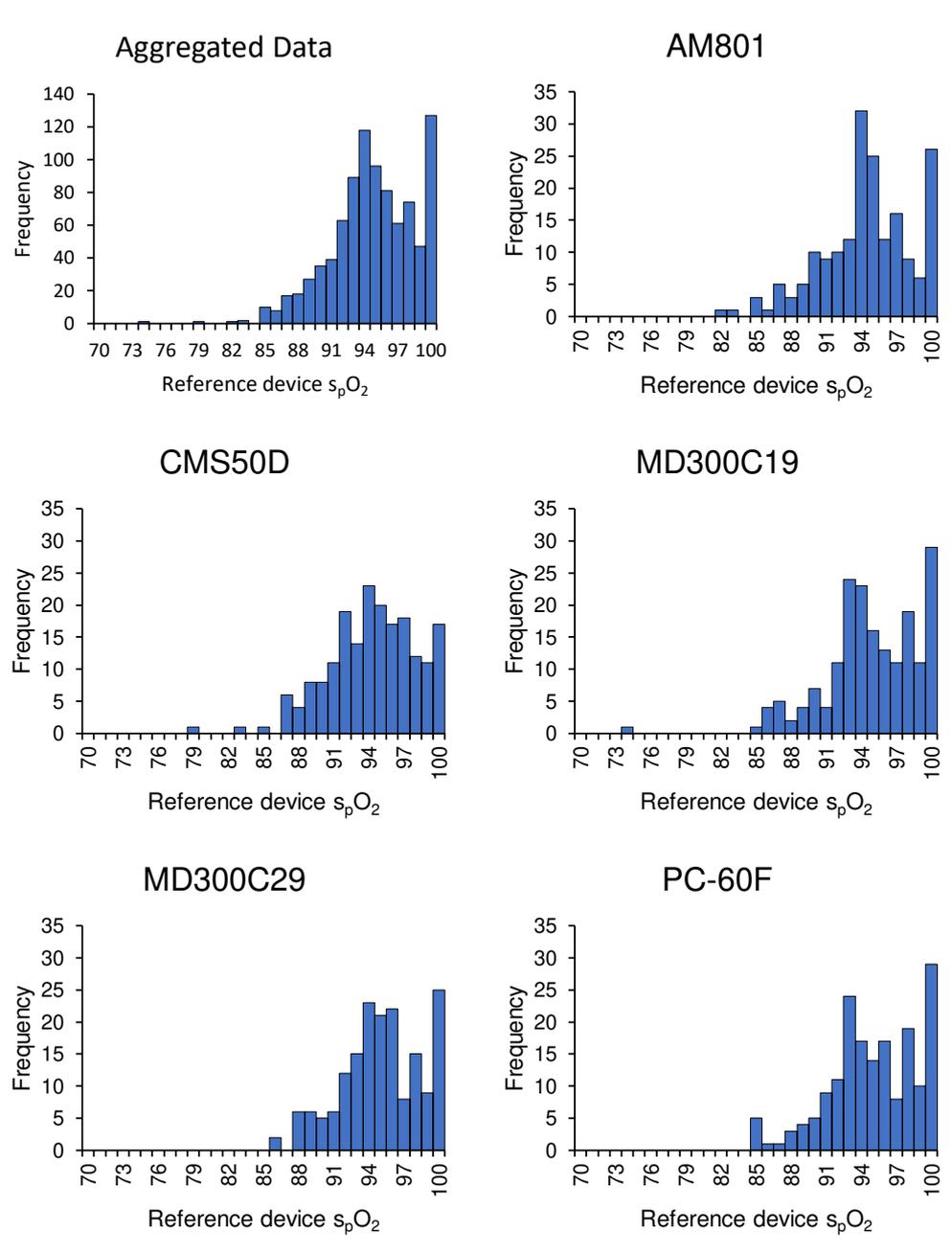


Figure 2 - The population distribution of s_pO_2 for the aggregated data set (top left) and for each PPO considered in isolation. These plots show the distribution of s_pO_2 as measured using the reference pulse oximeter.

Bland-Altman Plots

Bland-Altman plots are a standard visualisation of the agreement between two diagnostic measurements. These plots, for each of the five investigated PPOs, are shown in Figure 3.

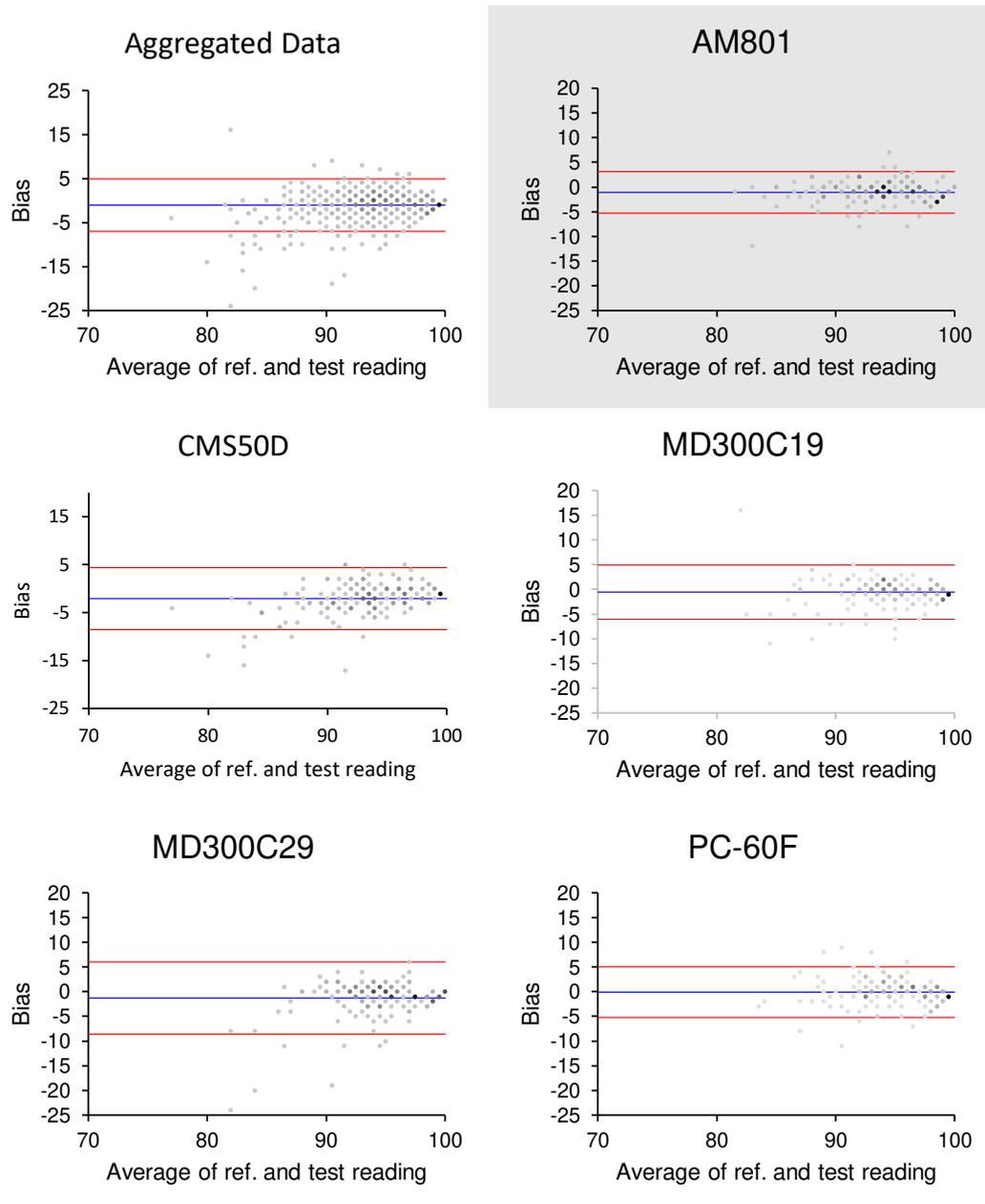


Figure 3 – Bland-Altman plots showing the agreement between the PPO devices and the reference device. The vertical axis shows PPO bias (the difference between the PPO reading and the reference device reading). The blue line shows the mean bias, the red lines show the ± 1.96 standard deviation points, such that 95% of the points fall between these lines. Pulse oximeters give measurements to a precision of 0 decimal places, this means that many points are overlain in plots like these ones. Greyscale is used to represent the number of overlain points, with darker markers where many points are overlain, and lighter markers where fewer points are overlain.

Results for Individual PPOs

The multiple regression analysis results for the individual PPOs are give in the tables below.

AM801

	Bias		False Positive		False Negative	
	Sig.	Coefficient	Sig.	Odds Ratio	Sig.	Odds Ratio
Gender	0.551	-0.211 (-0.909 to 0.487)	0.566	1.501 (0.375 to 6.007)		
Age	0.158	-0.015 (-0.035 to 0.006)	0.080	1.051 (0.994 to 1.111)	0.416	1.063 (0.918 to 1.230)
Smoking status	0.513	0.330 (-0.665 to 1.325)	0.838	1.218 (0.183 to 8.101)	0.469	0.023 (0.000 to 601.4)
PAD status	0.127	-0.705 (-1.612 to 0.202)	0.631	1.443 (0.323 to 6.446)	0.361	141.7 (0.003 to 5.90m)
FSP	0.052	-0.294 (-0.591 to 0.002)	0.102	1.926 (0.879 to 4.220)	0.112	0.126 (0.010 to 1.622)
Ref. s_pO_2	0.014	-0.111 (-0.198 to -0.023)	<0.001	0.369 (0.210 to 0.650)	0.045	22.05 (1.068 to 455.1)

CMS50D

	Bias		False Positive		False Negative	
	Sig.	Coefficient	Sig.	Odds Ratio	Sig.	Odds Ratio
Gender	0.545	-0.0334 (-1.420 to 0.752)	0.585	1.325 (0.482 to 3.647)	0.141	53.67 (0.268 to 10.7k)
Age	0.093	-0.027 (-0.060 to 0.005)	0.830	0.996 (0.964 to 1.030)	0.305	1.041 (0.964 to 1.124)
Smoking status	0.699	0.283 (-1.158 to 1.724)	0.517	1.588 (0.391 to 6.443)	0.270	0.049 (0.000 to 10.46)
PAD status	0.098	1.174 (-0.220 to 2.568)	0.290	0.456 (0.106 to 1.952)	1.000	597m undefined
FSP	0.109	-0.343 (-0.763 to 0.077)	0.332	0.768 (0.451 to 1.308)	0.100	0.134 (0.012 to 1.466)
Ref. s_pO_2	0.319	0.068 (-0.066 to 0.202)	<0.001	0.616 (0.485 to 0.783)	0.123	10.46 (0.531 to 205.9)

MD300C19

	Bias		False Positive		False Negative	
	Sig.	Coefficient	Sig.	Odds Ratio	Sig.	Odds Ratio
Gender	0.129	-0.652 (-1.497 to 0.192)	0.249	0.449 (0.115 to 1.750)	0.999	230m undefined
Age	0.316	-0.013 (-0.037 to 0.012)	0.870	0.996 (0.948 to 1.046)	1.000	0.765 undefined
Smoking status	0.018	1.451 (0.255 to 2.646)	0.669	0.621 (0.070 to 5.529)	1.000	6.902 undefined
PAD status	0.268	-0.607 (-1.686 to 0.472)	0.152	3.293 (0.646 to 16.79)	1.000	3.94m undefined
FSP	0.020	-0.428 (-0.788 to -0.068)	0.333	1.351 (0.734 to 2.487)	0.998	0.000 undefined
Ref. s_pO_2	<0.001	-0.255 (-0.355 to -0.156)	0.039	0.715 (0.519 to 0.983)	0.996	207m undefined

MD300C29

	Bias		False Positive		False Negative	
	Sig.	Coefficient	Sig.	Odds Ratio	Sig.	Odds Ratio
Gender	0.403	0.518 (0.704 to 1.741)	0.538	0.689 (0.211 to 2.253)	0.267	0.000 (0.000 to 463.8)
Age	0.976	0.001 (-0.035 to 0.036)	0.633	0.990 (0.952 to 1.031)		
Smoking status	0.083	1.502 (-0.201 to 3.206)	0.998	0.998 (0.166 to 5.986)	0.999	0.000 undefined
PAD status	0.011	-2.010 (-3.456 to -0.474)	0.375	1.895 (0.462 to 7.771)	1.000	33.6m undefined
FSP	0.001	-0.886 (-1.412 to -0.360)	0.215	1.385 (0.828 to 2.315)	0.442	5.168 (0.078 to 341.0)
Ref. s_pO_2	0.054	-0.167 (-0.337 to 0.003)	0.405	0.901 (0.705 to 1.151)	0.299	987.0 (0.002 to 447m)

PC-60F

	Bias		False Positive		False Negative	
	Sig.	Coefficient	Sig.	Odds Ratio	Sig.	Odds Ratio
Gender	0.332	0.394 (-0.406 to 1.193)	0.949	1.060 (0.180 to 6.221)	0.510	2.942 (0.119 to 73.03)
Age	0.196	-0.016 (-0.041 to 0.008)	0.052	1.082 (0.999 to 1.172)	0.262	1.067 (0.953 to 1.196)
Smoking status	0.176	-0.770 (-1.889 to 0.348)	0.728	1.499 (0.154 to 14.63)	0.999	0.000 undefined
PAD status	0.028	1.143 (0.126 to 2.160)	0.152	0.227 (0.030 to 1.725)	1.000	0.240 undefined
FSP	0.254	-0.210 (-0.571 to 0.152)	0.151	1.957 (0.783 to 4.892)	0.359	0.302 (0.023 to 3.894)
Ref. s_pO_2	<0.001	-0.318 (-0.418 to -0.217)	0.043	0.666 (0.449 to 0.987)	0.084	1.588 (0.940 to 2.681)

FIGURE LEGENDS IN SUPPLEMENTAL MATERIAL

Figure S1 – The frequency distribution across a range of relevant reference device s_pO_2 for i) data pooled from all test oximeters (top left), ii) data for each of the individual test oximeters (other plots)

Figure S2 – Bland-Altman plots are a standard visualisation of the agreement between two diagnostic measurements. The vertical axes show bias (difference between test oximeters and reference oximeter measurements), and the horizontal axes show the average of two measures. The blue lines denote the mean bias, the red lines show ± 1.96 standard deviations from the mean, such that 95% of the points fall between these lines. Pulse oximeter measurements have a precision of 0 decimal places, so many points are overlain in these plots. Greyscale intensity is used to represent the number of overlain points, darker markers show many overlain points, and lighter markers show fewer overlain points. The top left plot shows the results where data is pooled for all test oximeters. Data for each of the individual test oximeters are shown in the other plots.

Table S3 – Five tables summarising the effect of six different subject characteristics on test device performance in terms of i) measurement bias evaluated by multiple regression analysis (column 1), ii) false positive evaluated by logistic regression analysis (column 2) and iii) false negative logistic regression analysis (column 3). The two sub-columns under “Bias” present the p-values (left) and coefficients (right) from each regression analysis. The sub-columns under “false positives” and “false negatives” show the p-value (left) and odds ratio (right) from the binary logistic regression analyses. Significant values are highlighted in bold and the magnitude of coefficient indicates the size of the effect. The p-values that are statistically significant are highlighted in bold) and the 95% confidence intervals are shown in parentheses below the coefficient and odds ratio values.