Anaesthesia, 1991, Volume 46, pages 202–206

APPARATUS

Potential errors in pulse oximetry

I. Pulse oximeter evaluation*

A. C. RALSTON, R. K. WEBB AND W. B. RUNCIMAN

Summary

There is no absolute reference for oxygen saturation, although multiwavelength in vitro oximeters are accepted as the 'gold standard'. Regardless of whether fractional or functional saturation is used by manufacturers to calibrate their oximeters, evaluation against fractional saturation is recommended since this is the clinically relevant variable. The use of standard notation and comparisons based on bias and precision is recommended. The accuracy of pulse oximetry is intrinsically limited by the use of only two wavelengths, and is dependent on the initial calibration population. The empirical algorithms used to convert the signal to its 'readout value' and the quality control of hardware may both be important sources of variability between oximeters. Change in blood temperature may introduce errors in pulse oximeter and in vitro oximeter saturation readings, but these will be clinically insignificant. Changes in blood pH should not decrease pulse oximetry accuracy.

Key words

Equipment; pulse oximeter.
Measurement.

Many recent papers have dealt with the accuracy of pulse oximetry and comparisons of different brands under various conditions, but there are many potential sources of error that need to be considered for the evaluation and appropriate use of pulse oximeters. There is widespread confusion about what pulse oximeters actually measure and which reference oximeter to use. The use of different notations and statistical methods have made comparisons between studies difficult.

This paper is one of a series: two companion papers (which compare and rank, under conditions of poor perfusion, 20 pulse oximeters with finger probes)[1] and 10 with ear, nose or forehead probes (in press). The series aims to review all the important sources of error in pulse oximetry and to discuss their clinical significance. Other papers consider the effects of changes in saturation and signal quality; the effects of dyshaemoglobins, dyes, other pigments and extraneous factors; and present a clinical overview of the significance of all the issues discussed.

Theory of pulse oximetry

Two wavelengths of light are transmitted from light emitting diode (LED) sources and pass through an arterial bed, usually in a finger or earlobe. The attenuated light is received by a detector and converted into an electrical signal which is analysed by a microprocessor to give pulse rate and oxygen saturation readings. The two wavelengths used are differentially absorbed by oxygenated haemoglobin (HbO_2) and de-oxygenated, or reduced, haemoglobin (Hb). Assuming there are no other major haemoglobin species, the amount that one wavelength is attenuated compared to the other gives the fraction of haemoglobin saturated by oxygen $[HbO_2]/([HbO_2]+[Hb])$, where the square brackets denote concentration.

Both wavelengths are also absorbed by venous blood and tissues. The attenuation of the light is analysed over a full pulse beat to make the saturation measurement independent of these factors. The total absorption of the light has a constant component from the tissue and from steadily flowing venous blood, and a changing component as a result of pulsation of arterial blood. The constant component is subtracted from the total, so that the net absorption of each wavelength can be attributed to arterial blood only. The two LEDs are cycled on and off 480 times per second (for a mains power frequency of 60 Hz) or 400 times per second (for a mains power frequency of 50 Hz), with only one being on at a time.[2] This enables a single detector to be used to sample first one wavelength and then the other. The detector measures the background level of ambient light after the two LED 'on' periods. This is subtracted from the transmitted LED signals so that

A.C. Ralston, BSc, BAppSc, Medical Physicist, R. K. Webb, MB, BS, FFARACS, Senior Staff Specialist, W.B. Runciman, BSc(Med), MB, BCh, FFARACS, PhD, Professor and Head, Department of Anaesthesia and Intensive Care, Royal Adelaide Hospital, North Terrace, Adelaide, South Australia, Australia 5000.
Correspondence should be addressed to Professor W.B. Runciman please.
*Combined study with Australian Patient Safety Foundation, GPO Box 400, Adelaide, South Australia 5001.
Accepted 15 July 1990

0003–2409/91/030202+05 $03.00/0

changing background levels are not included in the signal absorption calculations. This method can fail when the ambient light is almost in phase with the LED cycle, and can lead to erroneous saturation values. The pulse oximeter detector cycles run at a multiple of the local mains power frequency to reduce this interference.

The saturation values that are displayed are not instantaneous but are averages taken over 3 to 10 seconds to help reduce the effect of pressure wave variations due to motion of the subject. To reject fleeting high or low readings, weighting is used by some manufacturers to adjust the contribution of values according to how close they are to the previous average.

These methods are explained in more detail in review articles by Tremper and Barker[3] and Wukitsch *et al.*[2]

Reference oximeters

There must be a direct oxygen saturation measurement for comparison to determine experimentally pulse oximeter accuracy. The 'gold standard' for this is generally accepted to be a multiwavelength *in vitro* oximeter. This instrument transmits light of various wavelengths across a sample of blood. The principle is the same as for pulse oximeters except that the attenuation of the beam from the entire blood sample is measured, in contrast to pulse oximeters which analyse only the changing blood volume within a volume of tissue.

Examples of commercially available oximeters are the Instrumentation Laboratory IL 282 and 482 Co-oximeters (Lexington, MA, USA), and the Radiometer OSM-2 and OSM-3 Hemoximeters (Copenhagen, Denmark). These instruments actually measure slightly different quantities in a blood sample, and there is debate which is more appropriate to use as a reference (see below).

The Radiometer OSM-2 uses two wavelengths of light to determine the percentage of HbO_2 relative to the amount of Hb, and ideally calculates $[HbO_2]/([HbO_2]+[Hb])$. This saturation is known as the functional saturation, but blood, even from 'normal' volunteers, usually contains small amounts of other haemoglobin species, such as carboxyhaemoglobin (HbCO) and methaemoglobin (MetHb), and these affect the absorption of both wavelengths. The saturation measured by the Radiometer is true *functional* saturation only if there is no HbCO or MetHb in the blood sample. If blood is drawn into capillary tubes coated with sodium dithionite to reduce the haemoglobin, the HbCO content can also be measured in a second analysis.

The IL 482 and 282 models[4] use four wavelengths of light (535.0, 585.2, 594.5, 626.6 nm) to detect the concentrations of HbO_2, HbCO, Hb and MetHb, and give the oxygen saturation as a percentage of the total haemoglobin, i.e. as a percentage of the sum of the four species. This saturation is known as the *fractional* saturation, which is displayed by the Radiometer OSM-3 using six wavelengths. Pulse oximeters are empirically calibrated by the manufacturer against an *in vitro* oximeter. It is therefore important to examine the realistic accuracy limits of *in vitro* oximeters.

The IL 482 operator's manual[4] claims an accuracy of 1% (95% confidence limits) and precision of 0.5% for HbO_2 measurements for samples with O–10% MetHb and a haemolysed pH of 7.0–7.4. Outside these MetHb and pH ranges the pH sensitivity of methaemoglobin can cause significant changes in absorption at all four wavelengths. Accuracy is also compromised by the presence of high lipid levels (which can cause scattering of light), bilirubin, intravenous dyes, sulphaemoglobin and fetal haemoglobin. (The latter four are discussed briefly elsewhere in this series.)

It is not feasible to validate the claim of 1% accuracy, since we know of no quality control sample of accurately known or measured saturation that can be used to check this. Haemoglobin in titrated blood can be fully saturated with oxygen only when there are no other haemoglobin species in the blood, and cannot be easily partially saturated for daily multipoint calibration. Nevertheless, it seems reasonable to accept the accuracy claim, given the high degree or reproducibility of results (within 1%) and the standing of the manufacturers in this area.

Notation and performance criteria

Hundreds of letters and papers are published on the subject of pulse oximetry, and many of them use quite different notation to describe measured and actual quantities. Researchers involved in clinical trials have used a variety of ways to analyse data, which makes comparisons between trials difficult without using the original data. Functional and fractional saturations as measured by a cooximeter and the approximation to functional saturation measured by a two-wavelength Hemoximeter have all been used as the reference quantity. Pulse oximeters are initially calibrated against functional or fractional saturation depending on the manufacturer.

Zijlstra and Oeseburg[5] recently published a letter on the definition and notation of saturation terms, but unfortunately have a common misconception about what pulse oximeters actually measure. They state that pulse oximeters measure functional saturation, $[HbO_2]/([HbO_2]+[Hb])$, and should only be compared to the functional saturation calculated from the HbO_2 and Hb measurement on a co-oximeter. In fact pulse oximeters only measure a ratio of transmitted red and infrared light intensities, and relate this to a 'look-up' table of empirical oxygen saturation values.[2] The values in the table depend on the manufacturer's aim of approximating functional or fractional saturation, but will in reality be neither of these unless the dyshaemoglobin levels in a subject's arterial blood are exactly the same as the average values of those used for the creation of the 'look-up' table.

The clinician needs to know the fractional saturation, since a high functional saturation can mask life-threatening situations with low fractional saturation (e.g. with high carbon monoxide levels). We have therefore used and recommend the use of fractional saturation as a reference for clinical trials. Pulse oximeters cannot measure this quantity (just as they cannot measure functional saturation), but using fractional saturation as the reference provides the clinician with a realistic assessment of the magnitude of the errors of pathophysiological significance which is likely to be found for the group of patients under consideration.

We have used the symbol Sao_2 to denote functional saturation, as first adopted by physiologists, and $\%HbO_2$ to represent fractional oxygen saturation. Many groups have adopted a wide variety of symbols which we shall not attempt to catalogue here, except to illustrate this with the use of Sao_2 and $FSao_2$ by Barker *et al.*[6] to denote fractional

and functional saturation respectively, compared to the So_2 and $FHbO_2$ used by Zijlstra and Oeseburg[5] to denote functional and fractional respectively.

We agree with Zijlstra and Oeseburg[5] that the use of the symbol Spo_2 (with the 'p' being for pulse oximetry) is not ideal, since it implies the existence of a new quantity measurable with pulse oximetry, when in fact pulse oximetry aims to approximate functional or fractional saturation. However, the use of the term Spo_2 to denote saturation as indicated by a pulse oximeter is almost universal and we will continue to use it to avoid further confusion.

Comparisons

The presentation of comparative data and subsequent ranking of different brands of pulse oximeters also varies widely. The three basic methods of statistical analysis are regression analysis (correlation coefficient, slope and intercept), the limits of agreement method of Bland and Altman[7] and the use of bias and precision.

The disadvantages of regression analysis are that high correlation between two variables does not necessarily indicate good agreement even with a slope of near unity and an intercept of near zero, and the results of studies using regression analysis cannot be readily compared to one another.[3,7]

Bland and Altman devised a method of comparing two indirect methods of clinical measurement[7] and used pulse oximetry measurements and cooximeter measurements as an example. Some groups have taken up this method, but it ignores the status of the multiwavelength cooximeter as well accepted 'gold standard'. Such cooximeters can measure fractional oxygen saturation with a high degree of reproducibility and avoid many of the inherent problems of *in vivo* measurements with all their uncertainties. Hence we consider cooximetry to be not just another method of indirect measurement of questionable accuracy, but to be close enough to actual saturation for clinical purposes, except in a few specific instances, namely the presence of high levels of fetal haemoglobin, bilirubin and intravenous dyes. Thus this method, however useful in other cases, is not applicable to pulse oximetry assessment when a co-oximeter is used as the reference. It may, however, be appropriate when comparing one pulse oximeter to another.

The use of bias and precision is helpful in getting a clear picture of a pulse oximeter's performance and how this compares to other units or other studies. Bias is the average of the differences between the pulse oximeter readings and the cooximeter readings, and precision is the standard deviation of these differences around the average. These quantities may need to be given separately for different ranges of conditions, for example high or low saturation ranges.

A unit may be very precise, so that the results are highly reproducible with a low scatter, but have a high bias so that the results are not centred on the true values. Conversely a unit may be highly accurate, having a low bias, but have poor precision, with values swinging widely from side to side of the true value. Good examples of these cases are given elsewhere.[1] Which of these qualities is more important depends on the clinical application, and ranking of units should rely on subsequent weighting of these. Our companion paper on pulse oximetry accuracy under conditions of poor perfusion[1] discusses the use of various combinations of criteria in more detail.

Effects of temperature

A pulse oximeter measures the saturation of blood at a temperature in between core and peripheral temperatures, whereas commercially available *in vitro* oximeters measure haemoglobin saturation at 37.0°C. Many factors change simultaneously in the blood, including oxygen saturation when a volume of blood is sealed from air and subjected to a change in temperature.[8] Thus the saturation measured by the *in vitro* oximeter at 37°C may differ from the actual saturation at the pulse oximeter probe site, especially if the patient is cold as the result of induced hypothermia for surgery, or hyperthermic due to fever.

Siggaard-Andersen *et al.*[9] state that an oxygen saturation of 90% at 37°C will increase by about 1% when the blood is cooled to room temperature. This effect is even less for lower saturations, with 50% saturation increasing to about 50.5% for the same decrease in temperature. Thus the differences in saturation between blood at 37°C, core body temperature and peripheral temperature are clinically negligible and need not be corrected for.

Siggaard-Andersen *et al.*[9] also showed that a change in temperature shifts the absorption spectra of HbO_2 and Hb, causing a significant change in HbO_2 extinction coefficient in going from 37°C to 17°C, with a lesser effect on Hb. Values are only given for wavelengths up to 650 nm, so it is not possible to estimate from these data what effect this might have on the accuracy of oxygen saturation measurements made by a pulse oximeter. However, in our intensive care unit[10] and recovery room[1] studies there was no correlation between core temperature, skin temperature or core-skin temperature difference and the differences between Spo_2 readings and IL 482 Co-oximeter $\%HbO_2$ measurements made at 37°C, suggesting that systematic errors from this source are not of clinical significance.

In summary, if a patient is at a temperature other than 37°C this may affect the accuracy of pulse oximetry by shifting the absorption spectra of the haemoglobin species, but clinical data has shown no evidence of this. *In vitro* oximeters have no temperature compensation but the effect of this on *in vitro* measurements is less than 1% error. It is thus reasonable to conclude that overall errors resulting from high or low body temperatures are unlikely to be clinically significant.

Dissociation curves

Oxyhaemoglobin dissociation curves are frequently given in physiology texts and research papers on oximetry. They present a series of curves for varying conditions. For a given Po_2 one can directly read off the graph what the saturation value is for a given temperature, pH, or other variable. However, it is difficult to determine from these graphs what the effect of changing conditions will be on the saturation in a sealed system, such as a blood vessel or syringe. In this case the Po_2 is not kept constant, and will change in response to any change in saturation.

A decrease in temperature for example, from 37°C to 25°C shifts the dissociation curve to the left and raises the oxygen saturation. The increase in oxygen saturation, or

oxygen attached to haemoglobin, means that oxygen is taken out of solution and the Po_2 is lowered. The lower Po_2 opposes the increase in saturation, and an equilibrium point is found on the 25°C dissociation curve. The resultant change in saturation is less than it would be if the Po_2 was fixed.

Using the example given in the previous section on the effects of temperature, blood in a sealed system at body temperature and with a saturation of 50% will increase its saturation to 50.5% when cooled to room temperature.[9] If the Po_2 of this blood was fixed at about 30 mmHg, for the same decrease in temperature the saturation would increase from 50% to approximately 85%.[8] Hence these graphs can only indicate the direction but not the magnitude of the effect of changing variables *in vivo* or in sealed blood samples.

Effects of pH

A change in the pH of arterial blood shifts the oxygen dissociation curve and thereby changes $\%HbO_2$,[8] but the extinction coefficients of the haemoglobin species are not altered and so the accuracy of an Spo_2 reading is not affected.

In vitro oximeters break open the red blood cells by lysis before measuring saturation, and thereby cause a mixing of the intra- and extracellular fluids, the pH values of which differ by 0.2.[11] Any change in pH will result in new $\%HbO_2$ and Po_2 values. Thus the $\%HbO_2$ of lysed blood measured *in vitro* is probably slightly different from that of the arterial blood measured by a pulse oximeter *in vivo*. However, pulse oximeters are calibrated against *in vitro* oximeters and so the change in saturation caused by lysing is included in pulse oximeter 'look up' table conversion factors.

Probe reliability

What is quoted as being a single wavelength for an LED, such as the ones used in pulse oximeter probes, is actually the peak wavelength of a narrow range. Ninety-five percent of the LED outputs falls within a range of about ±30–40 nm. The absorption coefficients of the haemoglobin species within this range vary somewhat, but since oximeters are calibrated empirically this variation should not matter unless the wavelength spectrum changes after calibration.

The peak wavelength shifts with a change in ambient temperature, increasing by about 0.1 nm for every degree Celsius above 25°C. Thus a change in ambient temperature, which is approximately skin temperature, from 25°C to 35°C would shift the peak wavelength about 1 nm, causing no significant change in absorption coefficients. Severinghaus[12] mentions that age and the increase in LED intensity in response to low light transmission that some pulse oximeters employ can shift the peak wavelength, but does not say by how much.

A much greater variation in peak wavelength of up to 15 nm is caused by differences in manufacture between diodes.[13] Pulse oximeter manufacturers either reject all diodes outside the acceptable wavelength range or build into the pulse oximeter a device to correct for different wavelength output.[3]

Reusable probes eventually need replacement, usually because of damage to wiring. We have seen no confirmed reports of saturation errors from this, except a case of false desaturation when a disposable probe had been re-used several times.[14,15]

There is a need for portable test equipment that could be used regularly to confirm the integrity of a pulse oximeter and its probe. Munley *et al.*[16] have designed and built such a test object which comprises a hollow dummy finger and a rotating core with a thin slab of pigment. This, surprisingly, causes quite different saturation values for dummy fingers of different pigment concentrations. These authors found a range of Spo_2 readings from 93.5% to 34% for the same core placed in different dummy fingers. The static dummy finger should in theory have no effect on the the Spo_2, which is dependent only on the rotating pigmented core. The Spo_2, in practice is apparently highly dependent on the pigmentation of the dummy finger, which may be related to the similarly unexpected effects of nail polish. Whether this is a problem of electronics or software, it is a paradox that needs to be addressed by manufacturers if errors due to the light absorption of static pigmentation are to be eliminated.

Pulse oximeters are in general robust enough for frequent transportation and many have self checks of the electronics and software. The largest source of error in probe manufacture is the variation between LEDs, which is determined by the quality control checks by the manufacturer, and may be a point to consider when brands are compared before purchase on the basis of price.

Signal processing

Clinical studies show large differences in the bias and precision of the many brands of pulse oximeter,[1,3,10] even though they use similar or identical hardware. These differences are almost certainly due to the different algorithms used in processing the measured light intensity signals to obtain a final saturation value. Most algorithms and calibration 'look-up' tables are confidential, so the smaller and newer companies have to start more or less from scratch in their software design. The better models have sophisticated methods which reject noise and motion artefacts, which take frequent running averages of saturation throughout a pulse pressure wave that are carefully weighted to avoid showing sudden changes,[2] and which take many factors into account to determine when to indicate no reading if the signal quality is too poor. These differences between brands will eventually disappear as companies use only proven optimum algorithms, and the less sophisticated models drop out of the market.

The calibration of pulse oximeters is empirical, and thus relies on the use of a suitable population for building a 'look-up' table that gives the same answers as *in vitro* saturation measurements. Whether functional or fractional saturation is chosen as the reference parameter, the levels of dyshaemoglobins and other red and infrared absorbers in the calibration subjects is crucial to the accuracy of the machine on patients. For example, using healthy, non-smoking volunteers with negligible dyshaemoglobin levels for the calibration means that a pulse oximeter is less accurate when used on day surgery patients who had their last cigarettes over breakfast. It is clearly impossible to match the dyshaemoglobin levels of the calibration and patient populations exactly, which is an intrinsic limitation to the accuracy of pulse oximetry in its present form. The

obvious way to avoid this problem is to use more than two wavelengths in the pulse oximeter probe, and thus enable the accurate detection of more than two light absorbing components in the blood.

Conclusions

Assessment of accuracy is best obtained by comparison of pulse oximetry readings with *in vitro* oximeter measurements, although these are based on the same spectrophotometric principles as pulse oximeters and are prone to some of the same errors. These are discussed in a later paper in this series. Some researchers compare pulse oximetry readings to functional saturation, others to fractional saturation. Some pulse oximeters are empirically calibrated against functional saturation, others against fractional saturation by using volunteers with average dyshaemoglobin levels. A pulse oximeter can consistently provide a good measure of either of these only if the blood of the subject has the same dyshaemoglobin levels as the average levels of the population group used for creating a calibration 'look-up' table. Fractional saturation is the more useful information for assessing the available oxygen in a patient's arterial blood, and is the quantity we recommend as the reference in clinical trials of pulse oximeters. Ranking of pulse oximeters should be based on the weight put on bias and precision, which may differ for different clinical applications. However, high precision with a moderate bias would seem more desirable than low bias and moderate precision in the context in which pulse oximetry is usually used.

The effect of temperature differences between the pulse oximeter site and the reference *in vitro* oximeter, and the effect of a change in blood temperature on the absorbance spectra of Hb and HbO_2, are likely to be clinically insignificant.

The pH, and hence saturation, of a blood sample may change with the lysis process required for an *in vitro* measurement, but this is taken into account by the initial calibration of a pulse oximeter against these measurements. A change in blood pH does not affect pulse oximeter accuracy.

Pulse oximeter probes are prone to the same wear and tear and variation in manufacturing standards as all monitoring devices. Intrinsic problems stem from the shift in the output spectra of the LEDs with a change in temperature, the inability of the two wavelength system to provide accurate measurements of oxygen saturation in the presence of more than two red or infrared absorbers, and the mismatch of the dyshaemoglobin levels of the calibration and patient populations.

Wavelength reproducibility is an area that will be improved by future technical advances. Similarly, noise and motion artefact rejection algorithms will improve as more complex signal processing techniques can be stored on microchips. The striking differences between the performance of different pulse oximeters will probably decrease as companies adopt the progressively improved signal processing methods, but the problems of absolute calibration in the presence of dyshaemoglobins and other pulsatile absorbers can only be achieved by the use of more than two wavelengths.

References

1. Clayton DG, Webb RK, Ralston AC, Duthie D, Runciman WB. A comparison of the performance of twenty pulse oximeters under conditions of poor perfusion. *Anaesthesia* 1991; **46:** 3–10.
2. Wukitsch MW, Petterson MT, Tobler DR, Pologe JA. Pulse oximetry: analysis of theory, technology, and practice. *Journal of Clinical Monitoring* 1988; **4:** 290–301.
3. Tremper KK, Barker SJ. Pulse oximetry. *Anesthesiology* 1989; **70:** 98–108.
4. Instrumentation Laboratory IL 482 Operator's Manual, Lexington, MA.
5. Zijlstra WG, Oeseburg B. Definition and notation of hemoglobin oxygen saturation. *IEEE Transactions on Biomedical Engineering* 1989; **36:** 872.
6. Barker SJ, Tremper KK, Hyatt J. Effects of methemoglobinemia on pulse oximetry and mixed venous oximetry. *Anesthesiology* 1989; **70:** 112–7.
7. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1:** 307–10.
8. Roughton FJW. Transport of oxygen and carbon dioxide. In: *Handbook of physiology*, Vol. 3: *Respiration*. Washington, D.C.: *American Physiological Society* 1964; **3:** 767–825.
9. Siggaard-Andersen O, Norgaard-Pedersen B, Rem J. Hemoglobin pigments. Spectrophotometric determination of oxy-, carboxy-, met-, and sulfhemoglobin in capillary blood. *Clinica Chimica Acta* 1972; **42:** 85–100.
10. Emergency Care Research Institute. Pulse oximeters. *Health Devices* 1989; **18:** 185–230.
11. Lambertsen CJ. Transport of oxygen and carbon dioxide by the blood. In: *Medical physiology*, Vol. 2. Saint Louis: C. V. Mosby, 1974: 1399–422.
12. Severinghaus JW. Pulse oximetry uses and limitations. ASA 40th Annual Refresher Course Lectures and Clinical Update Program. American Society of Anesthesiologists Inc. 1989.
13. Pologe JA. Pulse oximetry: technical aspects of machine design. *International Anesthesiology Clinics* 1987; **25:** 137–53.
14. Racys V, Nahrwold ML. Reusing the Nellcor pulse oximeter probe: is it worth the savings? *Anesthesiology* 1987; **66:** 713.
15. Kao YJ, Badgwell JM. Reusing the Nellcor pulse probe: clarification. *Anesthesiology* 1987; **67:** 866.
16. Munley AJ, Sik MJ, Shaw A. A test object for assessing pulse oximeters. *Lancet I* 1989: 1048–9.