Differential Pulse Oximeter Accuracy, Occult Hypoxemia Prevalence, and Clinical Outcomes by Patient Race/Ethnicity

August 2023

U.S. Department of Veterans Affairs
Veterans Health Administration
Health Services Research & Development Service

Recommended citation: Parr NJ, Beech EH, Young S. Differential Pulse Oximeter Accuracy, Occult Hypoxemia Prevalence, and Clinical Outcomes by Patient Race/Ethnicity: A Systematic Review. Washington, DC: Evidence Synthesis Program, Health Services Research and Development Service, Office of Research and Development, Department of Veterans Affairs. VA ESP Project #09-199; 2023.

Appendix

SEARCH STRATEGIES

SYSTEMATIC REVIEWS

Search Date: 2/23/2023		Search Statement	Results
MEDLINE **Origination to 2/23/2023**	1	exp Oximetry/ or exp Photoplethysmography/ or (oximet* or PPG or photoplethysmography or ((saturation adj1 (oxygen or oxyhemoglobin or oxyhaemoglobin)) or (blood adj1 (gas or oxygen or arterial)) or PtcO2 or TcPCO2 or SpO2 or %SpO2 or Sp O2)).ti,ab,kf.	126417
	2	Skin Pigmentation/ or exp Ethnicity/ or Minority Groups/ or (((cutaneous or skin) adj3 (color* or colour* or pigment* or tone* or type)) or race or racial or ethnic* or ((minorit* or biracial or multiracial or african* or afro-american* or asian* or asiatic or black* or caucasian* or hispanic* or indian* or indigenous or latin* or native* or nonwhite* or (pacific adj3 islander*) or white) adj2 (patient* or participant* or adult*)) or ((persons or people) adj3 (color or colour))).ti,ab,kf.	433474
	3	((over?estimat* or under?estimat* or detect* or classify* or rule in or rule out or identif*) adj3 (hypox* or occult)).ti,ab,kf.	7296
	4	(accura* or inaccura* or agreement or precision or evaluat* or reliab* or bias or concordance or performance or validat* or error* or fail* or discrepanc* or disparit* or vary or varied or varies or variab* or sensitivity or specificity or over?estimat* or under?estimat* or detect* or classify* or rule in or rule out or identif*).ti,ab,kf.	13177197
	5	exp Pediatrics/ or (pediatric* or paediatric* or neonat* or infant* or hyperbaric or wearable).ti,ab,kw.	1150414
	6	2 or 3	1150414
	7	1 AND 4 AND 6	1362
	8	7 not 5	1119
	9	(systematic review.ti. or meta-analysis.pt. or meta-analysis.ti. or systematic literature review.ti. or this systematic review.tw. or pooling project.tw. or (systematic review.ti,ab. and review.pt.) or meta synthesis.ti. or meta-analy*.ti. or integrative review.tw. or integrative research review.tw. or rapid review.tw. or umbrella review.tw. or consensus development conference.pt. or practice guideline.pt. or drug class reviews.ti. or ("cochrane database of systematic reviews" or acp journal club or health technology assessment winchester england or evidence report technology assessment summary or "jbi database of systematic reviews and implementation reports").jn. or (clinical guideline and management).tw. or ((evidence based.ti. or evidence-based medicine/ or best practice*.ti. or evidence synthesis.ti,ab.) and (((review.pt. or diseases category/ or behavior.mp.) and behavior mechanisms/) or therapeutics/ or evaluation studies.pt. or validation studies.pt. or guideline.pt. or pmcbook.mp.)) or (((systematic or systematically).tw. or critical.ti,ab. or study selection.tw. or ((predetermined or inclusion) and criteri*).tw. or exclusion criteri*.tw. or main outcome measures.tw. or standard of care.tw. or standards of care.tw.) and ((survey or	572025

Search Date: 2/23/2023		Search Statement	Results
		surveys).ti,ab. or overview*.tw. or review.ti,ab. or reviews.ti,ab. or search*.tw. or handsearch.tw. or analysis.ti. or critique.ti,ab. or appraisal.tw. or (reduction.tw. and (risk/ or risk.tw.) and (death or recurrence).mp.)) and ((literature or articles or publications or publication or bibliography or bibliographies or published).ti,ab. or pooled data.tw. or unpublished.tw. or citation.tw. or citations.tw. or database.ti,ab. or internet.ti,ab. or textbooks.ti,ab. or references.tw. or scales.tw. or papers.tw. or datasets.tw. or trials.ti,ab. or meta-analy*.tw. or (clinical and studies).ti,ab. or treatment outcome/ or treatment outcome.tw. or pmcbook.mp.))) not (letter or newspaper article).pt.	
	10	8 and 9	42
	11	limit 9 to english language	41
	12	limit 10 to last 7 years	26
Cochrane Database of Systematic Reviews **2005 to February 22, 2023**	1	MeSH descriptor: [Oximetry] explode all trees	1169
	2	MeSH descriptor: [Photoplethysmography] this term only	107
	3	(oximet* or PPG or photoplethysmography or ((saturation adj1 (oxygen or oxyhemoglobin or oxyhaemoglobin)) or (blood adj1 (gas or oxygen or arterial)) or PtcO2 or TcPCO2 or SpO2 or %SpO2 or Sp O2)):ti,ab,kw	12798
	4	{OR #1-#3}	12956
	5	MeSH descriptor: [Skin Pigmentation] this term only	374
	6	MeSH descriptor: [Ethnicity] explode all trees	2978
	7	MeSH descriptor: [Minority Groups] this term only	513
	8	(((cutaneous or skin) NEAR/3 (color* or colour* or pigment* or tone* or type)) or race or racial or ethnic* or ((minorit* or biracial or multiracial or african* or afro-american* or asian* or asiatic or black* or caucasian* or hispanic* or indian* or indigenous or latin* or native* or nonwhite* or (pacific adj3 islander*) or white) NEAR/2 (patient* or participant* or adult*)) or ((persons or people) NEAR/3 (color or colour))):ti,ab,kw	34578
	9	{OR #5-#7}	3640
	10	((over?estimat* or under?estimat* or detect* or classify* or rule in or rule out or identif*) NEAR/3 (hypox* or occult)):ti,ab,kw	4076
	11	(accura* or inaccura* or agreement or precision or evaluat* or reliab* or bias or concordance or performance or validat* or error* or fail* or discrepanc* or disparit* or vary or varied or varies or variab* or sensitivity or specificity or over?estimat* or under?estimat* or detect* or classify* or rule in or rule out or identif*):ti,ab,kw	994866
	12	MeSH descriptor: [Pediatrics] explode all trees	918
	13	(pediatric* or paediatric* or neonat* or infant* or hyperbaric or wearable):ti,ab,kw	118779
	14	#12 OR #13	118792
	15	#9 OR #10	7713
	16	#4 AND #11 AND #15	378
	17	#16 NOT #14	309

Search Date: 2/23/2023		Search Statement	Results
	18	limit to last 7 years	1
		Total	1
		Total after deduplication	1

PRIMARY STUDIES

Search Date: 2/23/2023		Search Statement	Results
MEDLINE **Origination to 2/23/2023**	1	exp Oximetry/ or exp Photoplethysmography/ or (oximet* or PPG or photoplethysmography or ((saturation adj1 (oxygen or oxyhemoglobin or oxyhaemoglobin)) or (blood adj1 (gas or oxygen or arterial)) or PtcO2 or TcPCO2 or SpO2 or %SpO2 or Sp O2)).ti,ab,kf.	126556
	2	exp Skin Pigmentation/ or exp Ethnicity/ or exp Minority Groups/ or (((cutaneous or skin) adj3 (color* or colour* or pigment* or tone* or type)) or race or racial or ethnic* or ((minorit* or biracial or multiracial or african* or afro-american* or asian* or asiatic or black* or caucasian* or hispanic* or indian* or indigenous or latin* or native* or nonwhite* or (pacific adj3 islander*) or white) adj2 (patient* or participant* or adult*)) or ((persons or people) adj3 (color or colour))).ti,ab,kf.	434193
	3	((over?estimat* or under?estimat* or detect* or classify* or rule in or rule out or identif*) adj3 (hypox* or occult)).ti,ab,kf.	7300
	4	(accura* or inaccura* or agreement or precision or evaluat* or reliab* or bias or concordance or performance or validat* or error* or fail* or discrepanc* or disparit* or vary or varied or varies or variab* or sensitivity or specificity or over?estimat* or under?estimat* or detect* or classify* or rule in or rule out or identif*).ti,ab,kf.	13196427
	5	exp Pediatrics/ or (pediatric* or paediatric* or neonat* or infant* or hyperbaric or wearable).ti,ab,kw.	1151790
	6	2 OR 3	441424
	7	1 AND 4 AND 6	1364
	8	7 NOT 5	1120
	9	limit 8 to English language	1070
CINAHL **Origination to 2/23/2023**	1	((MH "Oximetry+") OR (MH "Oximeters+")) OR TI (oximet* OR photoplethysmography OR PPG OR ((saturation adj1 (oxygen OR oxyhemoglobin OR oxy haemoglobin)) OR (blood adj1 (gas OR oxygen OR arterial)) OR PtcO2 OR TcPCO2 OR SpO2 OR %SpO2 OR Sp O2) OR AB (oximet* OR photoplethysmography OR PPG OR ((saturation adj1 (oxygen OR oxyhemoglobin OR oxy haemoglobin)) OR (blood adj1 (gas OR oxygen OR arterial)) OR PtcO2 OR TcPCO2 OR SpO2 OR %SpO2 OR Sp O2)	10,601
	2	((MH "Skin Pigmentation") OR (MH "Ethnic Groups+") OR (MH "Minority Groups")) OR TI (((cutaneous OR skin) adj3 (color* OR colour* OR pigment* OR tone* OR type*)) OR ethnic* OR race OR racial OR ((minorit* or biracial or multiracial or african* or afro-american* or asian* or asiatic or black* or caucasian* or hispanic* or indian* or indigenous or latin* or native* or	249,307

Search Date: 2/23/2023		Search Statement	Results
		nonwhite* or (pacific adj3 islander*) or white) adj2 (patient* or participant* or adult*)) OR ((people OR persons) adj3 (color OR colour))) OR AB (((cutaneous OR skin) adj3 (color* OR colour* OR pigment* OR tone* OR type*)) OR ethnic* OR race OR racial OR ((minorit* or biracial or multiracial or african* or afro-american* or asian* or asiatic or black* or caucasian* or hispanic* or indian* or indigenous or latin* or native* or nonwhite* or (pacific adj3 islander*) or white) adj2 (patient* or participant* or adult*)) OR ((people OR persons) adj3 (color OR colour)))	
	3	((overestimat* or over-estimat* OR underestimat* OR under-estimat* OR detect* or classify* or rule in or rule out or identif*)) adj3 (hypox* or occult))	0
	4	TI ((accura* or inaccura* or agreement or precision or evaluat* or reliab* or bias or concordance or performance or validat* or error* or fail* or discrepanc* or disparit* or vary or varied or varies or variab* or sensitivity or specificity or overestimat* or over-estimat* or underestimat* or under-estimat* or detect* or classify* or rule in or rule out or identif*)) OR AB ((accura* or inaccura* or agreement or precision or evaluat* or reliab* or bias or concordance or performance or validat* or error* or fail* or discrepanc* or disparit* or vary or varied or varies or variab* or sensitivity or specificity or overestimat* or over-estimat* or underestimat* or under-estimat* or detect* or classify* or rule in or rule out or identif*))	2487269
	5	(MH "Pediatrics+") OR TI ((pediatric* or paediatric* or neonat* or infant* or hyperbaric or wearable)) OR AB ((pediatric* or paediatric* or neonat* or infant* or hyperbaric or wearable))	349837
	6	S2 OR S3	249307
	7	S1 AND S4 AND S6	121
	8	S7 NOT S5	97
	9	limit 8 to English language	97
Scopus **Origination to 2/23/2023**	1	TITLE-ABS-KEY (oximet* OR ppg OR photoplethysmography OR (saturation W/1 (oxygen OR oxyhemoglobin OR oxyhaemoglobin)) OR (blood W/1 (gas OR oxygen OR arterial)) OR ptco2 OR tcpco2 OR spo2 OR %spo2 OR "sp o2")	249433
	2	TITLE-ABS-KEY (((cutaneous OR skin) W/3 (color* OR colour* OR pigment* OR tone* OR type)) OR race OR racial OR ethnic* OR ((minorit* OR biracial OR multiracial OR african* OR afro-american* OR asian* OR asiatic OR black* OR caucasian* OR hispanic* OR indian* OR indigenous OR latin* OR native* OR nonwhite* OR (pacific W/3 islander*) OR white) W/2 (patient* OR participant* OR adult*)) OR ((persons OR people) W/3 (color OR colour)))	874035
	3	TITLE-ABS-KEY ((over?estimat* OR under?estimat* OR detect* OR classify* OR "rule in" OR "rule out" OR identif*) W/3 (hypox* OR occult))	10188
	4	TITLE-ABS-KEY (accura* OR inaccura* OR agreement OR precision OR evaluat* OR reliab* OR bias OR concordance OR performance OR validat* OR error* OR fail* OR discrepanc*	36072897

Search Date: 2/23/2023		Search Statement	Results
		OR disparit* OR vary OR varied OR varies OR variab* OR sensitivity OR specificity OR over?estimat* OR under?estimat* OR detect* OR classify* OR "rule in" OR "rule out" OR identif*)	
	5	TITLE-ABS-KEY (pediatric* OR paediatric* OR neonat* OR infant* OR hyperbaric OR wearable)	2260722
	6	2 OR 3	884101
	7	1 AND 4 AND 6	2691
	8	7 NOT 5	2162
	9	limit 8 to English language	2023
		Total	3190
		Total after deduplication	2226

STUDIES EXCLUDED DURING FULL-TEXT SCREENING

Citation	Exclude Reason
Occult Hypoxemia Is More Common in Black Patients than in White Patients About to Undergo ECMO for Respiratory Failure. *Critical Care Alert.* 2022;30(3):1-3.	Ineligible publication type
Impact of Skin Color on Spo2 Detection of Hypoxemia. In. Vol 43. Alisa Veijo, California: American Association of Critical-Care Nurses; 2023:80-80.	Ineligible publication type
Adams A, Cho HJ. Ensuring Progress Toward Racial Equity in Pulse Oximetry. *JAMA Internal Medicine.* 2022;182(12):1329.	Ineligible publication type
Anonymous. Racial Bias in Pulse Oximetry Measurement. *The New England journal of medicine.* 2021;385(26):2496.	Ineligible publication type
Arefin MS, Dumont AP, Patil CA. Monte Carlo based Simulations of Racial bias in Pulse Oximetry. Paper presented at: Progress in Biomedical Optics and Imaging - Proceedings of SPIE2022.	Ineligible publication type
Baek HJ, Shin J, Cho J. The Effect of Optical Crosstalk on Accuracy of Reflectance-Type Pulse Oximeter for Mobile Healthcare. *Journal of healthcare engineering.* 2018;2018:3521738.	Ineligible comparator
Balmaceda J, D Gerber E, J Arnold M, A Williams D, Snyder K, M Pandya S. RACIAL BIAS IN PULSE OXIMETRY IN ACUTE HYPOXEMIC RESPIRATORY FAILURE. *CHEST.* 2022;162(4):A1168-A1168.	Ineligible publication type
Barker SJ, Wilson WC. Racial effects on Masimo pulse oximetry: a laboratory study. *Journal of clinical monitoring and computing.* 2022.	Ineligible population
Bickler P, Tremper KK. The pulse oximeter is amazing, but not perfect. *Anesthesiology.* 2022;136(5):670-671.	Ineligible publication type
Bota GW, Rowe BH. Continuous monitoring of oxygen saturation in prehospital patients with severe illness: The problem of unrecognized hypoxemia. *Journal of Emergency Medicine.* 1995;13(3):305-311.	Ineligible intervention
Brownscombe JJ, Loane H, Honan B. COVID-19 highlights the need for action on pulse oximeter accuracy in people with dark skin. *Medical Journal of Australia.* 2022;216(10):539.	Ineligible publication type
Bunch D. Pulse Oximeters May Overestimate Blood Oxygen Levels in Ethnic Groups. *AARC Newsroom.* 2022:8-8.	Unable to locate full text
Burnett GW, Stannard B, Wax DB, et al. Self-reported Race/Ethnicity and Occult Hypoxemia: Reply. *Anesthesiology.* 2022;137(3):371-372.	Ineligible publication type
Cabanas AM, Fuentes-Guajardo M, Latorre K, Leon D, Martin-Escudero P. Skin Pigmentation Influence on Pulse Oximetry Accuracy: A Systematic Review and Bibliometric Analysis. *Sensors (Basel, Switzerland).* 2022;22(9).	Outdated or ineligible systematic review
Cahan C, Decker MJ, Hoekje PL, et al. Agreement between noninvasive oximetric values for oxygen saturation. *CHEST.* 1990;97(4):814-819.	Ineligible outcome
Crooks C, West J, Card T, Shaw D, Simmonds M, Fogarty A. Reply to: Inaccuracy of pulse oximetry in darker-skinned patients is unchanged across 32 years. *The European respiratory journal.* 2022;59(6).	Ineligible publication type
Crooks CJ, West J, Morling JR, et al. Pulse oximeter measurement error of oxygen saturation in patients with SARS-CoV-2 infection stratified by smoking status. *European Respiratory Journal.* 2022;60(5).	Ineligible outcome
Dabbous A, Bijjani A, Nader T, Dahdah S, Tarraf S, Baraka A. Incidence of postoperative hypoxemia as detected by pulse oximetry. *Middle East journal of anaesthesiology.* 1992;11(4):321-329.	Unable to locate full text

Citation	Exclude Reason
Ebmeier SJ, Barker M, Bacon M, et al. A Two Centre Observational Study of Simultaneous Pulse Oximetry and Arterial Oxygen Saturation Recordings in Intensive Care Unit Patients. *Anaesthesia and Intensive Care.* 2018;46(3):297-303.	Duplicate
Emery JR. Skin pigmentation as an influence on the accuracy of pulse oximetry. *Journal of Perinatology.* 1987;7(4):329-330.	Ineligible population
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. *Anesthesia and analgesia.* 2007;105(6 Suppl):S18-S23.	Ineligible population
Ferrari M, Quaresima V. Racial discrepancies in oximetry: where do we stand? The gold standard choice. *Anaesthesia.* 2022;77(4):492-492.	Ineligible publication type
Ferrari M, Quaresima V, Scholkmann F. Pulse oximetry, racial bias and statistical bias: further improvements of pulse oximetry are necessary. *Annals of intensive care.* 2022;12(1):19.	Ineligible publication type
Gokhale SG, Daggubati V, Alexandrakis G. Innovative technology to eliminate the racial bias in non-invasive, point-of-care (POC) haemoglobin and pulse oximetry measurements. *BMJ Innovations.* 2022.	Ineligible intervention
Gottlieb ER, Ziegler J, Rush B. Ensuring Progress Toward Racial Equity in Pulse Oximetry-Reply. *JAMA Internal Medicine.* 2022;182(12):1329-1330.	Ineligible publication type
Gottlieb ER, Ziegler J, Morley K, Rush B, Celi LA. Assessment of Racial and Ethnic Differences in Oxygen Supplementation Among Patients in the Intensive Care Unit. *JAMA Internal Medicine.* 2022;182(8):849-858.	Ineligible outcome
Gadrey SM, Mohanty P, Haughey SP, et al. Overt and Occult Hypoxemia in Patients Hospitalized With COVID-19. *Critical care explorations.* 2023;5(1):e0825.	Ineligible intervention
Harlan EA, Colon Hidalgo D, Valley TS. Addressing racial bias in pulse oximetry. *Chest Physician.* 2022;17(8):18-18.	Ineligible publication type
Hassan EA, Mohamed SN, Hamouda EH, Ahmed NT. Clinical evaluation for the pharyngeal oxygen saturation measurements in shocked patients. *BMC Nursing.* 2022;21(1).	Ineligible comparator
Hess D. Detection and monitoring of hypoxemia and oxygen therapy. Paper presented at: Respiratory Care2000.	Unable to locate full text
Hobensack M, Phan NM. Racial Bias in Pulse Oximetry: Sjoding M, Dickson R, Iwashyna T, et al. N Engl J Med. 2020;383(25):2477-2478. doi:10.1056/NEJMc2029240. *Journal of Emergency Medicine (0736-4679).* 2021;60(2):262-263.	Duplicate
Holder AL, Wong A-KI. The Big Consequences of Small Discrepancies: Why Racial Differences in Pulse Oximetry Errors Matter. *Critical Care Medicine.* 2022;50(2):335-337.	Ineligible publication type
Hunasikatti M. Racial bias in accuracy of pulse oximetry and its impact on assessments of hypopnea and T90 in clinical studies. *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine.* 2021;17(5):1145.	Ineligible publication type
Jacobs JW, Abels E. The potential for unnecessary medical interventions due to inaccurate pulse oximetry measurements. *Heart & lung : the journal of critical care.* 2022.	Ineligible publication type
Jouffroy R, Jost D, Prunet B. Prehospital pulse oximetry: a red flag for early detection of silent hypoxemia in COVID-19 patients. *Critical care (London, England).* 2020;24(1):313.	Ineligible outcome

Citation	Exclude Reason
Keating L, Christian C, Strykowski RK, Lee C, Strek ME, Adegunsoye A. PULSE OXIMETRY INACCURATELY REFLECTS SEVERITY OF LUNG FUNCTION IMPAIRMENT ACROSS RACIAL AND DIAGNOSTIC SUBGROUPS OF INTERSTITIAL LUNG DISEASE. *CHEST.* 2022;162(4):A2618-A2619.	Ineligible publication type
Keller MD, Harrison-Smith B, Patil C, Arefin MS. Skin colour affects the accuracy of medical oxygen sensors. *Nature.* 2022;610(7932):449-451.	Ineligible publication type
Knight MJ, Subbe CP, Inada-Kim M. Racial discrepancies in oximetry: where do we stand? *Anaesthesia.* 2022;77(2):129-131.	Ineligible publication type
Lahri S. Early detection of hypoxia in covid-19. *Pan African Medical Journal.* 2020;35:1-2.	Ineligible publication type
Lee WW, Mayberry K, Crapo R, Jensen RL. The accuracy of pulse oximetry in the emergency department. *The American journal of emergency medicine.* 2000;18(4):427-431.	Ineligible outcome
Lipchak D, Chupov A. Sensorex: The Challenges for Engineering Implementation of Low-Cost Non-Invasive Pulse Oximeter Applicable to Diverse Patient Population. Paper presented at: International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices, EDM2021.	Ineligible publication type
Longcoy J, Patwari R, Hasler S, et al. Racial and Ethnic Differences in Hospital Admissions of Emergency Department COVID-19 Patients. *Medical Care.* 2022;60(6):415-422.	Ineligible intervention
Louw A, Cracco C, Cerf C, et al. Accuracy of pulse oximetry in the intensive care unit. *Intensive Care Medicine.* 2001;27(10):1606-1613.	Ineligible outcome
Melton JD, Heller MB, Kaplan R, Mohan-Klein K. Occult hypoxemia during aeromedical transport: Detection by pulse oximetry. *Prehospital and Disaster Medicine.* 1989;4(2):115-120.	Ineligible intervention
Mendelson Y, Kent JC, Shahnarian A, Welch GW, Giasi RM. Evaluation of the Datascope ACCUSAT pulse oximeter in healthy adults. *Journal of clinical monitoring.* 1988;4(1):59-63.	Ineligible outcome
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. *South African Medical Journal.* 2021;111(10):950-956.	Ineligible intervention
Noninvasive blood gas monitoring: a review for use in the adult critical care unit. Technology Subcommittee of the Working Group on Critical Care, Ontario Ministry of Health. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* 1992;146(5):703-712.	Outdated or ineligible systematic review
Norton HL. Variation in pulse oximetry readings: melanin, not ethnicity, is the appropriate variable to use when investigating bias. *Anaesthesia.* 2022;77(3):354-355.	Ineligible intervention
Ochoa-Gutierrez V, Guerrero-Zuniga S, Reboud J, Pazmino-Betancourth M, Harvey AR, Cooper JM. Changes in Oxygenation Levels During Moderate Altitude Simulation (Hypoxia-Induced): A Pilot Study Investigating the Impact of Skin Pigmentation in Pulse Oximetry. *Advances in experimental medicine and biology.* 2022;1395:391-396.	Ineligible population
Palmer SJ. Racial bias in pulse oximetry: a significant problem in healthcare. *British Journal of Healthcare Assistants.* 2022;16(9):430-433.	Ineligible publication type

Citation	Exclude Reason
Philip KEJ, Tidswell R, McFadyen C. Racial bias in pulse oximetry: more statistical detail may help tackle the problem. *BMJ (Clinical research ed).* 2021;372:n298.	Ineligible publication type
Pierson DJ. Pulse oximetry versus arterial blood gas specimens in long-term oxygen therapy. *Lung.* 1990;168 Suppl:782-788.	Ineligible publication type
Ploen L, Pilcher J, Beckert L, Swanney M, Beasley R. An investigation into the bias of pulse oximeters. Paper presented at: Respirology2016.	Ineligible publication type
Ralston AC, Webb RK, Runciman WB. Potential errors in pulse oximetry. III: Effects of interferences, dyes, dyshaemoglobins and other pigments. *Anaesthesia.* 1991;46(4):291-295.	Ineligible publication type
Ralston AC, Webb RK, Runciman WB. Potential errors in pulse oximetry III: Effects of interference, dyes, dyshaemoglobins and other pigments. *Anaesthesia.* 1991;46(4):291-295.	Ineligible publication type
Saunders NA, Powles AC, Rebuck AS. Ear oximetry: accuracy and practicability in the assessment of arterial oxygenation. *The American review of respiratory disease.* 1976;113(6):745-749.	Ineligible population
Schallom M, Prentice D, Sona C, Arroyo C, Mazuski J. Comparison of nasal and forehead oximetry accuracy and pressure injury in critically ill patients. *Heart Lung.* 2018;47(2):93-99.	Ineligible intervention
Shimada Y, Nakashima K, Fujiwara Y, et al. Evaluation of a new reflectance pulse oximeter for clinical applications. *Medical & Biological Engineering & Computing.* 1991;29(5):557-561.	Ineligible comparator
Sinaki FY, Ward R, Abbott D, et al. Ethnic disparities in publicly-available pulse oximetry databases. *Communications medicine.* 2022;2:59.	Ineligible intervention
Sjoding M, Iwashyna TJ, Valley TS. More on Racial Bias in Pulse Oximetry Measurement. Reply. *The New England journal of medicine.* 2021;384(13):1278.	Ineligible publication type
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). *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde.* 2019;109(3):154-158.	Ineligible comparator
Smyth RJ, D'Urzo AD, Slutsky AS, Galko BM, Rebuck AS. Ear oximetry during combined hypoxia and exercise. *Journal of Applied Physiology.* 1985;60(2):716-719.	Ineligible population
Stannard B, Levin MA, Lin H-M, Weiner MM. Regional cerebral oximetry is consistent across self-reported racial groups and predicts 30-day mortality in cardiac surgery: a retrospective analysis. *Journal of clinical monitoring and computing.* 2021;35(2):413-421.	Ineligible intervention
Stewart KG, Rowbottom SJ. Inaccuracy of pulse oximetry in patients with severe tricuspid regurgitation. *Anaesthesia.* 1991;46(8):668-670.	Ineligible comparator
Thrush D, Hodges MR. Accuracy of pulse oximetry during hypoxemia. *South Med J.* 1994;87(4):518-521.	Ineligible population
Tobin MJ, Jubran A. Pulse oximetry, racial bias and statistical bias. *Annals of intensive care.* 2022;12(1):2.	Ineligible publication type
Tobin MJ, Jubran A. Inaccuracy of pulse oximetry in darker-skinned patients is unchanged across 32 years. *The European respiratory journal.* 2022;59(6).	Ineligible publication type
Tobin MJ, Jubran A. Unreliable pulse oximetry in dark-skin patients: a plea for algorithm disclosure. *Annals of Intensive Care.* 2022;12(1).	Ineligible publication type

Citation	Exclude Reason
Todd B. Pulse Oximetry May Be Inaccurate in Patients with Darker Skin. *American Journal of Nursing.* 2021;121(4):16.	Ineligible publication type
Valbuena VSM, Merchant RM, Hough CL. Racial and Ethnic Bias in Pulse Oximetry and Clinical Outcomes. *JAMA Internal Medicine.* 2022;182(7):699-700.	Ineligible publication type
Van de Louw A, Cracco C, Cerf C, et al. Accuracy of pulse oximetry in the intensive care unit. *Intensive care medicine.* 2001;27(10):1606-1613.	Ineligible outcome
Ward E, Katz MH. Confronting the Clinical Implications of Racial and Ethnic Discrepancy in Pulse Oximetry. *JAMA Internal Medicine.* 2022;182(8):858-858.	Ineligible publication type
Zeballos RJ, Weisman IM. Reliability of noninvasive oximetry in black subjects during exercise and hypoxia. *The American review of respiratory disease.* 1991;144(6):1240-1244.	Ineligible population

RISK OF BIAS ASSESSMENTS

OBSERVATIONAL STUDIES (QUADAS)

Study Name or Author Year	Patient Selection	Patient Categorization	Index Test	Reference Standard	Flow and Timing	Overall Rating
Abrams 2022	Unclear Does not describe how participants were enrolled in the study.	High Used race (Black patients examined separately) as a proxy for pigmentation and method of determination was not reported.	Low Index test was conducted prior to interpretation of results from reference test.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low ABG reading was obtained 1 minute after PO reading. All patients received the same reference standard. All patients were included in analysis.	Moderate
Adler 1998	Low All adult patients requiring ABG measurement in the ER were eligible for enrollment. Included consecutive patients with vital signs and complete data on oxygen saturation measurements.	Low Used the Munsell color system and categorized patients into 3 groups.	Low Index test was conducted prior to interpretation of results from reference test.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low PO measurement was taken at the time of ABG sampling, and ABG samples were taken immediately to the lab.	Low
Bangash 2022	Low Retrospective study that included patients for whom paired measurements and ethnicity were recorded.	High Used self-reported race (Black, Asian, White, Other) as a proxy for pigmentation.	Unclear Unclear whether the index test was conducted prior to the results of the reference test being automatically entered into the HER.	Low Reference test results were automatically entered into the EHR upon analysis.	Low Included paired measurements occurring within a 20 minute time period.	Moderate
Blanchet 2023	Low Appears that all patients meeting criteria during the study period were eligible and exclusions were appropriate.	Unclear Used the Fitzpatrick scale to characterize skin color, but do not report who made the assessment and only report that 96% were light skin (types 1 and 2).	Low Index test was conducted prior to interpretation of results from reference test.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low PO values were collected simultaneously with ABG samples. Appears all patients were included in analysis.	Unclear
Bothma 1996	Low	Low	Low	Unclear	Low	Low

Study Name or Author Year	Patient Selection	Patient Categorization	Index Test	Reference Standard	Flow and Timing	Overall Rating
	Included consecutive patients meeting criteria and exclusions were appropriate.	Patients were selected for the study when they subjectively appeared darkly pigmented. Pigmentation was objectively quantified with a portable EEL reflectance spectrophotometer (calculated mean reflectance). They compare their mean value to that of another study that included a group of pigmented volunteers.	Index test was conducted prior to interpretation of results from reference test.	Does not state whether reference test results were interpreted without knowledge of index test results.	PO measurements occurred at the same time as blood sample, which was immediately taken for analysis. The reference standard test (CO-oximeter) was the same for all patients. Appears that all patients were included in analysis.	
Burnett 2022	Low Included all patients who received an anesthetic with at least 1 ABG sample during the time period of the study.	High Used self-reported race/ethnicity (White, Other, Hispanic/Latinx, Black, Asian) as a proxy for pigmentation.	Low Used the mean SpO2 during a 5 minute interval starting 10 min before ABG time.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low Used the mean SpO2 during a 5 minute interval starting 10 min before ABG time. The reference standard test was the same for all patients. Appear to include all data where both measurements were identified during the specified time period.	Moderate
Cecil 1988	Low The only patient selection criterion was that the patient required an ABG test. Data collection occurred consecutively as requests for blood gas analysis were received from the attending physician.	High A subjective assessment of skin pigmentation was made using a scale of I to 3 for light, medium, and dark pigment levels, respectively.	Low The PO probe was placed while the blood sample was taken.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low The PO probe was placed while the blood sample was taken and was analyzed within 15 minutes. Included all data with the exception of identified outliers.	Moderate
Chelsey 2022	Low Included all critically ill patients with paired measurements within 10 minutes of each other during the study period.	High Self-reported race/ethnicity obtained from medical record data.	Unclear Unknown whether the index test was conducted prior to the interpretation of the	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low Paired SpO2 with SaO2 measurements via ABG obtained within 10 min of each other. The 2 sites used different blood gas analyzers, but these devices	Moderate

Study Name or Author Year	Patient Selection	Patient Categorization	Index Test	Reference Standard	Flow and Timing	Overall Rating
			reference standard results.		are regularly calibrated. Some analyses only included Black and White patient groups due to sample size, otherwise all patients appear to be included in analyses.	
Crooks 2023	Low Retrospective study that included patients admitted to a level 3 intensive treatment unit bed with COVID-19 infection during the study period.	High Race/ethnicity determined from EHR data.	Low Unknown whether the index test was conducted prior to the interpretation of the reference standard results, but ward oximetry measurements are recorded routinely electronically using NerveCentre.	Low Reference test results were automatically entered into the EHR upon analysis.	Low Used last recorded observations and blood tests prior to the time of ITU admission. Timing of these measurements relative to one another is unknown. Method of ABG measurement not recorded. All patients included in analysis although patients without ethnicity data were excluded from multivariate analysis.	Moderate
Ebmeier 2018	Low Included consecutive patients admitted to each ICU who had routine PO and ABG measurement as part of routine clinical care. Additional exclusions are appropriate.	Low The Fitzpatrick scale was used to categorize skin color at the bedside by the study investigator. The 6 numerical ratings were divided into 3 categories.	Low Index test was conducted prior to interpretation of results from reference test.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low The SpO2 value recorded was the first value displayed after the blood entered the syringe for the ABG sample. ABG analysis was the same for all patients. A small number of patients were excluded (2 opted out of the study and 8 had measurements taken within 2 hours of ICU admission).	Low
Escourrou 1990	Unclear No information provided on selection of patients.	High States only that 5 patients were "moderately pigmented but not Black;" no information was provided on how this judgment was made.	Unclear Oximeters used continuously display saturation. ABG was measured at rest and exercise and analyzed immediately; it is unclear whether continuous PO measurements were observed with	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Unclear The time between ABG and PO measurement is not reported. ABG was sampled and analyzed the same for all patients. A satisfactory pulse output could not be obtained in 5 patients who were excluded. In 3 patients the transcutaneous value	Moderate

Study Name or Author Year	Patient Selection	Patient Categorization	Index Test	Reference Standard	Flow and Timing	Overall Rating
			knowledge of the results of the initial ABG test.		was not obtained at exercise due to poor signal.	
Fawzy 2022	Low	High	Unclear	Unclear	Low	Moderate
	Retrospective study that included patients for whom specified data were available during the study period.	Used self-reported race/ethnicity (Asian, Black/African American, White and Hispanic or non-Hispanic).	Retrospective data and timing/conduct of measurements is unknown.	Retrospective data and timing/conduct of measurements is unknown.	SaO2 and Sp02 readings occurred within 10 minutes of each other. All ABG samples were analyzed via CO-oximetry using an ABL brand device. Flow diagrams shows reasons for exclusion of patients from each analysis and exclusions are appropriate.	
Gabrielczyk 1988	Unclear	High	Low	Unclear	Low	Moderate
	No information provided on selection of patients.	Only report that 4 patients had "racially pigmented skin."	Index test was conducted prior to interpretation of results from reference test.	Does not state whether reference test results were interpreted without knowledge of index test results.	Blood was drawn after the PO reading was obtained. All patients had ABG measurement done with the same CO-oximeter. Measurement pairs from all patients were included in analysis.	
Harskamp 2021	Low	Low	Low	Low	Low	Low
	Enrolled consecutive adult patients admitted to the ICU with a catheter for ABG samples. Exclusions listed are appropriate.	Used the Fitzpatrick classification scale, as assessed by the site investigator. Considered Fitzpatrick type IV-VI to be dark skin type, in accordance with FDA guidance.	PO readings were blinded for SaO2.	The intensive care personnel analyzing blood samples were blind to PO readings.	PO readings were done directly (within 10 minutes) after blood sample was taken. The same equipment was used to analyze all blood samples. Included all valid SpO2 measurements.	
Henry 2022	Low	High	Low	Unclear	Low	Moderate
	Retrospective study that included adults meeting criteria from 4 self-identified racial groups with paired measurements.	Four self-identified racial groups.	There were zero minutes of separation between tests so the results from the blood sample would not have been analyzed before the PO reading was taken.	Does not state whether reference test results were interpreted without knowledge of index test results.	PO readings and blood samples were done simultaneously (*ie*, zero minutes of separation). Method of blood gas analysis not reported. Appear to include all patients in analysis.	

Study Name or Author Year	Patient Selection	Patient Categorization	Index Test	Reference Standard	Flow and Timing	Overall Rating
Hinkelbein (2007) acrylic	Low	High	Low	Unclear	Low	Moderate
	All adult ICU patients meeting criteria (mechanically ventilated, Caucasian patients with white skin color, plethysmographic waveform was of poor quality, other factors that can affect PO readings) with an arterial line within the study period who agreed to participate were included. Excluded non-White patients (n=2) due to "possible severe interference with the measurement."	Do not report how determination of skin color/race was made for inclusion in study and white patients may have represented a range of skin pigmentation.	PO readings were taken before blood sample.	Does not state whether reference test results were interpreted without knowledge of index test results.	Blood sample was taken directly after PO readings. Same device type was used for all ABG measurements. All study patients were included in analysis.	
Hinkelbein 2007 (nail polish)	Low	High	Low	Unclear	Low	Moderate
	All ICU patients meeting criteria (mechanically ventilated, 5 fingers without nail polish, Caucasian patients with white skin color, plethysmographic waveform was of poor quality, other factors that can affect PO readings) within the study period who agreed to participate were included. Excluded non-White patients (n=2).	Do not report how determination of skin color/race was made for inclusion in study and White patients may have represented a range of skin pigmentation.	Blood sample was drawn at the same time as PO readings.	Does not state whether reference test results were interpreted without knowledge of index test results.	Blood sample was taken at the time of PO readings. Same device type was used for all ABG measurements. All study patients were included in analysis.	
Jubran 1990	Unclear	High	Low	Unclear	Unclear	Moderate
	Unclear whether consecutive patients meeting criteria were included.	The skin of each of the Black patients was inspected by one of the investigators and subjectively graded as light, moderately dark, or very dark.	Each PO reading was taken at the same time as the paired blood sample.	Does not state whether reference test results were interpreted without knowledge of index test results. Since the index test values were set a priori according to the study protocol, investigators would have been aware of these	Blood sample was taken when each target PO reading was reached. Arterial oxygen saturation was measured with a CO-oximeter for all patients. It is unclear whether all patients were included in analysis.	

Study Name or Author Year	Patient Selection	Patient Categorization	Index Test	Reference Standard	Flow and Timing	Overall Rating
				values, but the publication does not report who analyzed the blood samples.		
Lee 1993	Unclear Unclear whether consecutive patients meeting criteria were included.	High Made note of the patient's race (Chinese, Malay, Indian) but it was not reported how this was determined.	Low SpO2 reading was taken when blood sample was drawn.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low SpO2 reading was taken when blood sample was drawn. All SaO2 readings were done with the same blood gas analyzer device type. Appears that jaundiced patients were excluded from analysis looking at effect of race.	Moderate
McGovern 1996	Unclear Patients were recruited from another study; details are not provided.	High Reports that all included subjects were White but does not report how this was determined or whether there was variation in skin color within the group.	Low Blood samples were evaluated after the exercise test.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low Arterial blood was collected at baseline and at each workload. SpO2 and SaO2 were recorded in the last 15s of each 1 min exercise period. All blood samples were analyzed with the same CO-oximeter. Measurements were obtained in all patients and data from each patient was included in the analysis.	Moderate
Muñoz 2008	Low Included all patients under assessment for long-term oxygen therapy at an outpatient center who had both measurements taken during the study period who met criteria. Excluded patients with factors that could affect measurements and exclusions are appropriate.	High All patients were Caucasian; do not report how race determination was made or whether there was variation in skin color/pigmentation within this group.	Low States that SaO2 and SpO2 were obtained simultaneously.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low States that SaO2 and SpO2 were obtained simultaneously. ABG was analyzed with the same CO-oximeter for all patients. All patients meeting criteria included in analysis.	Moderate
Nguyen 2022	Low Included all consecutive patients admitted to the	High Examine the effect of ethnic group, but do not	Low	Unclear Does not state whether reference test results were	Low SpO2 was measured continuously but the	Moderate

Study Name or Author Year	Patient Selection	Patient Categorization	Index Test	Reference Standard	Flow and Timing	Overall Rating
	ICU for acute respiratory failure (COVID-19 and non-COVID-19) during 2 different time periods who met criteria. Exclusion criteria included factors known to affect PO readings and exclusions are appropriate.	explain how this was determined; report % Black patients in table.	SpO2 reading was taken when ABG was performed.	interpreted without knowledge of index test results.	measurement was recorded at the time the ABG was performed. SaO2 was measured by the same type of blood gas analyzer device for all readings. All included patients included in analysis.	
Pilcher 2020	Low Patients 16 or older who were to have an ABG measurement as part of routine clinical care in hospital wards and outpatient clinics were recruited. Exclusions appear appropriate.	Low Based on modified Fitzpatrick scale with patient skin color classified as either: Light (Type I to Type II), Medium (Type III to Type IV) or Dark (Type V to Type VI).	Low SpO2 reading was taken when blood was drawn for ABG.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low The SpO2 value recorded was the first value on the oximeter when blood was first observed to enter the collection vial. Models of ABG analyzers appear to have varied. Excluded measurement paired with ABG samples identified to be venous or unusable. Participants in which there was a reported concern with oximeter accuracy were not excluded from analyses. Analyzed all participants that were included.	Low
Ries 1985	Unclear Included pulmonary patients referred to the laboratory for clinical exercise testing. Unclear if the study group represents all referred patients during a time period. Does not note any exclusions.	High Report skin pigmentation for a subset of participants. Skin pigmentation was assessed using a semiquantitative scale of light (1) to dark (4). Do not describe how this judgment was made or by whom.	Low Ear oximetry readings were recorded by a separate observer simultaneously with each arterial blood sample.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	High Ear oximetry readings were recorded simultaneously with each arterial blood sample. Blood samples were analyzed using 2 different blood gas analyzers that were regularly calibrated as well as by CO-oximeter. Skin pigment data was only collected for the subset of patients who were tested with both oximeter types.	High
Ries 1989	Unclear Included patients referred to the laboratory for evaluation of arterial blood gas changes during	High Skin color quantified by the Munsell color system. The 5YR hue chart was used as it	Low SpO2 readings and withdrawal of blood	Unclear Does not state whether reference test results were interpreted without	Low Readings from the ear oximeters were taken simultaneously with the withdrawal of each arterial	Moderate

Study Name or Author Year	Patient Selection	Patient Categorization	Index Test	Reference Standard	Flow and Timing	Overall Rating
	exercise. Unclear if the study group represents all referred patients during a time period. Does not note any exclusions.	corresponded most closely to the range of skin colors. A technician selected the tile from the hue chart which best matched the skin at the probe placement site. However, the 4 lightest colors were each considered their own group while the darker colors were grouped together in a single group (range of skin colors was not evenly distributed across the groups).	sample occurred simultaneously.	knowledge of index test results.	blood sample. All blood samples were analyzed with a blood gas analyzer that was regularly calibrated and each sample was checked with a CO-oximeter. Appears measurements from all patients are included but not all patients have measurements for both oximeters.	
Seitz 2022	Low Retrospective study that included all patients meeting criteria with paired measurements and race documented as Black or White during the study period. Excluded patients with COVID-19. Patients with other values for race were not included due to an inadequate number of patients for comparison.	High Race (Black or White) was used as a surrogate for skin pigmentation.	Low SpO2 values were directly transferred from the bedside monitor into the institutional data warehouse every 1 minute.	Unclear Retrospective data and timing/conduct of measurements is unknown.	Low Measurements occurred within 10 minutes of each other and the SpO2 value closest in time was used. SaO2 measurement were done using the same CO-oximeter. All patients included in study were included in analysis.	Moderate
Sjoding 2020	Unclear Retrospective study; does not clearly describe how patients were selected for inclusion.	High Race (non-Hispanic Black or non-Hispanic White) was used as a surrogate for skin pigmentation.	Unclear Retrospective data and timing/conduct of measurements is unknown.	Unclear Retrospective data and timing/conduct of measurements is unknown.	Low Measurements occurred within 10 minutes of each other. SaO2 was directly measured by CO-oximetry for all patients. Appear to include all study patients in analysis.	Moderate
Sudat 2022	Low Appear to include all patients meeting their criteria during the study period. Excluded patients	High Race (Black or White) was used as a	Unclear Retrospective data and timing/conduct of	Unclear Retrospective data and timing/conduct of	Low Paired each SaO2 measurement with the nearest recorded SpO2 for the same person, truncated	Moderate

Study Name or Author Year	Patient Selection	Patient Categorization	Index Test	Reference Standard	Flow and Timing	Overall Rating
	who were not non-Hispanic Black or non-Hispanic white. Second cohort included ED patients with COVID-19. Excluded visits with no documented SpO2.	surrogate for skin pigmentation.	measurements is unknown.	measurements is unknown.	at ± 10m from the earlier of the ABG specimen time or result time. No information provided on ABG measurement. Appear to include all patients meeting their criteria in the respective cohort analyses.	
Valbuena 2022 ECMO	Low Retrospective study including adult patients with ARDS or COVID-19 on ECMO for respiratory failure during the study period. Included patients with relevant data (blood gas samples had to meet certain criteria for timing).	High Race/ethnicity was used as a surrogate for skin pigmentation.	Unclear Retrospective data and timing/conduct of measurements is unknown.	Unclear Retrospective data and timing/conduct of measurements is unknown.	Low No data on the timing of measurements were available. No information provided on SaO2 measurement. For hypoxemia analyses, only included race/ethnicity categories that met their calculated sample size threshold.	Moderate
Valbuena 2022 VHA	Low Retrospective study including all SpO2 and SaO2 data available for hospital stays, with some exclusions for indicators of critical illness (to capture a general hospital sample, not ICU). Valid records in the database require core identifiers, including race and ethnic origin, to be present.	High Race (Black, Hispanic, or White) was used as a surrogate for skin pigmentation.	Unclear Retrospective data and timing/conduct of measurements is unknown.	Unclear Retrospective data and timing/conduct of measurements is unknown.	Low Included pairs of measurements occurring with 10 minutes of one another. Information on whether the SaO2 reading was measured by CO-oximetry was not available. Do not include Hispanic patients in all analyses.	Moderate
Wang 1987	Unclear Unclear whether consecutive patients meeting criteria were included.	High All patients were characterized as 'pigmented;' do not report how this judgment was made or on variation within the sample.	Low Index test was conducted prior to interpretation of results from reference test.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low Arterial blood was drawn simultaneously with PO readings. Oxygen saturation was measured with a CO-oximeter for some samples and a blood gas analyzer for others. Either an ear or finger PO reading was obtained from all patients, but not both.	Moderate

Study Name or Author Year	Patient Selection	Patient Categorization	Index Test	Reference Standard	Flow and Timing	Overall Rating
Wiles 2022	Low Retrospective study that included all consecutive patients meeting criteria with both Spo2 and SaO2 measurements available; exclusions were appropriate.	High Race/ethnicity was used as a surrogate for skin pigmentation.	Low Used the mean value of SpO2 readings that occurred during the 4 minute period prior to ABG analysis.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low Used the mean value of SpO2 readings that occurred during the 4 minute period prior to ABG analysis. Arterial oxygen saturation was determined by CO-oximetry for all patients. All study patients included in analysis.	Moderate
Wiles 2022 letter	Low Retrospective study that included critically ill patients with COVID-19 pneumonitis receiving invasive mechanical ventilation with pairs of SpO2 and SaO2 measurements during the study period. Other exclusions are appropriate.	High Categorized patients by ethnic group (South Asian, Black, White, and other).	Low SpO2 measurements performed automatically by bedside monitoring system. Used the mean value of SpO2 readings that occurred during the 4 minute period prior to ABG analysis.	Unclear Does not state whether reference test results were interpreted without knowledge of index test results.	Low Used the mean value of SpO2 readings that occurred during the 4 minute period prior to ABG analysis. Same CO-oximeter device type used for all SaO2 measurements. Appear to include all study patients in analysis.	Moderate
Wong 2021	Low Retrospective study that included patients from all units available in the data sets with SpO2 measurements within the specified range and with self-identified race/ethnicity classified as Asian, Black, Hispanic, or White.	High Classified patients by race/ethnicity, not skin pigmentation.	Unclear Retrospective data and timing/conduct of measurements is unknown.	Unclear Retrospective data and timing/conduct of measurements is unknown.	Low Each ABG-measured SaO2 was matched with the closest SpO2 value recorded within the previous 5 minutes. Do not report how blood gas analysis was conducted. All included patients included in analysis; excluded patients from subgroup analysis when data on corresponding characteristics were missing.	Moderate

Abbreviations. ABG=arterial blood gas, PO=pulse oximeter, ER=emergency room, HER=electronic health record, ITU=intensive treatment unit, ICU=intensive care unit, FDA=Food and Drug Administration, ARDS=acute respiratory distress syndrome, ECMO=extracorporeal membrane oxygenation.

OBSERVATIONAL STUDIES (QUIPS)

Study Name or Author Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Overall Risk of Bias
Burnett 2022	Low Included all patients who received an anesthetic with at least 1 ABG sample during the time period of the study.	Low Retrospective cohort study. Appear to include all data where both measurements were identified during the specified time period.	High Used self-reported race/ethnicity obtained from medical record data as a proxy for pigmentation.	Low Occult hypoxemia defined as SaO2 less than 88% despite SpO2 greater than 92%. SaO2 measurement was the same for all patients and changes in SpO2 devices were accounted for by year measurement was taken.	Moderate Used a multivariable model which controlled for all of the collected demographic, comorbidity, and operative variables. Do not account for all potential confounding variables (*eg*, socioeconomic factors).	Low GEE modeling was used to determine if race was an independent predictor of occult hypoxemia. No evidence of selective reporting of results.	Moderate
Chelsey 2022	Low Included all critically ill patients with paired measurements within 10 minutes of each other during the study period.	Low Retrospective cohort study. Some analyses only included Black and White patient groups due to sample size, otherwise all patients appear to be included in analyses.	High Used self-reported race/ethnicity (White, Other, Hispanic/Latinx, Black, Asian) as a proxy for pigmentation.	Low Occult hypoxemia defined as SaO2 < 88% when pulse oximeter oxygen saturation was between 92–96%. The 2 sites used different blood gas analyzers, but these devices are regularly calibrated. Different pulse oximeters were used at the 2 study sites but there were no differences between sites when compared.	Moderate Multivariable model controlled for a limited number of potential confounders (age, sex, hemoglobin).	Low Used multivariable logistic regression model to examine the association between self-reported race and occult hypoxemia. No evidence of selective reporting of results.	Moderate
Crooks 2023	Low Retrospective study that included patients admitted to a level 3 intensive treatment unit bed with COVID-19	Low Retrospective cohort study. Patients without race data were not included in multivariate analysis	High Used self-reported race/ethnicity (White, Black/mixed, Indian/Pakistani,	High Used last recorded observations and blood tests prior to the time of ITU admission. Timing of these	Moderate Model adjusted for age, sex, weekend ICU admission, and vaccination status. Other potential confounding	Low Utilized a multivariate model. No evidence of selective reporting.	High

Study Name or Author Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Overall Risk of Bias
	infection during the study period.	but otherwise all patients were retained.	Other) as a proxy for pigmentation.	measurements relative to one another is unknown. Do not report device types for PO and ABG measurements but these measurements are automatically uploaded to the EHR. Don't provide detail on measurement of other clinical variables (*eg*, mean respiratory rate).	variables (*eg*, socioeconomic factors, comorbidities) were not controlled for.		
Fawzy 2022	Low Included patients for whom specified data were available during the study period.	Low Retrospective cohort study. Flow diagrams shows reasons for exclusion of patients from each analysis and exclusions are appropriate.	High Used self-reported race/ethnicity (Asian, Black/African American, White and Hispanic or non-Hispanic) as a proxy for pigmentation.	Moderate All ABG samples were analyzed via CO-oximetry using an ABL brand device. PO device type and reading location was not reported and likely varied between sites. Delayed treatment recognition was defined as those patients with a predicted SaO2 of 94% or less before a measured SpO2 of 94% or less or oxygen treatment initiation. Unrecognized treatment eligibility was defined as those patients with a predicted SaO2 of 94% or less who did not initiate treatment with oxygen or have a	Moderate Model was adjusted for covariates that captured disease severity or an underlying comorbidity or had a known or theoretical association with PO accuracy, including demographic characteristics along with time-varying clinical and laboratory variables. Do not account for all potential confounding variables (*eg*, socioeconomic factors).	Moderate The difference in time to recognition of treatment eligibility between patients of racial and ethnic minority groups and White patients was estimated using a Cox proportional hazards model. Among individuals with delayed recognition of treatment eligibility, Wilcoxon rank sum tests were used to compare the distributions of length of delayed recognition between groups. Only patients with complete data on covariates were included in adjusted model. No evidence of selective reporting of results.	Moderate

Study Name or Author Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Overall Risk of Bias
				recorded SpO2 of 94% or less at any time.			
Henry 2022	Low Retrospective study that included adults meeting criteria from 4 self-identified racial groups with paired measurements.	Low Appear to include all patients in analysis.	High Patients were categorized into 4 self-identified racial groups.	Moderate PO readings and blood samples were done simultaneously (*ie*, zero minutes of separation). Method of blood gas analysis and pulse oximetry not reported and may have differed between study sites. Occult hypoxemia was defined as SaO2 less than 88% despite a normal SpO2 (*ie*, ≥ 92%). Hospital-free days were counted as the number of days alive and out of hospital following the index time through 28 days of follow-up.	Moderate Adjusted for select potential confounders in both analysis of occult hypoxemia (age, mean arterial pressure less than 65 mmHg or the use of continuous infusions of IV vasopressors at the time of SaO2 assessment, and presence of COPD or home oxygen use) and analysis of treatment outcomes (age, sex, COPD or home oxygen use, index location [ICU vs surgical], and acuity of illness).	Low GEE was used to account for multiple observations. No evidence of selective reporting.	Moderate
Seitz 2022	Low Retrospective study that included all patients meeting criteria with paired measurements and race documented as Black or White during the study period. Excluded patients with COVID-19. Patients with other values for race were not included due to an	Low Retrospective study and all patients included in study were included in analysis.	High Race (Black or White) was used as a surrogate for skin pigmentation	Low Measurements occurred within 10 minutes of each other and the SpO2 value closest in time was used. SaO2 and SpO2 measurements were done using the same devices for all patients. Occult hypoxemia was defined as SaO2 <	High Did not examine occult hypoxemia with a multivariate model that accounts for potential confounders.	Low Compared rate of occult hypoxemia between groups (t test?). No evidence of selective reporting.	Low

Study Name or Author Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Overall Risk of Bias
	inadequate number of patients for comparison.			88% with SpO2 values of 92–96%.			
Sudat 2022	Low Appear to include all patients meeting their criteria during the study period. Excluded patients who were not non-Hispanic Black or non-Hispanic White. Second cohort included ED patients with COVID-19. Excluded visits with no documented SpO2.	Low Retrospective study; appear to include all patients meeting their criteria in the respective cohort analyses.	High Race (Black or White) was used as a surrogate for skin pigmentation.	Moderate Paired each SaO2 measurement with the nearest recorded SpO2 for the same person, truncated at ± 10m from the earlier of the ABG specimen time or result time. No information provided on ABG or PO measurement. Treatment outcomes (time spent in ED, hospital admission, dexamethasone administration and timing, oxygen supplementation and timing, return to the hospital after discharge home) from EHR data. Defined hypoxemia as an SaO2 < 90%.	Moderate Important differences were noted between NHB and NHW groups at baseline (homelessness, insurance types, comorbidities), but these and additional demographic and clinical covariates were included in the model. Do not account for all possible confounders.	Low Used G-computation to build 2 counterfactuals to assess the possible impacts of differential SpO2 measurement error on COVID-19-related outcomes for NHB patients. Computed the mean difference between the predicted outcome with the observed SpO2 values and the predicted outcome with SpO2 values shifted by the measurement difference from the initial PO bias analysis. Also compared prevalence of OH between groups and reported p value. No evidence of selective reporting.	Moderate
Valbuena 2022 ECMO	Moderate Retrospective study including adult patients with ARDS or COVID-19 on ECMO for respiratory failure during the study period. Included patients with relevant data (blood gas samples had to meet certain criteria for timing). For	Low Retrospective study using registry data.	High Race/ethnicity was used as a surrogate for skin pigmentation.	High No data on the timing of measurements were available. No information provided on SaO2 or SpO2 measurement. Occult hypoxemia was defined as low arterial oxygen saturation (SaO2 ≤ 88%) on arterial	High Only adjusted for sex and measured SpO2.	Low Multivariable analyses were performed by logistic regression for each race and ethnicity group compared with White patients to examine the relationship between these variables with the odds of occult hypoxemia. No evidence of selective reporting.	High

Study Name or Author Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Overall Risk of Bias
	hypoxemia analyses, only included race/ethnicity categories that met their calculated sample size threshold (N ≥ 400). Excluded a large number of patients for this reason, including groups where patients may have had more pigmented skin (*eg*, North African).			blood gas measurement despite a pulse oximetry reading in the range of 92% to 96%.			
Valbuena 2022 VHA	Low Retrospective study including all SpO2 and SaO2 data available for hospital stays, with some exclusions for indicators of critical illness (to capture a general hospital sample, not ICU). Valid records in the database require core identifiers, including race and ethnic origin, to be present.	Low Retrospective study and all patients included in study were included in analysis, although Hispanic patients were not included in all analyses.	High Race (Black, Hispanic, or White) was used as a surrogate for skin pigmentation.	Moderate Included pairs of measurements occurring with 10 minutes of one another. No information on device/location for PO and SaO2 measurements was available. Occult hypoxemia was defined as defined as arterial SaO2 <88% despite a SpO2 reading of ≥92%.	Moderate Models were adjusted for patient level characteristics that included age, sex, patient comorbidities, supplemental oxygen, and diagnoses on admission. Do not account for all potential confounders.	Low Fit a multivariable logistic regression model to predict the odds of occult hypoxemia. No evidence of selective reporting.	Moderate
Wong 2021	Low Retrospective study that included patients from all units available in the data sets with SpO2 measurements within the specified range and with self-identified	Low Retrospective study. All included patients included in analysis; excluded patients from subgroup analysis when data on corresponding characteristics were missing.	High Classified patients by race/ethnicity, not skin pigmentation.	Moderate Each ABG-measured SaO2 was matched with the closest SpO2 value recorded within the previous 5 minutes. Do not report how blood gas analysis was conducted or	Moderate Adjusted only for age, sex, SOFA score.	Low Multivariate logistic regression was used for assessing binary end points, multivariate ordinal regression for numerical end points, and multivariate linear models for continuous end points, using analysis of variance to	Moderate

Study Name or Author Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Overall Risk of Bias
	race/ethnicity classified as Asian, Black, Hispanic, or White.			device/location of PO measurement. Occult hypoxemia was defined as SpO2 > 88% but SaO2 < 88%. Clinical outcomes (in-hospital mortality, length of stay, organ dysfunction [SOFA scores], laboratory values) extracted from HER.		test for the impact of hidden hypoxemia while adjusting for other covariates. Calculated relative risk of OH by race/ethnicity. No evidence of selective reporting.	

Abbreviations. ABG=arterial blood gas, GEE=generalized estimating equation, ITU=intensive treatment unit, PO=pulse oximeter, HER=electronic health record, ICU=intensive care unit, COPD=chronic-obstructive pulmonary disease, ED=emergency department.

SYSTEMATIC REVIEWS (ROBIS)

Study Name or Author Year	Study Eligibility Criteria	Identification and Selection of Studies	Data Collection and Study Appraisal	Synthesis and Findings	Overall Risk of Bias
Shi 2022	Low Reasonable and clearly defined eligibility criteria.	Low Multiple databases searched. Conducted searches of clinical trial registries. Hand-searched reference lists of included studies and relevant reports. Dual independent study selection.	Low A single reviewer abstracted data and assessed risk of bias, checked by another reviewer. Risk of bias was assessed using appropriate criteria.	Low Appears all data were included, as appropriate. Performed both pre-planned and post-hoc sensitivity analyses.	Low

CHARACTERISTICS OF INCLUDED STUDIES

Study Name or Author Year	Country Setting	Participant Characteristics	Pulse Oximetry *Location* CO-oximetry	Skin Pigmentation or Race/Ethnicity Factors Measurement	Outcome Types
Abrams 2002 N=294	US Clinical	Prospective liver transplant candidates with cirrhosis and controls with an indication for ABG measurement Mean age: 53.3 % male: 54.4 % non-Hispanic White: NR	Nellcor N-200 *Finger* Radiometer ABL520	Race (Black, White) NR	Bias
Adler 1998 N=298	US Clinical	NR Mean age: 60 % male: 51 % non-Hispanic White: NR	Nellcor D-25 *Finger* 4-wavelength spectrophotometer or CO-oximeter (Radiometer OSM3)	Skin pigmentation (dark, intermediate, light) Munsell color system	Bias
Bangash 2022 N=18,069	UK Clinical	NR Median age: 63 % male: 57.9 % non-Hispanic White: 81.2	NR NR	Race (White, Asian, Black, Other) Self-reported	Bias, prevalence of occult hypoxemia
Blanchet 2023 N=193	Canada Clinical	Stable adults in the intensive care unit with an arterial catheter in place Median age: 66.3 years % male: 79.3	Nellcor N-600 *Finger* Radiometer ABL 800Flex OSM-3	96.2% light skin (types 1 and 2) Fitzpatrick scale	Bias, root mean square error, prevalence of occult hypoxemia, observations
Bothma 1996 N=100	US Clinical	Critically ill patients wrth arterial lines in situ	Simed S100e *Finger* Nihon Koden *Finger* Ohmeda 3740 *Ear* Instrumentation Laboratories IL482 CO-oximeter System	Subjects (darkly pigmented patients) had a mean reflectance of 19.9 Portable EEL Reflectance spectrophotometer	Bias
Burnett 2022 N=46,253	US Clinical	Patients receiving anesthetic Median age: 57± 21 years % male: 54.5 % White: 47.8	Primarily Nellcor prior to 2011 and Masimo after GEMStat Premier 3000	NR	Bias, prevalence of occult hypoxemia, OR observations

Study Name or Author Year	Country Setting	Participant Characteristics	Pulse Oximetry *Location* CO-oximetry	Skin Pigmentation or Race/Ethnicity Factors Measurement	Outcome Types
Cecil 1988 N=152	US Clinical	ABG determination was needed Median age: 64.1 ± 18.7 years % male: 53.9 % White: 89.5	Nellcor N-100; Ohmeda 3700 *Finger* IL 282 CO-Oximeter	For Black patients, the subjective estimate of pigmentation level ranged from 1 to 3 (2.5 ± 0.6). A subjective assessment of skin pigmentation was made using a scale of I to 3 for light, medium, and dark pigment levels, respectively.	Bias
Chelsey 2022 N=7,693	US Clinical	Critically ill patients admitted to ICUs Median age: 64 years % male: 58.9 % White: 60	Covidien Nellcor oximeter and OxiMax disposable finger sensors; Masimo pulse oximeter and LNCS disposable sensors *Finger* Multi-wavelength CO-oximeter (ABL90 and GEM 4000)	NR	Bias, intrasubject variability in measurement error, prevalence of occult hypoxemia, OR both
Crooks 2023 N=748	UK Clinical	Patients admitted to a level 3 intensive treatment unit bed with COVID-19 infection Median age: 56 years % male: 62.5 % White:56.1	NR	NR	Bias, prevalence of occult hypoxemia, treatment delivery
Ebmeier 2018 N=394	Australia and New Zealand	ICU patients Median age: 62.5 ± 15.1 % female: 36.6 % European: 80.6	Marquette Rac-4A and Masimo (NZ); Philips IntelliVue MP70 monitors with Philips Adult Reusable SpO2 sensors (Australia) *Ear, finger, or toe* Alpha-stat method using Radiometer ABL 800 FLEX	Dark 3.1% Medium 34.0% Light 61.9% The Fitzpatrick Scale	Bias, limits of agreement, association with oxygen saturation
Escourrou 1990 N=101	France Research	Patients with COPD, emphysema, sarcoidosis, pulmonary fibrosis, Hodgkin's disease, primary pulmonary hypertension, cirrhosis of the liver, restrictive pulmonary disease, miscellaneous	Ohmeda Biox 3700 *Ear* Criticare CSI 501 + *Ear* Nellcor N200	5 patients were moderately pigmented but not black NR	Bias, standard error

Study Name or Author Year	Country Setting	Participant Characteristics	Pulse Oximetry *Location* CO-oximetry	Skin Pigmentation or Race/Ethnicity Factors Measurement	Outcome Types
		diseases, dyspnea of unknown origin Age: 17 – 81 years % male: 71.3	Sampled using a microsampler and analyzed using a Corning 178 gas analyzer. Measured SaO2 was determined by a spectrophotometric method.		
Fawzy 2022 N=1216 (accuracy) 6673 (treatment eligibility analysis)	US Clinical	For sample with concurrent SaO2 and SpO2 measurements: Median 59.8 years % male: 58.3 % White: 37.8	NR	NR	Bias, prevalence of occult hypoxemia, treatment delivery
Gabrielczyk 1988 N=21	UK Clinical	Hypothermic patients after cardiac surgery Median age: 59.5 years	Nellcor N-100 *Finger* Hemoglobin saturation (Radiometer OS2), blood gas estimation (ABL30 Radiometer)	Racially pigmented skin 19% (4 patients and 14 paired obs) NR	Bias
Harskamp 2021 N=35	Netherlands Clinical	ICU patients primarily admitted for respiratory failure due to COVID-19 or other pulmonary diseases Median age: 69 years % female: 40	Direct-to-consumer pulse oximeters: AFAC FS10D, AGPTEK FS10C, ANAPULSE ANP 100, Cocobear, Contec CMS50D1, HYLOGY MD-H37, Mommed YM101, PRCMISEMED F4PRO, PULOX PO-200 and Zacurate Pro Series 500 DL Included a hospital-grade SpO2 monitor (Philips M1191BL sensor glove) as a clinical index test ABL90 Flex Plus	Fitzpatrick IV-VI skintype 14.3% Fitzpatrick classification scale	Bias, root mean square difference, mean absolute error
Henry 2022 N=26,603	US Clinical	Patients admitted to the ICU or undergoing surgery during inpatient hospitalization Median age: 64 years % male: 58.4 % White: 92.1	NR	NR	Bias, prevalence of occult hypoxemia, OR Obs, treatment delivery
Hinkelbein 2007 acrylic N=46	Germany Clinical	Critically ill and mechanically ventilated (for at least 24	Siemens (monitor SC1281 and module SIREM connected to a re-usable Nellcor DS-100A Durasensor finger probe) or	Only included patients with white skin color (Caucasian race)	Bias

Study Name or Author Year	Country Setting	Participant Characteristics	Pulse Oximetry *Location* CO-oximetry	Skin Pigmentation or Race/Ethnicity Factors Measurement	Outcome Types
		hours via a tracheal tube or cannula) patients Median age: 58.1 years	Philips (IntelliVue MP70 attached to the finger probe M1191A) device Radiometer Copenhagen System ABL625		
Hinkelbein 2007 nail polish	Germany Clinical	Critically ill and mechanically ventilated (for at least 24 hours via a tracheal tube) patients Median age: 59 ± 14.2 % female: 38	Siemens monitor SC1281 and module SIREM, connected to a reusable finger sensor probe Nellcor DS100-A Durasensor *Finger* Radiometer Copenhagen ABL System	Only included patients with white skin color (Caucasian race)	Bias
Jubran 1990 N=54	US Clinical	Patients in the MICU receiving mechanical ventilation Median age: 53 years % female: 55.6 % White: 46.3	Nellcor *Finger, ear, or toe* Ohmeda-Biox 3700 *Ear or finger* Arterial O2 saturation was measured with a CO-oximeter	NR The skin of each of the Black patients was inspected by one of the investigators and subjectively graded as light, moderately dark, or very dark	Bias, frequency of inaccurate readings (*ie*, bias > 4%)
Lee 1993 N=33	Singapore Clinical	MICU patients Median age: 54.4 years % Chinese: 66.7 % Malay: 18.2 % Indian: 15.2	Nellcor, Simed, Critikon *NR* Nova Stat Profile 3 pH/blood gas analyser (This represents functional oxygen saturation and not fractional oxygen saturation as would be given with a CO-oximeter.)	NR	Bias
McGovern 1996 N=8	US Research	Patients in stable condition with severe COPD Median age: 63.2 ± 9.6 % male: 100	Ohmeda 3700 *Finger* IL 482 CO-oximeter	All subjects were White NR	Bias, bias after correction for COHb level
Muñoz 2008 N=846	Spain Clinical	Patients under assessment for long-term home oxygen therapy in a specialized outpatient clinic Median age: 68.4 ± 12.2 % male: 70	Minolta Pulsox-7 IL 682 CO-oximeter	All were Caucasian NR	Bias, impact of arterial oxygen tension (Pa, O2)

Study Name or Author Year	Country Setting	Participant Characteristics	Pulse Oximetry *Location* CO-oximetry	Skin Pigmentation or Race/Ethnicity Factors Measurement	Outcome Types
Nguyen 2022 N=55	France Clinical	Patients with acute respiratory distress syndrome (ARDS) with and without COVID-19. All patients were mechanically ventilated Median age: 60 ± 17 years % male: 60 % Black: 31%	Masimoset *Finger (all but 3) or ear* ABL800 Radiometer	NR	Bias, Correlations, agreement, concordance between relative changes in SpO2 and SaO2
Pilcher 2020 N=400	Australia and New Zealand Clinical	Patients who were to have an ABG measurement as part of routine clinical care Median age: 64 ± 15.2 % male: 53	Most common were Nonin Avant 9700, Masimo Rainbow Radical 7, and Nonin Avant 4000) *Ear or finger* The model used was recorded for each patient	Fitzpatrick Score 1 11% II 49.5% III 31.8% IV 7.5% V 0.3% VI 0% Based on modified Fitzpatrick scale with patient skin color classified as either: Light (Type I to Type II), Medium (Type III to Type IV) or Dark (Type V to Type VI)	Bias
Ries 1985 N=116	US Clinical	Pulmonary patients referred for evaluation of ABG changes during exercise NR	Hewlett-Packard 47201A and Biox IIA *Ear* IL813 and 513 blood gas analyzers. The SaO2 was measured both directly with the CO-oximeter and calculated from the measured PaO2, PaCO2, and pH, assuming a normal body temperature and P50	Only reported skin pigmentation in Phase 3 patients (simultaneous testing of both ear oximeters): Moderate skin pigmentation 21.7% Light skin pigmentation 78.3% Skin pigmentation was assessed using a semiquantitative scale of light (1) to dark (4)	Bias
Ries 1989 N=187	US Clinical	Pulmonary patients referred to the laboratory for clinical exercise testing NR	Ohmeda Biox III and Hewlett-Packard 47201A *Ear*	Group 1, value of 8 (very light); group 2, value of 7 (light); group 3, value of 6 (average); and group 4,	Bias, prevalence of technical problems

Study Name or Author Year	Country Setting	Participant Characteristics	Pulse Oximetry *Location* CO-oximetry	Skin Pigmentation or Race/Ethnicity Factors Measurement	Outcome Types
			Blood gas analyzers	value less than or equal to 5 (moderately dark to very dark). These value scores corresponded to rows of lightness (value) on the 5YR hue chart Skin color quantified by the Munsell color system. The 5YR hue chart was used as it corresponded most closely to the range of skin colors observed. A technician of normal color vision selected the tile from the 5YR hue chart which best matched the skin at the placement site for the ear probes.	
Seitz 2022 N=1,024	US Clinical	Critically ill adults receiving mechanical ventilation. Excluded patients with COVID-19 Median age (Black patients): 54 Median age (White patients): 58 % female (Black patients): 47 % female (White patients): 43	Nellcor Werfen GEM Premier 5000 blood gas analyzer	NR	Bias, prevalence of occult hypoxemia, obs
Sjoding 2021 N=10,001	US Clinical	One cohort was patients receiving supplemental oxygen. The other cohort was patients in ICUs. NR % White (UM cohort): 82.8 % White (multicenter cohort): 87.5	NR CO-oximetry	NR	Bias
Sudat 2022 N=11,237	US Clinical	All hospital patients who self-identified as non-Hispanic White or non-Hispanic Black (Cohort 1); all	NR NR	NR	Bias (cohort 1), prevalence of occult hypoxemia, obs

Study Name or Author Year	Country Setting	Participant Characteristics	Pulse Oximetry *Location* CO-oximetry	Skin Pigmentation or Race/Ethnicity Factors Measurement	Outcome Types
		adults who visited the ED with COVID-19 and self-identified as NHW or NHB (cohort 2) Median age (Cohort 2): 57.3 years % female (Cohort 2): 52.3 % NHW (Cohort 1): 81.1 % NHB (Cohort 1): 17.2 % NHW (Cohort 2): 87.5 % NHB (Cohort 2): 12.5			
Valbuena 2022 N=28,531	US Clinical	Inpatients in general care (medical and surgical not ICU) Median age (NHW): 69 years Median age (NHB): 66 Median age (Hispanic/Latino): 68 % male: 96	NR	NR	Bias, root mean square error, probability obs,
Valbuena 2022 ECMO N=372	US Clinical	Patients with ARDS or COVID-19 in respiratory failure and about to undergo extracorporeal membrane oxygenation % male: 68 % White: 50	NR	NR	Bias, prevalence of occult hypoxemia, OR obs
Wang 1987 N=31	Singapore Clinical	Most (27) patients admitted to the hospital with acute symptoms. In 4 outpatients, ABG was performed as part of lung function assessment Median age: NR % male: 96.8	Biox III *Finger and ear* NR	All patients were characterized as “pigmented” NR	Bias
Wiles 2022 N=194	UK Clinical	Patients who were critically ill with COVID-19 pneumonitis who received non-invasive respiratory	B1x5 M/P monitoring system using NellcorTM reusable or disposable probes *NR*	NR	Bias, prevalence of occult hypoxemia, Obs

Study Name or Author Year	Country Setting	Participant Characteristics	Pulse Oximetry *Location* CO-oximetry	Skin Pigmentation or Race/Ethnicity Factors Measurement	Outcome Types
		support CPAP, low-flow/high-flow oxygen therapy) within the first 7 days of critical care admission. Median age: 62 ± 12.4 % White: 69.6	RAPIDpoint 500 analyser		
Wiles 2022 Letter N=178	UK Clinical	Patients with COVID-19 pneumonitis who were receiving mechanical ventilation of their lungs Median age: 60 ± 14 years % male: 75 % White ethnic origin: 70.8	B1x5 M/P monitoring system using NellcorTM reusable or disposable probes *NR* RAPIDpoint 500 analyser	NR	Bias, prevalence of occult hypoxemia, obs
Wong 2021 N=NR	US Clinical	Patients with SpO2 within the range of 88%to 100% Median age: 62.2% % female: 42.9 % Black: 29.6	NR	NR	Bias, prevalence of occult hypoxemia, RR both

Abbreviations. ABG=arterial blood gas; NR=not reported, NHW=non-Hispanic White, NHB=non-Hispanic Black, CPAP=continuous positive airway pressure.

PEER REVIEW COMMENTS AND RESPONSES

Comment #	Reviewer #	Comment	Author Response
Are the objectives, scope, and methods for this review clearly described?			
1	1	Yes	
2	2	Yes	
3	3	Yes	
4	4	Yes	
5	5	Yes	
6	6	Yes	
Is there any indication of bias in our synthesis of the evidence?			
7	1	No	
8	2	No	
9	3	No	
10	4	No	
11	5	No	
12	6	No	
Are there any published or unpublished studies that we may have overlooked?			
13	1	No	
14	2	No	
15	3	No	
16	4	Yes - Okunlola et al 2021 Respiratory Care is important for putting pulse oximeter errors in perspective.	Thank you for this suggestion. We have incorporated this recommendation paper into the Discussion.
17	5	No	
18	6	No	
Additional suggestions or comments can be provided below.			
19	1	None	
20	2	None	

Comment #	Reviewer #	Comment	Author Response
21	3	p.v. line 5/6 and line 11. Why do the first two key findings use the mitigating word “likely”? The data which has been graded as moderate strength shows that pulse oximeters often/typically overestimate Black patients’ O2 sat and that occult hypoxemia is more common among Black than White patients. Using mitigating language, while scientifically accurate, detracts from the very important message of this review. I suggest re-writing these top two key findings either by just removing the word “likely” or re-phrasing the sentences (for example, Pulse oximeters are more likely to overestimate Black or African American than White patients’ blood oxygen saturation level……)	Use of “likely” (or “may”) correspond to strength of evidence ratings (moderate or low, respectively), which we have clarified in the Methods section.
22	3	p.vi. line 30-32. “In Black or African American patients, pulse oximeters appear to overestimate blood oxygen saturation by an average of 150% compared with CO-oximetry in arterial blood...” I do not understand what is meant by 150%. This would benefit from more explanation. This same phrasing is used in the main findings section on p8. Lines 28-30, where is again is unclear to me what this means and would benefit from more explanation.	Thank you. This was a typographical error on our part that has been corrected.
23	3	On p9, line 30, I don’t understand how the addition of 68,455 paired observations with a mean bias of 1.4 to the Shi 2022 review of 2,646 paired observations with mean bias of 0.31, resulted in an updated estimate that is basically the same as the original mean bias described in the Shi paper (0.31). P9, line 48. Again, I don’t understand how the updated estimate of mean bias changed so little from Shi’s estimate, and I really don’t understand how the lower bound of the confidence interval moved lower than for either the Shi 2022 review of the recent studies. I assume that my lack of understanding is just because I don’t understand the math involved (I chose not to read Tang et al’s paper that is cited as the reference for the method of pooling the estimates because most other clinicians reading this ESP report also will not read the Tang paper). Provided the math is right and it is my lack of math knowledge that is the problem, I recommend explaining somewhere how the math has created these non-intuitive situations. This could be explained in the discussion – or it could be pre-empted by more explanation in the “synthesis” section.	Thank you for pointing this out. The reason the overall (updated) estimate is quite similar to the Shi estimate is because the precision estimates/standard deviations from the studies included in Shi were much smaller than those in the more recent (and much larger) studies. This results in Shi’s estimate being more heavily weighted in the final meta-analysis that pools pre-2021 and recent evidence, regardless of the fact that the more recent evidence includes a much larger amount of data.

Comment #	Reviewer #	Comment	Author Response
24	3	Three minor comments: 1) P12. Line 8. Final word on this line is “few”, I think it is supposed to be “fewer”. 2) P12, line 21. “No or delayed recognition of eligibility..” is hard to parse. I think this sentence means “Failure to recognize eligibility or delayed recognition of eligibility…” 3) P13, line 29-29. Cites ref 28 as having proposed use of a correction factor, whereas the authors of this paper explicitly state that a correction factor cannot be used. It would be a more accurate reflection of this paper to place the citation at the end of the sentence ie after stating that this approach “may have limited efficacy because the accuracy and variability of pulse oximetry readings is influenced by a number of factors that cannot be accounted for in a single correction factor.”	Thank you. We have corrected these issues.
25	4	This is an important and timely review that accurately describes a probable increased rate of missed diagnosis of hypoxemia in black patients monitored with current pulse oximeter technology. I think the authors have done a good job of capturing the concept that pulse oximeters can be biased to read too high in black patients. This issue has been observed multiple times, in multiple clinical scenarios, and with varying methodologies. Based on the weight of the published evidence, there is basis for concern and review.	Thank you for this comment.
26	4	I think the review needs to take several additional steps to increase its value. The first is that the reason for clinical errors in pulse oximeter function are not well understood, and contradict some very good clinical laboratory evidence that the magnitude of intrinsic pulse oximeter errors is smaller than observed clinically. The review by Okunlola points out this issue clearly This point is important because it emphasizes that factors in the clinical environment can amplify a small bias in the technology. Merely fixing this bias by technological means is unlikely to erase the structural issues with black patients having poor access to care, presenting to the hospital sicker, and possibly receiving a lower quality of care in the hospital. Structural racism will not yield to minute improvements in pulse oximeter technology.	Thank you for this feedback. We have elaborated on several of these points, particularly in the Discussion section.

Comment #	Reviewer #	Comment	Author Response
		The other missed opportunity is to emphasize that any medical technology is imperfect and needs to be interpreted with clinical judgement. Emphasizing that clinicians owe black patients more skepticism and investigation based on pulse oximeter readings is important at the present time. The problem is that clinical decision making is often based on a lab value threshold—e.g. administering supplemental oxygen at a saturation threshold of 89% or lower, not recognizing that if the pulse oximeter reads 89% the true saturation may be 85%-96% depending on signal quality, due to such factors as low perfusion, misapplied probes, patient movement, clinician inexperience, etc. This would have much more impact than waiting for manufacturers to fix an error that is probably in the 1-2% range.	
27	4	One problem that you want to fix is on page 20, line 30. The data do not indicate that pulse oximeters overestimate saturation by 150%-- that is clearly a wrong conclusion. The right analysis is rate of missed hypoxemia, assuming a threshold and range, as done by Sjoding and colleagues from the Univ. of Michigan. Mean bias or Arms values do not capture the clinically important ranges of readings or errors.	Thank you. This was a typographical error on our part that has been corrected.
28	5	None	
29	6	This report is extremely well written. Complex information is presented in a lucid and comprehensive manner. I deeply appreciated the detailed explanations of systematic review methodology and synthetic algorithmic evaluation.	Thank you for this comment.
30	6	Ln. 28 pg. vi: Please indicate whether any of the studies come from hospital systems comparable to VHA in any way.	Where such studies were not already highlighted, we have attempted to emphasize when a study or evidence base is comparable to the VA setting (eg, from integrated health systems).
31	6	Ln 30ff pg. vi: Are these patients comparable to Veterans in any way? Veterans are sicker, may have more toxic exposures etc. This should be considered to ensure the patient populations are truly comparable. If they are close in comparison, or are not representative or comparable to Veterans, it should be pointed out. I note that you bring out VA evidence a little later in the review. Perhaps moving it up in the review would be better.	We agree that Veterans are a unique population with a greater burden of chronic disease than the general population. At the same time, the complex and dynamic array of factors (including skin pigmentation level) that likely result in differential pulse oximeter accuracy make it challenging to identify specific patient characteristics (outside of skin pigmentation level, or by potentially imprecise proxy, race/ethnicity) that are relevant to use to evaluate the comparability of non-VA study samples

Comment #	Reviewer #	Comment	Author Response
			with the VA patient population. In our view, the most feasible way of addressing this (important) consideration is to emphasize studies conducted in the VA patient population, which we have done.
32	6	Ln 39 pg vii: It's not just structural "inequities," the disparities in this report are terrific examples of structural racism. The biases and unequal outcomes described in pulse oximetry readings here are racist (i.e., skin-tone based, poor correlation between science of device development and real-world use, etc.), maintained by structural factors in US healthcare to which VHA is not immune. That the historic component of this problem was well understood for decades yet no action was taken to address it is symptomatic of structural encasements that render the suffering of Black patients invisible. White Supremacy, therefore, is at the heart of this problem. What remains troubling is that I could not find the word 'racism' used to describe the underpinnings of this issue anywhere in this synthesis. By not doing so, the case is being (inadvertently) made that the lungs of Black or African Americans are somehow different biologically than those of White people, and therefore are more diseased. The ultimate point of this review may well be the dismantling of racially-tailored clinical practices in favor of identification and evaluation of more precise biomarkers. Yet, context of the likely structural conditions (i.e., the harms of clinicians using phenotypic race to determine treatment decisions) are merely hinted at in this review. Structural racism is not neutral, it is a harm to Veteran patients. More precision regarding how race is being defined and operationalized here would be helpful (e.g., self-identified, etc.). In addition, what was it about the COVID-19 pandemic that heightened awareness of the discrepancies in pulse oximetry? Was it more Black patients with poor lung function in the ER? Or similar? Because of the historicity of the problems, the harms are not new. Yet, something about COVID-19 brought this up and out.	Thank you for raising these important points. We have revised the section in which "inequities" were mentioned. Regarding the point about race definitions, we have added a section to the Discussion about the use of race/ethnicity (versus objective skin pigmentation measures) in accuracy studies. It is likely that there were many factors related to the pandemic that broadened awareness of biases in pulse oximeter accuracy, potentially including the ones you have mentioned. Another contributing factor may be the exponential increase in accuracy data that became available due to, or at least contemporary with, the pandemic.
33	6	Because systematic reviews are often predicated on assessing bias in published studies, and that this research was undertaken to examine racial bias, more selective wording or terminology	Thank you. We have attempted to better distinguish the various meanings of "bias" used throughout the review.

Comment #	Reviewer #	Comment	Author Response
		would be advisable. At times, I was confused as to which bias the authors were referring: risk of bias in systematic reviews versus bias in pulse oximetry readings.	
34	7	Well written with appropriate methods that are well explained. The tables are useful, though a figure could perhaps aid the reader in visually understanding the magnitude of differences better, as the main table has a lot of numeric data, perhaps akin to a forest plot in a meta analysis or systematic review.	Thank you for this comment.
35	7	Since the principal mechanistic argument is that skin pigmentation affects the wavelength of light being perceived by the photometer in the SpO2 probe, it would be helpful if a bit more discussion about how this was operationalized in these studies was done. For example, in the Shi paper that is referenced, did they use something semi-quantitative like the Fitzpatrick scale, or was based on the investigators in the study or something else? This would help clinicians who are familiar with the ways that skin tone is semi-quantitatively measured understand how that factored into the research that derived that SpO2 may be unreliable.	Thank you. We have added considerable discussion of this topic in the Discussion section.
36	7	page vi line 32: " an average of 150% compared with CO-oximetry in arterial blood (pooled mean bias = 1.54, 95% CI [0.99, 2.10])" wouldn't that be 54% overestimation? also CI crosses 1, is this significant?	Thank you. This was a typographical error on our part that has been corrected.
37	8	The report is quite thoughtful, rigorous, and extensive. It substantially exceeds any other synthesis yet available, and I hope it will be published. My compliments to the team for their care.	Thank you for this comment.
38	8	My major advice is clarifying the Table headings on page 9 -- moving between things where the null value is 0 and the null value is 1 was a little hard for me, given how often the Bias and Precision (for which the null value is 0) happened to have values in the "plausible aOR range of 1.xx". Similarly, in the footnote to that table, consider defining k.	Thank you for raising these points. We have simplified the table for clarity. *k* was previously defined in the table footnote.