Accuracy of Heart Rate Measurement with Wrist-Worn Wearable Devices in Various Skin Tones: a Systematic Review

Daniel Koerber1 · Shawn Khan1 · Tahmina Shamsheri2 · Abirami Kirubarajan1,3 · Sangeeta Mehta1A,5

Received: 5 August 2022 / Revised: 21 October 2022 / Accepted: 1 November 2022 / Published online: 14 November 2022
© W. Montague Cobb-NMA Health Institute 2022

Abstract
Background  Wearable consumer technology allows for the collection of a growing amount of personal health data. Through the analysis of reflected LED light on the skin, heart rate measurement and arrhythmia detection can be performed. Given that melanin alters skin light absorption, this study seeks to summarize the accuracy of cardiac data from wrist-worn wearable devices for participants of varying skin tones.

Methods  We conducted a systematic review, searching Embase, MEDLINE, CINAHL, and Cochrane for original studies that stratified heart rate and rhythm data for consumer wearable technology according to participant race and/or skin tone.

Results  A total of 10 studies involving 469 participants met inclusion criteria. The frequency-weighted Fitzpatrick score for skin tone was reported in six studies (n = 293), with a mean participant score of 3.5 (range 1–6). Overall, four of the ten studies reported a significant reduction in accuracy of heart rate measurement with wearable devices in darker-skinned individuals, compared to participants with lighter skin tones. Four studies noted no effect of user skin tone on accuracy. The remaining two studies showed mixed results.

Conclusions  Preliminary evidence is inconclusive, but some studies suggest that wearable devices may be less accurate for detecting heart rate in participants with darker skin tones. Higher quality evidence is necessary, with larger sample sizes and more objective stratification of participants by skin tone, in order to characterize potential racial bias in consumer devices.

Keywords  Racial bias · Wearable devices · Photoplethysmography · Heart rate · Heart rhythm

Introduction
There is growing interest in using consumer wrist-worn wearable technology to collect personal health data, including heart rate and rhythm. Several companies including Apple, Samsung, and Fitbit have been approved by the Food and Drug Administration to market these sensors as a form of early detection of arrhythmias such as atrial fibrillation. The associated hardware in modern devices relies on two components: photoplethysmography (PPG), which passively monitors heart rate and rhythm, and electrical sensors, which require the user to remain still and touch the device with their opposite hand throughout the detection period [1].

Given that the electrical sensors require active user engagement to measure heart rate and rhythm, the majority of wearable heart rate and rhythm data is collected using PPG. This technology relies on LEDs projecting light onto the skin, and photodetectors measuring the quantity of reflected light [1, 2]. Hemoglobin in dermal blood vessels drives the bulk of light absorption, and thus changes in blood pressure and flow rate affect light absorption, which is then detected by PPG optical sensors [2, 3]. Proprietary algorithms then convert light absorption into heart rate and rhythm data.

While safe and inexpensive, the accuracy of PPG is impaired by factors that impede light transmission, such as elevated body mass index, tattoos, and higher hair follicle density. Given the potential for racial bias in consumer technology, this study seeks to summarize the accuracy of cardiac data from wrist-worn wearable devices for participants of varying skin tones.
Hypoxemia (pulse oximetry > 92% and arterial oxygen saturation < 88%), compared to 3.6% white patients, and in the second cohort (n = 37,308), 17.0% Black patients had undetected hypoxemia, compared to 6.2% white patients. Similarly, a retrospective study of 7126 patients with COVID-19 found that pulse oximetry overestimated arterial oxygen saturation among Asian, Black, and Hispanic patients compared to white patients, leading to a delay in the initiation of guideline-based therapies.

While the inaccuracy of transmission PPG used in pulse oximeters is well established in darker skin tones, it remains unclear whether this holds true for the reflectance PPG used in consumer wrist-worn devices. Our systematic review aims to review current literature on the accuracy of cardiac data of these devices in populations of various skin tones.

**Methods**

**Search Strategy**

The authors systematically searched four databases from database inception to Nov 5, 2021: (1) the Ovid versions of MEDLINE and MEDLINE Daily including e-publications, in progress, and non-indexed citations; (2) Embase Classic and Embase; (3) CINAHL; and (4) Cochrane CENTRAL. The search included all original studies in any language that evaluated race in measuring cardiovascular health data by consumer wearable technologies (Table 1). No exclusions were applied on the basis of language or country of origin. The complete search strategy can be found in Table 2.

**Study Screening**

This review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Title and abstract screening, full-text review, and data extraction were performed independently by two investigators (SK and DK), with a third reviewer (AK) resolving discrepancies. Backwards searching was also performed by screening the reference lists of all included studies to identify relevant articles. Findings regarding race outcomes were extracted from included studies and qualitatively analyzed.

**Terminology**

In this review, we use the term “white” to refer to the classification of people of European descent with largely lighter skin tones. We use the term “Black” to refer to the classification of people of African and/or Caribbean descent with darker skin tones.

The race-related terms “white” and “Black” are contextual categories shaped by sociocultural forces and how one self-identifies, encompassing various skin tones and ethnic backgrounds. Consequently, several studies use the Fitzpatrick phototype scale to describe skin tone. The Fitzpatrick scale is a semi-quantitative means of describing skin tone, ranging from types 1 to 6, where type 1 skin refers to skin that always burns and does not tan, and type 6 refers to skin that never burns and always tans darkly.

**Assessing Methodological Quality**

The Risk of Bias Assessment Tool for Nonrandomized Studies (RoBANS) tool was used to assess the methodological quality of included studies. Studies were appraised independently by two parallel reviewers (SK, DK).

**Results**

The search strategy yielded 58 results from MEDLINE, 147 from Embase, 65 from CINAHL, and 386 from Cochrane CENTRAL, totaling 656 articles. After removing duplicate articles, 581 records remained. There were 25 records after title and abstract screening; the full texts of these papers were reviewed, and 6 articles met the inclusion criteria. An additional two studies were found through hand searching the literature, and 2 studies through backwards searching, yielding a total of 10 included studies. The results of the systematic search are presented in a PRISMA flowchart in Fig. 1 and the search strategy is presented in Table 2.

Table 1 outlines the characteristics of included observational studies. The 10 studies included a total of 469 participants, with a mean age of 34 years (range 20–62); a total of 38% (range 0–67%) of participants were women. All studies were observational cohort studies. The majority of studies were conducted in the USA (n = 5), with the remaining studies arising from Australia (n = 1), Canada (n = 1), France...
### Table 1: Characteristics of included observational studies

| Study       | Country  | Sample size | % Male | Mean age, years | Race/skin tone | Consumer wearable device(s) | Magnitude of HR/ECG discrepancies | Relevant Findings                                                                 |
|-------------|----------|-------------|--------|----------------|----------------|-------------------------------|-----------------------------------|-----------------------------------------------------------------------------------|
| Bent 2020¹  | USA      | 53          | 40%    | 25.6           | Fitzpatrick (FP)| Empatica E4, Apple Watch, Fitbit Charge, Garmin Vivosmart 3, Xiaomi MiBand, Biovotion Everion | No HR/ECG differences across skin tones | There was no statistically significant difference in HR accuracy across skin tones |
| Etiwy 2019² | USA      | 80          | 81%    | 62             | White (n=61), Black (n=14), Other (n=5) | Apple Watch, Fitbit Blaze, Garmin Forerunner 235, TomTom Spark Cardio | No HR/ECG differences across skin tones | Skin color did not impact the accuracy of HR measurements                          |
| Hermand 2019³ | France  | 70          | 48%    | 20             | Mean FP = 3.4 ± 2.2 | Polar OH1 monitor | Not reported | HR accuracy was lower in patients with darker skin (p < 0.001)                     |
| Hochstadt 2021⁴ | Israel | 24          | 80%    | 41             | FP 1–4 (NR), 5 (n=2), 6 (n=2) | Cardiac Sense | Linear regression coefficient = 0.98, p < 0.001 | R-R measurement accuracy was reduced in patients with darker skin (FP 5 or 6) |
| Pasadyn 2019⁵ | USA      | 50          | 68%    | 29             | Asian (n=7), Black (n=1), Hispanic (n=1), White (n=38), Other (n=3) | Apple Watch III, Fitbit Iconic, Garmin Vivosmart, TomTom Spark 3 | p = 0.01 | Skin color altered the accuracy of HR measurements                                 |
| Ray 2021⁶    | Canada   | 18          | 33%    | 27             | FP 1 (n=1), 2 (n=2), 3 (n=5), 4 (n=4), 5 (n=4), 6 (n=2) | Moto, LG, Polar, Huawei, TicWatch, Misfit, Fossil | 50% less HR readings for FP 6 than for lighter skin tones | Four of the seven watches report significantly fewer high-confidence readings for darker skin tones; no difference in accuracy |
| Sañudo 2019⁷ | Spain    | 45          | 100%   | 24             | FP 2 (n=15), 3 (n=15), 5 (n=15) | Apple Watch | No HR/ECG differences across skin tones | The 3 skin types analyzed had no influence on HR values measured during exercise |
Table 1 (continued)

| Study          | Country  | Sample size | % Male | Mean age, years | Race/skin tone | Consumer wearable device(s)                                      | Magnitude of HR/ECG discrepancies | Relevant Findings                                                                 |
|----------------|----------|-------------|--------|-----------------|----------------|-----------------------------------------------------------------|-----------------------------------|--------------------------------------------------------------------------------|
| Shcherbina 2017<sup>8</sup> USA | 60       | 48%         | 38     | Apple Watch, Basis Peak, Fitbit Surge, Microsoft, Band, Mio Alpha 2, PulseOn, Samsung Gear S2 | NR                          | Device error in HR measurement was higher in patients with darker skin tone; data quantifying the impact was not reported |
| Spierer 2015<sup>9</sup> USA    | 47       | 57%         | 28     | Omron HR500U (OHR), Mio Alpha (MA) | Mean average error FP1 = 3 bpm FP6 = 16 bpm | Mio Alpha: Error rates increased linearly with darker skin (p < 0.01); this trend was not seen with the OHR |
| Wallen 2015<sup>10</sup> Australia | 22       | 50%         | 25     | Apple Watch, Fitbit Charge HR, Samsung Gear S, Mio Alpha | FP > 4 (r = 0.94) (p < 0.05) | No statistical differences based on skin color, except for the Apple Watch, where HR accuracy for the FP > 4 group was statistically lower than the FP < 4 group |

*BPM*, beats per minute; *HR*, heart rate; *ECG*, electrocardiography; *FP*, Fitzpatrick rating scale; *MDE*, mean directional error; *NR*, not reported
The most common wearable device brands that were studied included the Apple Watch \((n = 6)\), Fitbit \((n = 5)\), Mio Alpha \((n = 3)\), and Garmin \((n = 3)\). Of note, since only a few studies provided information on which generation of a given device was used for testing, this review groups the devices by manufacturer. Of the ten studies, five compared the relative accuracy of different manufacturers. Of those, all five found the Apple Watch to be more accurate than the other commercial devices \([2, 9–12]\). For instance, Etiwy et al. found the Apple Watch to follow ECG data with a correlation coefficient of 0.80, compared to 0.52 in Garmin devices \([9]\).

As a gold standard comparison for accuracy, either an ECG \((n = 7)\) or a chest strap \((n = 3)\) was used, two modalities that are known to correlate well \([13]\). Similar to ECGs, chest straps use electrodes to measure electrical activity and determine heart rate. None of the included studies compared the detection of atrial fibrillation or other arrhythmias in participants with various skin tones. Similarly, none of the studies assessed oxygen saturation or other vital signs.

In three of the included studies, skin tone was the primary variable considered in the accuracy of wrist-worn heart rate monitors \([1, 2, 14]\). For the seven other studies, skin tone was one of several covariates considered in the accuracy analysis.

The Fitzpatrick scale was used to classify participant skin tones in eight of the ten studies. In the two studies that did not use the Fitzpatrick scale, one classified participants as white, Black, or other, and another classified participants based on ethnic background, including Asian, Black, and Hispanic — precluding aggregate analysis of skin tone outcomes. Three studies outlined the proportion of participants of each skin tone. The Fitzpatrick scale was used to classify participant skin tones in eight of the ten studies. In the two studies that did not use the Fitzpatrick scale, one classified participants as white, Black, or other, and another classified participants based on ethnic background, including Asian, Black, and Hispanic — precluding aggregate analysis of skin tone outcomes. Three studies outlined the proportion of participants of each skin tone.
Fig. 1 PRISMA flowchart

Fig. 2 Risk of Bias Assessment Tool for 10 included non-randomized studies (RoBANS)
Fitzpatrick skin type, three studies solely reported the mean Fitzpatrick score, one study grouped participants as Fitzpatrick < 4 and Fitzpatrick > 4, and one study only reported skin tone data for dark-skinned individuals (Fitzpatrick 5 or 6), which pertained to 4 of the 24 participants. Only three of the ten studies included participants with Fitzpatrick 6 skin tones.

Six studies reported an average Fitzpatrick score, and across these studies the frequency-weighted mean score was 3.5 (range 1–6) (n = 293).

Four of the ten studies reported statistically significant reductions in accuracy of heart rate data in participants with darker skin tones [3, 10, 11, 15]. Four studies found no difference in accuracy between participants of different skin tones [1, 2, 9, 14]. The remaining two studies observed mixed results across different wearable devices [12, 16]. No studies reported a higher accuracy of heart rate measurement in participants with darker skin tones.

Hermand et al. reported lower heart rate accuracy in patients with darker skin (p < 0.001) [15]. Pasadyn et al. similarly found that the accuracy of heart rate detection was slightly lower in non-white participants (p = 0.01) [10]. However, the magnitude of measurement inaccuracy was not quantified in either study. Shcherbina et al. noted that smart watch device error was higher for darker skin tones, but the degree of this effect was not reported [11]. Hochstadt et al. noted a linear regression coefficient of 0.98 (p < 0.001) when comparing PPG to ECG data in patients with darker skin (Fitzpatrick 5 or 6), suggesting that darker skin tone reduced accuracy, albeit with a relatively small impact [3].

An equal number of studies did not support this relationship. Bent et al. found equivalent accuracy for all devices tested across various skin tones [2]. Etiwy et al. and Sanudo et al. similarly found that skin color did not influence heart rate accuracy [1, 9]. Ray et al. found that various WearOS watches systematically underestimated the reliability of HR readings taken from dark skin, despite no substantial differences in error, leading to significantly fewer recorded data points in patients with dark skin [14].

The remaining two studies had mixed findings. Wallen et al. noted lower heart rate accuracy in participants with a Fitzpatrick scale score > 4 with the Apple Watch, but not in the other studied smartwatches [12]. Spierer et al. found that one of the two devices tested (Mio Alpha) had a higher mean average error in Fitzpatrick 6 patients compared to Fitzpatrick 1 patients (16 beats/min compared to 3 beats/ min, respectively) [16].

**Discussion**

This systematic review of 10 studies and 469 participants summarizes the accuracy of heart rate measurement of wearable devices across diverse skin tones. This review identified a relative scarcity of studies that consider the interactions of skin tone when characterizing smart watch device accuracy in cardiac outcomes, resulting in inconclusive findings.

There has been increased interest in the use of wearable devices to measure heart rate and detect arrhythmias, such as atrial fibrillation. Their use is supported by a growing body of studies which have demonstrated accuracy in recording cardiac data. One study showed utility in consumer devices for post-discharge monitoring of tachycardia in ICU patients, noting high sensitivity (99%) with low-to-moderate specificity (70%) [17]. A study by Banerjee et al. found that the sensitivity and specificity of consumer wrist-worn devices were 92% and 88% respectively for atrial fibrillation detection [18].

While some studies have posited that artificial intelligence algorithms may play a role in optimizing data measurement in outpatient and emergency cardiology, including the detection of abnormal heart rates and rhythms, other studies have identified potential racial biases within machine learning [19]. As such, further scrutiny of the tools we use and the data they from which they are derived is necessary in order to reduce bias in medicine. This review highlights the importance of research and development studies enrolling diverse participants, and that validation studies must ensure that devices are tested in a range of skin types.

While the racial limitations of transmission PPG have been consistently demonstrated in pulse oximetry, this systematic review demonstrated that this relationship is less evident in studies performed to date on the reflectance PPG used in wearable devices. One potential reason for this difference in accuracy is the adoption of green light in newer wrist-worn devices, compared to the red light used in pulse oximetry. Fallow et al. identified that green wavelengths (520 nm) displayed greater accuracy in heart rate measurement regardless of skin type when compared to other wavelengths, both at rest and during exercise [20]. Further research may identify the utility of green wavelengths, instead of infrared wavelengths, in reducing racial bias in pulse oximetry.

Overall, this is the first systematic review of the accuracy of cardiac data by wearable devices based on race and/or skin tone. The study used rigorous research methodology including the search of multiple research databases and screening of publications by two independent reviewers in duplicate. However, the included studies described mixed results and as such, it remains unclear whether wearable devices have reduced accuracy in heart rate and arrhythmia detection in people with darker skin tones. This review highlights the importance of research and development studies enrolling diverse participants, and that validation studies must ensure that devices are tested in individuals with a range of skin types.
There are limitations to this systematic review, largely related to the preliminary nature of the evidence base. The identified studies in this review were not blinded, included small sample sizes, evaluated primarily young subjects, and used different wearable device brands and models. Many of the studies that reported a significant interaction between skin tone and cardiac data accuracy did not quantify the magnitude of error. Furthermore, discrepancies existed in how skin tone was reported, such as categorizations of skin tones as either light or dark, or by race, ethnicity, or Fitzpatrick scale subgroups. Some studies grouped participants by Fitzpatrick scale ranges given a shortage of participants at the extremes of skin tone (i.e., Fitzpatrick scale 1 and 6). This heterogeneity precluded meta-analysis and may contribute to the variability in reported results. Further, despite being used as a gold standard in three of the ten included studies, chest straps have never been validated across different skin tones. As a result, the authors recommend that future work should report accuracy of wearable device data stratified by race and/or skin tone, and the use of spectrocolorimeters whenever possible, to provide more objective skin color measurements than the Fitzpatrick scale [21, 22].

**Conclusion**

Early evidence of racial bias in wrist-worn wearables is mixed, but some studies demonstrate reduced accuracy of heart rate data in users with darker skin tones. Further higher quality evidence is needed, involving a greater proportion of patients with darker skin tones, as well as objective measurements of pigmentation, to better characterize potential racial bias in the accuracy of heart rate measurements.

**Author Contribution** All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Daniel Koerber and Shawn Khan. The first draft of the manuscript was written by Daniel Koerber and Shawn Khan, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Declarations**

**Conflict of Interest** The authors declare no competing interests.

**References**

1. Sañudo B, De Hoyos M, Muñoz-López A, Perry J, Abt G. Pilot study assessing the influence of skin type on the heart rate measurements obtained by photoplethysmography with the Apple Watch. J Med Syst. 2019;43.
2. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med [Internet]. 2020;2/13. 2020;3:18. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32047863
3. Hochstadt A, Havakuk O, Chorin E, Schwartz AL, Merdler I, Laufer M, et al. Continuous heart rhythm monitoring using mobile photoplethysmography in ambulatory patients. J Electrocardiol [Internet]. 2020/05/04. 2020;60:138–41. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32361502
4. Nelson BW, Low CA, Jacobson N, Arean P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med [Internet]. 2020/07/03. 2020;3:90. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32513085
5. Khan Y, Han D, Pierre A, Ting J, Wang X, Lochner CM, et al. A flexible organic reflectance oximeter array. Proc Natl Acad Sci U S A. 2018;115:E11015–24.
6. Sjöding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. N Engl J Med. 2020;383:2477–8.
7. Fawzy A, Wu TD, Wang K, Robinson ML, Farha J, Bradke A, et al. Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with COVID-19. JAMA Intern Med. 2022;21224.
8. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol. 1988;124:869–71.
9. Etiwy M, Akhrass Z, Gillinov L, Alashi A, Wang R, Blackburn G, et al. Accuracy of wearable heart rate monitors in cardiac rehabilitation. Cardiovasc Diagn Ther [Internet]. 2019/07/06. 2019;9:262–71. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31275816
10. Pasadyn SR, Soudan M, Gillinov M, Houghtaling P, Phelan D, Gillinov N, et al. Accuracy of commercially available heart rate monitors in athletes: a prospective study. Cardiovasc Diagn Ther [Internet]. 2019/09/27. 2019;9:379–85. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31555543
11. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christie JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med [Internet]. 2017/05/26. 2017;7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28538708
12. Wallen MP, Gomersall SR, Keating SE, Wisloff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One [Internet]. 2016/05/28. 2016;11:e0154420. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27232714
13. Schaffarczyk M, Rogers B, Reer R, Gronwald T. Validity of the polar H10 sensor for heart rate variability analysis during resting state and incremental exercise in recreational men and women. Sensors. 2022;22:6536.
14. Ray I, Liaqat D, Gabel M, E. de L. Skin tone, confidence, and data quality of heart rate sensing in WearOS smartwatches. IEEE Int Conf Pervasive Comput Commun Work other Affil Events. 2021;213–9.
15. Hermand E, Cassirame J, Ennequin G, Hue O. Validation of a photoplethysmographic heart rate monitor: polar OH1. Int J Sport Med [Internet]. 2019/06/13. 2019;40:462–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31189190
16. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol [Internet]. 2015/06/27. 2015;39:264–71. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26112379
17. Kroll RR, McKenzie ED, Boyd JG, Shept P, Howes D, Wood M, et al. Use of wearable devices for post-discharge monitoring of ICU patients: a feasibility study. J Intensive Care [Internet]. 2017/12/05. 2017;5:64. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29201377
18. Banerjee S, Banerjee R, Zhuang L, Persen K. Wrist worn wearable for detecting atrial fibrillation. J Am Coll Cardiol. 2020;75.
19. Kirubarajan A, Taher A, Khan S, Masood S. Artificial intelligence in emergency medicine: a scoping review. J Am Coll Emerg Physicians Open [Internet]. 2021/01/05. 2020;1:1691–702. Available from: https://www.ncbi.nlm.nih.gov/pubmed/33392578
20. Fallow BA, Tarumi T, Tanaka H. Influence of skin type and wavelength on light wave reflectance. J Clin Monit Comput [Internet]. 2013/02/12. 2013;27:313–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23397431
21. Ly BCK, Dyer EB, Feig JL, Chien AL, Del Bino S. Research techniques made simple: cutaneous colorimetry: a reliable technique for objective skin color measurement. J Invest Dermatol [Internet]. The Authors; 2020;140:3–12.e1. Available from: https://doi.org/10.1016/j.jid.2019.11.003
22. Pershing LK, Tirumala VP, Nelson JL, Corlett JL, Lin AG, Meyer LJ, et al. Reflectance spectrophotometer: the dermatologists’ sphygmomanometer for skin phototyping? J Invest Dermatol [Internet]. Elsevier Masson SAS; 2008;128:1633–40. Available from: https://doi.org/10.1038/sj.jid.5701238

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.