Hemoglobin Accuracy Report of SDK 5.9.1

Executive Summary

Goal

This document assesses the accuracy of Hb evaluations in SDK 5.9 [iOS and Android] rPPG by comparing them with invasive blood test results, using data collected from India and Italy.

Results

- The Hemoglobin measured by Binah's SDK was found to be **within the accuracy target (±1.5 g/dl) in 79.4%** of the measurements for iOS and Android and the following confounding factors (see appendix):
    - Both female and male
    - All skin tones (Fitzpatrick I to VI)
    - Ages 18 to 81
    - BMI from light to morbid obesity
    - Distances close and far from the face
    - Luminance from dark to brighter surroundings
    - Similar performance on all devices used for recordings
    - Similar performance in several countries with different ethnicities

Conclusions

This report describes the results of accuracy studies conducted in India and Italy. Binah.ai's SDK 5.9 Hb evaluations were correlated with regular blood test results for both iOS and Android operating systems. The correlation factor was r=0.622 and 79.4% of the measurements for both operating systems were within the target error range.

Introduction

Hemoglobin (Hb) is a family of closely related proteins responsible for oxygen transportation in the body. Each Hb molecule contains four iron atoms, allowing it to bind up to four oxygen (O_2) molecules. It attaches to red blood cells, facilitating oxygen delivery from the lungs to organs and tissues via blood flow. Additionally, Hb interacts with carbon dioxide (CO_2), carbon monoxide (CO), and nitric oxide (NO), which play key biological roles. It also contributes to the disc-like shape of red blood cells, aiding their smooth passage through blood vessels.[1–3]

Hb levels are usually assessed by a blood test. Low Hb levels indicate reduced oxygen in the blood, which can lead to anemia. Symptoms of anemia range from mild fatigue, weakness, and lethargy to severe cases involving syncope, shortness of breath, and decreased exercise tolerance. Hb is measured in grams per decilitre (g/dL), with normal levels ranging from 13-18 g/dL in males and 12-15 g/dL in non-pregnant females.[4,5]

Currently, Hb assessment requires an invasive blood test. Replacing this method with a non-invasive, cost-effective, and remote alternative would greatly enhance accessibility, affordability, and usability.

PPG (photoplethysmography) is a non-invasive, simple, and low-cost tool that can reflect blood flow and blood volume changes in blood vessels. The PPG waveform comprises a pulsatile ('AC') physiological waveform attributed to blood volume changes with each heartbeat and is superimposed on a slowly varying ('DC') baseline with various lower frequency components attributed to respiration, sympathetic nervous system activity, and thermoregulation. PPG technology has been used in a wide range of commercially available medical devices to measure blood pressure, oxygen saturation, cardiac output, and to assess autonomic function.[6] Hb can also be evaluated by devices using PPG technology. FDA approved such a device in a range of 8-17 gr/dl with a measurement error of ±1 gr/dl.[7]

Camera-based approaches make it possible to derive remote PPG (rPPG) signals, and therefore might enable a remote and non-invasive measurement of blood parameters. Binah.ai’s Hb algorithm uses the rPPG signal recorded from facial skin tissue. The algorithm extracts face video images, produces an rPPG signal, analyzes the data using AI, and provides the end-user with a Hb measurement in real-time.

This report describes the results of accuracy studies conducted in Israel, India, South Africa, Italy, and Nepal that compare Binah.ai’s Hb evaluations with the results of regular, invasive blood tests.

Methods

Binah.ai's Hb measurements were compared to the Hb values received in regular blood tests of all subjects.

Measurement set-up:

In all sites, each participant was instructed to sit as stable as possible. Recordings were conducted in a testing room, with controlled and fixed artificial ambient light.

For rPPG measurements, a mobile device was placed on a stand in front of the participant. The participant's face filled most of the frame's area (distance of about 20-40 cm) and was positioned in the center of the frame. The camera was set at the forehead's level and perpendicular to the face.
Participants were instructed to look at the screen throughout the recording.
Participants were instructed to take off their glasses and to avoid any movement, including talking, and were required to sit still with their feet flat on the floor. Each recording lasted 60 seconds.

Blood tests were conducted as close as possible to the rPPG measurement to minimize timing discrepancies.

Statistical analysis:

Accuracy was calculated using the following parameters:

$$AE\ (Absolute\ Error) = \ App_i - Ref_i$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (App_i - Ref_i)^2}{N}}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |App_i - Ref_i|$$

When,
N is the number of data points.
App is the measurement of the Binah.ai's application.
Ref is the reference results.
i is the index number of measurements.

For this report, Binah.ai's SDK 5.9 was compared to invasive blood test results.
The measurements were recorded in several locations in Israel, India, Italy, South Africa, and Japan using the mobile device models listed below:

iOS: iPhone 13 Pro, iPhone 13 Pro Max, iPhone 14, iPhone 14 Pro max, iPhone 14 Plus.
Android: Samsung S21 Ultra, Samsung S23 Ultra, Pixel 6 Pro, Google pixel 8a, Xiaomi 14 CIVI.

Results

Measurement disposition

Number of subjects/measurements with reported Hb: 284/499

Number of Unique Subjects and Measurements by country and Hb distribution

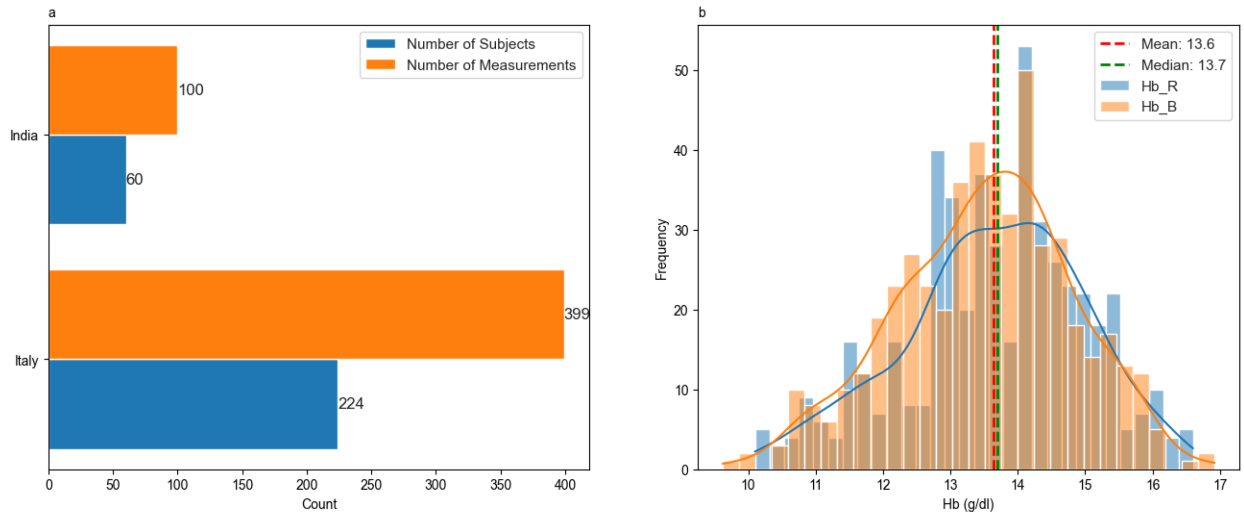


Figure 1:
a. **Number of Unique Subjects and Measurements (with reference values) for each country.**
b. **Distribution of Hb measured by reference device and Binah.ai's application.** The mean and median lines are calculated for the reference values.

Demographics Data:

Subjects/Measurements	Age (mean ± std)	BMI (mean ± std)	Sex (F/M)	
284/ 499	53.7 ± 14.3	27.6 ± 6.0	162 / 122	
Fitzpatrick Skin Tone (I/II/III/IV/V/VI)	Beard (No/Yes)	Glasses (No/Yes)	Face cream (No/Yes)	
25 / 150 / 61 / 23 / 20 / 5	218 / 66	208 / 76	212 / 72	
Distance (mean ± std)	Luminance (mean ± std)	Angle yaw (mean ± std)	Angle roll (mean ± std)	Angle pitch (mean ± std)
0.26 ± 0.04	121.0 ± 77.8	4.9 ± 4.0	2.2 ± 1.9	8.2 ± 6.5

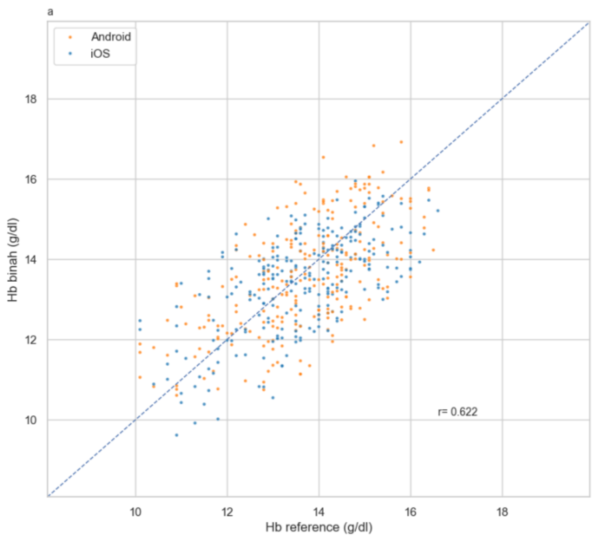
Table 1: Demographic data for experiments using Binah.ai application.
* Fitzpatrick skin tone classifications are I- Pale white, II- white, III- Darker white, IV- Light brown, V- Brown, VI- Dark brown or black. ** Skin tone, beard, glasses, and face cream information does not exist for all subjects

Accuracy Data:

OS	Unique Subjects	Measurements	MAE±std	Ref Range
iOS	253	253	0.9 ± 0.6	10.0 - 17.0
Android	246	246	1.0 ± 0.7	10.0 - 16.0

Table 2: Accuracy data (MAE±std) when Binah.ai's and the reference device's measurements are compared in the presented Hb range (Ref range).
MAE - Mean Absolute Error, std - Standard Deviation

Correlation and Bland-Altman plots



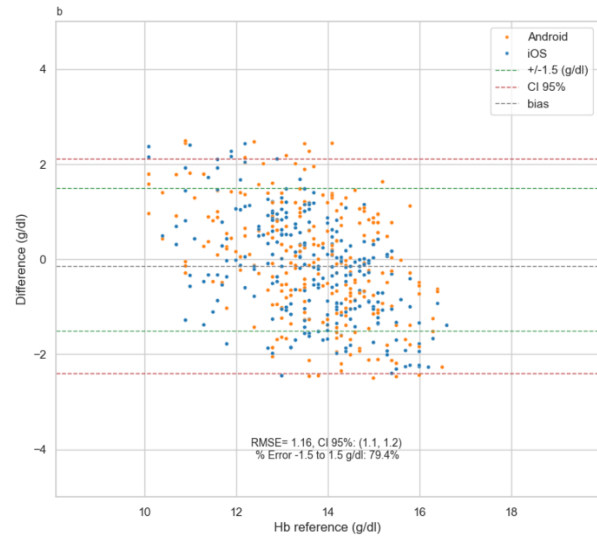


Figure 2:
a. Correlation plot by operating system - Binah.ai's Hb estimations versus invasive blood test results were found to be correlated with an r=0.622 for both operating systems (Android and iOS).
b. **Bland-Altman plot by operating system** - Comparison between Hb measurements of the two methods (Binah.ai and the blood test results) in the presented reference range.

The "Bias" gray dashed line stands for the mean difference between measurements of Binah.ai and the blood tests results, the "Error" green dashed lines of ±1.5 g/dl represent the value of the accuracy criterion, the "Limits of agreement" red dashed lines mark the limit of 95% of the samples

Conclusions

This report describes the results of accuracy studies conducted in India and Italy. Binah.ai's Hb evaluations were correlated with regular blood test results for both iOS and Android operating systems. The correlation factor was r=0.622 and 79.4% of the measurements for both operating systems are within the target error range.

References

1. Schechter, A. N. Hemoglobin research and the origins of molecular medicine. *Blood, J. Am. Soc. Hematol.* **112**, 3927–3938 (2008).
2. Farid, Y., Bowman, N. S. & Lecat, P. Biochemistry, hemoglobin synthesis. (2019).
3. Pandya, N. K. & Sharma, S. Capnography and pulse oximetry. in *StatPearls [Internet]* (StatPearls Publishing, 2021).
4. Turner, J., Parsi, M. & Badireddy, M. Anemia. in *StatPearls [Internet]* (StatPearls Publishing, 2022).
5. Guralnik, J. M., Ershler, W. B., Schrier, S. L. & Picozzi, V. J. Anemia in the elderly: a public health crisis in hematology. *ASH Educ. Progr. B.* **2005**, 528–532 (2005).
6. Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **28**, (2007).
7. Masimo. Masimo brochure - Noninvasive and Continuous Hemoglobin (SpHb®) Monitoring. https://www.masimo.com/siteassets/us/documents/pdf/plm-10283b_brochure_total_hemoglobin_sphb_us.pdf.

Appendix

Hemoglobin error by Age and BMI

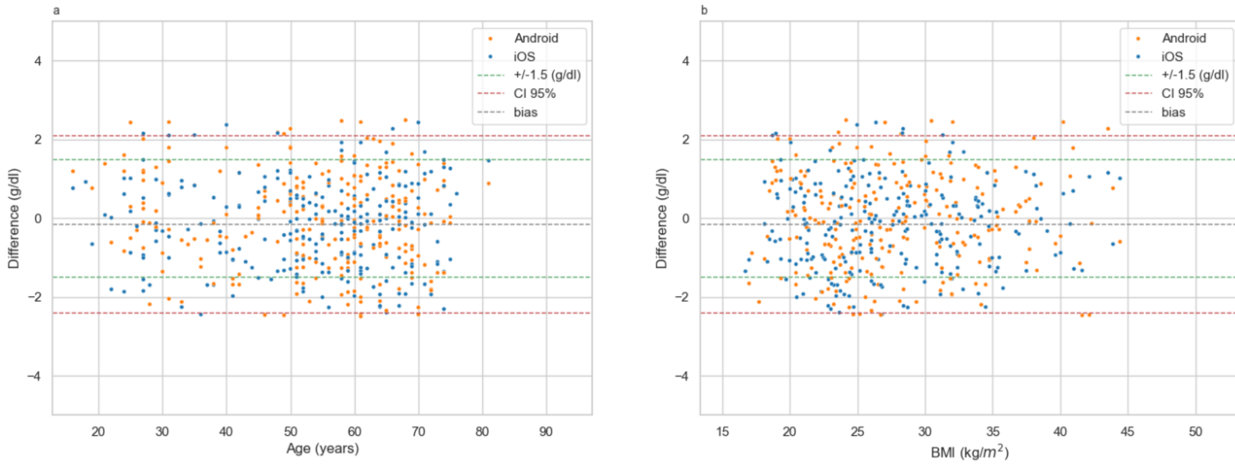


Figure 4:

a. **Bland-Altman plot by age** - Comparison between Hb measurements obtained by Binah.ai and the reference device for both operating systems (Android and iOS) within the presented age range.

b**. Bland-Altman plot by BMI -** Comparison between Hb measurements obtained by Binah.ai and the reference device for both operating systems (Android and iOS) within the presented BMI range.

The "Bias" gray dashed line stands for the mean difference between measurements of Binah.ai and the blood tests results, the "Error" green dashed lines of ±1.5 g/dl represent the value of the accuracy criterion, the "Limits of agreement" red dashed lines mark the limit of 95% of the samples

Hemoglobin error by skin tone with Gender and OS

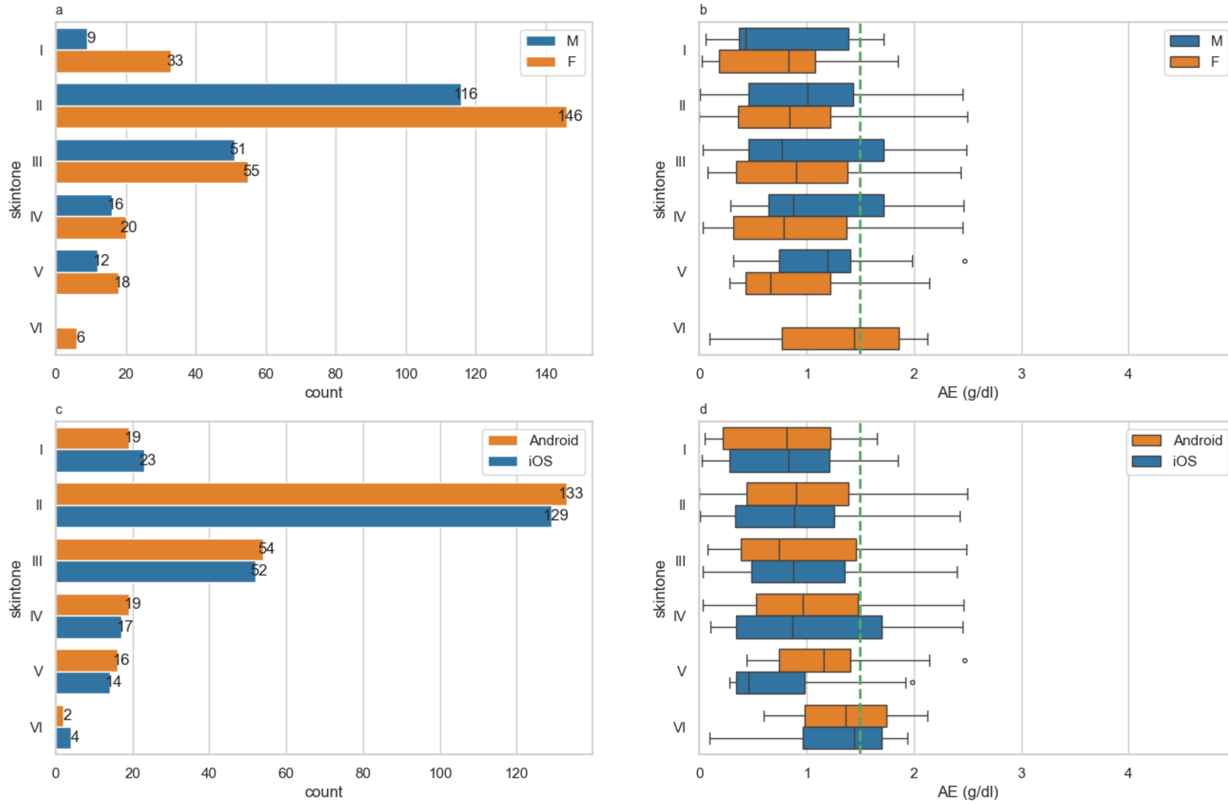


Figure 5:

a. **Number of measurements by Fitzpatrick skin tone and sex** (female and male).

b. **Box plot by Fitzpatrick skin tone and Sex** – Hb measurements obtained by Binah.ai versus the reference device for both sexes (female and male) across all presented skin tones.

The green dashed "Error" lines set at 1.5 g/dl represents the accuracy criterion's value.

c. **Number of measurements by Fitzpatrick skin tone and operating system** (Android and iOS).

d. **Box plot by Fitzpatrick skin tone and operating system** - Hb measurements obtained by Binah.ai versus the reference device presented by OS (Android and iOS) across all skin tones.

The green dashed "Error" lines set at 1.5 g/dl represents the value of the accuracy criterion's value.

Hb error by distance and luminance

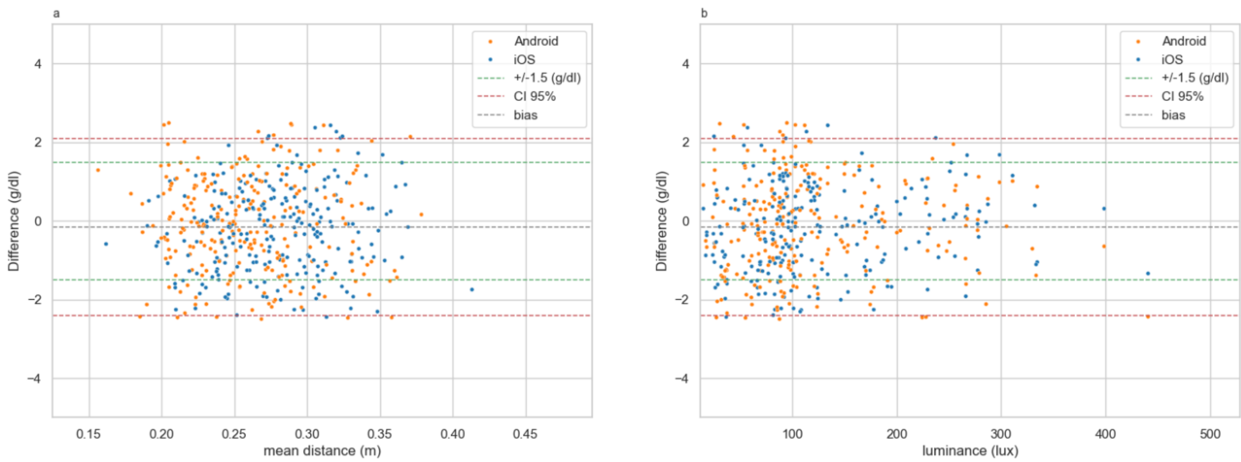


Figure 6:
a. **Bland-Altman plot by distance (m)** - Comparison between Hb measurements obtained by Binah.ai and the reference device for both operating systems (Android and iOS) within the presented distance range.
b**. Bland-Altman plot by luminance (lux)** - Comparison between Hb measurements obtained by Binah.ai and the reference device for both operating systems (Android and iOS) within the presented luminance range.

The "Bias" gray dashed line stands for the mean difference between measurements of Binah.ai and the blood tests results, the "Error" green dashed lines of ±1.5 g/dl represent the value of the accuracy criterion, the "Limits of agreement" red dashed lines mark the limit of 95% of the samples

Hb error by face Angles

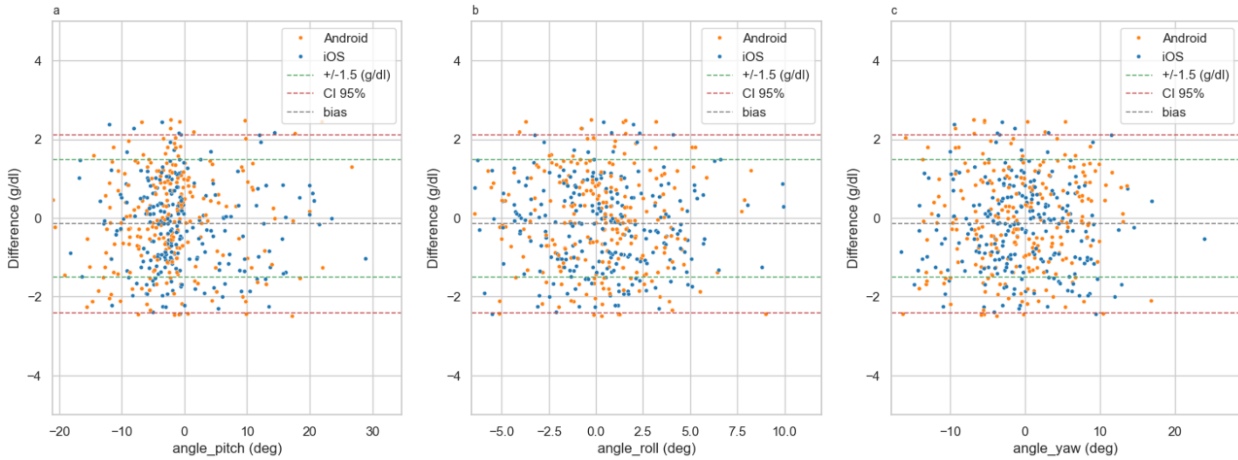


Figure 7:
a. **Bland-Altman plot by pitch angle (deg)** - Comparison between Hb measurements obtained by Binah.ai and the reference device for both operating systems (Android and iOS) within the presented pitch angle range.
b. **Bland-Altman plot by roll angle (deg)** - Comparison between Hb measurements obtained by Binah.ai and the reference device for both operating systems (Android and iOS) within the presented roll angle range.
c**. Bland-Altman plot by yaw angle (deg)** - Comparison between Hb measurements obtained by Binah.ai and the reference device for both operating systems (Android and iOS) within the presented yaw angle range.

The "Bias" gray dashed line stands for the mean difference between measurements of Binah.ai and the blood tests results, the "Error" green dashed lines of ±1.5 g/dl represent the value of the accuracy criterion, the "Limits of agreement" red dashed lines mark the limit of 95% of the samples

Boxplot of Hb error by Mobile Device Models

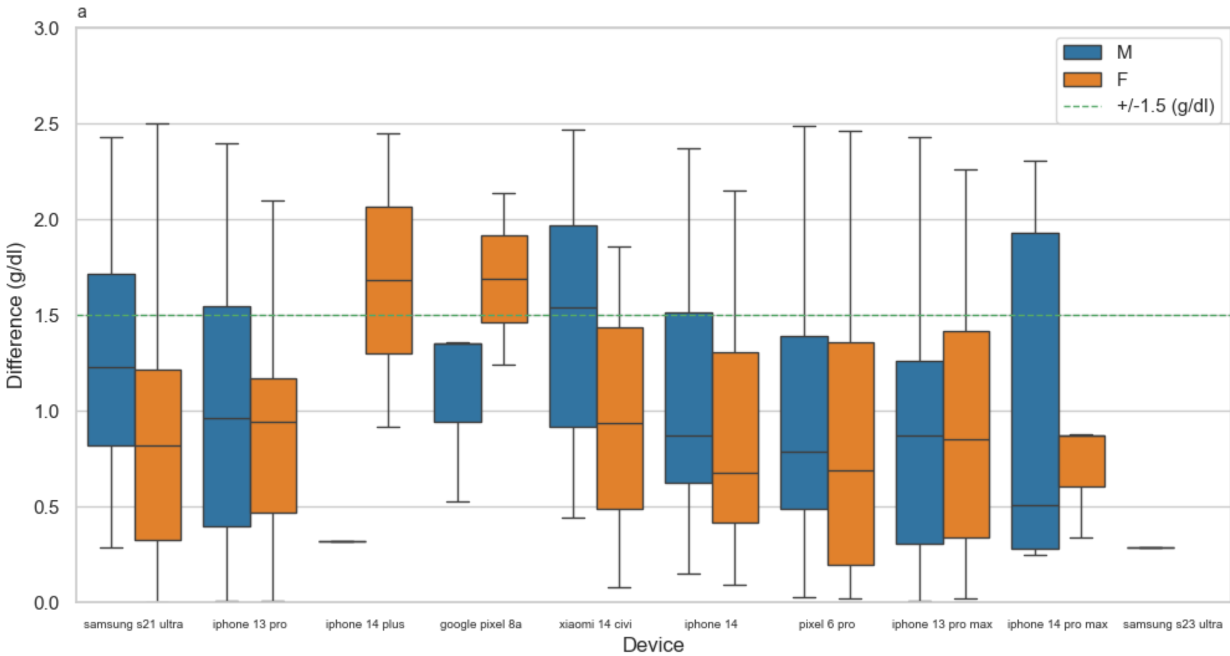


Figure 11:

Error by Mobile device models - Hb measurements obtained by Binah.ai versus the reference device for both sexes (female and male) presented by mobile device models.

The green dashed "Error" line set at 1.5 g/dl

Boxplot of Hb error by Country

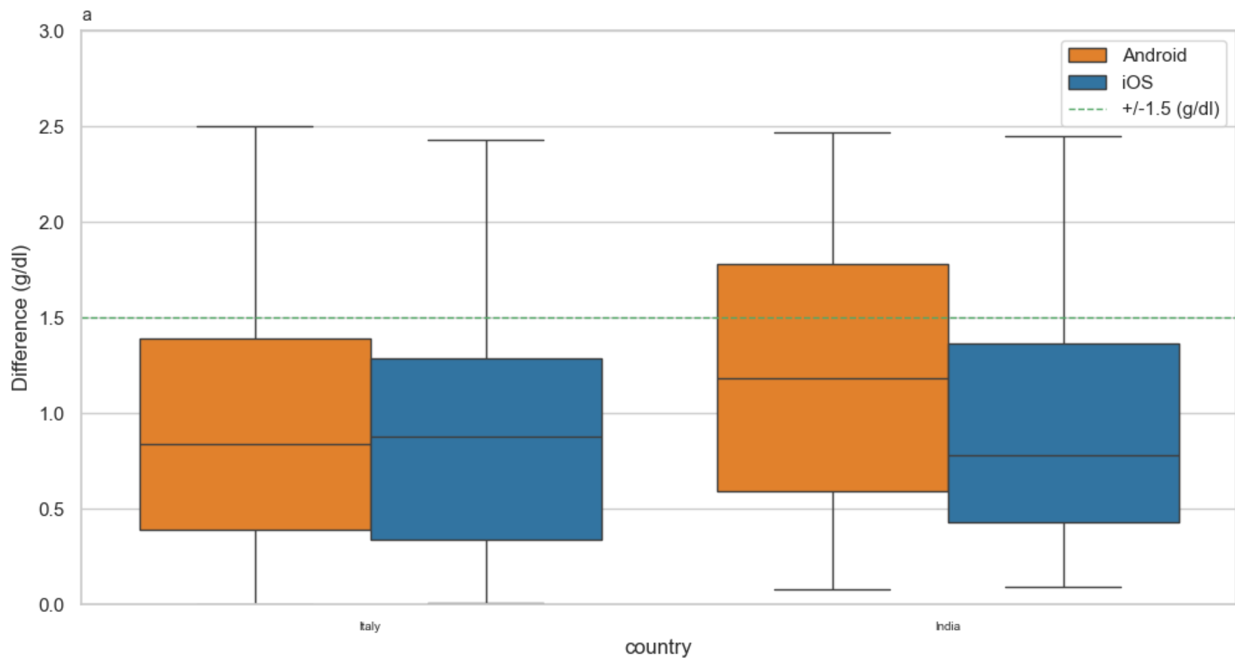


Figure 12:
Error by country - Hb measurements obtained by Binah.ai versus the reference device for both operating systems (Android and iOS), presented by country.
The green dashed "Error" line set at 1.5 g/dl