Overview of the Repeatability, Reproducibility, and Agreement of the Biometry Values Provided by Various Ophthalmic Devices

JOS J. ROZEMA, KRISTIEN WOUTERS, DANNY G.P. MATHYSEN, AND MARIE-JOSÉ TASSIGNON

• PURPOSE: To present an overview of the measurement errors for various biometric devices, as well as a meta-analysis of the agreement between biometric devices using the Pentacam, Orbscan, and IOL Master as a reference.

• DESIGN: Meta-analysis of the literature.

• METHODS: The meta-analysis is based on data from 216 articles that compare a total of 24 different devices with the reference devices for the following 9 parameters: mean, steep and flat curvature of the anterior and posterior cornea; central corneal thickness; anterior chamber depth; and axial length. After the weighted average difference between devices has been determined, the "two one-sided t test" was used to test for equivalence between devices within certain thresholds defined by the measurement errors and the influence of these differences on the calculated refraction.

• RESULTS: In only 17 of the 70 comparisons a device was equivalent with the reference device within the thresholds set by the measurement error. More lenient thresholds, based on a change in calculated refraction of ±0.25 diopter, increased this number to a maximum of 25/50 comparisons (excluding pachymetry). High degrees of inconsistency were seen in the reported results, which could partially explain the low agreement between devices.

• CONCLUSION: As a rule, biometry measurements taken by different devices should not be considered equivalent, although several exceptions could be identified. We therefore recommend that clinical studies involving multiple device types treat this as a within-subject variable to avoid bias. The follow-up of individual patients using different devices should be avoided at all times. (Am J Ophthalmol 2014;158:1111–1120. © 2014 by Elsevier Inc. All rights reserved.)

R ELIABLE BIOMETRY IS AN ESSENTIAL PART OF ANY ophthalmic practice. For this reason, a wide range of instruments have been put on the market, each purporting to have the highest degree of accuracy. But although these devices have undoubtedly been developed and tested according to the state of the art, they will all suffer from errors and sensitivities intrinsic to the measurement process. These present themselves as variations between consecutive acquisitions, and between acquisitions made by different observers or devices.

For individual clinics this large variety in available equipment is inconsequential, as they usually own only a few instruments operated consistently by a limited number of experienced individuals. This ensures that the best possible biometry is obtained in the clinical follow-up of each patient. But for research projects, often involving any number of clinics working together, each with their own choice of biometry devices, these equipment differences may form an important confounding factor to the study design.

Typically this problem is handled in 2 ways: either only clinics that own a particular "preferred" device are selected for participation, or equipment differences are simply ignored. Both approaches have their disadvantages, however, as selection may exclude clinics that could otherwise make interesting contributions to the study, and ignoring the differences may introduce systematic errors into the analysis. Hence, before ignoring equipment differences and inserting measurements from different devices into a single database, it is imperative that their agreement and measurement errors have been validated statistically. But although this has already been done quite extensively in the literature for many parameters and many combinations of devices, the results in these reports are often contradictory to each other and authors have very different opinions on what may be considered an acceptable error.

It is therefore the aim of this work to present an overview of the measurement errors of the biometry devices available in the literature, as well as of the agreement between these devices and 3 of the most commonly used clinical instruments: the Pentacam Scheimpflug camera (Oculus Optikgeräte, Wetzlar, Germany) and Orbscan scanning slit corneal topographer (Bausch & Lomb, Bridgewater, New Jersey, USA) for the anterior segment biometry, and the IOL Master (Carl Zeiss, Jena, Germany) for axial length

measurements. Any systematic differences identified will then be compared to a number of clinical standards, in order to verify which of these differences may have an actual clinical impact.

METHODS

• LITERATURE SEARCH: The literature search was performed in 2 parts: first, a search for reports on the measurement errors of biometric devices, followed by a second, independent search for reports comparing the measurements of biometry devices with those of the reference instruments. Given the number of parameters and number of devices involved in this work, the literature search was performed through combinations of keywords.

For the study on the measurement errors these keywords were structured as "[*Device*] [*Parameter*] [*repeatability/reproducibility/measurement error*]," with [*Device*] the name of the biometry device and [*Parameter*] the parameter of interest. For the meta-analysis of the comparison between devices, the structure of the keywords was "[*Device 1*] [*Device 2*] [*Parameter*]," with [*Device 1*] being Pentacam, OrbScan, or IOL Master and [*Device 2*] the name of another device (eg, Sirius, Galilei, ...) or a generic technological name (eg, Placido, ultrasound, ...).

The literature search was performed in June 2014 using PubMed and Google Scholar. It included both English and non-English references published after the year 2000, thus avoiding results obtained using outdated versions of the devices or their software. Since biometric devices are typically calibrated using eye models that are close to the epidemiologic average (eg, the Gullstrand eye model), only articles using healthy adult eyes were considered in this analysis. The only exceptions were papers comparing corneal or anterior chamber parameters in cataract patients, as these measurements are not influenced by the condition of the lens. For the comparison of axial length measurements, however, cataract patients were excluded.

Finally a selection was made based on the way the remaining articles presented their data. For the study on measurement errors, only those reporting the standard deviation of the repeated measurements, or parameters that can be calculated back to the standard deviation, were included. For the meta-analysis of the comparison, both the average and the standard deviation of the difference between devices (or the 95% limits of agreement) had to be reported, along the lines suggested by Altman and Bland[1] and McAlinden and associates.[2]

It should be noted that for the Orbscan a distinction was made between comparative articles that use the "acoustic factor" (AF) to report the central corneal thickness and articles that do not. This AF typically has a value of 0.92 (although different values are used as well) and aims to cor-

TABLE 1. Thresholds to Assess Clinical Equivalence of Measurements Provided by Ophthalmic Biometry Devices, Defined by Effect on Calculated Parameters

Parameter Unit	$K_{a,m}$ D	$K_{p,m}$ D	CCT μm	ACD mm	L mm
Measurement error E_{meas}					
E_{meas}[a]	0.152[b]	0.093[b]	8.12	0.085	0.046
Change in calculated refraction ΔS_{calc}					
0.125 D	0.122	0.122	–	0.095	0.046
0.250 D	0.244	0.244	–	0.192	0.093
Change in IOL power (SRK/T)[c]					
0.125 D	0.118	–	–	–	0.042
0.250 D	0.235	–	–	–	0.082
Change in IOL power (Haigis)[c]					
0.125 D	0.087	–	–	0.300	0.037
0.250 D	0.174	–	–	0.600	0.074

ACD = anterior chamber depth; CCT = central corneal thickness; D = diopter; ΔS_{calc} = change in calculated refraction; E_{meas} = total measurement error; IOL = intraocular lens; $K_{a,m}, K_{a,s}, K_{a,f}$ = mean, steep, and flat anterior keratometry; $K_{p,m}, K_{p,s}, K_{p,f}$ = mean, steep, and flat posterior keratometry; L = axial length; SRK/T = Sanders-Retzlaff-Kraff Theoretical IOL calculation formula.

[a]Derived from the measurement errors in Table 2, averaged over all devices except those designated outliers in the text.

[b]For steep and flat axes different values were used, determined in the same manner ($K_{a,s}$: 0.185 D; $K_{a,f}$: 0.178 D; $K_{p,s}$: 0.062 D; $K_{p,f}$: 0.062 D).

[c]Using IOL constants of the Alcon AcrySof SA60AT IOL (A = 118.4; Haigis $a_0 = -0.111, a_1 = 0.249, a_2 = 0.179$), targeting emmetropia (available at http://www.augenklinik.uni-wuerzburg.de/ulib/c1.htm, accessed August 8, 2014).

rect for an overestimation of the corneal thickness by the Orbscan. Also, for the anterior chamber depth (ACD) there were articles that included the cornea into the measurement value and other articles that did not. However, these articles were not considered separately owing to there being an insufficient number of articles to warrant a separate analysis.

• MEASUREMENT ERROR: The measurement error E_{meas} may be defined by $E_{meas} = (E^2_{repeat} + E^2_{reprod})^{1/2}$, which is composed of the repeatability E_{repeat}, the standard deviation of multiple measurements performed by a single operator, and the reproducibility E_{reprod}, the standard deviation between measurements performed at different sessions or by different operators. Although in principle there could also be calibration differences between devices of the same type, we were unable to include this factor into our analysis, as very little information on this type of error is available.[3]

To obtain a global estimate of the repeatability and reproducibility of the various biometry devices, the numerical values found in the literature review were tabulated