



Published in final edited form as:

*Optom Vis Sci.* 2016 July ; 93(7): 720–730. doi:10.1097/OPX.0000000000000869.

## Variability and Errors of Manually Digitized Goldmann Visual Fields

**Michael P. Barry, MS, Ava K. Bittner, OD, PhD, FAAO, Liancheng Yang, MS, Rebecca Marcus, BA, Mian Haris Iftikhar, MD, and Gislin Dagnelie, PhD, FAAO**

Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, (MPB), Nova Southeastern University, College of Optometry, Ft. Lauderdale, Florida (AKB), Department of Ophthalmology (LY, MHI, GD), Department of Neuroscience (RM), Johns Hopkins University, Baltimore, Maryland

### Abstract

**Purpose**—Goldmann visual fields (GVFs) are useful for tracking changes in areas of functional retina, including the periphery, in inherited retinal degeneration patients. Quantitative GVF analysis requires digitization of the chart coordinates for the main axes and isopter points marked by the GVF operator during testing. This study investigated inter- and intra-digitizer variability among users of a manual GVF digitization program.

**Methods**—Ten digitizers were trained for one hour, then digitized 23 different GVFs from inherited retinal degeneration patients in each of three testing blocks. Digitizers labeled each

---

Corresponding author: Gislin Dagnelie, Johns Hopkins Hospital, Wilmer Woods 355-A, Baltimore, MD 21287, [gislin@jhu.edu](mailto:gislin@jhu.edu). This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies. Results in this manuscript were previously presented at the ARVO 2013 Annual Meeting on May 5<sup>th</sup>, 2013 in Seattle, WA and at the 2013 annual meeting of the American Academy of Optometry on October 24<sup>th</sup>, 2013 in Seattle, WA.

#### SUPPLEMENTAL DIGITAL CONTENT

Nineteen additional Goldmann visual fields can be found online as Supplemental Digital Content 1–19, available at [LWW insert link].

#### SUPPLEMENTAL DIGITAL CONTENT FIGURE LEGENDS

- Supplemental Digital Content Figure 1 (PDF). GVF 1.
- Supplemental Digital Content Figure 2 (PDF). GVF 2.
- Supplemental Digital Content Figure 3 (PDF). GVF 3.
- Supplemental Digital Content Figure 4 (PDF). GVF 4.
- Supplemental Digital Content Figure 5 (PDF). GVF 7.
- Supplemental Digital Content Figure 6 (PDF). GVF 8.
- Supplemental Digital Content Figure 7 (PDF). GVF 9.
- Supplemental Digital Content Figure 8 (PDF). GVF 10.
- Supplemental Digital Content Figure 9 (PDF). GVF 11.
- Supplemental Digital Content Figure 10 (PDF). GVF 12.
- Supplemental Digital Content Figure 11 (PDF). GVF 14.
- Supplemental Digital Content Figure 12 (PDF). GVF 17.
- Supplemental Digital Content Figure 13 (PDF). GVF 18.
- Supplemental Digital Content Figure 14 (PDF). GVF 19.
- Supplemental Digital Content Figure 15 (PDF). GVF 20.
- Supplemental Digital Content Figure 16 (PDF). GVF 21.
- Supplemental Digital Content Figure 17 (PDF). GVF 22.
- Supplemental Digital Content Figure 18 (PDF). GVF 24.
- Supplemental Digital Content Figure 19 (PDF). GVF 25.

isopter as seeing or non-seeing, and its target size. Isopters with the same test target within each GVF were grouped to create isopter groups.

**Results**—The standard deviation of isopter group area showed an approximate square-root relationship with total isopter group area. Accordingly, the coefficient of variation for isopter group area decreased from 68% to 0.2% with increasing isopter group area. A bootstrap version of ANOVA did not reveal a significant effect of digitizers on isopter group area. Simulations involving random sampling of digitizers showed that 5–7 digitizers would be required to catch 95–99% of labeling errors and isopter misses, on the basis of data discrepancies, with 99% probability.

**Conclusions**—These data suggest that any minimally trained digitizer would be capable of reliably determining any isopter area, regardless of size. Studies using this software could either use 5–7 minimally trained digitizers for each GVF, three digitizers who demonstrate low frequencies of errors on a practice set of GVFs, or two digitizers with an expert reader to adjudicate discrepancies and catch errors.

### Keywords

Goldmann visual field; digitize; variability; error; FieldDigitize

---

Diseases such as retinitis pigmentosa<sup>1–4</sup> (RP) and Leber congenital amaurosis<sup>5,6</sup> (LCA) often present with pronounced visual field loss, as one of the hallmarks of these inherited retinal degenerations is loss of peripheral vision. Studies of treatments that aim to slow or reverse the progression of such retinal degenerative diseases have included manual kinetic visual field testing, such as with the Goldmann perimeter, to demonstrate efficacy.<sup>7–13</sup> Goldmann visual fields (GVFs) have also been used to evaluate the safety of treatments in patients.<sup>14,15</sup> In recent years, semi-automated kinetic perimetry has been introduced as a potential alternative for Goldmann perimetry. semi-automated kinetic perimetry can provide measures of seeing and non-seeing areas as solid angles, in steradians or degrees squared. To date, no kinetic visual field method will allow digitization of drawn isopter contours or translation from solid angles into areas of functional retina capable of detecting kinetic stimuli. These retinal area values can be tracked over time and patterns can be evaluated across subjects.

Calculation of areas from manual Goldmann perimetry data has been addressed<sup>16–23</sup> and used<sup>5,7,8,10,14,18,21,24–37</sup> many times in the past several decades. Popular methods include the use of software that can analyze scanned images of GVFs (e.g. Adobe Photoshop,<sup>23,34,37</sup> Engauge Digitizer,<sup>32</sup> GIMP,<sup>27</sup> JHU's FieldDigitize,<sup>7,10,24</sup> MathWorks MATLAB,<sup>14</sup> NIH's ImageJ,<sup>18,31</sup> Scion Image,<sup>34,36</sup> UCL's Retinal Area Analysis Tool,<sup>28</sup> UTHSCSA ImageTool<sup>29,33</sup>), or handheld devices to convey position information to computer software.<sup>5,20,21,23,26</sup> To our knowledge, however, none of these previous studies have explored the variability of calculated retinal areas or the error rates of digitizers. Here, the term “error” refers to an omission or erroneous classification of one isopter: mislabeling an isopter target size, mislabeling an isopter as seeing or non-seeing, or missing an isopter during digitization. Zahid et al.<sup>23</sup> did recently examine inter- and intra-digitizer variabilities for two digitizing techniques, but neither of these techniques tracked errors in processing

isopters or translated planimetric map areas to retinal surface areas, which requires calibration for each GVF.

In order to use digitized GVF data, one must understand how much information is potentially lost in the manual digitization process. This is particularly important because the digitization process is labor-intensive and repetitive, and skilled GVF operators are in high demand, and digitization is often relegated to less well-trained staff, or even to research assistants who may not be available for the entire duration of a clinical trial. We therefore designed and conducted a study to determine the amount of variability in the output values of the digitizing process, specifically when digitizers are mouse-clicking on scanned images, and the likelihood of errors in the output if digitizers are not experienced GVF operators. If inter-digitizer variance tends to outweigh intra-digitizer variance, then researchers and clinicians would need to take particular care by choosing reliable digitizers and avoid making changes in digitizers over time. If errors are too common, researchers and clinicians may also need to employ a larger pool of digitizers and/or add an expert reader who is more qualified to check for errors. This study investigated the variability and error rates in digitized data produced by digitizers who received minimal training and digitized a variety of GVFs with a wide range of measured vision in patients with RP or LCA obtained by many GVF test operators.

## METHODS

### Study Materials

GVFs used in this study were collected in 27 patients with moderate to severe visual field loss due to RP or LCA participating in a multi-center clinical trial across 5 countries and had been acquired by experienced GVF operators trained to adhere to a strict set of procedures. All GVFs contained one or more seeing (i.e., encompassing a seeing area in the field) or non-seeing (i.e., encompassing a scotoma) isopters, whose presence was confirmed at multiple visits. Use of refractive correction, isopter colors, and markings on the GVF charts were consistent with the clinical trial procedures.

### Digitizers

Ten normally-sighted Johns Hopkins University students and employees attended a one hour training session that explained the meaning of GVF isopters, the color and naming conventions, and the use of the FieldDigitize v4.0 software, described below. Eight of the digitizers were JHU undergraduates (ages 18–22), and the other two were employees of the Wilmer Eye Institute (ages 20–40). None of the digitizers had prior experience with GVF testing or with digitization. The digitizers were all asked to use the software to digitize the same sets of GVFs over a period of two months.

This research was approved by the Johns Hopkins Institutional Review Board and adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from all digitizers prior to performing study procedures.

## Digitizing Process

GVPs were scanned and presented on a PC screen to digitizers. The program used for presentation and digitizing, FieldDigitize v4.0, was developed at the Lions Vision Center of the Wilmer Eye Institute at the Johns Hopkins School of Medicine. The program provides retinal area calculations based on the relationships between planimetric visual field maps and retinal surface area, as previously documented by Dagnelie.<sup>16</sup> An earlier version of this program was used by Bittner et al. to evaluate the test-retest repeatability of GVPs obtained in RP subjects,<sup>24</sup> and again to examine the effects of acupuncture therapy on the visual function of RP subjects.<sup>7</sup>

During digitization, users mouse-clicked on 5 cardinal points in the GVP grid to establish the horizontal and vertical scales. These points were the center of the GVP, 90° eccentricity at 0° azimuth, 70° eccentricity at 90° azimuth, 90° eccentricity at 180° azimuth, and 70° eccentricity at 270° azimuth. The use of 5 rather than 3 calibration points allowed for the detection of distortions and correction for rotations that might have occurred during scanning. Digitizers then identified each isopter as seeing or non-seeing and selected the size of the test light (I-4e through V-4e). Digitizers clicked on only those points along the isopter lines that had been recorded by the GVP operators as transition points from non-seeing to seeing. Digitizers were instructed to complete all isopters in a GVP before transitioning to the next GVP.

## Data Collection

GVPs were made available to digitizers in blocks of 25 on three separate dates: November 29, 2012, December 06, 2012, and December 12, 2012. Digitizers were instructed to digitize all GVPs in each block before proceeding to the next block. The same GVPs were used in all three blocks, but were given different file names and different presentation orders. GVPs 6 and 23 were not consistently presented and were mistakenly replaced by new GVPs in block 2 and/or 3, so data from these GVPs and their replacements were eliminated from the analysis. GVP identifier numbers used in this report reflect the order used for block 1. General statistics for all GVPs are shown in Table 1. All isopters across all GVPs were given unique identifier numbers for analyses.

## Definitions

For this study, we used the term isopter strictly to refer to a closed set of connected line segments that were drawn by a GVP operator.<sup>35, 38–40</sup> GVPs maps used in this study often incorporated multiple isopters per test target size to describe the GVP landscape, which included “non-seeing” isopters surrounding scotomatous areas and “seeing” isopters surrounding islands of vision. The test target luminance used in the GVPs was 4e in all cases, so only the target size varied. Isopters that were measured with a single test target size in one GVP were combined to create a grouping variable we defined as an isopter group. Each possible combination of GVP and target size was a distinct isopter group, thus an isopter group was comprised of all isopters determined with the same test target size within a GVP. These isopter groups represented the total seeing region for the measured eye for one test target size. We determined isopter group areas in mm<sup>2</sup> by summing the area values generated by the digitization program for all of the seeing and non-seeing isopters within

each isopter group. Non-seeing isopters were assigned negative area values in this study. It should be noted that in some previous publications, isopter groups have been referred to as isopters; <sup>14,20,23,27,30–32,37</sup> the two terms should not be confused.

Errors in this study had three types: mislabeled isopter target sizes, isopters mislabeled as seeing or non-seeing, and isopters that were missed during digitization. We used the term distinct errors to refer to all errors of the same type made for the same isopter in the same block, without regard for which digitizer made the error. If a digitizer simultaneously mislabeled both an isopter's target size and its seeing or non-seeing status, these errors would be recorded as two distinct errors, one for each error type. Distinct errors with mislabeled target sizes could therefore have instances with conflicting reports of seeing/non-seeing status, and seeing/non-seeing distinct errors could have conflicting reports of target size. If all digitizers were to make the same distinct error, the data provided by digitizing alone would not be able to suggest that any digitizer made an error of that type for that isopter, and only review by an expert reader would reveal the error.

### Data Analysis

Data were analyzed using Microsoft Excel 2013 and R versions 3.0.2 and 3.1.0. All plots were made using R 3.1.0. Only data from the 23 GVFs that were presented in every block to every digitizer were analyzed. Isopter measurements were matched to isopter identifier numbers by inspection, based on reported area, target size and seeing value parameters, and screenshots of digitized fields. Errors were flagged by comparing the reported data to accepted data for each isopter. Accepted data were attributes of each isopter as determined by the authors, specifically presence in a GVF, target size, and whether the isopter enclosed a seeing or non-seeing area.

Researchers <sup>14, 20, 21, 23–26, 30, 35, 37</sup> have tended to value the total seeing area for each test light size over the areas enclosed by individual isopters. Observing this preference, our variability analyses focused on isopter groups. Only the target size accepted by the authors for each isopter was used for determining isopter groups, so isopter labeling errors made by digitizers had no effect on isopter group calculations. Standard deviations (SDs) and coefficients of variation (CVs) for area measurements were calculated for isopter groups across all measurements, within digitizers, and between digitizers. SDs and CVs of isopter group areas were analyzed as functions of overall mean area values by linear regression, using the least squares method on log-log values. Confidence intervals for lines of best fit were determined by resampling data with replacement and running the same analyses 10,000 times. This bootstrap method was employed to avoid the normality requirement of classic regression analysis, thus avoiding the need for another data transformation step aimed at normalizing the distribution.

Effects of digitizers on isopter group area measurements were analyzed using a bootstrap variation of an analysis of variance (ANOVA). Because area measurement variances were not uniform across isopter groups, classic ANOVA techniques could not be applied. Instead, variability was analyzed using components of a two-factor ANOVA model with interactions and bootstrap datasets. Just as for a typical ANOVA, the total sum of squared deviations (SS) was partitioned into SSs for digitizer and isopter group factors, the interactions of these

factors, and residuals. Previously,<sup>41</sup> the typical F-statistic has been used with resampling methods for determining the significance of data. Following this, we calculated an analogue of the F-statistic by dividing SSs of the factors and the interaction by the SS of the residuals. This analogue only differs from the F-statistic by a factor equal to a ratio of the involved degrees of freedom, which remained constant throughout our analyses. This scale factor would be necessary only for comparison with the F-distribution, which we did not use in our analyses, and was therefore ignored. We used these ratios of SSs to determine the relative contribution of variance by each factor. To determine the significance of the observed digitizer-residuals and interaction-residuals SS ratios, the distributions of these ratios, under the assumption of no digitizer effects, were generated by bootstrap resampling.

The bootstrap datasets were produced by resampling isopter group areas with replacement within isopter groups and across all digitizers 10,000 times. Isopter group areas were thus randomly assigned to digitizers in each bootstrap dataset. The analysis described above was then performed on each dataset, and the distributions of SS ratios were generated. The p-values for the observed true SS ratios were determined by calculating what proportion of the generated ratios was greater than the observed ratios.

Distinct errors were analyzed to investigate how many digitizers would be required to catch nearly all errors when error detection is restricted to investigating discrepancies in data. Relatively strict requirements for error detection were chosen: detect at least 95% or 99% of errors in a dataset with at least 99% probability. For each sample size of 2 through 8, we randomly sampled digitizer IDs from our pool of 10 without replacement 20,000 times. We then counted the number of distinct errors that were made by digitizers in each sample, and how many of those digitizers made each distinct error. We considered any distinct error made by all the sampled digitizers to be undetected. We calculated the percentage of errors that were detected for each sample, and then evaluated the detection percentages at the 1<sup>st</sup> percentile for each sample size.

## RESULTS

### Variability

As a first step in analyzing the sources of variability in calculated isopter group areas we examined the distributions of overall (up to 30 for each isopter group), intra-digitizer (up to 3 per isopter group for each of 10 digitizers), and inter-digitizer (up to 10 per isopter group) SD values. The simple statistics of these SDs, across all 65 isopter group areas, are summarized in Table 2. SDs between digitizers tended to be higher on average than SDs within digitizers. Intra-digitizer SDs, however, had a much greater range than inter-digitizer SDs. Overall SDs tended to show an approximate square-root relationship with isopter group area. This relationship is plotted on a log-log scale in Figure 1, with regression details.

Next we computed relative SDs by dividing each SD by the overall mean. Simple statistics of these relative SDs are shown in Table 3. In the case of overall SDs, the relative SD is equivalent to the CV. CVs tended to be quite small, typically less than 10%, except for isopter groups with very small retinal areas, i.e.  $< 1 \text{ mm}^2$ . In no case did a CV reach or

exceed 100%. CVs are plotted against isopter group areas in Figure 2, showing an approximate inverse square root relationship.

We then examined whether the variability in our data was significantly affected by either the individual characteristics of the digitizers or by the way in which each digitizer processed specific isopters. Our ANOVA with bootstrap datasets revealed p-values for digitizer and digitizer-isopter group interaction SSs to residual SS ratios of 0.51 and 0.25, respectively. These probabilities are those of the observed or greater ratios being selected at random from the simulated distributions. Thus, no statistically significant effect of digitizer ( $p=0.51$ ) or interaction effect ( $p=0.25$ ) was found.

To investigate the relative contribution of variability added by the digitizing process to the analysis of GVFs from RP patients, we compared our data to those provided by Bittner et al.<sup>24</sup> Bittner et al. had previously used a single test operator to administer the GVF test twice within a single study visit for one eye in RP patients. Isopters were digitized using the same software as the present study, and areas for isopter groups were calculated. SDs calculated for these isopter group areas thus combine variabilities introduced by the RP patient, the GVF operator, and the GVF digitizer. These test-retest SDs, when compared to the SDs of isopter group area in the present study, were found to be much greater than the SDs of digitizing, especially for larger isopter group areas. This is clearly shown in Figure 3: The slope of test-retest SDs versus isopter group size on a log-log scale was found to be more than 40% greater than that of digitizing SDs.

## Errors

There were a total of 293 instances of errors and 155 distinct errors observed in this study. Of the distinct errors, 62 were missed isopters [mean 4.4 and range 0–31 across GVFs; 8 GVFs (35%) without any missed isopters], 25 were mislabeled target sizes [mean 1.9 and range 0–34 across GVFs; 14 GVFs (61%) without any mislabeled target sizes], and 68 were incorrect seeing/non-seeing labels [mean 4.5 and range 0–38 across GVFs; 11 GVFs (48%) without any mislabeled seeing areas]. These distinct error counts per GVF reflect the cumulative number of distinct errors in a GVF made across all 30 times that GVF was processed. Figures 4–7 show GVFs for which there were 10 or more errors of one type; the remaining 19 GVFs can be found online as Supplemental Digital Content 1–19, available at [\[LWW insert link\]](#).

Errors made by specific digitizers are shown in Table 4. Digitizers 2, 4, 6, 8, and 9 contributed substantially more errors than other digitizers in isopter misses ( 20 or 17% of attempted isopters), target-size errors ( 10 or 8%), or seeing/non-seeing errors ( 10 or 8%).

If these five digitizers are removed from analysis, the number of distinct errors falls from 155 to 35 (23%). Table 5 shows 17 distinct errors that were made more than once within the same presentation block by the five more accurate digitizers; these are errors that may not be detected by comparing the work of two minimally-trained digitizers alone. As shown, these errors are confined to GVFs 5, 13, 16 (Figures 4, 5, and 7).

Based on our simulations of randomly sampling digitizers, at least 5 digitizers were required to detect 95% percent of errors, and at least 7 digitizers were required to detect 99% of errors. Error detection here is based solely on discrepancies among digitizers' data, and reported detection percentages were achieved in at least 99% of simulations with the given sample sizes.

Rather than random selection, one could screen potential digitizers by providing a practice sample set of previously classified GVFs to determine error rates, in order to help achieve a desired minimum error rate. If we were to select the two digitizers with the lowest gross error rates observed in this study (i.e., digitizers 1 and 3), only 2 errors would not be detected by comparing their data. Including the next most accurate digitizer (i.e., digitizer 5) in the analysis would ensure that all errors in our sample would be detected.

## DISCUSSION

We found that the SDs of digitized isopter group area measurements decreased with decreasing isopter group area. However, CVs for isopter group areas increased as isopter group areas decreased, which would imply that the variability of digitized areas may be problematic for small isopters. The largest CV observed in this study was 68%; thus the SD of the isopter group area measure was not observed to ever exceed its average. We therefore interpret these data to suggest that isopters of any size can be digitized to produce meaningful data, at least within the range of isopter sizes tested here (  $0.085 \text{ mm}^2$ , i.e.,  $1.5 \text{ degree}^2$  at  $45^\circ$  eccentricity, equivalent to  $1.2 \text{ degree}^2$  in the fovea) and using the described or comparable methods. This means that there should be no loss of isopter data during the digitization process, and GVF test operators should be encouraged to make very detailed assessments of tiny seeing and non-seeing areas.

When examining the variability across and within digitizers, we found that inter-digitizer SDs were slightly higher on average than intra-digitizer SDs. This could raise concern about the effect of individual digitizers on digitized data; however, the range of intra-digitizer SDs was much greater than that of inter-digitizer SDs. When using a bootstrap version of ANOVA, we found no significant effect of digitizers on isopter group area. This implies that, when dealing with minimally-trained digitizers and no concern for misses or labeling errors, one may not need to consider which digitizer performed the digitization for any given isopter group. Further, the comparison of digitization variability data with GVF test-retest data from Bittner et al.<sup>24</sup> reveals that the variability introduced to manual GVF data by digitizing is minor relative to the variability contributed by the RP patient and the GVF operator. We suggest that with improved marking of isopters by GVF operators and/or more training of digitizers in isopter detection and identification, an identical error made by 2 reliable digitizers is expected to occur less frequently than once in 98 isopters. No such errors occurred in this study in any block after removal of error-prone digitizers and of GVFs with problematic isopters (e.g. #5, 13, 16).

One of the limitations of this study may be the education level of the digitizers and whether these results are generalizable to any possible digitizer. All digitizers in this study were affiliated with an institution for research and higher education, and our results may have



been different with random sampling from a larger literate population. Another limitation is that the GVF's used in this study were obtained from GVF operators who were trained to obtain and record the GVF's according to a specific protocol, so perhaps if the operators had not been instructed to mark the GVF's in a particular manner (e.g., if the operators used the same color for each isopter, or shading to indicate non-seeing areas), the digitization error rates might have been different. This study also did not address any possible labeling errors made by GVF operators.

It may be possible to prevent some errors related to GVF misinterpretation if the GVF operator who tested the patient can digitize their GVF results immediately after the testing. However, it may not be feasible for clinical GVF operators to take on the additional work of digitizing for research purposes, and clinical trials generally do not favor such an arrangement, for fear that data acquisition errors may go undetected. In the case of retrospective studies, it may not be practical or possible to recruit the GVF operators as digitizers. Furthermore, for well-marked and documented GVF's, there is no reason to expect that the operator's digitizing variability would significantly differ from that of any other digitizer with comparable digitizing experience.

Digitizers in this study did make numerous errors in isopter labeling, and missed isopters when digitizing. Those who wish to digitize GVF's accurately should therefore employ some mechanism for catching these errors. One simple solution is to have an expert reader monitor the work of the digitizers. This reader would need to compare all digitized data to the original GVF's to make sure that areas seem appropriate, all isopters were labeled properly, and no isopters were missed. Alternatively, one could adopt a system that uses multiple digitizers for any given GVF and that flags discrepancies among digitizers' data. A reviewer would then be required to identify and resolve any flagged errors. Our recommendations for structuring groups for error detection are summarized in Table 6.

Based on our pool of digitizers, leaving error detection to flagging data discrepancies would require a large number of digitizers or a screening process for digitizers. Assuming it is desirable to catch nearly every error that is produced, our simulations for the selection of digitizers in this study suggest two possible options: (1) randomly choose five to seven minimally trained digitizers to digitize all GVF's; or (2) screen multiple digitizers to find three who demonstrate low error rates and select those to digitize all GVF's. These recommendations assume that each digitizer only digitizes each GVF once, and that digitizer training is comparable to that used for this study. Because error detection from data comparisons alone may require great redundancy in digitizer work, supervisors may choose to consider whether this is a time- and cost-effective approach. If training of GVF operators and digitizers is enhanced in order to avoid problems such as those seen with GVF's 5, 13, and 16, then using just two digitizers may be sufficient to catch most errors. More than two digitizers may be desired in dedicated reading centers, however, to add tolerance for cases in which a digitizer may unexpectedly be unable to work. Thus the findings of this study indicate that manual digitization of GVF's by a single minimally trained digitizer is not sufficient to catch all errors, but that a small team of 2–3 digitizers with error checking by an expert reader will assure the production of accurate and reliable digitized GVF data.

In this paper we have not considered the inherent inaccuracies in the data contributed by the GVF operator (who needs to manually move the stimulus at constant speed and note the location at which the patient signals seeing or losing the stimulus) and by the patient (who needs to maintain a constant response criterion and reaction time). Clinical trials (such as the one from which the GVFs for the current study were obtained) will try to minimize operator variability through standardized training across participating centers, and patient variability through repeated baseline tests. Semi-automated kinetic perimetry may appear to take the operator variability out of the equation, but the operator still has to instruct the machine which areas of complex fields require further refinement, especially in patients with advanced disease. Thus even in semi-automated kinetic perimetry the operator remains a significant contributor to variability. Also, as noted earlier, current semi-automated kinetic perimetry software does not convert solid angles or chart areas into retinal areas, so digitization cannot be avoided even when using semi-automated kinetic perimetry. As long as this situation persists, judicious choice of a digitization and review strategy for kinetic perimetry data from clinical trials will remain critically important.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors would like to thank Claire Barnes and Ronald Schuchard for providing their counsel throughout the study and manuscript development.

This research was supported by the following grants from: NIH K23 EY018356, to AKB; QLT Inc. (Vancouver, BC, Canada), to GD; and NIH T32 EY07143, to the Johns Hopkins Visual Neuroscience Training Program for MPB.

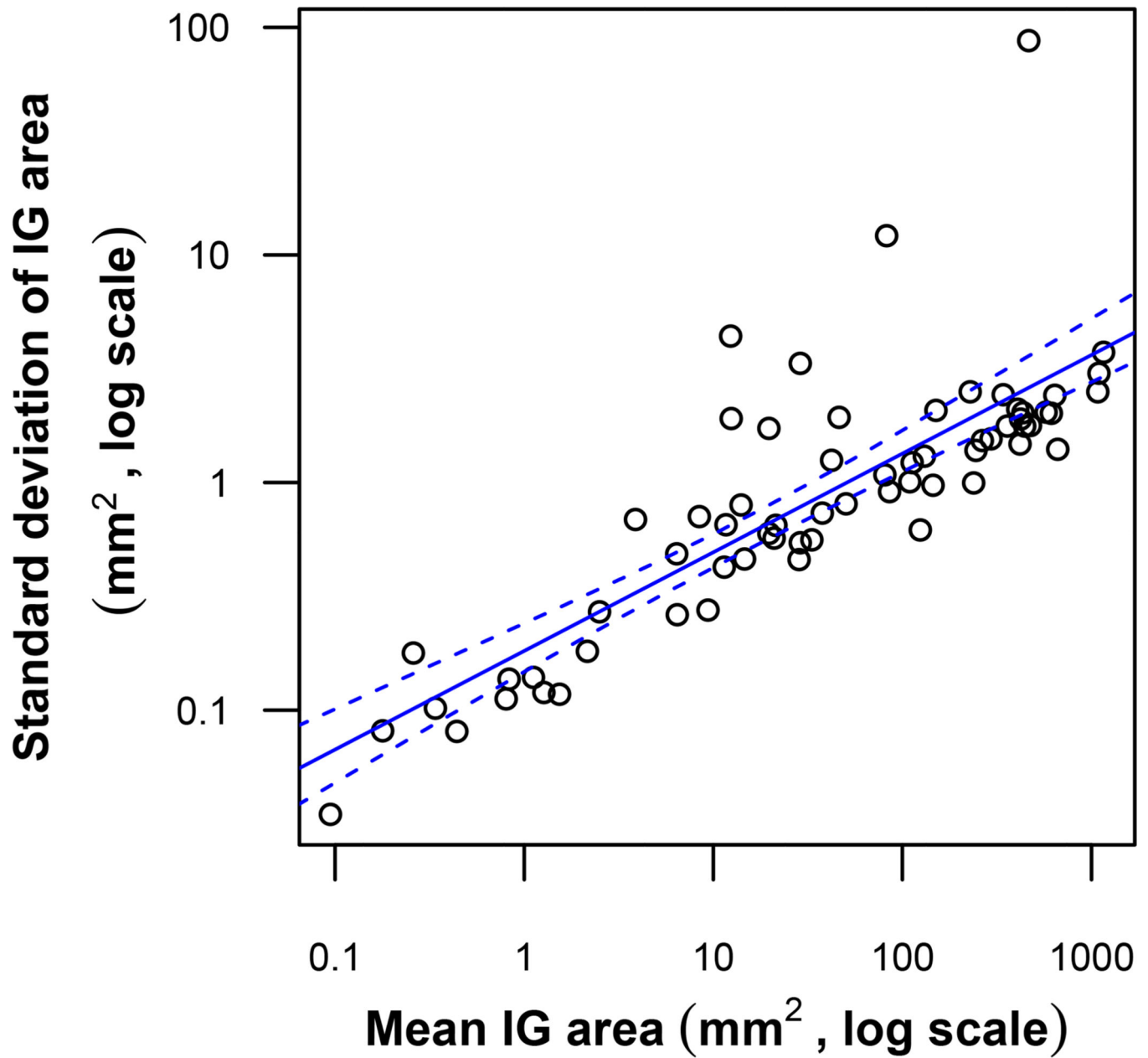
Under a licensing agreement between QLT, Inc. and the Johns Hopkins University, GD is entitled to licensing royalties on the software described in this manuscript. QLT funded the study described in this manuscript. Also, GD, AKB, and MPB, while employed at Johns Hopkins University, were paid as consultants to QLT, for advice unrelated to the subject matter of this study.

## REFERENCES

1. Derby H. A case of retinitis pigmentosa, the parents of the patient being first cousins: read before, and the patient exhibited to, the Boston Society for Medical Improvement. Communicated for the Boston Medical and Surgical Journal. *Boston Med Surg J.* 1865; 72:149–151.
2. Hulke, JW. *A Practical Treatise on the Use of the Ophthalmoscope.* London: Churchill; 1861.
3. Webster D. The aetiology of retinitis pigmentosa, with cases. *Trans Am Ophthalmol Soc.* 1878; 2:495–505.
4. Wilson H. Ophthalmoscopic notes. *Dublin Q J Med Sci.* 1865; 40:72–78.
5. Jacobson SG, Aleman TS, Cideciyan AV, Roman AJ, Sumaroka A, Windsor EA, Schwartz SB, Heon E, Stone EM. Defining the residual vision in leber congenital amaurosis caused by RPE65 mutations. *Invest Ophthalmol Vis Sci.* 2009; 50:2368–2375. [PubMed: 19117922]
6. Leber T. Ueber Retinitis pigmentosa und angeborene Amaurose. *Graefe's Arch Clin Exper Ophthalmol.* 1869; 15(3):1–25.
7. Bittner AK, Gould JM, Rosenfarb A, Rozanski C, Dagnelie G. A pilot study of an acupuncture protocol to improve visual function in retinitis pigmentosa patients. *Clin Exp Optom.* 2014; 97:240–247. [PubMed: 24773463]

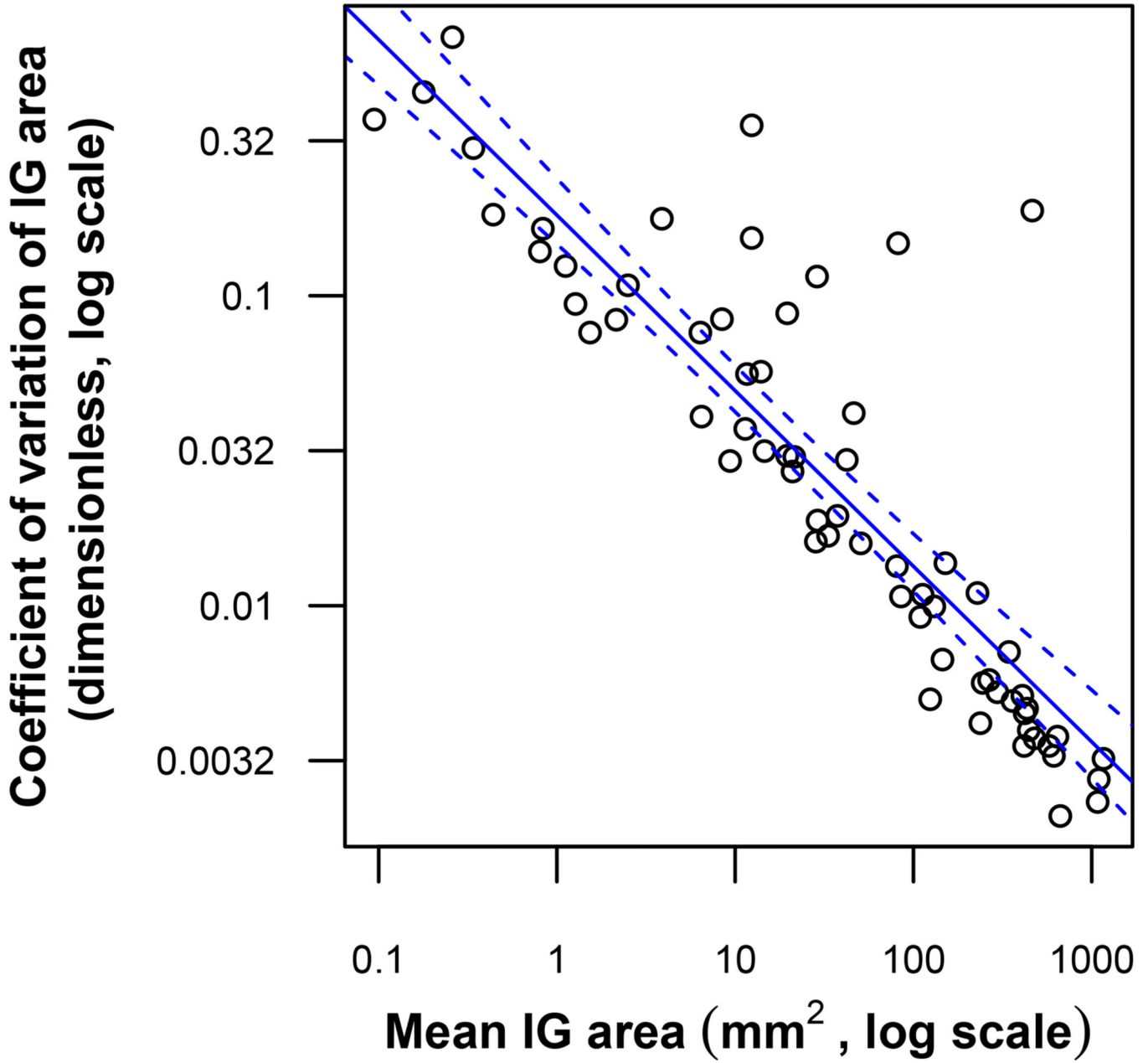
8. Clemson CM, Tzekov R, Krebs M, Checchi JM, Bigelow C, Kaushal S. Therapeutic potential of valproic acid for retinitis pigmentosa. *Br J Ophthalmol.* 2011; 95:89–93. [PubMed: 20647559]
9. Ivandic BT, Ivandic T. Low-level laser therapy improves vision in a patient with retinitis pigmentosa. *Photomed Laser Surg.* 2014; 32:181–184. [PubMed: 24527959]
10. Koenekoop RK, Sui R, Sallum J, van den Born LI, Ajlan R, Khan A, den Hollander AI, Cremers FP, Mendola JD, Bittner AK, Dagnelie G, Schuchard RA, Saperstein DA. Oral 9-cis retinoid for childhood blindness due to Leber congenital amaurosis caused by RPE65 or LRAT mutations: an open-label phase 1b trial. *Lancet.* 2014; 384:1513–1520. [PubMed: 25030840]
11. Rotenstreich Y, Belkin M, Sadetzki S, Chetrit A, Ferman-Attar G, Sher I, Harari A, Shaish A, Harats D. Treatment with 9-cis beta-carotene-rich powder in patients with retinitis pigmentosa: a randomized crossover trial. *JAMA Ophthalmol.* 2013; 131:985–992. [PubMed: 23700011]
12. Schatz A, Rock T, Naycheva L, Willmann G, Wilhelm B, Peters T, Bartz-Schmidt KU, Zrenner E, Messias A, Gekeler F. Transcorneal electrical stimulation for patients with retinitis pigmentosa: a prospective, randomized, sham-controlled exploratory study. *Invest Ophthalmol Vis Sci.* 2011; 52:4485–4496. [PubMed: 21467183]
13. Testa F, Maguire AM, Rossi S, Pierce EA, Melillo P, Marshall K, Banfi S, Surace EM, Sun J, Acerra C, Wright JF, Wellman J, High KA, Auricchio A, Bennett J, Simonelli F. Three-year follow-up after unilateral subretinal delivery of adeno-associated virus in patients with Leber congenital Amaurosis type 2. *Ophthalmology.* 2013; 120:1283–1291. [PubMed: 23474247]
14. Bhalla S, Joshi D, Bhullar S, Kasuga D, Park Y, Kay CN. Long-term follow-up for efficacy and safety of treatment of retinitis pigmentosa with valproic acid. *Br J Ophthalmol.* 2013; 97:895–899. [PubMed: 23603755]
15. Ruether K, Pung T, Kellner U, Schmitz B, Hartmann C, Seeliger M. Electrophysiologic evaluation of a patient with peripheral visual field contraction associated with vigabatrin. *Arch Ophthalmol.* 1998; 116:817–819. [PubMed: 9639460]
16. Dagnelie G. Conversion of planimetric visual-field data into solid angles and retinal areas. *Clin Vision Sci.* 1990; 5:95–100.
17. Linstone FA, Heckenlively JR, Solish AM. The use of planimetry in the quantitative analysis of visual fields. *Glaucoma.* 1982; 4:17–19.
18. Maamari RN, D'Ambrosio MV, Joseph JM, Tao JP. The efficacy of a novel mobile phone application for Goldmann ptosis visual field interpretation. *Ophthal Plast Reconstr Surg.* 2014; 30:141–145.
19. Odaka T, Fujisawa K, Akazawa K, Sakamoto M, Kinukawa N, Kamakura T, Nishioka Y, Itasaka H, Watanabe Y, Nose Y. A visual field quantification system for the Goldmann Perimeter. *J Med Syst.* 1992; 16:161–169. [PubMed: 1460401]
20. Pe'er J, Zajicek G, Barzel I. Computerised evaluation of visual fields. *Br J Ophthalmol.* 1983; 67:50–53. [PubMed: 6848135]
21. Ross DF, Fishman GA, Gilbert LD, Anderson RJ. Variability of visual field measurements in normal subjects and patients with retinitis pigmentosa. *Arch Ophthalmol.* 1984; 102:1004–1010. [PubMed: 6743076]
22. Weleber RG, Tobler WR. Computerized quantitative analysis of kinetic visual fields. *Am J Ophthalmol.* 1986; 101:461–468. [PubMed: 3963107]
23. Zahid S, Peeler C, Khan N, Davis J, Mahmood M, Heckenlively JR, Jayasundera T. Digital quantification of Goldmann visual fields (GVFs) as a means for genotype-phenotype comparisons and detection of progression in retinal degenerations. *Adv Exp Med Biol.* 2014; 801:131–137. [PubMed: 24664690]
24. Bittner AK, Iftikhar MH, Dagnelie G. Test-retest, within-visit variability of Goldmann visual fields in retinitis pigmentosa. *Invest Ophthalmol Vis Sci.* 2011; 52:8042–8046. [PubMed: 21896857]
25. Christoforidis JB. Volume of visual field assessed with kinetic perimetry and its application to static perimetry. *Clin Ophthalmol.* 2011; 5:535–541. [PubMed: 21573042]
26. Iannaccone A, Rispoli E, Vingolo EM, Onori P, Steindl K, Rispoli D, Pannarale MR. Correlation between Goldmann perimetry and maximal electroretinogram response in retinitis pigmentosa. *Doc Ophthalmol.* 1995; 90:129–142. [PubMed: 7497885]

27. Ito S, Yoneoka Y, Hatase T, Fujii Y, Fukuchi T, Iijima A. Prediction of postoperative visual field size from preoperative optic chiasm shape in patients with pituitary adenoma. *Adv Biomed Eng.* 2015; 4:80–85.
28. Lenassi E, Saihan Z, Cipriani V, Le Quesne Stabej P, Moore AT, Luxon LM, Bitner-Glindzicz M, Webster AR. Natural History and Retinal Structure in Patients with Usher Syndrome Type 1 Owing to MYO7A Mutation. *Ophthalmology.* 2014; 121:580–587. [PubMed: 24199935]
29. Lorenz B, Strohmayer E, Zahn S, Friedburg C, Kramer M, Preising M, Stieger K. Chromatic pupillometry dissects function of the three different light-sensitive retinal cell populations in RPE65 deficiency. *Invest Ophthalmol Vis Sci.* 2012; 53:5641–5652. [PubMed: 22807296]
30. Nowomiejska K, Vonthein R, Paetzold J, Zagorski Z, Kardon R, Schiefer U. Comparison between semiautomated kinetic perimetry and conventional Goldmann manual kinetic perimetry in advanced visual field loss. *Ophthalmology.* 2005; 112:1343–1354. [PubMed: 15996734]
31. Ogura S, Yasukawa T, Kato A, Usui H, Hirano Y, Yoshida M, Ogura Y. Wide-field fundus autofluorescence imaging to evaluate retinal function in patients with retinitis pigmentosa. *Am J Ophthalmol.* 2014; 158:1093-8.e3. [PubMed: 25062603]
32. Patel DE, Cumberland PM, Walters BC, Russell-Eggitt I, Cortina-Borja M, Rahi JS. Study of Optimal Perimetric Testing In Children (OPTIC): Normative Visual Field Values in Children. *Ophthalmology.* 2015; 122:1711–1717. [PubMed: 26072348]
33. Paunescu K, Wabbels B, Preising MN, Lorenz B. Longitudinal and cross-sectional study of patients with early-onset severe retinal dystrophy associated with RPE65 mutations. *Graefes Arch Clin Exp Ophthalmol.* 2005; 243:417–426. [PubMed: 15565294]
34. Ramirez AM, Chaya CJ, Gordon LK, Giaconi JA. A comparison of semiautomated versus manual Goldmann kinetic perimetry in patients with visually significant glaucoma. *J Glaucoma.* 2008; 17:111–117. [PubMed: 18344756]
35. Roman AJ, Cideciyan AV, Schwartz SB, Olivares MB, Heon E, Jacobson SG. Intervisit variability of visual parameters in Leber congenital amaurosis caused by RPE65 mutations. *Invest Ophthalmol Vis Sci.* 2013; 54:1378–1383. [PubMed: 23341016]
36. Ushida H, Kachi S, Asami T, Ishikawa K, Kondo M, Terasaki H. Influence of Preoperative Intravitreal Bevacizumab on Visual Function in Eyes with Proliferative Diabetic Retinopathy. *Ophthalm Research.* 2013; 49:30–36.
37. Zahid S, Khan N, Branham K, Othman M, Karoukis AJ, Sharma N, Moncrief A, Mahmood MN, Sieving PA, Swaroop A, Heckenlively JR, Jayasundera T. Phenotypic conservation in patients with X-linked retinitis pigmentosa caused by RPGR mutations. *JAMA Ophthalmol.* 2013; 131:1016–1025. [PubMed: 23681342]
38. Renfrew S. The regional variations of extrafoveal perception of form in the central visual fields (photopic vision) with special reference to lesions of the visual pathways. *Br J Ophthalmol.* 1950; 34:577–593. [PubMed: 14791920]
39. Rowe FJ, Rowlands A. Comparison of Diagnostic Accuracy between Octopus 900 and Goldmann Kinetic Visual Fields. *BioMed Res Int.* 2014; 2014:11.
40. Traquair HM. Essential Considerations in Regard to the Field of Vision: Contraction or Depression? *Br J Ophthalmol.* 1924; 8:49–58. [PubMed: 18168359]
41. Anderson M, Braak CT. Permutation tests for multi-factorial analysis of variance. *J Stat Computat Simulat.* 2003; 73:85–113.

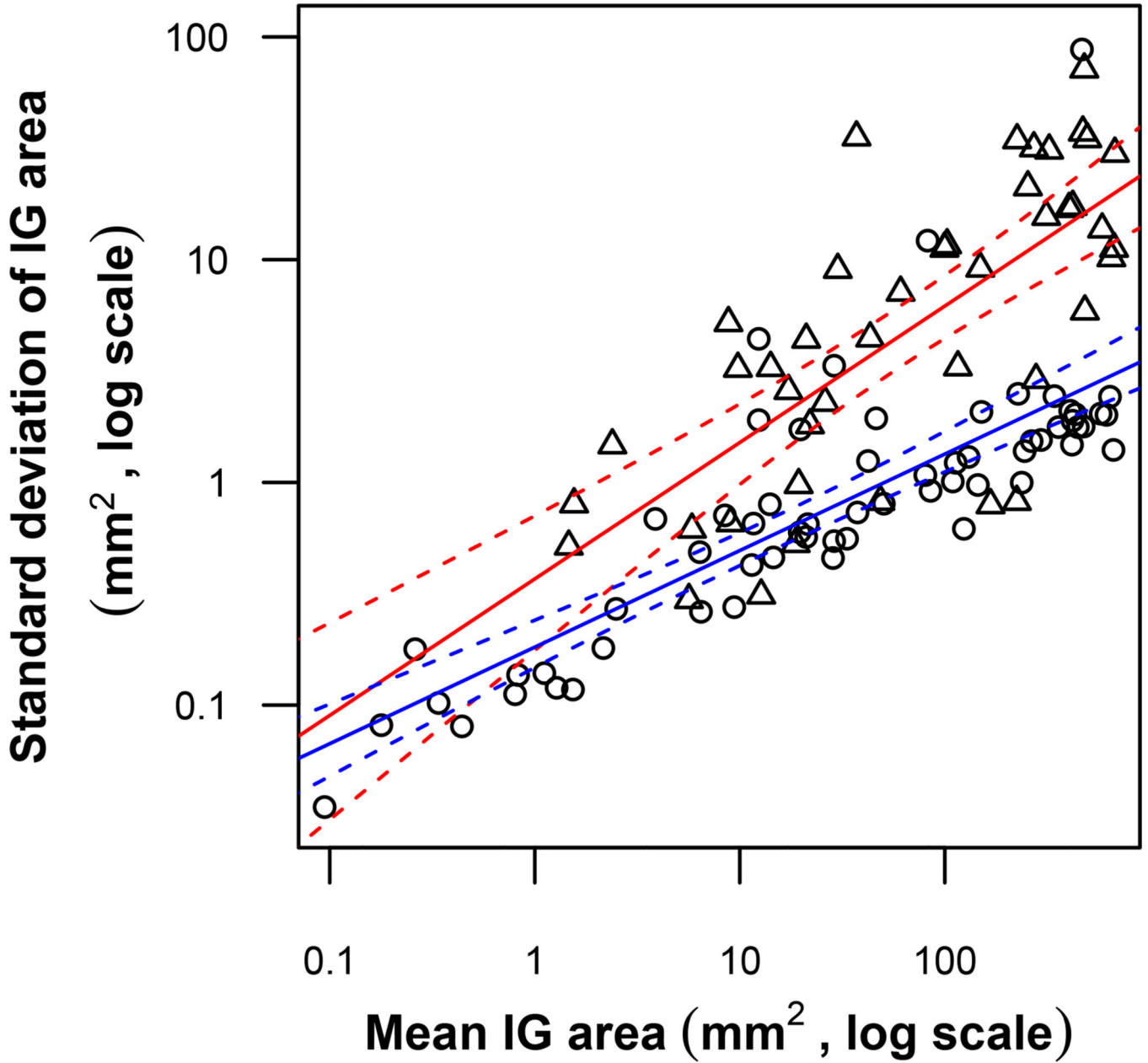


**Figure 1.**

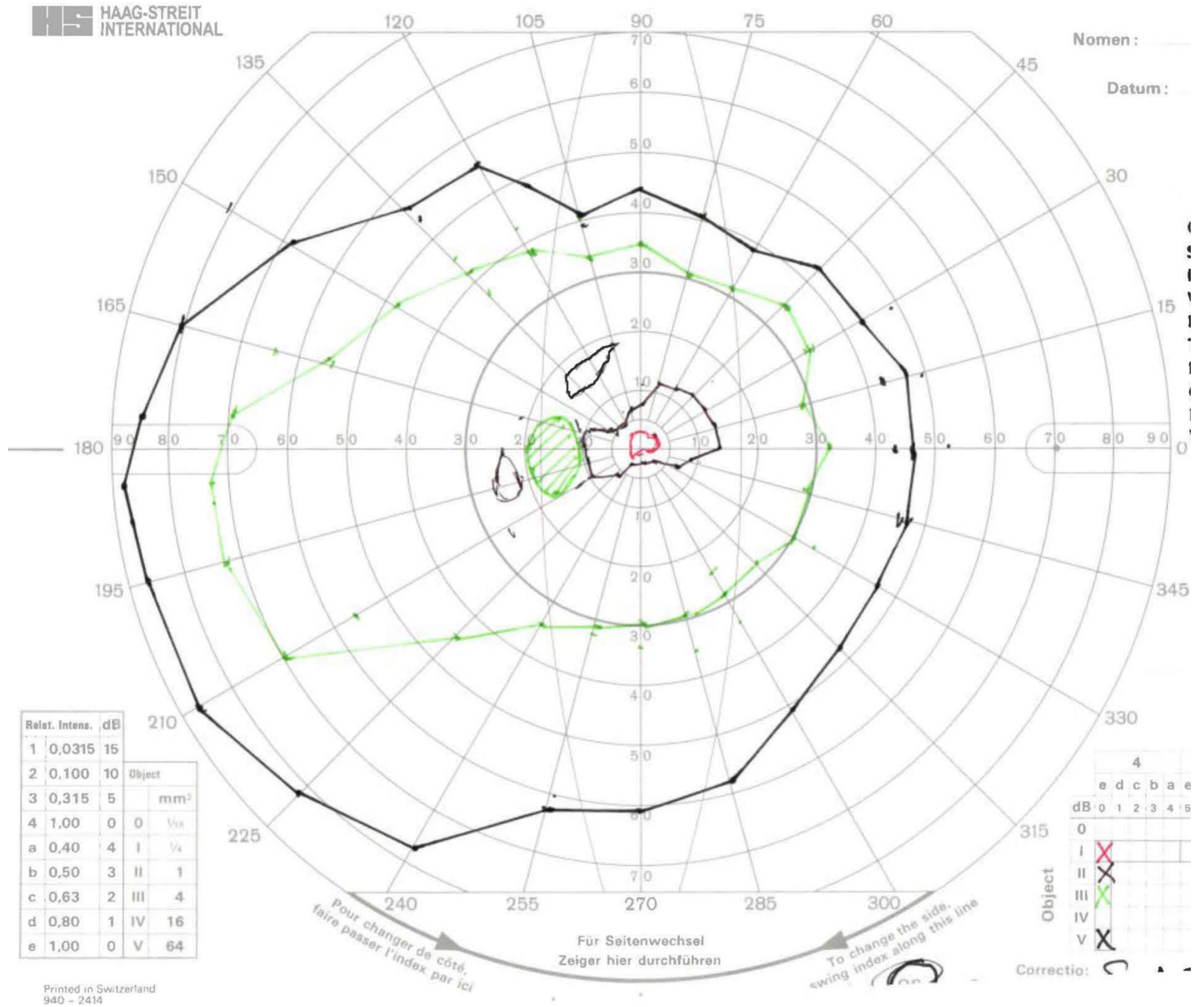
Scatterplot of SD vs. average area for measurements of all 65 Isopter Group (IG) areas. The solid line is the line of best fit as determined by the least squares method. The area between the dashed lines represents the 95% confidence interval of this regression line. The line of best fit, on the log-log scale, has an intercept of  $-0.74$ , slope of  $0.43$ , and  $R^2$  value of  $0.67$ . This relationship can be represented as  $SD \approx 10^{-0.74} area^{0.43} \approx 0.18 area^{0.43}$ .



**Figure 2.** Scatterplot of CV vs. average area for measurements of Isopter Group (IG) areas. The solid line is the line of best fit as determined by the least squares method. The area between the dashed lines represents the 95% confidence interval of this regression line. The line of best fit, on the log-log scale, has an intercept of  $-0.74$ , slope of  $-0.57$ , and  $R^2$  value of  $0.78$ . This relationship can be represented as  $CV \approx 10^{-0.74} area^{-0.57} \approx 0.18 area^{-0.57}$ .



**Figure 3.** Comparison of GVF test-retest variability and digitizer variability with respect to Isopter Group (IG) area. The circles and their line of best fit (lower line) for IG digitizer variability are copied from Figure 1. The overlaid triangles and their line of best fit (higher line) show GVF test-retest data provided by Bittner et al.<sup>24</sup> Each triangle represents data from digitizing the same IG from the same eye using two GVFs that were obtained by the same GVF operator on the same day. Dashed lines represent the 95% confidence intervals for each regression line. The triangles' line of best fit, on the log-log scale, has an intercept of  $-0.43$ , slope of  $0.61$ , and  $R^2$  value of  $0.53$ . This relationship can be represented as  $Operator\ SD \approx 10^{-0.43} area^{0.61} \approx 0.37 area^{0.61}$ .



**Figure 4.** GVF 5 (34 target-size errors, 38 seeing/non-seeing errors). Most problems with this GVF resulted from the mistaken identification of the II-4e seeing isopters as V-4e non-seeing isopters due to the similar colors used for these two test targets. Errors were made by all digitizers except Digitizer 3. When these target-size errors occurred, the superior-most II-4e seeing area's target size was always mislabeled, and the central II-4e seeing area's target size was only mislabeled when all three II-4e were similarly mislabeled. Seeing/non-seeing errors coincided with target-size errors, with few exceptions. Isolated seeing/non-seeing errors followed the same occurrence pattern as target-size errors.



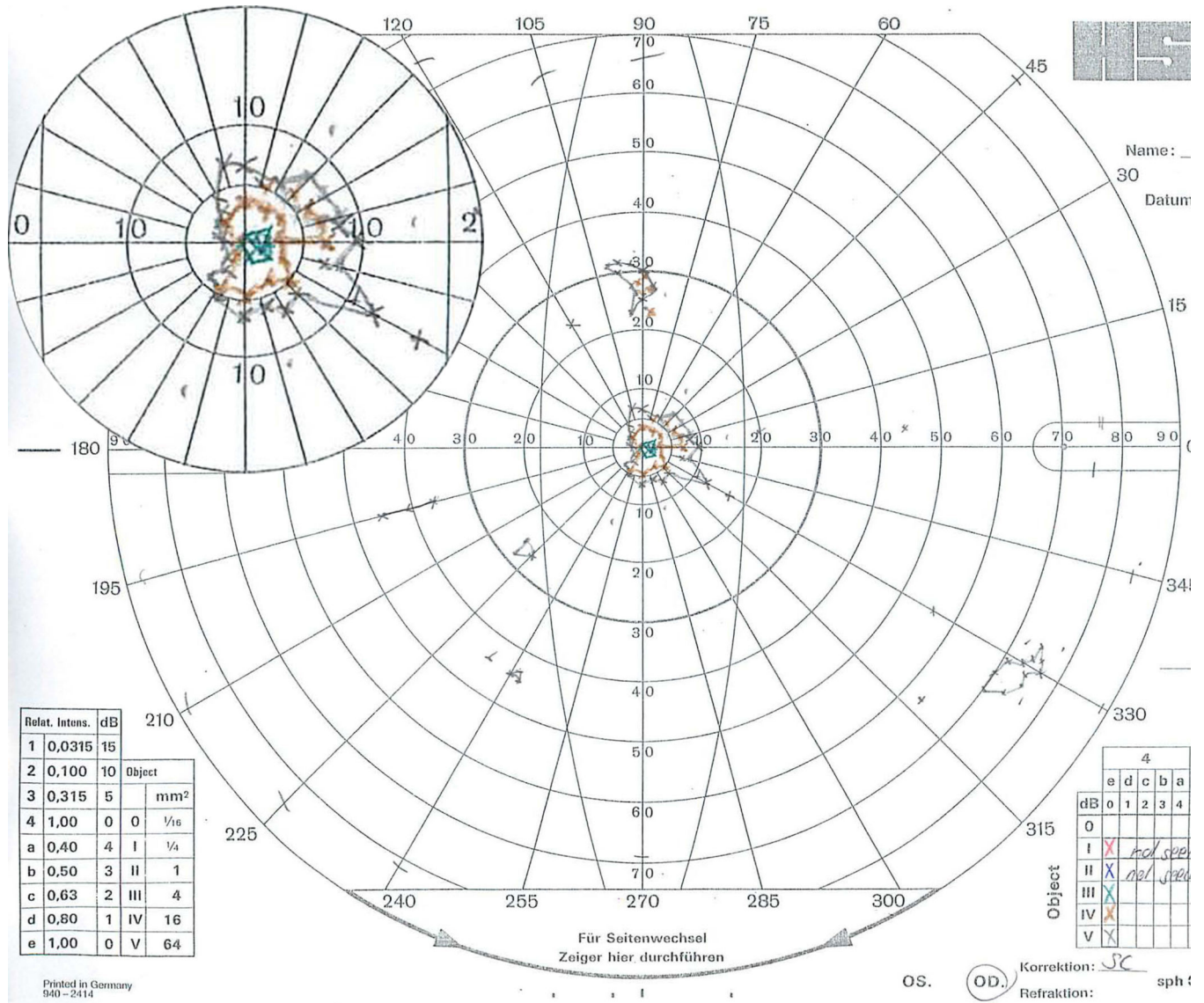


Figure 5.

GVF 13 (31 isopter misses). The two small V-4e seeing areas in the inferior portion of the left hemifield were missed 12 (superior isopter, ~27° inferior-nasal from fixation, along the 225° meridian) and 15 (inferior isopter, ~45° inferior-nasal from fixation, along the 240° meridian) times. These very small isopters likely blended with the markings of the GVF sheet and the operator's markings. Except for Digitizer 6, who only missed the superior left V-4e seeing isopter, all other isopter misses for this field were also accompanied by missing the inferior left V-4e seeing isopter. The top-left inset, provided for clarity and covers no isopters, shows a 200% magnified view of the central 20°.

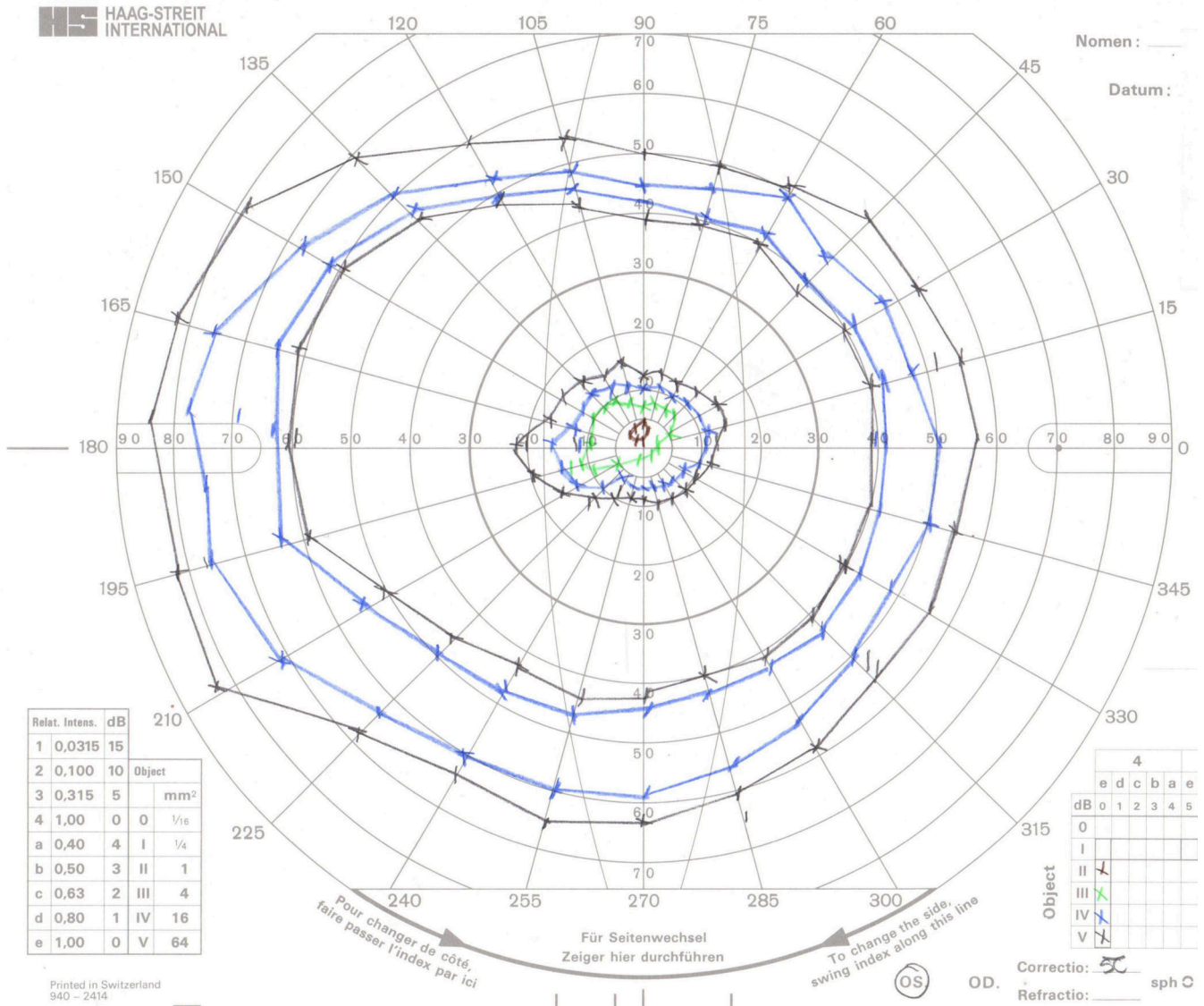
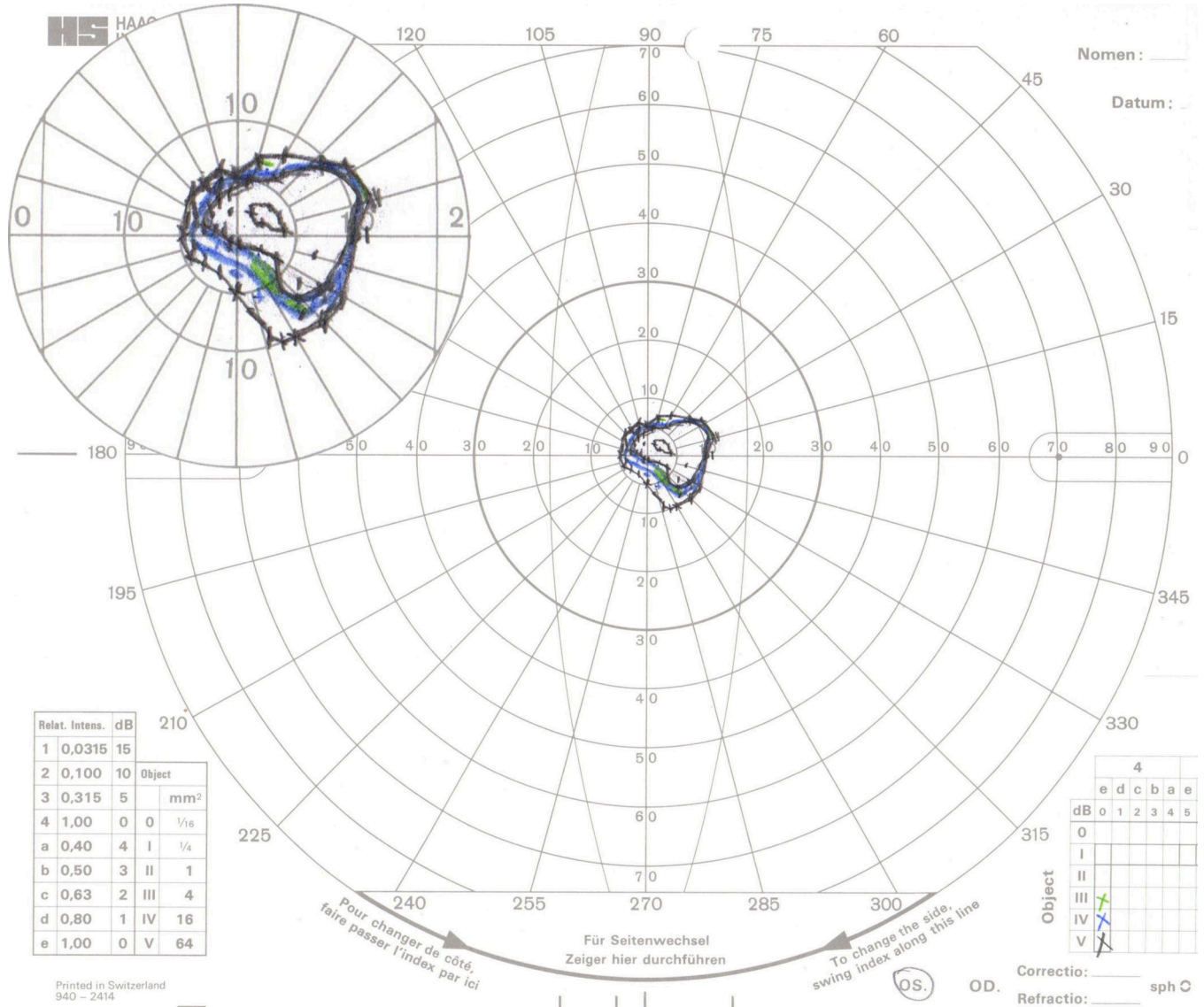


Figure 6.

GVF 15 (24 seeing/non-seeing errors). Digitizers 2, 4, and 8 consistently made seeing/non-seeing errors with this field containing a central seeing island, mid-peripheral ring scotoma and far peripherally seen ring of vision. No other digitizers made any seeing/non-seeing errors here. In these cases, the V-4e and IV-4e non-seeing isopters and/or the central V-4e and IV-4e seeing isopters were mislabeled. Digitizer 2 consistently mislabeled the central seeing isopters as non-seeing. Digitizer 4 consistently mislabeled the non-seeing isopters. Digitizer 8 consistently mislabeled both the non-seeing isopters and the central seeing isopters.



**Figure 7.**

GVF 16 (31 isopter misses, 15 seeing/non-seeing errors). This was the most complicated GVF presented to the digitizers. All isopters were within the central 20° of this field, and that area is shown with 200% magnification in the top-left inset. The III-4e seeing area (~5° inferior-nasal from fixation, along the 300–315° meridians) was missed 9 times: consistently by Digitizer 10, 2 of 3 times by Digitizers 4 and 9, and 1 of 3 times by Digitizers 2 and 6. Missing the true III-4e seeing isopter was accompanied by mislabeling the IV-4e non-seeing isopter as the III-4e seeing isopter for Digitizers 4, 6, and 9. No other target-size errors were made for this GVF. The IV-4e non-seeing isopter was missed 18 of 30 times, while the IV-4e seeing isopter was never missed. The V-4e non-seeing and central seeing (~3° superior-nasal from fixation, along the 15–30° meridians) isopters were each missed twice. Missing the V-4e non-seeing isopter was accompanied by mislabeling the central seeing isopter as non-

seeing. The two inner V-4e isopters were each additionally mislabeled as seeing or non-seeing four other times.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Characteristics of GVF's used for digitizing. Excluded GVF's were not included in analyses because they were not consistently provided to digitizers due to an administrative error.

**Table 1**

GVF Characteristics										
	Count	Isopter count per GVF			Isopter retinal area (mm <sup>2</sup> )			Percent non-seeing isopters per GVF		
		Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Presented GVF's	27	5.3	2	9	99	0.085	700	16%	0%	67%
Analyzed GVF's	23	5.1	2	9	110	0.085	700	17%	0%	67%
Excluded GVF's	4	6	5	7	94	0.087	440	16%	0%	29%

**Table 2**

Distribution of SDs of Isopter Group (IG) area measurements within IGs. (Note: IG area = total seeing area in a GVF recorded with a particular target size)

SDs across IGs (mm <sup>2</sup> )				
SD type, within IGs	Mean	Median	Minimum	Maximum
Overall (all measurements)	2.7	1.0	0.035	88
Between digitizers	1.7	0.70	0.022	51
Within digitizers	1.4	0.52	0.0050	280

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

Distribution of relative SDs of Isopter Group (IG) area measurements, within IGs (dimensionless, SD divided by overall mean for IG area, expressed as %).

<b>Relative SDs across IGs</b>				
<b>Relative SD type, within IGs</b>	<b>Mean</b>	<b>Median</b>	<b>Minimum</b>	<b>Maximum</b>
Overall (CV)	7.4%	2.0%	0.21%	68%
Between digitizers	5.4%	1.7%	0.14%	76%
Within digitizers	4.9%	1.0%	0.036%	69%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4**

Error types and instance counts for each digitizer.

Errors made by digitizers			
Digitizer number	Isopters missed	Target size errors	Seeing/non-seeing errors
1	8	2	2
2	35	8	21
3	3	0	0
4	20	10	36
5	2	6	6
6	25	5	4
7	8	4	5
8	5	4	30
9	5	11	13
10	9	2	4

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



Distinct errors made by 2 or more digitizers in the same block, after excluding digitizers 2, 4, 6, 8, and 9.

**Table 5**

Common errors from the 5 most accurate digitizers						
Error type	Field	Target	Seeing value	Reported target	Reported seeing value	Instances by better digitizers
Miss	13	V-4e	1	NA	NA	2
Miss	13	V-4e	1	NA	NA	4
Miss	13	V-4e	1	NA	NA	2
Miss	13	V-4e	1	NA	NA	2
Miss	16	IV-4e	-1	NA	NA	3
Miss	16	IV-4e	-1	NA	NA	3
Miss	16	IV-4e	-1	NA	NA	3
Target size	5	II-4e	1	V-4e	-1	2
Target size	5	II-4e	1	V-4e	-1	3
Target size	5	II-4e	1	V-4e	-1	2
Target size	5	II-4e	1	V-4e	-1	3
Seeing/non-seeing	5	II-4e	1	V-4e	-1	2
Seeing/non-seeing	5	II-4e	1	V-4e	-1	2
Seeing/non-seeing	5	II-4e	1	V-4e	-1	3
Seeing/non-seeing	5	II-4e	1	II-4e	-1	2
Seeing/non-seeing	5	II-4e	1	V-4e	-1	2
Seeing/non-seeing	5	II-4e	1	V-4e	-1	3

Recommended work group structures for detecting errors in GVF digitization. The missed error percentages for random digitizer selection represent the maximum observed within 99% of this study's simulations. The missed errors for screened digitizers were those observed for the best digitizers in this study. Performance of expert readers was not evaluated in this study.

**Table 6**

Digitizing work group structures and errors missed						
Error detection method	Supervisor tasks	Redundant digitizing	Digitizer selection method	Digitizers	Errors missed	
Expert reader review	Check all GVF data for errors	No	Random	1+	-	
Flag digitizing discrepancies	Investigate discrepancies only	Yes	Random	5	5%	
				7	1%	
			8	0%		
			Screened	2	15% (2/13)	
				3	0% (0/21)	