

Comparison of Snellen and Early Treatment Diabetic Retinopathy Study charts using a computer simulation

Reuben R. Shamir¹, Yael Friedman², Leo Joskowicz¹, Michael Mimouni³, Eytan Z. Blumenthal³

¹School of Engineering and Computer Science, the Hebrew University, Jerusalem 91120, Israel

²Department of Ophthalmology, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel

³Department of Ophthalmology, Rambam Medical Center, Haifa 31096, Israel

Correspondence to: Eytan Z. Blumenthal. Department of Ophthalmology, Rambam Medical Center P.O.B 9602, Haifa 31096, Israel. e_blumenthal@rambam.health.gov.il

Received: 2014-05-23 Accepted: 2015-03-30

Abstract

• **AIM:** To compare accuracy, reproducibility and test duration for the Snellen and the Early Treatment Diabetic Retinopathy Study (ETDRS) charts, two main tools used to measure visual acuity (VA).

• **METHODS:** A computer simulation was programmed to run multiple virtual patients, each with a unique set of assigned parameters, including VA, false –positive and false –negative error values. For each virtual patient, assigned VA was randomly chosen along a continuous scale spanning the range between 1.0 to 0.0 logMAR units (equivalent to 20/200 to 20/20). Each of 30 000 virtual patients were run ten times on each of the two VA charts.

• **RESULTS:** Average test duration (expressed as the total number of characters presented during the test \pm SD) was 12.6 ± 11.1 and 31.2 ± 14.7 characters, for the Snellen and ETDRS, respectively. Accuracy, defined as the absolute difference (\pm SD) between the assigned VA and the measured VA, expressed in logMAR units, was superior in the ETDRS charts: 0.12 ± 0.14 and 0.08 ± 0.08 , for the Snellen and ETDRS charts, respectively. Reproducibility, expressed as test –retest variability, was superior in the ETDRS charts: 0.23 ± 0.17 and 0.11 ± 0.09 logMAR units, for the Snellen and ETDRS charts, respectively.

• **CONCLUSION:** A comparison of true (assigned) VA to measured VA, demonstrated, on average, better accuracy and reproducibility of the ETDRS chart, but at the penalty of significantly longer test duration. These differences were most pronounced in the low VA range. The reproducibility using a simulation approach is in line with reproducibility values found in several clinical studies.

• **KEYWORDS:** Early Treatment Diabetic Retinopathy Study; Snellen; computer simulation; visual acuity testing; virtual patients

DOI:10.18240/ijo.2016.01.20

Shamir RR, Friedman Y, Joskowicz L, Mimouni M, Blumenthal EZ. Comparison of Snellen and Early Treatment Diabetic Retinopathy Study charts using a computer simulation. *Int J Ophthalmol* 2016;9(1):119–123

INTRODUCTION

Visual acuity (VA) has been chosen as an endpoint in countless clinical studies, including trials on diabetic retinopathy [1-2], macular degeneration [3], cataract surgery [4-5], endophthalmitis [6], refractive surgery [7] and others. Two commonly used tools for testing VA in both the clinical and research setting include the Snellen and the Early Treatment Diabetic Retinopathy Study (ETDRS) VA charts [8-9].

The Snellen VA chart, a frequently used chart for measuring VA, consists of letters whose size is determined by the visual angle they occupy. The main drawbacks of this chart include the following: different rows have significantly different number of characters (Table 1), spanning from 1-8 characters per row; there is an inconsistent decrease in letter size from one row to the next; and not all presented characters are equally legible [8]. The ETDRS chart was introduced in the 1980s [1], following recognition of the significant limitations of the Snellen chart, especially so in the low VA range. The main improvements incorporated into the ETDRS chart include: an equal number of characters per row (Table 1); an equal logarithmic decrement between successive rows; and the use of character types which are of relatively uniform legibility [8-9]. While it is generally agreed that the ETDRS chart has several distinct advantages over the Snellen chart, and has gained a role in clinical trials, its acceptance and penetration into the routine clinical encounter has been limited, perhaps related to the size and bulkiness of the chart, its cost, prolonged testing time and availability.

The reproducibility achieved during VA testing using each of these two commonly used charts has been previously studied [10-15]. These clinical studies have focused on reproducibility, a common surrogate for accuracy, because a method for determining true VA, as opposed to measured VA, does not exist.

Comparing Snellen and ETDRS Charts

Table 1 A description of each of the 8 Snellen and 11 ETDRS chart rows, spanning the VA range: 6/60-6/6

Charts	Visual acuity (decimal)	Visual acuity (logMAR)	Metric/Feet equivalent	No. of characters
Snellen	0.1	1.0	6/60; 20/200	1
	0.2	0.7	6/30; 20/100	2
	0.3	0.52	6/20; 20/70	3
	0.4	0.4	6/15; 20/50	4
	0.5	0.3	6/12; 20/40	5
	0.6	0.22	6/10; 20/30	6
	0.8	0.1	6/7.5; 20/25	7
	1.0	0.0	6/6; 20/20	8
ETDRS	0.1	1.0	6/60; 20/200	5
	0.125	0.9	6/48; 20/160	5
	0.16	0.8	6/38; 20/125	5
	0.20	0.7	6/30; 20/100	5
	0.25	0.6	6/24; 20/80	5
	0.32	0.5	6/19; 20/63	5
	0.40	0.4	6/15; 20/50	5
	0.50	0.3	6/12; 20/40	5
	0.63	0.2	6/9.5; 20/32	5
	0.80	0.1	6/7.5; 20/25	5
	1.0	0.0	6/6; 20/20	5

For this study we created a computer simulation model to evaluate and compare the Snellen and ETDRS charts. Virtual patients were used to help quantify accuracy, reproducibility and test duration for testing with each of the above VA charts. Developing a simulation model provides us with the unique possibility to isolate and study the contribution of individual factors on accuracy in VA testing. It also has the advantage of allowing us to test VA accuracy directly, which we defined as the absolute difference between the assigned VA and the measured VA, a comparison which would be impossible to measure by testing real patients.

Given the unique ability of computer simulations to study specific questions in a well defined and fixed setting, they have been used extensively in evaluating visual field (VF) algorithms [16-19] as well as VF progression tools [20-22]. In this study we exploit a computer simulation to provide data that can contribute towards developing a more refined and accurate test for VA.

MATERIALS AND METHODS

MATLAB, version 6.1 (The MathWorks, Natick, MA, USA) a high level language computing and modeling software platform was used to create a computer simulation that tested the VA of 30 000 virtual patients each carrying a unique set of parameters.

Visual Acuity Charts The Revised 2000 Series ETDRS charts (Precision Vision, La Salle, Illinois, USA) and the Traditional Snellen Eye Chart (Precision Vision, La Salle, Illinois, USA) were modeled into the simulation.

Visual Acuity Assignment Each of the virtual patients was randomly assigned a "true VA" from a continuous scale

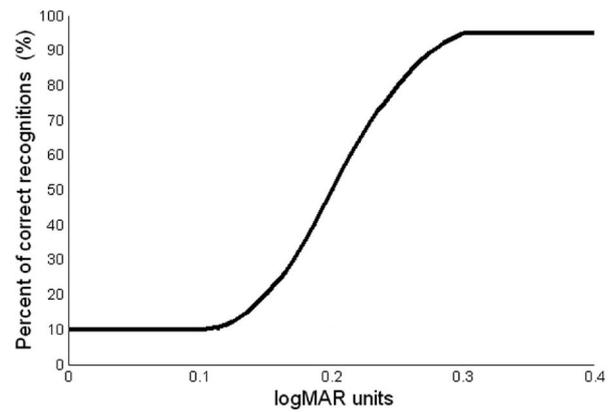


Figure 1 Probability of correctly determining the character of a certain logMAR unit.

spanning the range 1.0-0.0 logMAR units, equivalent to 0.1-1.0 decimal units, or 20/200-20/20 (Table 2). A frequency of seeing curve, analogous to the frequency of seeing curve described in relation to VF's [23], was developed to describe a patient's chance of correctly identifying a character of a given size, based on the patient's true VA. The curve graphs the character size in logMAR units along the x-axis, and the probability (from 0-100%) of correctly determining the character on the y-axis (Figure 1). Based on a study that determined that changes of 0.2 logMAR or greater can be consistently differentiated from no change [24], the curve width was randomly assigned to span the range between 0.15-0.25 logMAR units with a uniform distribution.

False Positive and False Negative Responses In order to best simulate real VA testing, false positive and false negative responses were incorporated into the simulation. A false positive refers to a scenario where a virtual patient recognizes a character that based on the VA frequency of seeing curve, should not have been identified correctly. This reflects a real life situation where a real patient would correctly name a character from random guessing. With ten characters to choose from, the false positive rate modeled for this simulation was 10% on average (range: 0-20%).

A false negative response is an incorrect recognition of characters that according to the patients VA frequency of seeing curve should have been identified correctly. False negative errors reflect real life situations such as absent mindedness as well as errors in hearing and documenting on the part of the examiner. False negative responses were incorporated at an average rate of 5% (range: 0-10%).

Test Termination Criteria Test termination criteria were established to avoid testing the virtual patient on each and every row, down to the bottom of the chart. In this study, the VA test terminates when <50% of the characters in the row have been correctly identified, similar to the termination criteria used in most clinical settings. Of note, while VA is often tested in the clinic up until a VA of 20/20 (1.0, 6/6) and

Table 2 “Virtual patient” characteristics: “intended values” compared to “assigned values” randomly generated for each of the 30 000 “virtual patients”

VA Range	<i>n</i> (virtual patients)	VA (logMAR units)	Width of “frequency of seeing curve” (logMAR units)	False negative rates (%)	False positive rates (%)
All VA (0.0-1.0)					
Intended (mean)	30000	0.5	0.2 (range: 0.15-0.25)	5 (range: 0-10)	10 (range: 0-20)
Assigned	30000	0.502±0.289	0.2002±0.0289	4.99±2.89	10.00±5.77
High VA (0.0-0.33)					
Intended (mean)	10000	0.165	0.2 (range: 0.15-0.25)	5 (range: 0-10)	10 (range: 0-20)
Assigned	9863	0.165±0.095	0.2002±0.0289	4.94±2.90	10.0 ±5.80
Moderate VA (0.33-0.67)					
Intended (mean)	10000	0.5	0.2 (range: 0.15-0.25)	5 (range: 0-10)	10 (range: 0-20)
Assigned	10087	0.500±0.097	0.2003±0.0288	5.01±2.89	10.00±5.76
Low VA (0.67-1.0)					
Intended	10000	0.835	0.2 (range: 0.15-0.25)	5 (range: 0-10)	10 (range: 0-20)
Assigned	10050	0.834±0.095	0.2001±0.0289	5.02±2.90	9.99±5.75

VA: Visual acuity; SD: Standard deviation.

Table 3 A comparison of the Snellen and ETDRS chart results

Population subgroups	Snellen			ETDRS			Statistical comparison ¹ : Snellen vs ETDRS		
	Accuracy	Reproducibility (TRV)	Duration	Accuracy	Reproducibility (TRV)	Duration	Accuracy	Reproducibility (TRV)	Duration
All VA (0.0-1.0)	0.119±0.139	0.232±0.167	12.6±11.1	0.077±0.076	0.107±0.085	31.2±14.7	<0.0001	<0.00001	<0.00001
High VA (0.0-0.33)	0.119±0.185	0.277±0.227	25.3±10.0	0.088±0.095	0.134±0.114	47.7±6.9	<0.0001	<0.0001	<0.00001
Moderate VA (0.33-0.67)	0.117±0.127	0.216±0.125	9.0±4.4	0.078±0.071	0.107±0.069	31.4±6.3	<0.0001	<0.0001	<0.00001
Low VA (0.67-1.0)	0.122±0.089	0.205±0.118	3.6±1.8	0.065±0.057	0.080±0.052	14.9±5.6	<0.0001	<0.0001	<0.00001

TRV: Test-retest variability; VA: Visual acuity. ¹Paired *t*-test.

no further, both the Snellen and ETDRS charts contain character lines extending far beyond this cutoff. In this study we chose to recruit "virtual patients" in the 6/60-6/6 VA range. However, each "virtual patient" was tested as far down the chart as he/she could "see", even beyond the 6/6 row, pending the "termination criteria".

Scoring Methods The algorithm by which a set of correct and incorrect character responses are translated into a "measured VA" is known as the scoring method. In our study we chose the "smallest character row for which >50% of the characters were correctly identified", a common scoring method that would be equally feasible, and minimize bias, between the two charts.

Based on the unique set of assigned parameters including assigned VA, frequency of seeing curve width, false positive and false negative values, the simulation can compute for each character size presented, whether the patient would identify it correctly. Each virtual patient underwent two separate VA tests, one using the Snellen and the other using the Original Series ETDRS chart. The testing order in a computer simulation is of no consequence. On each of the charts, the patient was presented every character of each row, one by one, starting from the top row and proceeding downwards, until the test termination criteria was fulfilled. Next, the scoring method was used to determine the measured VA.

Reproducibility In previous clinical trials evaluating the

Snellen and ETDRS VA charts, test-retest variability (TRV) has been used as a measure of reproducibility^[24]. Reproducibility values, expressed as TRV, were determined by running each virtual patient through the VA simulation test ten times. TRV was calculated as ±1.96 standard deviations^[23].

Statistical Analysis MATLAB, version 6.1 (The MathWorks, Natick, MA, USA), and JMP statistical software version 5.0 (SAS Institute, Cary, NC, USA) were used to analyze the data. Statistical analysis performed in this study included descriptive statistics and the paired *t*-test used for comparison of the Snellen and ETDRS chart results. *P*-values <0.05 were considered statistically significant.

RESULTS

Table 3 provides accuracy, reproducibility and test duration data, separately, for both the Snellen and ETDRS charts. In addition, each of these values, across the two charts, is compared for statistical significance. This analysis is provided for the group as a whole, as well as for each of the three VA sub-groups: high, moderate and low VA.

The most noticeable difference between the two charts relates to the length of the VA examination, as measured by the average number of characters presented during the test. The test duration, for Snellen and ETDRS charts, respectively, was found to be: 12.6 and 31.2 for the entire group, 25.3 and 47.7 for high VA subjects, 9.0 and 31.4 for moderate VA subjects and 3.6 and 14.9 for low VA subjects.

Reproducibility, as well as accuracy, both for the group as a whole, as well as in the sub-group analysis, was better for the ETDRS chart, as listed in Table 3. In the population sub-group analysis presented in Table 3, the 30 000 "virtual patients" were sub-divided into three sub-groups based on their assigned VA. This sub-group analysis was performed to determine whether the overall results found in this study equally pertain to patients with high, moderate and low VA.

In respect to accuracy, for the Snellen chart, the low VA group demonstrated, on average, a somewhat larger error, while, for the ETDRS chart, the inverse was found, with the best accuracy achieved for the low VA group. In respect to reproducibility, both charts demonstrated the tightest reproducibility for the low VA group.

Table 3 highlights the differences found between the two charts. The ETDRS chart was found to be, overall, more accurate and reproducible in both the analysis and sub-analysis comparisons. However, with a penalty of significantly longer test duration, found on average to be more than doubled.

DISCUSSION

Many studies evaluating diagnostic tools report reproducibility, rather than accuracy, since accuracy cannot be evaluated in a scenario lacking a superior gold standard, which the outcome of the test could be compared against. Hence, in this simulation study we receive a relatively rare opportunity to evaluate how reproducibility and accuracy might fare with each other. Table 3 demonstrates that mean accuracy was found to be roughly half the reproducibility, when reproducibility is presented as TRV.

It is of interest to compare the reproducibility found using our simulation approach against studies performed on real patients. Our current study, on "virtual patients", found an overall TRV of 0.23 and 0.11 logMAR units, for Snellen and ETDRS charts, respectively. In comparison, prior studies performed on real subjects have found TRV spanning 0.24-0.33 logMAR for Snellen and 0.09-0.20 logMAR for ETDRS charts^[14,25-26].

The ETDRS chart is often considered superior to the Snellen chart. While our data strengthen this conclusion, we highlight significant discrepancies in test duration, which might, as the sole differentiating factor, allow the longer test to simply be proportionally more accurate. This finding is supported by previous studies that have shown that the time taken to complete the ETDRS is longer than that taken to complete the Snellen chart^[27]. Test (or actually number of character) duration differences relates to the difference in the number of characters on each chart in the 6/60 to 6/6 range, specifically 36 in the Snellen chart and 55 in the ETDRS chart. In addition, the longer duration may be related to the different design of the VA chart as there are less characters in the top rows of the Snellen chart and the examination was

performed from the top row and proceeded downwards. It remains to be determined whether the ETDRS chart remains superior even after adjustments are made for test duration.

A simulation approach can analyze aspects of the chart layouts which cannot be studied on "real patients", and, in particular, address the concept of "accuracy". Nevertheless, the approach utilized in this study, harbors several limitations common to all simulation studies, primarily, that the human response cannot be reliably modeled in full. Of note, we did not address the known differential readability of different characters comprising the two character sets^[28]. We did not model a "learning effect", nor a "tiring effect", and, perhaps most illusive, a simulation approach is inherently unable to model unexpected deviations from the expected human response, which in the context of VA is expected to follow the 'S-shaped' frequency of seeing curve. We exchanged test duration for the number of letters presented because tests were virtually run in a computer simulation setting. As we chose to recruit 'virtual patients' in the 6/60-6/6 VA range, this study's conclusions cannot be applied to those with acuities outside this range. As each character is presented to the virtual patient one by one, this study did not model the crowding phenomenon encountered in real life charts.

The testing of VA, which provides the "vital signs" of an eye exam, is a relatively neglected field. A computer simulation approach may assist in evaluating strengths and weaknesses of various chart designs and testing algorithms, in the hope of improving the chart layout, and ultimately devising more sophisticated and accurate computerized methods for measuring VA in both the clinical and research setting.

ACKNOWLEDGEMENTS

Conflicts of Interest: Shamir RR, None; Friedman Y, None; Jaskowicz L, None; Mimouni M, None; Blumenthal EZ, None.

REFERENCES

- 1 Mitchell P, Bressler N, Tolley K, Gallagher M, Petrillo J, Ferreira A, Wood R, Bandello F; RESTORE Study Group. Patient-reported visual function outcomes improve after ranibizumab treatment in patients with vision impairment due to diabetic macular edema: randomized clinical trial. *JAMA Ophthalmol* 2013;131(10):1339-1347.
- 2 Sivaprasad S, Crosby-Nwaobi R, Heng LZ, Peto T, Michaelides M, Hykin P. Injection frequency and response to bevacizumab monotherapy for diabetic macular oedema (BOLT Report 5). *Br J Ophthalmol* 2013;97(9):1177-1180.
- 3 Christen WG, Glynn RJ, Manson JE, MacFadyen J, Bubes V, Schwartz M, Buring JE, Sesso HD, Gaziano JM. Effects of multivitamin supplement on cataract and age-related macular degeneration in a randomized trial of male physicians. *Ophthalmology* 2014;121(2):525-534.
- 4 Wilkins MR, Allan BD, Rubin GS, Findl O, Hollick EJ, Bunce C, Xing W; Moorfields IOL Study Group. Randomized trial of multifocal intraocular lenses versus monovision after bilateral cataract surgery. *Ophthalmology* 2013;120(12):2449-2455.el.

- 5 Reddy KP, Kandulla J, Auffarth GU. Effectiveness and safety of femtosecond laser-assisted lens fragmentation and anterior capsulotomy versus the manual technique in cataract surgery. *J Cataract Refract Surg* 2013;39(9):1297-1306.
- 6 Albrecht E, Richards JC, Pollock T, Cook C, Myers L. Adjunctive use of intravitreal dexamethasone in presumed bacterial endophthalmitis: a randomised trial. *Br J Ophthalmol* 2011;95(10):1385-1388.
- 7 Sia RK, Ryan DS, Edwards JD, Stutzman RD, Bower KS. The U.S. Army Surface Ablation Study: comparison of PRK, MMC-PRK, and LASEK in moderate to high myopia. *J Refract Surg* 2014;30(4):256-264.
- 8 Chen AH, Norazman FN, Buari NH. Comparison of visual acuity estimates using three different letter charts under two ambient room illuminations. *Indian J Ophthalmol* 2012;60(2):101-104.
- 9 Bailey IL, Lovie JE. New design principles for visual acuity letter charts. *Am J Optom Physiol Opt* 1976;53(11):740-745.
- 10 Vesely P, Synek S. Repeatability and reliability of the visual acuity examination on logMAR ETDRS and Snellen chart. *Cesk Slov Oftalmol* 2012;68(2):71-75.
- 11 Shah N, Laidlaw DA, Rashid S, Hysi P. Validation of printed and computerised crowded Kay picture logMAR tests against gold standard ETDRS acuity test chart measurements in adult and amblyopic paediatric subjects. *Eye (Lond)* 2012;26(4):593-600.
- 12 Shah N, Laidlaw DA, Shah SP, Sivasubramaniam S, Bunce C, Cousens S. Computerized repeating and averaging improve the test-retest variability of ETDRS visual acuity measurements: implications for sensitivity and specificity. *Invest Ophthalmol Vis Sci* 2011;52(13):9397-9402.
- 13 Kuo HK, Kuo MT, Tiong IS, Wu PC, Chen YJ, Chen CH. Visual acuity as measured with Landolt C chart and Early Treatment of Diabetic Retinopathy Study (ETDRS) chart. *Graefes Arch Clin Exp Ophthalmol* 2011;249(4):601-605.
- 14 Laidlaw DA, Abbott A, Rosser DA. Development of a clinically feasible logMAR alternative to the Snellen chart: performance of the "compact reduced logMAR" visual acuity chart in amblyopic children. *Br J Ophthalmol* 2003;87(10):1232-1234.
- 15 Ruamviboonsuk P, Tiensuan M, Kunawut C, Masayaanon P. Repeatability of an automated Landolt C test, compared with the early treatment of diabetic retinopathy study (ETDRS) chart testing. *Am J Ophthalmol* 2003;136(4):662-669.
- 16 Spenceley SE, Henson DB. Visual field test simulation and error in threshold estimation. *Br J Ophthalmol* 1996;80(4):304-308.
- 17 Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Development of efficient threshold strategies for frequency doubling technology perimetry using computer simulation. *Invest Ophthalmol Vis Sci* 2002;43(2):322-331.
- 18 Vesti E, Spry PG, Chauhan BC, Johnson CA. Sensitivity differences between real-patient and computer-stimulated visual fields. *J Glaucoma* 2002;11(1):35-45.
- 19 Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Properties of perimetric threshold estimates from full threshold, ZEST, and SITA-like strategies, as determined by computer simulation. *Invest Ophthalmol Vis Sci* 2003;44(11):4787-4795.
- 20 McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefes Arch Clin Exp Ophthalmol* 1995;233(12):750-755.
- 21 Vesti E, Johnson CA, Chauhan BC. Comparison of different methods for detecting glaucomatous visual field progression. *Invest Ophthalmol Vis Sci* 2003;44(9):3873-3879.
- 22 Gardiner SK, Crabb DP. Examination of different pointwise linear regression methods for determining visual field progression. *Invest Ophthalmol Vis Sci* 2002;43(5):1400-1407.
- 23 Anderson DR, Patella VM. *Automated static perimetry*. St. Louis: CV Mosby, 1990;10-35.
- 24 Rosser DA, Cousens SN, Murdoch IE, Fitzke FW, Laidlaw DA. How sensitive to clinical change are ETDRS logMAR visual acuity measurements? *Invest Ophthalmol Vis Sci* 2003;44(8):3278-3281.
- 25 Rosser DA, Laidlaw DA, Murdoch IE. The development of a "reduced logMAR" visual acuity chart for use in routine clinical practice. *Br J Ophthalmol* 2001;85(4):432-436.
- 26 Siderov J, Tiu AL. Variability of measurements of visual acuity in a large eye clinic. *Acta Ophthalmol Scand* 1999;77(6):673-676.
- 27 Lim LA, Frost NA, Powell RJ, Hewson P. Comparison of the ETDRS logMAR, 'compact reduced logMar' and Snellen charts in routine clinical practice. *Eye (Lond)* 2010;24(4):673-677.
- 28 Ricci F, Cedrone C, Cerulli L. Standardized measurement of visual acuity. *Ophthalmic Epidemiol* 1998;5(1):41-53.