



# City Research Online

## City St George's, University of London

**Citation:** Evans, B. E. W., Rodriguez-Carmona, M., Rauscher, F. G., Llapashtica, E., Koefoed, V. F., Ziemssen, F., Nitsche, R., Farini, A., Baldanzi, E., Gómez-Robledo, L., et al (2025). Multicenter evaluation of the color vision screener test. *Journal of the Optical Society of America A Optics, Image Science and Vision*, 42(5), B190-B198. doi: 10.1364/josaa.544985

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/34706/>

**Link to published version:** <https://doi.org/10.1364/josaa.544985>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

# 1 Multicenter evaluation of the Colour 2 Vision Screener test

3		
4	Benjamin E W Evans [1],	ORCID: 0000-0001-6611-4500
5	Marisa Rodriguez-Carmona [1],	ORCID: 0000-0002-7366-6159
6	Franziska G Rauscher [3][4],	ORCID: 0000-0003-0183-0340
7	Emsal Llapashtica [1],	ORCID: 0000-0003-1941-9971
8	Vilhelm F Koefoed [5][6],	ORCID: 0000-0001-7844-4881
9	Focke Ziemssen [1],	ORCID: 0000-0002-3873-0581
10	Rudolph Nitsche [1][3],	ORCID: 0009-0002-7205-1985
11	Alessandro Farini [7],	ORCID: 0000-0003-4088-7622
12	Elisabetta Baldanzi [7],	ORCID: 0000-0001-7671-9351
13	Luis Gómez Robledo [8],	ORCID: 0000-0002-6287-1632
14	Amanda Douglass [9][10],	ORCID: 0000-0002-6821-0034
15	Madeline Baker [9],	ORCID: 0009-0009-5978-3218
16	Roland Quast [11],	ORCID:
17	Sabine Roelcke [11],	ORCID:
18	Steven C. C. Ho [12],	ORCID: 0009-0003-1062-7739
19	John L Barbur [1]	ORCID: 0000-0002-2187-5004
20		
21	[1] The Henry Wellcome Laboratories for Vision Science, Centre for Applied Vision	
22	Research, School of Health and Medical Sciences, City St George's, University of	
23	London, UK.	
24	[2] Department of Ophthalmology, Leipzig University Medical Center, Liebigstrasse 10-14,	
25	04103 Leipzig, Germany	
26	[3] Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University,	
27	Härtelstrasse 16-18, 04107 Leipzig, Germany	
28	[4] Medical Informatics Center - Department of Medical Data Science, Leipzig University	
29	Medical Center, Härtelstrasse 16-18, 04107 Leipzig, Germany	
30	[5] Naval Refractive Surgery Center, San Diego, USA: NRSCSD	
31	[6] University of Bergen, Norway: UiB	
32	[7] ViOLa Visual Optics Lab - National Institute of Optics CNR	
33	[8] University of Granada, Granada, Spain	
34	[9] School of Medicine - Optometry, Faculty of Health, Deakin University, Australia	
35	[10] Department of Optometry and Vision Sciences, The University of Melbourne,	
36	Australia	
37	[11] Medizinisches Zentrum, Stuttgart Airport, Germany	
38	[12] Sunsmile Aeromedical, Hong Kong	
39		

40 [Abstract](#)

41 An international multicenter study was designed and carried out to evaluate the color vision  
42 screener (CVS) test for normal trichromats and congenital color deficient. Over 400  
43 participants from nine international Colour Assessment and Diagnosis (CAD) testing centers  
44 completed the CVS and the CAD test on calibrated visual displays. The CVS had a  
45 sensitivity and specificity [95 percent confidence intervals] of 1.00 [0.98-1.00] and 0.99 [0.97-  
46 1.00] with a positive and negative predictive index of 0.94 and 1.00 for an assumed  
47 prevalence of 8 percent. The CVS is quick, efficient, easy to use, and its sensitivity is  
48 equivalent to the optimal published Ishihara protocol.

49

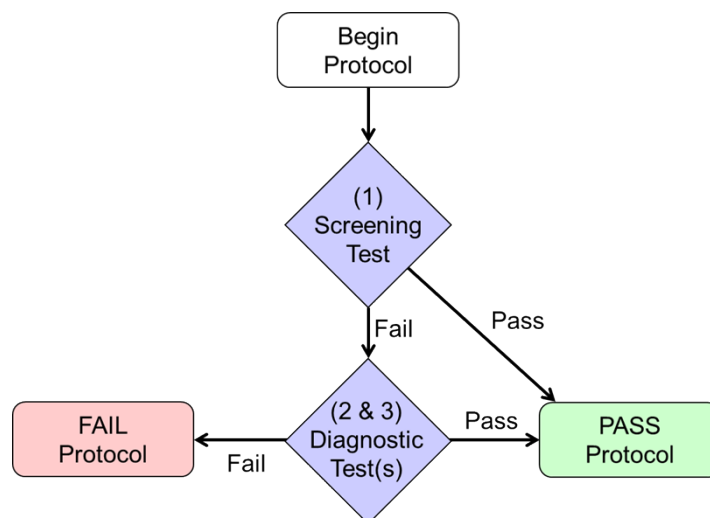
50 Keywords: Color Vision Assessment, Screening, Color Vision Deficiencies, Ishihara, CAD  
51 test, Color Vision

52

53 Introduction

54 The ability to establish an individual's type of color deficiency and to quantify the severity of  
55 color vision loss is of considerable value, both clinically and within occupational settings. [1–  
56 4] The ability to detect efficiently and to identify color signals can enhance visual  
57 performance [5–9], whilst reduced chromatic sensitivity can, particularly where no other  
58 redundant information is included, result in major accidents and, in the worse instance, loss  
59 of life. [10] The majority of individuals have normal trichromatic color vision – approximately  
60 8% of men and ~0.5% of women in Caucasian populations are reported to have congenital  
61 deutan and protan deficiencies whilst congenital tritan deficiencies, following an autosomal  
62 dominant inheritance pattern, are less common [11–14] – resulting in the need for an  
63 efficient screening test to rapidly detect those who require more time-consuming diagnostic  
64 assessment. [15,16] Currently, the Ishihara pseudoisochromatic plate test, a test designed to  
65 screen only for congenital red/green (RG) deficiencies, is often used to fulfil this screening  
66 requirement. [17]

67 As with other screening tests, the value of a color vision screener is determined primarily by  
68 its sensitivity and specificity, the probability that the test will correctly identify individuals with  
69 CVD or normal trichromatic color vision, respectively. [18] Diagnostic color assessment  
70 tests have a higher resource cost but aim to provide more information, such as the  
71 classification of color vision, the presence of combined acquired and congenital deficiency,  
72 and a measure of RG and YB loss. [19–21] Tests are combined and employed using  
73 different sets of rules or protocols to give clinical significance to the results of the test and/or  
74 to achieve a specific outcome. [4,22–24] The most widely used implementation, not specific  
75 to color vision, is screening followed by further diagnostic tests if the screening test is failed  
76 (Fig. 1).



78 *Fig. 1. The general stages in all color assessment protocols. The number and type of tests employed at each*  
79 *stage varies across protocols. Note that this procedure was not followed in this research as all participants were*  
80 *assessed with both the screening test and the gold standard reference measure.*

81 It is of value to outline the requirements for an 'ideal' method for 'full' color vision  
82 assessment. Ideal color vision assessment would fully isolate color signals, ensuring one  
83 can make use of only color cues, allow for the selective simulation of RG and YB chromatic  
84 mechanisms to classify accurately the class of any individual's color vision, quantify the  
85 severity of any RG or YB loss, and have a low resource cost (be inexpensive, quick to carry  
86 out, and easy to administer). Given the range of physiological properties present in the  
87 'normal' population (including variations in the L:M cone ratio, peak wavelength responsivity  
88 of cone photoreceptors, differences in photoreceptor pigment optical densities, and the  
89 effects of normal aging and pre-receptor filtering of light [25–27]), an 'ideal' test also needs  
90 to establish a 'normal' age-matched range, to account for the variation observed in the  
91 'normal' population. [28]

92 Diagnostic color threshold tests, such as the Colour Assessment and Diagnosis [29] (CAD)  
93 and Cambridge Colour Test [30] (CCT), meet several requirements for 'ideal' color vision  
94 assessment. However, both the CAD and CCT have a relatively high resource costs and  
95 take between 12 to 15 minutes to complete in the case of CAD. A potential solution,  
96 proposed previously in 2021, is a two-step color assessment protocol which utilizes the CAD  
97 test in combination with a quick and efficient Colour Vision Screener (CVS) test. [31]

98 The CAD and CVS tests have been described previously [31,32]. Both tests display  
99 moving, color-defined stimuli on a background of dynamic luminance contrast noise to mask  
100 any residual perceived luminance contrast signals. A short learning mode is built into the  
101 procedure of the CAD and CVS and must be passed before each test is undertaken. The  
102 CAD test uses 16 interleaved color directions specified in CIE 1931 (x,y) color space, with  
103 white point chromaticity coordinates of (0.305, 0.323), combined with a four alternative  
104 forced choice (AFC) procedure. The sampling of the hues is arranged to match as closely as  
105 possible the expected directions of deutan, protan, and tritan colour confusion bands. The  
106 CAD test outputs results in CAD units, where one CAD unit is based upon mean color  
107 thresholds measured in 330 healthy, young, normal trichromats [33]. These measured  
108 chromatic thresholds are directly proportional to the cone contrasts generated [34]. The  
109 CVS test uses a 2AFC and the chromaticity of the stimuli rotates through CIE (x,y) color  
110 space during each presentation (restricted to R/G, Y/B, and the suprathreshold regions of  
111 the corresponding color threshold ellipse) for the observer's age. The hue directions  
112 sampled in the CVS approximately match those employed in the CAD test. Suprathreshold  
113 stimuli, generated by adding an additional 150% chromatic contrast and a 45% luminance  
114 contrast component to YB CVS stimuli, are used to determine an individual's response

115 reliability. Measurements where <86% of suprathreshold stimuli are correctly identified are  
116 classified as 'unusable'. It is important to note that the CVS presents stimuli with different  
117 chromatic contrasts for different observers by utilizing the upper normal threshold limits  
118 established for normal aging when using the CAD test. The method establishes whether the  
119 subject's chromatic sensitivity falls within the 'normal' limits for the corresponding age. The  
120 CVS test currently exists in two parallel forms; one version built into the CAD test and  
121 designed to run on calibrated visual displays, and a second freely downloadable form as a  
122 standalone file designed for use on computers running the Windows operating system  
123 connected to uncalibrated visual displays that support the sRGB color mode. It should be  
124 acknowledged that colors are coded, and subsequently rendered, to be reproduced  
125 accurately on an ideal sRGB display, however, as with any production of color on a visual  
126 display, if the screen is uncalibrated one cannot know which color will actually be presented.  
127 This manuscript reports upon results obtained using fully calibrated visual displays,  
128 expanding upon preliminary results for the calibrated version of the test. [31]

129 A widely reported inaccuracy surrounding the CAD test can be attributed to a paper by  
130 Seshadri et al. [35] The 'web-based version of the CAD test' reported by Seshadri et al. is a  
131 ~90 second video showing the stimuli employed in the CAD test. Unfortunately, the fact that  
132 the web-video is simply representative of the stimuli used in the full CAD test is not reported  
133 in a number of publications which quote Seshadri et al., along with inaccurate statements  
134 surrounding the CAD tests' ability to detect YB loss. [36,37]

135 CAD systems, comprising of calibrated hardware and specific software, can be found  
136 worldwide at clinical and occupational centers. The resulting international network of CAD  
137 centers provides a valuable resource that can facilitate examining of a large number of  
138 participants in several separate locations using consistent testing conditions and identical  
139 hardware and software. This international consortium has been facilitated, in no small part,  
140 by the adoption of the CAD test across occupational environments [6,38]. The international  
141 network of CAD centers also enables a multicenter study methodology to be used to collect  
142 and analyze CAD (and CVS) data. A multicenter study, in which research is conducted in  
143 multiple centers following the same protocol, can confer several advantages over single  
144 center studies, including a larger sample size, a more diverse population, and increased  
145 generalizability [39–41].

146 This study aimed to evaluate the recently developed CVS test in an international  
147 collaborative multicenter study, carrying out the test with different examiners in different  
148 population groups. The aim was to establish the CVS's outcome and determine its suitability  
149 through comparison to the most popular color vision screening test for RG CVD, the Ishihara

150 pseudoisochromatic plate test [37,42]. This builds upon work introduced in 2021 [31] and is  
 151 the first of two papers evaluating the outcome of the CVS test on calibrated and uncalibrated  
 152 visual displays.

## 153 Methods

154 In 2019 the international consortium of CAD testing centers was formed, and all centers  
 155 were invited to participate in the validation of the CVS test. This invitation was re-extended  
 156 in 2021 following the disruption to research internationally caused by the COVID-19  
 157 pandemic. Over the course of the study, data collection was actively paused and resumed  
 158 to comply with local, national, and international guidelines. Center participation was  
 159 voluntary, there were no recruitment requirements for inclusion, and every center that  
 160 contributed results was accepted and included in the study. The participating centers and  
 161 researchers are shown in Table 1.

162 Each CAD center within the international consortium is equipped with standardized Advanced  
 163 and Optometric Test (AVOT) equipment, including a 10-bit visual display, a photometer, the  
 164 full CAD test, and programs for automatic calibration of the 10-bit visual display employed to  
 165 generate the visual stimuli. These items were used across all centers to ensure consistency.  
 166 Participating centers were provided with the CVS (v2.6.1 or v2.8) as an embedded option  
 167 within the CAD test. The CVS stimuli and psychophysical procedure, including the on-screen  
 168 instructions given to participants, were the same across centers.

169 *Table 1. The nine CAD testing centers who participated in the international multicenter study. Principal*  
 170 *researchers at each center are listed in alphabetic order.*

Center name	Location	Researcher(s)
City St George's, University of London	London, United Kingdom	Professor John Barbur, Dr Benjamin Evans, Dr Emsal Llapashtica, Dr Marisa Rodriguez-Carmona
Deakin University	Victoria, Australia	Ms Madeline Baker Miss Kate Coffey Dr Amanda Douglass
Leipzig University, Germany	Leipzig, Germany	Mr Rudolph Nitsche, Dr Franziska Rauscher, Professor Dr. med. Focke Ziemssen
Medizinisches Zentrum, Stuttgart Airport	Stuttgart, Germany	Professor Roland Quast, Dr Sabine Roelcke

Naval Refractive Surgery Center, San Diego, USA: NRSCSD	San Diego, USA	Dr Vilhelm F Koefoed
Sunsmile Aeromedical	Hong Kong	Dr Steven C. C. Ho
University of Bergen, Norway: UiB	Bergen, Norway	Dr Vilhelm F Koefoed
University of Granada	Granada, Spain	Professor Luis Gómez Robledo
ViOLa Visual Optics Lab - National Institute of Optics CNR	Florence, Italy	Dr Elisabetta Baldanzi, Professor Alessandro Farini

171

172 Local research and ethical approval was obtained at each center prior to any participant  
173 recruitment, and data collection and participant recruitment was independently managed at  
174 each center. Participants completed the CAD and CVS tests binocularly at a viewing  
175 distance of 1.4m and wore any refractive correction they habitually used for tasks at the  
176 tests' working distance. All examiners were familiar with carrying out the CAD test in routine  
177 clinical practice. The on-screen CVS instructions, always shown in English, ensured that  
178 each participant received standardized instructions, and participants were required to  
179 correctly identify all stimuli in the CAD and CVS test learning modes to ensure they  
180 understood the test procedure prior to taking the test. Informed consent was obtained from  
181 all participants, across all centers. Participants could withdraw from the study at any point  
182 and the study followed the tenets of the Declaration of Helsinki.

183 Data collected at centers were de-identified and securely transferred to the team at City St  
184 George's, University of London, in line with international data protection legislation. Records  
185 for 488 participants collected at nine CAD centers located in Europe, America, Asia, and  
186 Australia were received and exclusion criteria were applied. Exclusion criteria included  
187 duplicates, records with only CVS or CAD test data, participants above the age of 75 or  
188 below the age of 16, and participants with acquired color deficiency, as diagnosed by the  
189 CAD test.

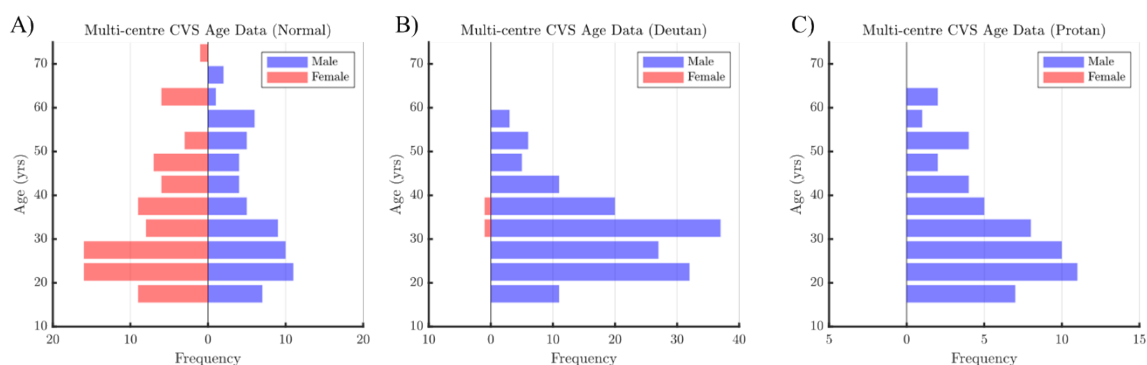
190 CAD data for each participant were used as a reference measure to determine the 'true  
191 status' of participants' color vision (i.e., 'normal trichromat', 'deutan', 'protan', etc.).  
192 Participant's CAD and CVS data were analyzed to determine the sensitivity, specificity, test  
193 accuracy (or efficiency), and positive and negative predictive value (PPV & NPV  
194 respectively) of the CVS. The CVS accuracy, PPV, and NPV were calculated for an  
195 assumed prevalence of 8 percent, which is in line with estimates for the prevalence of  
196 congenital RG CVD in Caucasian male populations [13]. Ninety-five percent confidence  
197 intervals were calculated for the sensitivity and specificity using the Wilson method [43] and

198 the outcome of the CVS was compared to the severity of loss, as determined by the CAD  
199 test.

200 A review of the literature was carried out to extract data for the Ishihara screening tests in  
201 populations of normal trichromats and individuals with congenital RG CVD. The search was  
202 conducted using Google Scholar and PubMed with the keywords (color vision assessment)  
203 AND (Ishihara) AND (anomaloscope). A pre-selection of papers was performed by screening  
204 titles and abstracts for relevance to the topic. Full-text articles meeting the inclusion criteria  
205 were then reviewed in detail. The inclusion criteria required studies with sample sizes of at  
206 least 140 participants, while exclusion criteria eliminated studies where a reference measure  
207 other than CAD or anomaloscopy was used, or where only participants who failed the  
208 Ishihara test completed the reference test to confirm the presence of any color vision  
209 deficiency. No artificial intelligence (AI) methodologies were used in the search or analysis  
210 process. Calculations for sensitivity, specificity, accuracy, PPV, and NPV were carried out  
211 using previously published data and compared to the CVS using the current cohort.

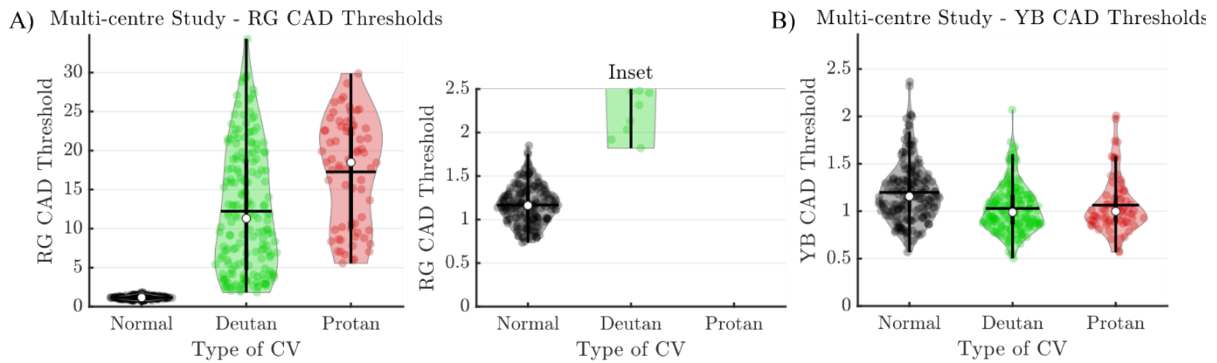
## 212 Results

213 Data from 180 participants with normal trichromatic color vision, 181 with deutan deficiency,  
214 and 69 with protan deficiency were analyzed following application of the exclusion criteria.  
215 Participants ranged from 16-71 years of age with a median [and interquartile range] of 30  
216 [23-40] years. Biological sex data were available for ~82% of the cohort, with the remaining  
217 data unavailable due to international data sharing limitations. The distribution of age for male  
218 and female normal trichromatic, deutan, and protan participants for this subset of the cohort  
219 is shown in Fig. 2. All included participants had YB CAD thresholds within the normal limits  
220 for their age and RG CAD thresholds within the expected range for participants' class of  
221 color vision deficiency, as shown in Fig. 3. No participant had an 'unusable' CVS response  
222 reliability as determined by the identification of suprathreshold stimuli throughout the test.



223

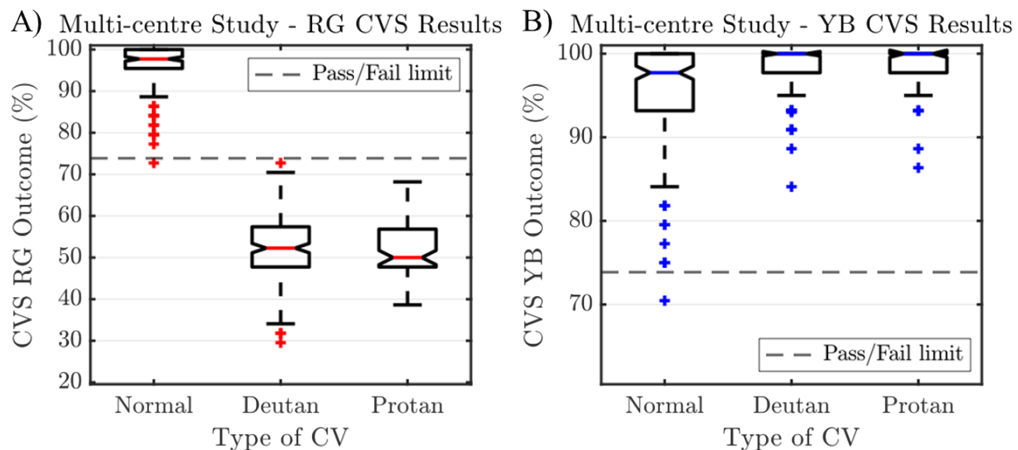
224 *Fig. 2. The age and biological sex distribution for the data collected across all centers. Biological sex data were*  
225 *available for approximately 82% of the cohort, and the distributions are split into A) participants with normal color*  
226 *vision, B) participants with a deutan deficiency, and C) participants with a protan deficiency.*



228

229 Fig. 3. Violin plots showing A) RG and B) YB CAD thresholds for all participants assessed in the multicenter  
 230 study. Violin plots combine a boxplot, a density trace, and the mean [horizontal black bar] into a single  
 231 graphic [44]. RG CAD thresholds ranged from 0.74 to 1.85 in normal trichromats, 1.82 to 34.30 in deutans, and  
 232 5.56 to 29.87 in protan participants. All normal trichromats, protans, and deutans had YB CAD thresholds within  
 233 the normal limits established for their age. All deutan participants have CAD thresholds outside the normal limits  
 234 for their age, and all normal trichromats have thresholds within the normal limits established for their age. The  
 235 overlap between the RG CAD thresholds for the least affected deuteranomalous subjects and the least sensitive  
 236 normal trichromats is due to the range of ages in the subject population and the normal age-adjusted limits  
 237 employed in the CAD test.

238 Across all centers and all participants, with assessments carried out by a range of  
 239 examiners, clinicians, and practitioners, one normal trichromat failed the RG component of  
 240 the CVS test, and one normal trichromat failed the YB component of the CVS test. All 181  
 241 deutans and 69 protans were correctly identified as having RG loss with normal YB  
 242 chromatic sensitivity by the CVS, passing the YB and failing the RG component of the CVS,  
 243 corresponding to a multicenter sensitivity and specificity [and 95% confidence intervals] of  
 244 1.00 [0.98 – 1.00] and 0.99 [0.97 – 1.00], respectively. For an assumed prevalence of 8  
 245 percent, the CVS has a PPV of 0.94 and an NPV of 1.00. At each individual center the RG  
 246 sensitivity was 1.00. The RG specificity was 1.00 at all but two centers and at both of these  
 247 centers one participant with normal RG chromatic sensitivity failed the RG component of the  
 248 CVS.



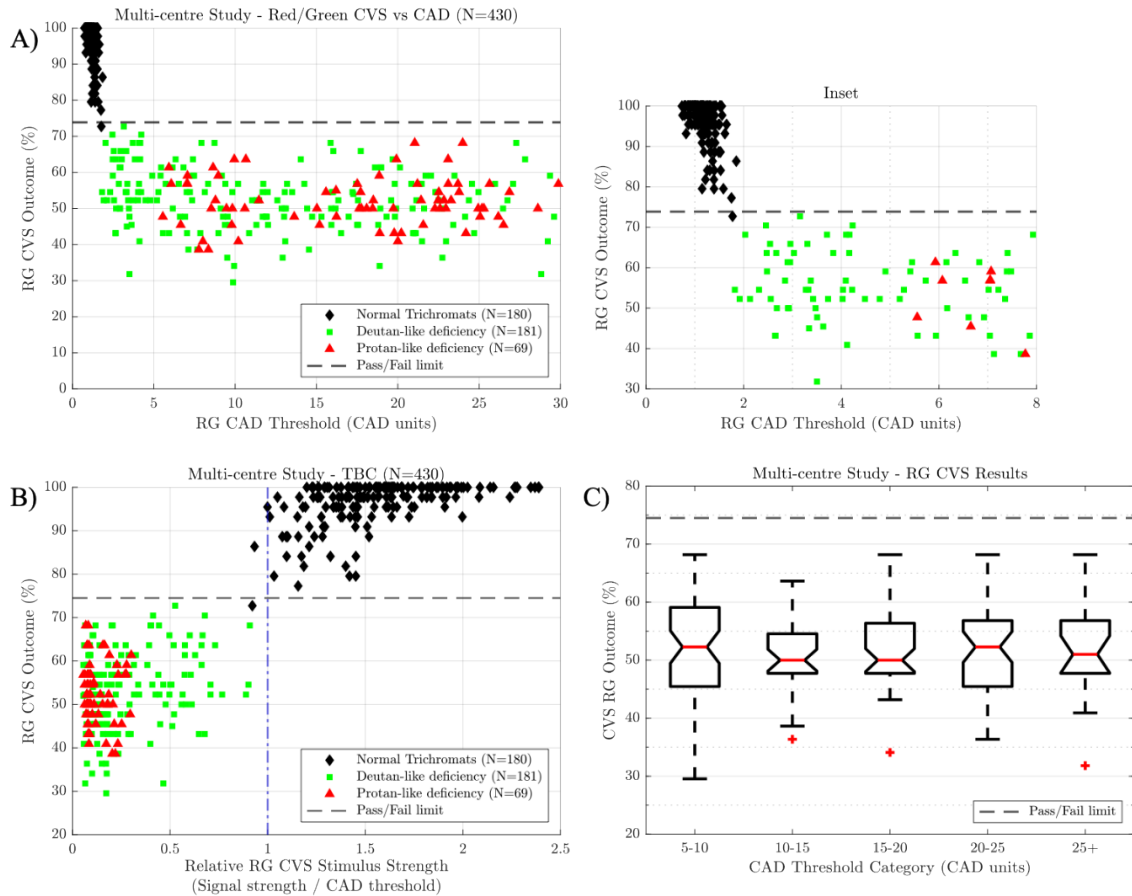
249

250 *Fig. 4. Notched boxplots showing (A) RG and (B) YB CVS outcomes for participants assessed as part of a*  
251 *multicenter CVS study. Values greater than  $q_3 + 1.5(q_3 - q_1)$  or less than  $q_1 - 1.5(q_3 - q_1)$  were classified as*  
252 *outliers (plotted as +). All deuterans and protans assessed were correctly classified by the CVS test. One normal*  
253 *trichromat was misdiagnosed for RG color vision and a different normal trichromat was misdiagnosed for YB*  
254 *color vision.*

255 Data for previously reported Ishihara pseudoisochromatic plate test data are shown in Table  
256 2. As previously reported across multiple studies, maximizing sensitivity occurs at a cost to  
257 specificity and vice-versa. The highest reported sensitivity of the Ishihara is statistically  
258 equivalent to the outcome of the CVS test when carried out on calibrated visual displays.

259 **[Table 2 here – landscape (see page 16)]**

260 The normal trichromat who failed the RG components of the CVS had a RG CAD threshold  
261 of 1.77 with an upper normal age-adjusted limit of 1.79. The relationship between the  
262 severity of RG color deficiency and the outcome of the CVS screener is shown for all  
263 participants in Fig. 5. Fig. 5B reveals the experimental agreement with the predicted test  
264 outcome - the probability of passing the CVS is proportional to an individual's sensitivity to  
265 color, as quantified by the CAD test. The least affected deutan participants, with the  
266 smallest CAD thresholds, have the highest probability within the group of congenital color  
267 deficient of passing the RG component of the CVS. The RG CVS outcome for individuals  
268 with RG CAD thresholds over 5 RG CAD units were statistically equivalent and reflect the  
269 expected spread around the change probability of a correct response (Fig. 5C).



270

271 *Fig. 5. (A) The outcome of 430 participants on the CVS test compared to their outcome on the reference CAD*  
 272 *test for RG stimuli. The CAD results are plotted in terms of CAD units; units based upon a standard young*  
 273 *observer who has a threshold of 1 CAD unit. These units are the standard output of the CAD test. An inset for*  
 274 *the one normal trichromat who failed had a RG CAD threshold of 1.77, and the normal CAD upper limit for their age (22 years) is 1.79. (B) The same data shown in (A) is plotted*  
 275 *using the relative RG CVS stimulus strength employed. The relative RG CVS stimulus strength is calculated by*  
 276 *dividing the signal strength used for each participant (which varies based on their age) by their measured RG*  
 277 *CAD threshold. For example, an observer with a RG CAD threshold two times larger than the stimulus strength*  
 278 *employed in the CVS test would have a relative RG CVS stimulus strength of 2, whereas an observer with a CAD*  
 279 *threshold which is half the signal strength employed would have a relative RG CVS stimulus strength of 0.5. The*  
 280 *least affected, or most sensitive, deutan observers will be shown stimulus strengths ~0.8x their threshold,*  
 281 *whereas the least affected, or more sensitive, protan observers are shown stimulus strengths ~0.25x their*  
 282 *threshold. (C) A selection of the data shown in (A) in notched boxplots, grouped in CAD threshold categories,*  
 283 *showcasing the equitability for the outcome of the RG CVS in individuals with CAD thresholds over 5 CAD units.*  
 284

## 285 Discussion

286 The CVS is a rapid screening test, typically taking 2-3 minutes to complete and is easy for  
 287 participants to understand and testers to administer. The present evaluation of this test in  
 288 430 participants across eight countries reveals that the test simultaneously achieves a high  
 289 sensitivity (1.00) and specificity (0.99). Table 2 demonstrates the trade-off between  
 290 sensitivity and specificity present in the most commonly used color screener, the Ishihara  
 291 pseudoisochromatic plate test. This compromise is further elucidated as the study sample  
 292 size increases. With a singular pass criterion, the CVS achieves simultaneously a  
 293 statistically equivalent sensitivity and specificity to the highest sensitivity and specificity

294 obtained in separate protocols for the most common screening test in current use, the  
295 Ishihara pseudoisochromatic plates.

296 Unlike the Ishihara test, the CVS can also screen for YB loss with high specificity (0.99  
297 [0.97-1.00]). Whilst the specificity of the YB CVS was high (only one normal trichromat failed  
298 the YB component of the CVS), the sensitivity of the YB CVS has yet to be experimentally  
299 established, not being determined here due to the exclusion criteria for this study. A  
300 potential challenge with doing so is the low prevalence of acquired color vision loss,  
301 particularly in younger participants, and confounding factors that may drive such acquired  
302 loss.

303 The PPV and NPV are the proportion of individuals who fail a test who are correctly  
304 diagnosed as having a CVD and the proportion of individuals who pass a test who are  
305 correctly diagnosed as not having a CVD, respectively. The PPV and NPV were calculated  
306 for a fixed prevalence of 8 percent to allow for a comparison between studies and provide a  
307 more accurate representation of the tests' expected performance in a male population. As  
308 with sensitivity and specificity, a good screening test maximizes both PPV and NPV. A high  
309 NPV ensures one is confident that individuals who pass have normal trichromatic color  
310 vision, which is of particular importance in occupational settings where one wishes to detect  
311 all CVDs at the screening stage. The NPV and the sensitivity are maximized when the pass  
312 protocol becomes more stringent. In practical terms, doing so maximizes the 'safety' of an  
313 occupational protocol at an increased resource cost, as more applicants need to complete  
314 further diagnostic testing to confirm the presence of any CVD. The alternative is to  
315 maximize specificity, ensuring all, or almost all, normal trichromats pass screening at the  
316 cost of allowing some CVDs to pass, potentially with moderate to severe CVD [23].

317 Whilst a multicenter approach confers several advantages it should be noted that the results  
318 of a large multicenter study are not necessarily applicable to less heterogeneous  
319 populations [45]. Within the context of multicenter color vision research, the age of  
320 participants is a key consideration (an older cohort would have a worse mean chromatic  
321 discrimination), along with the varying prevalence of CVD in different populations.

322 Participants assessed as part of the multicenter study were primarily young and of working  
323 age with a median [and IQR] age of 30 [23-40] years. The age-adjusted nature of the  
324 classification made by the CAD test ensures that any systematic inter-center differences in  
325 the median age of participants are taken into account by the CAD test when the class of  
326 CVD is determined. The incorporation of the normal age-matched limits into the CVS also  
327 means that whilst two participants of the same age at different centers will have been tested  
328 using the same stimuli, two participants of different ages within the same center will be

329 shown different stimuli and participants are screened based upon whether their RG and YB  
330 chromatic sensitivity is within the normal limits *for their age*, not an arbitrary fixed standard.

331 A larger proportion of male CVD participants is expected and consistent with the established  
332 prevalence of deutan and protan deficiencies in male and female populations. Specific to this  
333 multicenter study, CAD testing centers, by virtue of their position, are established,  
334 advertised, and used to assess those with known or suspected CVD and to determine  
335 whether individuals have the level of chromatic discrimination required to work in a specific  
336 occupational setting. Hence, the large prevalence of congenital RG CVD (58%) is not  
337 unexpected, and the study design minimizes the potential impact of varying CVD prevalence  
338 across centers located in different continents. The sensitivity and specificity can be  
339 calculated independent of prevalence, and, as previously, the PPV, NPV, and test accuracy  
340 (or test efficiency), were calculated for a fixed prevalence level of 8 percent.

341 The potential impact of the screening protocol employed upon the observed prevalence of  
342 CVD has been highlighted in a study by Arnegard et al. [46] in which 193 young Norwegian  
343 males were screened for congenital RG CVD. Screening was carried out with the 24-plate  
344 edition of the Ishihara test ( $\geq 3$  errors) and genetic testing using the Agena MassArray  
345 system. Genetic screening revealed a 10.4% prevalence of congenital RG CVD, yet the  
346 results of the Ishihara test only indicated a prevalence of 5.2% in the same sample. A similar  
347 discrepancy between the two screening methods was reported for a female sample. It  
348 should, however, be noted that several other studies have used the Ishihara test to identify  
349 the prevalence of RG CVD using larger samples ( $\geq 5000$  male participants) in European  
350 Caucasian populations and found prevalences  $\sim 8\%$ , including a study carried out in Norway  
351 with 9049 male participants [13,47].

352 One of the primary sources of error in multicenter studies can be a lack of protocol  
353 adherence across centers [39,48,49]. This limitation was minimized through the clear and  
354 concise user instructions displayed as part of the CAD and CVS software every time CAD or  
355 CVS testing is carried out. Both sets of instructions describe the test procedure centers  
356 should follow, and the CVS also provides on-screen instructions to all participants prior to  
357 starting the test, at the end of the learning mode, and at the end of the final CVS test,  
358 ensuring that the instructions provided to participants remained constant across all centers.  
359 The protocol adherence was not externally validated, but the participating centers were  
360 involved in color assessment in aviation and were inspected and had to comply with the  
361 requirements of their Civil Aviation Authorities. The CAD test also records the date of every  
362 display calibration check made by users for compliance with specified requirements.

363 The severity of loss measured by the CAD test for the cohort suggests the inclusion of both  
364 anomalous trichromats and dichromats, as shown by the 'double peak' within violin plots  
365 shown in Fig. 3. The secondary peak is likely attributable to dichromats with worse  
366 chromatic discrimination, and the primary peak attributable to anomalous trichromats. The  
367 maximum severity of loss capable of being measured and quantified by the CAD test is  
368 determined by the gamut of the visual display and the amplitude of the dynamic luminance  
369 contrast noise employed in the test. For the standard CAD test parameters, the maximum  
370 severity of loss is the same for deuterans and protans. Only one participant, a 49-year-old  
371 deutan participant, had a CAD threshold >30 RG CAD units. The least affected deuterans had  
372 a threshold of 1.82 RG CAD units whilst the least affected protan had a CAD threshold of  
373 5.56. This discrepancy has previously been reported and attributed, at least in part, to the  
374 difference in  $\delta\lambda_{\max}$  between each group [26,27,32].

375 The observed variation within presumed dichromats is of interest. This variation is likely due  
376 to factors that also drive the observed variation in the chromatic sensitivity of normal  
377 trichromats such as L:M cone ratio, small shifts in peak wavelength responsivity, differences  
378 in pigment optical density, and variation in pre-receptoral filtering such as the lens and  
379 macular pigment. Such changes contribute to the observed inter-subject variability in normal  
380 color vision. A potential limitation of the multicenter study procedure is the lack of testing via  
381 anomaloscopy, instead relying solely upon the CAD test for a reference measure. The  
382 agreement between CAD and the anomaloscope has previously been found to be high [50],  
383 and whilst the anomaloscope is renowned for its accuracy in distinguishing between protan  
384 and deutan observers - frequently employed as a gold standard for this purpose - the  
385 parameters of the match have poor agreement with an individual's chromatic discrimination  
386 thresholds, the latter being more relevant for occupational environments [51,52].

387 The CVS test results align with the predicted binomial outcome of the test. [31] The  
388 probability of correctly identifying CVS stimuli is determined by two factors, an individual's  
389 chromatic sensitivity and the stimuli shown during the test. As the stimuli's chromatic signal  
390 strength is based solely on an individual's age, the probability of a correct response is based  
391 upon how close an individual's chromatic sensitivity is to the limit for their age. Individuals  
392 with high chromatic sensitivity for their age – the majority of normal trichromats tested – have  
393 the highest *relative* CVS stimulus strength and, hence, the highest probability of correctly  
394 identifying the stimuli (i.e., the stimulus strength used is twice as large as their threshold).  
395 Individuals with lower chromatic sensitivity for their age – those with color deficiency – have  
396 the lowest *relative* CVS stimulus strength and subsequently the lowest probability of  
397 correctly identifying the stimuli. As shown in Fig. 5B, the normal trichromat who failed the  
398 RG component of the CVS had the lowest *relative* CVS RG stimulus strength of all normal

399 trichromats and hence had the smallest probability of a correct response for the stimulus  
400 strength employed in the CVS test. This is acceptable for occupational settings where failing  
401 the CVS indicates the need for further diagnostic testing.

402 A unique strength of the CVS test is that the signal strength displayed is proportional to the  
403 normal upper CAD limit for each participant's age. This property of the CVS allows the test  
404 to work with subjects of any age (from ~10 years of age). Additionally, when viewed from  
405 this relative frame of reference the separation between the least affected deuterans and  
406 protans is strikingly evident. No individual out of the 250 individuals with color deficiency  
407 assessed across all centers passed the CVS. Color deficient individuals with the highest probability of  
408 passing the RG CVS are the least affected deuterans, a small subsection of the deutan  
409 population. Such an outcome is of advantage and has benefits within occupational health  
410 domains.

411 Most participants with congenital CVD, ~78% of deuterans and ~99% of protans have RG  
412 CAD thresholds greater than or equal to 5 CAD units. [32] The outcome of the CVS was  
413 statistically equivalent across all participants within this threshold range. For relative  
414 stimulus strength participant's probability of correctly identifying the CVS stimuli is chance  
415 (50% for the 2AFC procedure), and, as with all CVS outcomes, are governed by a binomial  
416 distribution determined entirely by the probability of a correct response.

417 The 2AFC nature of the test carries advantages and limitations. The primary limitation is  
418 that there is a non-zero chance that an individual with CVD can pass the RG and YB  
419 components of the CVS. However, as one can observe from Table 2, such a limitation is  
420 present in current screening tests, including the Ishihara pseudoisochromatic plates. This  
421 limitation is offset by the CVS test's result being directly proportional to an individual's level  
422 of chromatic sensitivity. The probability of an individual with CVD passing the CVS is less  
423 than 0.1%. Individuals with the highest probability of passing are the least affected and have  
424 the highest chromatic sensitivity within the CVD group. The randomized nature of the CVS  
425 test presentations also presents an advantage by eliminating the possibility for an individual  
426 to memorize the test components.

## 427 Conclusion

428 A multicenter study was carried out to evaluate the outcome of the new CVS test in a large  
429 clinically relevant population containing normal trichromats and individuals with congenital  
430 CVD. Testing carried out by multiple examiners and across multiple locations revealed that  
431 the CVS has high sensitivity and specificity when carried out on spectrally calibrated visual  
432 displays. The outcome of the test achieves at least as good an outcome as the most

433 stringent protocol for RG color vision screening using the Ishihara pseudoisochromatic plate  
434 test. The new CVS test is, however, much more efficient since almost all normal trichromats  
435 pass. When the CVS is employed in a two-step color assessment protocol, the number of  
436 applicants needing a full color assessment is much reduced. The CVS test offers the  
437 potential to rapidly and efficiently screen for congenital CVD in clinical and occupational  
438 settings and integrates cleanly into a 'two-step' protocol for color vision assessment.

439

Gold standard	Publication	Test and pass protocol	N	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)	Assumed prevalence: 8%		
							Accuracy	PPV	NPV
CAD	-	CVS RG - All centers	430	0.58	1.00 (0.98 - 1.00)	0.99 (0.97 - 1.00)	0.99	0.94	1.00
Nagel & CAD	Rodriguez et al. (2021)	Ishihara 38 pl. 0 err pl. 1-25	1827	0.81	0.99 (0.99 - 1.00)	0.81 (0.76 - 0.84)	0.82	0.31	1.00
Nagel & CAD	Rodriguez et al. (2021)	Ishihara 38 pl. 0 err pl. 1-15	1827	0.81	0.99 (0.99 - 1.00)	0.89 (0.85 - 0.92)	0.90	0.44	1.00
Nagel & CAD	Rodriguez et al. (2021)	Ishihara 38 pl. ≤2 err pl. 1-17	1827	0.81	0.97 (0.96 - 0.98)	0.99 (0.97 - 1.00)	0.99	0.88	1.00
Nagel & CAD	Rodriguez et al. (2021)	Ishihara 38 pl. ≤4 err pl. 1-21	1827	0.81	0.96 (0.95 - 0.97)	0.99 (0.98 - 1.00)	0.99	0.94	1.00
Nagel	Birch (1997) & Birch and McKeever (1993)	Ishihara ≤8 errs pl. 2-17	872	0.46	0.81 (0.76 - 0.84)	1.00 (0.99 - 1.00)	0.98	1.00	0.98
Nagel	Birch (1997) & Birch and McKeever (1993)	Ishihara ≤6 errs pl. 2-17	872	0.46	0.94 (0.91 - 0.96)	0.95 (0.93 - 0.97)	0.95	0.64	0.99
Nagel	Birch (1997) & Birch and McKeever (1993)	Ishihara ≤3 errs pl. 2-17	872	0.46	0.99 (0.97 - 0.99)	0.94 (0.92 - 0.96)	0.94	0.59	1.00
Nagel	Birch (2010)	Ishihara 38 pl. ≤3 err pl. 2-17	486	1.00	0.98 (0.96 - 0.99)	-	-	-	-
Nagel	Birch (2010)	Ishihara 38 pl. ≤4 err pl. 2-17	486	1.00	0.95 (0.93 - 0.96)	-	-	-	-
Nagel	Aarnisalo (1979)	Ishihara 38 pl. 0 err pl. 1-25	150	0.33	1.00 (0.93 - 1.00)	0.67 (0.57 - 0.75)	0.70	0.21	1.00
Nagel	Aarnisalo (1979)	Ishihara 38 pl. ≤1 err pl. 1-25	150	0.33	0.96 (0.87 - 0.99)	0.95 (0.89 - 0.98)	0.95	0.63	1.00
Nagel	Aarnisalo (1979)	Ishihara 38 pl. ≤4 err pl. 1-25	150	0.33	0.84 (0.71 - 0.92)	1.00 (0.96 - 1.00)	0.99	1.00	0.99

Table 2. The number of participants assessed (N), prevalence of congenital RG CVD (Prev), sensitivity (Sens), and specificity (Spfc) for RG CVS data. The positive and negative predictive value (PPV and NPV respectively) have been calculated for a prevalence of 0.08, or 8% (the maximum prevalence observed in male populations). Equivalent statistics for published studies which employed the Ishihara pseudoisochromatic plates have been included to allow for a comparison of the CVS test and the Ishihara pseudoisochromatic plates.

440 **Funding.** The COLT Foundation; UK Civil Aviation Authority; Project PID2022-139056NB-  
441 I00, supported by MICIU/AEI /10.13039/501100011033 and ERDF, EU.

442 **Acknowledgments.** We wish to thank and acknowledge the COLT Foundation and the UK  
443 Civil Aviation Authority for their financial support. We also wish to thank Stuart Mitchell who  
444 has supported us from the UK CAA.

445 Some of this work was carried out under Project PID2022-139056NB-I00, supported by  
446 MICIU/AEI /10.13039/501100011033 and ERDF, EU.

447 **Disclosures.** The authors declare no conflicts of interest. Professor John Barbur designed  
448 all the Advanced Vision and Optometric Tests (AVOT), including the CAD test, but no longer  
449 has any commercial interests in these tests or current conflict of interest.

450 **Data availability.** Data underlying the results presented in this paper are not publicly  
451 available at this time but may be obtained from the authors upon reasonable request.

452

453

454   References

- 455   1.     M. O'Neill-Biba, S. Sivaprasad, M. Rodriguez-Carmona, J. E. Wolf, and J. L. Barbur,  
456         "Loss of chromatic sensitivity in AMD and diabetes: A comparative study," *Ophthalmic*  
457         *and Physiological Optics* **30**, 705–716 (2010).
- 458   2.     C. Feitosa-Santana, G. V. Paramei, M. Nishi, M. Gualtieri, M. F. Costa, and D. F.  
459         Ventura, "Color vision impairment in type 2 diabetes assessed by the D-15d test and  
460         the Cambridge Colour Test," *Ophthalmic and Physiological Optics* **30**, 717–723  
461         (2010).
- 462   3.     M. P. Simunovic, "Acquired color vision deficiency," *Surv Ophthalmol* **61**, 132–155  
463         (2016).
- 464   4.     A. Stockman, *Report on Colour Vision Standards for STU Officers* (2018).
- 465   5.     H. C. Walkey, A. Hurden, I. R. Moorhead, J. A. F. Taylor, J. L. Barbur, and J. A.  
466         Harlow, "Effective contrast of colored stimuli in the mesopic range: a metric for  
467         perceived contrast based on achromatic luminance contrast," *Journal of the Optical*  
468         *Society of America A* **22**, 17 (2005).
- 469   6.     J. Barbur, M. Rodriguez-Carmona, S. Evans, and N. Milburn, "Minimum Colour Vision  
470         Requirements for Professional Flight Crew," *Caa* 80 (2009).
- 471   7.     J. L. Barbur, M. Rodriguez-Carmona, J. Hickey, S. Evans, and A. Chorley, "Analysis  
472         of European colour vision certification requirements for air traffic control officers," *Civil*  
473         *Aviation Authority CAP* **1429**, (2016).
- 474   8.     P. Sumner and J. D. Mollon, "Chromaticity as a signal of ripeness in fruits taken by  
475         primates," *Journal of Experimental Biology* **203**, 1987–2000 (2000).
- 476   9.     B. C. Regan, C. Julliot, B. Simmen, F. Viénot, P. Charles-Dominique, and J. D.  
477         Mollon, "Fruits, foliage and the evolution of primate colour vision," *Philosophical*  
478         *Transactions of the Royal Society B: Biological Sciences* **356**, 229–283 (2001).
- 479   10.    J. D. Mollon and L. R. Cavonius, "The Lagerlunda Collision and the Introduction of  
480         Color Vision Testing," *Surv Ophthalmol* **57**, 178–194 (2012).
- 481   11.    M. Neitz and J. Neitz, "Molecular Genetics of Color Vision and Color Vision Defects,"  
482         *Archives of Ophthalmology* **118**, 691 (2000).
- 483   12.    J. Neitz and M. Neitz, "The genetics of normal and defective color vision," *Vision Res*  
484         **51**, 633–651 (2011).

- 485 13. J. Birch, "Worldwide prevalence of red-green color deficiency," *Journal of the Optical*  
486 *Society of America A* (2012).
- 487 14. L. N. Went and N. Pronk, "The genetics of tritan disturbances," *Hum Genet* **69**, 255–  
488 262 (1985).
- 489 15. D. M. Goodman and E. H. Livingston, "Screening Tests," *JAMA* **309**, 1185 (2013).
- 490 16. P. A. Davison and G. Scanlon, "Comparative analysis of four color vision screening  
491 tests benchmarked by anomaloscopy for detection and investigation of protanomaly  
492 and deuteranomaly," *Color Res Appl* (2024).
- 493 17. S. J. Dain, "Clinical colour vision tests," *Clin Exp Optom* 276–293 (2004).
- 494 18. D. G. Altman, "Practical statistics for medical research," *Stat Med* **10**, 1635–1636  
495 (1991).
- 496 19. J. Birch and M. Rodríguez-Carmona, "Occupational color vision standards: new  
497 prospects," *Journal of the Optical Society of America A* **31**, A55 (2014).
- 498 20. J. Birch, "Classification of anomalous trichromatism with the Nagel anomaloscope," in  
499 *Drum B (Ed) Colour Vision Deficiencies XI. Kluwer Aca- Demic Press, Netherlands*  
500 (1993), Vol. 56, pp. 19–24.
- 501 21. S. J. Dain, A. Casolin, J. Long, and M. R. Hilmi, "Color Vision and the Railways: Part  
502 1. The Railway LED Lantern Test," *Optometry and Vision Science* **92**, 138–146  
503 (2015).
- 504 22. T. H. Margrain, J. Birch, and C. G. Owen, "Colour vision requirements of firefighters,"  
505 *Occup Med (Chic Ill)* **46**, 114–124 (1996).
- 506 23. M. Rodríguez-Carmona, B. E. W. Evans, and J. L. Barbur, "Color vision  
507 assessment-2: Color assessment outcomes using single and multi-test protocols,"  
508 *Color Res Appl* **46**, 21–32 (2021).
- 509 24. V. F. Koefoed, T. Miles, J. B. Cason, and R. Troche, "Colour vision classification –  
510 comparing CAD and CIE 143:2001 International recommendations for colour vision  
511 requirements in transport," *Acta Ophthalmol* **98**, 726–735 (2020).
- 512 25. M. Rodríguez-Carmona and J. L. Barbur, "Variability in normal and defective colour  
513 vision," in *Colour Design*, J. Best, ed., 2nd ed. (Cambridge: Woodhead Publishing  
514 Limited, 2017), pp. 43–97.

- 515 26. P. B. M. Thomas and J. D. Mollon, "Modelling the Rayleigh match," *Vis Neurosci* **21**,  
516 477–482 (2004).
- 517 27. J. L. Barbur, M. Rodriguez-Carmona, J. A. Harlow, K. Mancuso, J. Neitz, and M.  
518 Neitz, "A study of unusual Rayleigh matches in deutan deficiency.," *Vis Neurosci* **25**,  
519 507–516 (2008).
- 520 28. Jennifer. Birch, *Diagnosis of Defective Colour Vision* (Butterworth-Heinemann, 2001).
- 521 29. M. Rodriguez-Carmona, J. Harlow, G. Walker, and J. Barbur, "The Variability of  
522 Normal Trichromatic Vision and the Establishment of the "Normal" Range," (2005).
- 523 30. J. D. Mollon and J. P. Reffin, "A computer-controlled colour vision test that combines  
524 the principles of Chibret and of Stilling," *Proceedings of the Physiological Society* **10–**  
525 **11**, 414 (1989).
- 526 31. J. L. Barbur, M. Rodriguez-Carmona, and B. E. W. Evans, "Color vision  
527 assessment-3. An efficient, two-step, color assessment protocol," *Color Res Appl* **46**,  
528 33–45 (2021).
- 529 32. J. L. Barbur and M. Rodriguez-Carmona, "Colour vision requirements in visually  
530 demanding occupations," *Br Med Bull* **122**, 51–77 (2017).
- 531 33. M. Rodriguez-Carmona, M. O'Neill-Biba, and J. L. Barbur, "Assessing the severity of  
532 color vision loss with implications for aviation and other occupational environments,"  
533 *Aviat Space Environ Med* **83**, 19–29 (2012).
- 534 34. M. Rodriguez-Carmona and J. L. Barbur, "Variability in normal and defective colour  
535 vision," in *Colour Design*, J. Best, ed. (Cambridge: Woodhead Publishing Limited,  
536 2012), pp. 43–97.
- 537 35. J. Seshadri, J. Christensen, V. Lakshminarayanan, and C. J. Bassi, "Evaluation of the  
538 new web-based "Colour Assessment and Diagnosis" test," *Optometry and Vision*  
539 *Science* **82**, 882–885 (2005).
- 540 36. A. French, K. Rose, E. Cornell, and K. Thompson, "The Evolution of Colour Vision  
541 Testing," *Australian Orthoptic Journal* **40**, 7–15 (2008).
- 542 37. A. Fanlo Zarazaga, J. Gutiérrez Vásquez, and V. Pueyo Royo, "Review of the main  
543 colour vision clinical assessment tests," *Archivos de la Sociedad Española de*  
544 *Oftalmología (English Edition)* **94**, 25–32 (2019).
- 545 38. T. Carter and J. Barbur, *Colour Vision Assessment for Maritime Navigational*  
546 *Lookout : Review for UK Maritime and Coastguard Agency (MCA)* (2015).

- 547 39. C. L. Meinert and S. Tonascia, "Single-center versus multicenter trials," in *Clinical*  
548 *Trials: Design, Conduct and Analysis* (1986), pp. 23–29.
- 549 40. L. J. Appel, "A primer on the design, conduct, and interpretation of clinical trials.," *Clin*  
550 *J Am Soc Nephrol* **1**, 1360–1367 (2006).
- 551 41. R. M. Lucas, A. L. Ponsonby, A. J. McMichael, I. van der Mei, C. Chapman, A.  
552 Coulthard, K. Dear, T. Dwyer, T. J. Kilpatrick, M. P. Pender, B. Taylor, P. Valery, and  
553 D. Williams, "Observational analytic studies in multiple sclerosis: Controlling bias  
554 through study design and conduct. The Australian Multicentre Study of Environment  
555 and Immune Function," *Multiple Sclerosis* **13**, 827–839 (2007).
- 556 42. J. Birch, "Efficiency of the Ishihara test for identifying red-green colour deficiency,"  
557 *Ophthalmic and Physiological Optics* **17**, 403–408 (1997).
- 558 43. L. D. Brown, T. T. Cai, and A. DasGupta, "Interval Estimation for a Binomial  
559 Proportion," *Statistical Science* **16**, 101–133 (2001).
- 560 44. J. L. Hintze and R. D. Nelson, "Violin Plots: A Box Plot-Density Trace Synergism  
561 Statistical Computing and Graphics Violin Plots: A Box Plot-Density Trace  
562 Synergism," *Source: The American Statistician* **52**, 181–184 (1998).
- 563 45. S. Seifirad and L. Alquran, "The bigger, the better? When multicenter clinical trials  
564 and meta-analyses do not work," *Curr Med Res Opin* **37**, 321–326 (2021).
- 565 46. S. Arnegard, R. C. Baraas, J. Neitz, L. A. Hagen, and M. Neitz, "Limitation of  
566 standard pseudoisochromatic plates in identifying colour vision deficiencies when  
567 compared with genetic testing," *Acta Ophthalmol* 1–8 (2022).
- 568 47. G. H. M. Waaler, "Über die Erblichkeitsverhältnisse der verschiedenen Arten von  
569 angeborener Rotgrünblindheit," *Z Indukt Abstamm Vererbungsl* **45**, 279–333 (1927).
- 570 48. A. Cheng, D. Kessler, R. Mackinnon, T. P. Chang, V. M. Nadkarni, E. A. Hunt, J.  
571 Duval-Arnould, Y. Lin, M. Pusic, and M. Auerbach, "Conducting multicenter research  
572 in healthcare simulation: Lessons learned from the INSPIRE network," *Advances in*  
573 *Simulation* **2**, 6 (2017).
- 574 49. M. K. Das, "Multicenter Studies: Relevance, Design and Implementation," *Indian*  
575 *Pediatr* **59**, 571–579 (2022).
- 576 50. B. E. W. Evans, M. Rodriguez-Carmona, and J. L. Barbur, "Color vision  
577 assessment-1: Visual signals that affect the results of the Farnsworth D-15 test,"  
578 *Color Res Appl* **46**, 7–20 (2021).

- 579 51. W. D. Wright, *Researches on Normal & Defective Colour Vision* (Kimpton, 1946).
- 580 52. L. M. Hurvich, "Color vision deficiencies," in *Visual Psychophysics* (Berlin: Springer-  
581 Verlag, 1972), Vol. 704, pp. 582–624.
- 582