Retinal Implantation of Electronic Vision Prostheses to Treat Retinitis Pigmentosa: A Systematic Review

Luke E. Hallum¹ and Steven C. Dakin²

¹ Department of Mechanical Engineering, University of Auckland, Auckland, New Zealand
² School of Optometry and Vision Science, University of Auckland, Auckland, New Zealand

Correspondence: Luke E. Hallum, Department of Mechanical Engineering, 5 Grafton Road, Auckland Central, Auckland 1010, New Zealand. e-mail: l.hallum@auckland.ac.nz

Received: January 15, 2020
Accepted: May 7, 2021
Published: August 12, 2021

Purpose: Retinitis pigmentosa (RP) is a hereditary disease causing photoreceptor degeneration and permanent vision loss. Retinal implantation of a stimulating electrode array is a new treatment for RP, but quantification of its efficacy is the subject of ongoing work. This review evaluates vision-related outcomes resulting from retinal implantation in participants with RP.

Methods: We searched MEDLINE and Embase for journal articles published since January 1, 2015. We selected articles describing studies of implanted participants that reported the postimplantation measurement of vision. We extracted study information including design, participants’ residual vision, comparators, and assessed outcomes. To assess the risk of bias, we used signaling questions and a target trial.

Results: Our search returned 425 abstracts. We reviewed the full text of 34 articles. We judged all studies to be at high risk of bias owing to the study design or experimental conduct. Regarding design, studies lacked the measures that typical clinical trials take to protect against bias (e.g., control groups and masking). Regarding experimental conduct, outcome measures were rarely comparable before and after implantation, and psychophysical methods were prone to bias (subjective, not forced choice, methods). The most common comparison found was between postimplantation visual function with the device powered off versus on. This comparison is at high risk of bias.

Conclusions: There is a need for high-quality evidence of efficacy of retinal implantation to treat RP.

Translational Relevance: For patients and clinicians to make informed choices about RP treatment, visual function restored by retinal implantation must be properly quantified and reported.

Introductions

Retinitis pigmentosa (RP) is a group of hereditary diseases causing photoreceptor degeneration. Vision loss owing to RP is progressive, and can result in total and permanent blindness. RP is a leading cause of blindness, afflicting 1 in 4000 people. In RP, the inner layers of the retina—comprising ganglion, bipolar, and amacrine cells—seem to be less vulnerable to degeneration than the outer layers, which comprise rod and cone photoreceptors. Because RP predominantly affects the outer layers of the retina, it has been proposed that retinal implants—which attempt to replace the function of photoreceptors by electrically stimulating the retina’s more viable inner layers—may be an effective treatment for RP (reviewed by Weiland et al.⁶). These vision prostheses typically comprise an electrode array that is implanted surgically either epiretinally (at the vitreoretinal interface), subretinally (between the photoreceptor layer and choroid), or suprachoroidally. That array is then coupled to an imaging light sensor device, such as a head-mounted camera, through a wired or wireless interface.

A recent review by Humayun et al.⁵⁵ describes the development of retinal implants and, concomitant with this, the development of new methods for
vision assessment in implantees. These new methods began with the intraoperative detection of single phosphenes, and, as we discuss elsewhere in this article, have recently expanded to include more conventional psychophysical tests (e.g., localization, spatial vision, and motion perception). As discussed by Humayun et al., vision tests in implantees have generally not demonstrated measurable outcomes using conventional clinical approaches (e.g., letter charts). Therefore, the field’s pioneers were required to develop novel methods to attempt to measure clinical outcomes, running alongside technological developments, for the specific purpose of determining what a retinal prosthesis could do for a patient. The absence of measurable outcomes using conventional clinical tests need not imply that implantation confers no benefit. To illustrate, patients with ultra-low vision, which is usually qualified (not quantified) in clinical terms such as “hand motion,” “light projection,” or “light perception,” do not exhibit measurable outcomes on conventional clinical tests, but nonetheless there are activities of daily living that seem to benefit from ultra-low vision (discussed by Humayun et al.). These considerations are important when evaluating vision-related outcomes of retinal implantation and ought to be considered alongside our analysis, presented in this article, of the studies we discovered through systematic review. We acknowledge the methodologic difficulty of testing a retinal prosthesis. At this stage of device development, our robust analysis is important for several reasons: it helps patients and clinicians to make informed treatment decisions; it helps to guide device design; and it improves the specificity of retinal implant candidacy criteria.

Four retinal implant devices have been approved for use in either North America or the European Economic Area, and there exist several preclinical experimental devices. Although these devices have received considerable attention in the media, quantification of their efficacy is the subject of ongoing work. For patients, clinicians, and funders to make informed decisions about experimental treatments for RP, it is vital that the efficacy of devices is properly quantified and reported. For patients, the primary outcome of interest will be the level of visual function that is restored after implantation. We, therefore, conducted a systematic review to evaluate vision-related outcomes resulting from retinal implantation of any device in participants with outer retinal degeneration. At the outset, we were primarily interested in the comparison of quantitative measures of vision made before and after implantation.

Methods

Data Sources and Study Selection

We searched MEDLINE and Embase for titles or abstracts containing any of our search strings. Search strings were all pairwise combinations of terms drawn from two sets. The first term was drawn from the following set: retina, retinal, epiretina, epiretinal, subretina, subretinal, transretina, transretinal. The second term was drawn from the following set: implant, implantation, prosthesis, prosthetic. An example search string is retinal implantation. We limited the search to journal articles published between January 1, 2015 and October 11, 2019 (date of last search: October 14, 2019). We screened abstracts resulting from the search to determine articles of primary interest. We then screened the full-text articles of these publications.

Study Eligibility Criteria

We included studies of participants who (a) had undergone retinal implantation to treat outer retina degeneration and (b) had received a postimplantation assessment of vision. Assessments included both “functional vision,” as well as “visual function”; the former required participants to perform real-world visual tasks (e.g., to recognize household objects), and the latter required participants to perform psychophysical tasks involving constrained visual stimuli (e.g., to detect sine-wave gratings). We excluded phosphene studies (e.g., a study by Shivdasani et al.) that, as opposed to presenting visual stimuli to implantees, electrically stimulated participants’ retinas with an acutely or chronically implanted electrode or electrode array. We included studies of any design, involving any device, including devices approved by the US Food and Drug Administration or an equivalent regulatory body, as well as experimental studies of devices yet to be approved.

Data Extraction and Quality Assessment

We extracted data from each study using a custom data extraction form (Appendix A) modelled on a recent systematic review and meta-analysis of cochlear implantation by Gaylor et al. To guide our assessment of risk of bias, we used the signaling questions of Gaylor et al.; in general, signaling questions are used to extract information about features of a study that are relevant to its risk of bias. After our initial
scoping of the review, we decided to add the following signaling questions:

- For interventional studies, were the outcome assessment methods comparable before and after implantation?
- Were the outcome assessment methods comparable with the device powered off and on?
- Do the psychophysical methods seem to be sound (e.g., were forced-choice methods used where appropriate)?

To guide our assessment of risk of bias, we also referred to the Cochrane Handbook of Systematic Reviews of Interventions, specifically chapters 8 and 25. To guide our reporting, we used methods described in the PRISMA statement.

Table 1. Retinal Prostheses

<table>
<thead>
<tr>
<th>Country</th>
<th>Site of SE</th>
<th>No. of SE</th>
<th>Example Studya</th>
<th>Known as</th>
<th>Light Sensor</th>
<th>Subtense of SEb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Suprachoroidal</td>
<td>20</td>
<td>Barnes et al., 201617</td>
<td>n.d.</td>
<td>Head-mounted camera</td>
<td>11.7° × 11.5°</td>
</tr>
<tr>
<td>France</td>
<td>Epiretinal</td>
<td>150</td>
<td>Muqit et al., 201919</td>
<td>IRIS II</td>
<td>Head-mounted camera</td>
<td>n.d.</td>
</tr>
<tr>
<td>Germany</td>
<td>Subretinal</td>
<td>1500</td>
<td>Stingl et al., 201510</td>
<td>Alpha IMS</td>
<td>Subretinal PDA</td>
<td>10.0° × 10.0°</td>
</tr>
<tr>
<td>Germany</td>
<td>Subretinal</td>
<td>1600</td>
<td>Stingl et al., 201731</td>
<td>Alpha AMS</td>
<td>Subretinal PDA</td>
<td>10.7° × 13.3°</td>
</tr>
<tr>
<td>Japan</td>
<td>Suprachoroidal</td>
<td>49</td>
<td>Fujikado et al., 20169</td>
<td>n.d.</td>
<td>Head-mounted camera</td>
<td>14.0° × 14.0°</td>
</tr>
<tr>
<td>US</td>
<td>Epiretinal</td>
<td>16</td>
<td>Yue et al., 201533</td>
<td>Argus I</td>
<td>Head-mounted camera</td>
<td>8.0° × 8.0°</td>
</tr>
<tr>
<td>US</td>
<td>Epiretinal</td>
<td>60</td>
<td>Ho et al., 201515</td>
<td>Argus II</td>
<td>Head-mounted camera</td>
<td>9.6° × 17.25°</td>
</tr>
</tbody>
</table>

Abbreviations: n.d., no data; PDA, photodiode array; SE, stimulating electrodes

dExample studies are identified by first author and year.

bWhere necessary, we used the conversion 1° visual subtense = 0.3 mm retinal distance. Additional sources: Stingl et al. 201350; Yanai et al., 200751; Luo and da Cruz, 2016.52
cElectrode array dimensions not reported in Muqit et al., 2019,39 the only IRIS II study discovered by our review.
Table 2. Localization-related Outcomes of Retinal Implantation

<table>
<thead>
<tr>
<th>Study and Country</th>
<th>Study Design</th>
<th>Follow-up (mo)</th>
<th>Participants</th>
<th>Mean Age (y)</th>
<th>Residual Vision</th>
<th>Comparator</th>
<th>Outcome Assessment Time (mo)</th>
<th>Reported Change in Localization (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ho 2015 multiple Clinical trial</td>
<td>36</td>
<td>30</td>
<td>58</td>
<td>BLP (n = 29) or no LP (n = 1)</td>
<td>Device off vs on</td>
<td>36</td>
<td>Improvement in 25 of 28 participants (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Luo 2015 UK Case series</td>
<td>0</td>
<td>5</td>
<td>60</td>
<td>BLP, acuity &gt; 2.9 logMAR</td>
<td>Device off vs on</td>
<td>0</td>
<td>Cohort improvement (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Stingl 2015 multiple Clinical trial</td>
<td>12</td>
<td>29</td>
<td>54</td>
<td>LPWP (n = 20) or no LP (n = 9)</td>
<td>Device off vs on</td>
<td>12</td>
<td>Improvement in 1 of 7 participants (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Yue 2015 multiple Case report</td>
<td>120</td>
<td>1</td>
<td>55</td>
<td>no LP</td>
<td>Device off vs on</td>
<td>120</td>
<td>Improvement (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Barnes 2016 Australia Clinical trial</td>
<td>10</td>
<td>3</td>
<td>55</td>
<td>BLP</td>
<td>Device off vs on</td>
<td>o.a.</td>
<td>Improvement in all participants (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Barry 2016 US Cohort study</td>
<td>40</td>
<td>3</td>
<td>n.d.</td>
<td>n.d.</td>
<td>Auditory feedback vs no feedback</td>
<td>6</td>
<td>Improvement in all participants (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Castaldi 2016 Italy Clinical trial</td>
<td>24</td>
<td>7</td>
<td>60</td>
<td>BLP, acuity &gt; 2.9 logMAR</td>
<td>Preimplantation vs.</td>
<td>60</td>
<td>Improvement in 17 of 21 participants (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>da Cruz 2016 multiple Clinical trial</td>
<td>60</td>
<td>30</td>
<td>58</td>
<td>BLP (n = 29) or no LP (n = 1)</td>
<td>Device off vs on</td>
<td>60</td>
<td>Improvement in 1 of 3 participants (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Fujikado 2016 Japan Clinical trial</td>
<td>12</td>
<td>3</td>
<td>56</td>
<td>Hand motion</td>
<td>Device off vs on</td>
<td>12</td>
<td>Improvement in 1 of 3 participants (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Hafed 2016 Germany Cohort study</td>
<td>0</td>
<td>2</td>
<td>n.d.</td>
<td>LP</td>
<td>NA</td>
<td>0</td>
<td>n.d.</td>
<td></td>
</tr>
<tr>
<td>Caspi 2017 US Cohort study</td>
<td>0</td>
<td>8</td>
<td>65</td>
<td>n.d.</td>
<td>Head-only scanning vs eye-head</td>
<td>0</td>
<td>Improvement in 6 of 8 participants (&lt;0.016)</td>
<td></td>
</tr>
<tr>
<td>Parmeggiani 2017 Italy Case series</td>
<td>12</td>
<td>3</td>
<td>48</td>
<td>BLP, acuity &gt; 2.9 logMAR</td>
<td>Preimplantation vs.</td>
<td>12</td>
<td>Improvement in all 3 participants (n.d.)</td>
<td></td>
</tr>
<tr>
<td>Petoe 2017 Australia Clinical trial</td>
<td>10</td>
<td>3</td>
<td>55</td>
<td>BLP</td>
<td>Device scrambled vs. on</td>
<td>o.a.</td>
<td>Improvement in 1 of 3 participants (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Stingl 2017 multiple Clinical trial</td>
<td>12</td>
<td>15</td>
<td>55</td>
<td>LPWP (n = 14) or no LP (n = 1)</td>
<td>Device off vs on</td>
<td>12</td>
<td>Improvement in 7 of 8 participants (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Edwards 2018 UK Clinical trial</td>
<td>12</td>
<td>6</td>
<td>53</td>
<td>Nonuseful or no LP</td>
<td>Device off vs on</td>
<td>o.a.</td>
<td>Improvement in 5 of 6 participants (n.d.)</td>
<td></td>
</tr>
<tr>
<td>Endo 2018 Japan Clinical trial</td>
<td>12</td>
<td>1</td>
<td>42</td>
<td>Hand motion (LE) LP (RE)</td>
<td>Preimplantation vs.</td>
<td>12</td>
<td>Improvement (n.d.)</td>
<td></td>
</tr>
<tr>
<td>He 2019 US Cohort study</td>
<td>0</td>
<td>5</td>
<td>63</td>
<td>BLP</td>
<td>Zoom vs no zoom</td>
<td>0</td>
<td>Cohort improvement (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Hsu 2019 Taiwan Case report</td>
<td>9</td>
<td>1</td>
<td>54</td>
<td>LP</td>
<td>Device off vs on</td>
<td>9</td>
<td>Improvement (n.d.)</td>
<td></td>
</tr>
<tr>
<td>Muqit 2019 multiple Clinical trial</td>
<td>6</td>
<td>10</td>
<td>59</td>
<td>acuity &gt; 2.3 logMAR</td>
<td>Device off vs on</td>
<td>6</td>
<td>Cohort improvement (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Rizzo 2019 Italy Case series</td>
<td>24</td>
<td>33</td>
<td>57</td>
<td>n.d.</td>
<td>Postimplantation</td>
<td>24</td>
<td>Improvement (n.d.)</td>
<td></td>
</tr>
<tr>
<td>Schafffrath 2019 multiple Clinical trial</td>
<td>12</td>
<td>47</td>
<td>56</td>
<td>LP</td>
<td>Device off vs on</td>
<td>12</td>
<td>Improvement in 16 of 35 participants (&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Takchidi 2019 Russia Clinical trial</td>
<td>6</td>
<td>2</td>
<td>57</td>
<td>n.d.</td>
<td>Device off vs on</td>
<td>6</td>
<td>Improvement in both participants (n.d.)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: BLP, bare-light perception; HM, hand motion; LE, left eye; logMAR, logarithm of the minimum angle of resolution; LP, light perception; LPWP, light perception without projection; NA, not applicable; n.d., no data; n.s., not statistically significant; o.a., overall; RE, right eye.

Table notes:

aStudies are identified by first author and year.
bWhere possible, we have adopted the reported design labels (e.g., “case series”).
cIf fewer participants underwent vision testing, the number of participants tested is specified in column “Reported change in localization.”
dComparisons are made within-participant unless otherwise stated.
eThe outcome is localization accuracy, unless stated otherwise. Where assessments at multiple times were available, we took the most recent, or an average over all (marked “o.a.”).
fCohort improvement for prehension, but not localization.
gParticipant 1 and participant 3 show auditory feedback improves localization accuracy; participant 1 and participant 2 show deterioration of accuracy in the absence of feedback.
hNine-dot fixation. See running text.
iTwo of 8 participants took the device-on but not device-off test.
jLocalization precision.
kRegarding visual outcomes, this study was primarily concerned with the effect of fibrosis and schisis. Data from only 2 participants are reported in detail, both of whom show “device off vs. on” improvement; preimplantation data are shown for 1 participant. The mean age is for the cohort completing study (n = 20), not cohort completing outcome assessment (n = 10).
forced-choice test using a high-contrast wedge stimulus appearing at one of four locations around a central fixation point. Stimuli were typically large, for example, subtending 20° of visual angle, and were presented for long durations (e.g., the study by Barnes et al.17). The second test type involved patients indicating the location of a large (diameter approximately equal to 10°) high-contrast square or circle appearing at a random location on a touch screen.18 Stimuli persisted on the screen until the participant’s touch landed (e.g., see da Cruz et al.19). Studies by Luo et al.20 and He et al.21 used a variation on the second test type by requiring participants to localize and reach to a high-contrast physical object (plastic blocks measuring $3.1 \times 3.1 \times 5.7$ cm or drinking cups). Rather than using prehension or pointing, Hafed et al.22 measured the eye movements of two implanted participants. Although their depiction of one participant’s pattern of fixation on a nine-dot calibration pattern (dot diameter = 2.4°) provides a qualitative assessment of localization accuracy, this accuracy was not quantified.

We discovered only three studies reporting localization accuracy both before and after implantation.23,24,25 Accuracy refers to a participant’s ability to point to a target in an unbiased fashion (i.e., point- ing may be variable, but, on average, pointing is to the correct location, and not, e.g., slightly to the left). This measure is in contrast with measures of precision, which quantify the reliability of a participant’s pointing (e.g., a participant always points slightly to the left, but this behavior is consistent). These three studies,23,24,25 following participants for up to 24 months, all reported preimplantation versus postimplantation vision improvement for all participants (Table 2). Most discovered studies compared localization accuracy when the device was powered off and compared that with accuracy when the device was on. Many of these studies were interventional (i.e., clinical trials) and followed participants (in one case, for 10 years), but others were observational (i.e., cohort studies). Most of these studies reported off versus on vision improvements in a majority of participants (Table 2). One clinical trial compared localization accuracy with the device on with accuracy after scrambling the phosphene map;10; normal versus scrambled vision improvement was seen in only one (of three) participants. One cohort study compared localization accuracy with and without auditory feedback, for all three participants, the presence of feedback either increased accuracy or its absence caused accuracy to decrease. Another cohort study integrated both eye and head movements into phosphene activation; localization precision was increased in six of eight participants when comparing eye-and-head with head-only scanning.

Several studies of localization accuracy requiring participants to point to targets (e.g., Hsu et al.28) did not report whether feedback was provided. This omission complicates the interpretation of results; without feedback, low accuracy may reflect participant bias as opposed to inaccurate light localization (as is the case with all psychophysical estimation tasks). Participants can use feedback to adjust for any inherent bias.26 Indeed, studies by Barry and Dagnelie26 and Caspi et al.27 are the only ones that explicitly addressed this complication; the latter (which reported precision, not accuracy) notes that pointing accuracy “should be a function of the position of electrodes on the retina, inherent point bias, and any differences in eye and [head-mounted] camera positions not taken into account.” It is unclear how to interpret many of the reported localization accuracies; if feedback was absent, does inaccuracy in performance indicate inherent bias or inaccurate localization?

**Spatial Vision and Motion Perception**

Our review discovered 10 studies that assessed implanted participants’ vision using sine- and square-wave grating stimuli.7,8,10,17,19,23,29,30,31,32 This outcome was typically assessed using a two- or four-alternative forced-choice test, presenting a grating to the participant who was required to report its orientation. In all cases, stimulus duration was long, enabling scanning of stimuli with the head-mounted camera or implanted photodiode array: Ho et al.29 reported a grating stimulus duration of 5 seconds, presumably the same duration used by da Cruz et al.19; Stingl, Edwards, and colleagues7,30,31 all appear to have used “no strict time limits... participants were encouraged to provide prompt answers”; Barnes, Petoe, and colleagues10,17 allowed participants to scan indefinitely before responding; and, Castaldi and colleagues23 used grating stimuli presented for a fixed duration of 1s. All studies measured each participant’s performance with the implanted device powered off, and compared that to device-on performance. However, none of the clinical trials compared grating acuity before implantation to grating acuity after implantation. Our review discovered reasonable consistency of experimental design across these studies (e.g., clinical trials with outcome assessment at multiple follow-ups, and the use of forced-choice methods). However the outcomes reported varied widely (e.g., one study reported spatial frequency discrimination threshold in units of cycles per degree, while another reported the percentage of their cohort with grating acuity better than
2.9 logMAR). All grating studies reported measurable grating acuity in implantees. Studies by Ho and colleagues\textsuperscript{29} and da Cruz and colleagues\textsuperscript{19} reported nine of 27 (33%; three-year follow-up) and eight of 21 (38%; five-year follow-up) participants, respectively, with grating acuity better than 2.9 logMAR. Petoe and colleagues\textsuperscript{10} (see also Barnes and colleagues\textsuperscript{17}) reported grating acuity of 0.124 cycles per degree in one participant (in this participant, device-off discrimination threshold was unmeasurably low, while device-on thresholds were unmeasurably low in the cohort’s remaining two participants); Stingl, Edwards, and colleagues reported grating acuities up to 3.3 cycles per degree \textsuperscript{7,30,31}; Castaldi and colleagues\textsuperscript{23} reported cohort performance on a grating detection task at rates well above chance; and Takakhidi and colleagues\textsuperscript{32} reported grating acuity less than 2.9 logMAR in both implanted study participants.

Taken together, the outcomes of the above-mentioned grating studies are encouraging. However, the interpretation of these results is not straightforward. It is difficult to separate the effect on performance of the following two factors: first, spatial vision, that is, perception arising from retinal patterns of activation; and, second, the combined effects of light perception and scanning. Indeed, when gratings are presented for long durations, participants are able to scan the stimulus, and a scanning participant can use integrated light perception to discriminate gratings of different orientations. We illustrate this concept in Figure 2. We are not suggesting that light perception and scanning exclusively account for the outcomes of the above-mentioned grating studies. Rather, we emphasize that these outcomes conflate spatial vision and light perception to an unknown extent. In other words, if implanted participants are allowed to scan grating stimuli, then caution must be exercised when using the outcomes of those experiments to draw conclusions about the restoration of spatial vision in the absence of any scanning. While the stimulus duration reported by Castaldi and colleagues\textsuperscript{23} was relatively short (1 second), their detection task involved “stationary, low spatial frequency” gratings which participants may have learned to scan. Grating studies by Stingl, Edwards, and colleagues\textsuperscript{7,30,31} involved only a small number of trials and/or reported each participant’s best achieved grating acuity,\textsuperscript{30,31} leaving the results vulnerable to false-positive findings.

The image processing implemented by a retinal prosthesis limits the efficacy of scanning; gratings beyond the passband of a regional averaging filter (e.g., see simulations by Chen, Hallum, and colleagues\textsuperscript{57,58}), or a similar low-pass filter, are greatly attenuated, and therefore no amount of scanning will enable the participant to discriminate these relatively high spatial frequency gratings. However, scanning relatively low spatial frequency gratings (i.e., within the passband) is an effective strategy. The participant tested by Barnes et al.\textsuperscript{17} scanned relatively low spatial frequency gratings for long periods (on average, 44.5 seconds per trial) to achieve grating acuities of between 0.09 and 0.124 cyc/deg, that is, acuities within the passband of their device. In Figure 2, we illustrate the effect of regional averaging, or a similar low-pass filter. As the averaged region increases in size (dashed circle superimposed on the stimulus in Figure 2), the modulation of the photodiode output is decreased. In other words, the efficacy of scanning is decreased as the averaged region increases. We also note that there is a potentially useful comparison to be made between the functional role of involuntary fixational eye movements and voluntary scanning, which contribute to normal and prosthetic vision, respectively.

Our review discovered four studies that used either drifting gratings or dot stimuli to test participants’
motion perception. These tests are relatively strong for motion perception. Participants performed no better than chance in two of these studies. Stingl et al. reported that a minority of participants passed a dot motion test at the most recent follow-up; however, these experiments were underpowered. Several studies of participants’ motion perception used a drifting, high-contrast bar as the stimulus, requiring participants to estimate its direction of motion. Although the outcomes of these moving bar studies are encouraging, the interpretation of results is not straightforward; analogous to the grating studies described elsewhere in this article, in moving bar studies, it is difficult to separate the effect on performance of two factors: motion perception and light perception. To illustrate, consider an implanted participant fixating the right side of the stimulus monitor; this participant could use the time of arrival, there, of light to estimate direction of motion at rates above chance. We illustrate this concept in more detail in Figure 3. We are not suggesting that light perception exclusively accounts for the outcomes of these moving bar studies. Rather, we emphasize that these outcomes conflate motion perception and light perception to an unknown extent. We note that participants with implants may have difficulty stabilizing fixation at a point on the stimulus monitor (in moving bar studies, participants were “instructed to maintain eye and camera [head] fixation on the center of a 19-inch touchscreen monitor”); this instability would add uncertainty to timing information, rendering it less useful in the performance of the task. Uncertainty could also be added to timing information by introducing an interval, randomized across trials, between the beginning of a trial and the onset of bar motion. However, none of the studies we discovered reported the use of such an interval. We also note that, if a participant with an implant adopted a strategy of fixating a peripheral location on the stimulus monitor (as we illustrate in Fig. 3A), then to do so would be to go against the experimental instructions. If a participant fixated the center of the stimulus monitor (as instructed), then timing information would not be useful in performing the task.

### Other Measures of Vision

Our review discovered that a wide range of other tests of vision are being used to assess the outcome of retinal implantation (several of which were questionnaire-based). Many of these other tests attempted to quantify real-world vision and included the following:

- Landolt C optotype identification
- “Door task”: patients attempted to walk to and touch a “simulated door,” that is, a large piece of contrasting felt on a wall
- “Line task”: “patients followed a white line painted on black tiles”
- Navigation in a controlled environment
- Sock sorting
- Walking direction discrimination task
- Tableware recognition
Tasks modelled on activities of daily living, such as those listed, are potentially useful in understanding the practical efficacy of implantation. To illustrate, here we describe some outcomes relating to navigation (door and line tasks, navigation in a controlled environment, and walking direction discrimination)\(^\text{9,19,29,33,37,38}\) and object recognition (tableware recognition, clockface reading, letter acuity).\(^\text{7,9,17,30,31}\) da Cruz et al.\(^\text{19}\) reported improved performance on their door and line tasks at all postimplantation follow-ups to 5 years. Their reported door task success rate at the 5-year follow-up, averaged across a cohort of 20 participants, was approximately 50\% (device on) versus approximately 20\% (device off). Their reported line-task success rate (5-year follow-up, \(n = 20\)) was approximately 70\% (device on) versus approximately 20\% (device off). However, line following outcomes reported by Fujikado et al.\(^\text{9}\) were mixed. Two of their three participants showed improved performance (device off versus on), that is, less deviation from the walked line, at two of six and three of six follow-up sessions, respectively. However, in their third participant, the device's power status had no effect on performance at any of the four follow-ups (all participants were followed for 12 months). Garcia et al.\(^\text{37}\) reported the performance of four implanted participants navigating a controlled environment, concluding that the implanted device did not provide sufficiently reliable visual information to improve navigation. On the first of two tasks (“path reproduction”), the cohort with implants (\(n = 4\)) showed no improvement when navigating with the device on as compared with device-off navigation. However, on the second task (“path completion”), two of the implanted participants showed device-on versus off improvement comparable in magnitude with that of a control group (control participants were normally sighted, but wore goggles restricting their field of view, and decreasing acuity to 1.6 logarithm of the minimum angle of resolution [logMAR]). Dagnelie et al.\(^\text{38}\) reported the performance of 27 implanted participants navigating a controlled environment, or discriminating the walking direction of another person. On the former task (“sidewalk tracking”), the performance of approximately two-thirds of participants improved (fewer “out of bounds” deviations from the sidewalk) with the device on versus off; conversely, approximately one-third of participants performed better with the device off versus on. Overall, that cohort’s performance improved with the device on. On the latter “walking direction” two-alternative forced-choice task, the performance of 18 participants was significantly above chance (50\% correct) with the device powered on. Only six participants performed significantly above chance with the device powered off.

Several of the studies we discovered reported outcomes concerning the recognition of objects.\(^\text{7,9,17,30,31}\) Fujikado et al.\(^\text{9}\) tested participants’ ability to discriminate a bowl from chopsticks. Two of their three participants showed improved performance (device off versus on) at one of seven and five of six follow-up sessions, respectively. In their third participant, the device’s power status (on versus off) had no effect on performance at any of the five follow-ups. In a similar task, Edwards et al.\(^\text{7}\) reported that their five implanted participants were able to locate household objects (knife, fork, spoon, and plate) on a tabletop when the device was powered on, but not when the device was off. Two participants were able to name, as well as locate, objects in follow-up sessions 3 months after implantation. Naming was not possible with the device powered off. Similar outcomes were reported by the same group in a larger cohort.\(^\text{31}\) Barnes et al.\(^\text{17}\) reported Landolt-C acuity in one implanted participant (of a cohort of three). They found a performance improvement (device off versus on) of 3.24 versus 2.65 logMAR. Stingl et al.\(^\text{31}\) also reported mixed acuity outcomes; they were able to measure Landolt-C acuities of 20/546 and 20/1111 (1.44 and 1.74 logMAR, respectively) in two of their 13 participants with implants.

Taken together, the outcomes of these studies involving tasks modelled on activities of daily living, are encouraging. However, a key question remains: To what extent did participants perform these tasks using scanning and light perception and to what extent did they use spatial vision or motion perception? Data are consistent with the interpretation that the implant may afford participants improved light perception, and this improvement in turn affords improved performance on a diverse range of visual outcomes.

**Discussion**

The studies discovered by our review indicate that retinal implantation of an electrode array to treat RP may be effective in restoring some degree of light perception. However, outcomes relating to spatial vision and motion perception, although encouraging,
were open to more than one interpretation. In most discovered studies, neither participants nor assessors were masked to experimental conditions, and control groups were not used as comparators. In clinical trials, outcome assessment was rarely comparable before and after implantation. For these reasons, all discovered studies were judged to be at high risk of bias. Our findings signal an urgent need for the development of objective protocols for vision assessment to be used at multiple sessions both before and after implantation.

We have applied a robust analysis to studies discovered by our review. We acknowledge that a high risk of bias is a likely finding, given that we reviewed several feasibility studies of devices in development, and we do not mean for our analysis to mitigate the interdisciplinary achievement represented by these human trials. However, at this early stage of device development, our robust analysis—specifically, identifying risk of bias and scope for improved study design—is important for two reasons. First, for patients and clinicians to make informed choices about RP treatment, vision restored by retinal implantation must be quantified and reported properly. Second, proper quantification and reporting could be used to guide device design, and to improve the specificity of retinal implant candidacy criteria. Improved design and candidacy criteria will help to speed the development of proven, effective devices.

We focused our analysis of studies using two key questions. First, is there evidence of retinal implants restoring light perception? This question is important; conclusive evidence that participants with no light perception before treatment can perceive light after treatment would help to justify the implantation of a stimulating electrode. In patients with low vision, light perception is typically assessed in the clinic via qualitative methods. In experimental participants, light perception is typically assessed using quantitative, forced-choice methods and multiple trials; an observer either discriminates which of two temporal intervals contains a light increment, or which of two or four spatial locations contains a light increment. Second, is there evidence of retinal implants restoring spatial vision or motion perception? This question is centrally important; an affirmative answer would largely justify the implantation of multi-electrode arrays, as opposed to a single electrode. In a typical grating acuity assessment, an observer discriminates briefly presented spatial patterns defined by sine or square wave variations of luminance (gratings) that are different in spatial frequency and/or orientation. Gratings are, typically, briefly presented (\(<200\) ms) to minimize the effect of eye movements on performance. Because the mean luminance of gratings does not vary as a function of spatial frequency and/or orientation, grating acuity assessment measures spatial vision. Motion perception is a rigorous assessment of retinal implant efficacy because it requires both a spatial and a temporal analysis of visual information; implant recipients able to discriminate grating stimuli, or random dot stimuli, that drift in different directions but are otherwise identical, would also provide evidence to justify the implantation of multi-electrode arrays. By analogy, cochlear implant scientists, engineers, and clinicians have devoted much effort to demonstrating the clinical benefit of multichannel devices as compared with single-channel devices (e.g., Tyler et al.*). Retinal implantation of multi-electrode arrays has yet to meet the same standards of quality of evidence. We note that the restoration of some spatial vision and motion perception is not the only justification for implanting a multi-electrode array; because the effects of RP can vary across the retina, implanting an array, as opposed to a single electrode, potentially increases the chance of contacting retina suitable for stimulation and, in turn, restoring some light perception.

Most of the studies discovered by our review measured a participant’s vision with the device powered off and compared that to performance measured with the device powered on. This comparison is problematic for several reasons. First, it does not assess the efficacy of retinal implantation as compared with no treatment, although it invites this interpretation. Indeed, retinal implantation of an electrode array may have a deleterious effect on residual vision. Second, the comparison is vulnerable to bias because in none of the studies we discovered were participants or outcome assessors masked to the power status of the device (power off versus on). A lack of masking is, for example, common in surgical trials and can bias outcomes:

If participants are not blinded, knowledge of group assignment may affect their behaviour in the trial and their responses to subjective outcome measures. For example, a participant who is aware that he is not receiving active treatment may be less likely to comply with the trial protocol. Those aware that they are receiving or not receiving therapy are more likely to provide biased assessments of the effectiveness of the intervention—most likely in opposite directions—than blinded participants. Similarly, blinded clinicians are much less likely to transfer their attitudes to participants or to provide differential treatment to the active and placebo groups than are unblinded clinicians.

In the context of retinal implantation, the lack of masking of participants to experimental variables...
may alter their behavior in a direction that favors the device when powered on. The lack of masking of outcome assessors renders them more likely to provide differential interaction with participants contingent upon the power status of the device in a fashion that may favor the device when powered on. Unmasked studies wherein the outcomes assessed are subjective are especially vulnerable to bias. All vision measurements are subjective; however, forced-choice methods are criterion free and, it stands to reason, less vulnerable to bias. We concede that masking participants is methodologically difficult where stimulation elicits clear, suprathreshold sensations. Nonetheless, participant masking deserves greater consideration; we discovered only one study (by Petoe et al.) that discussed the potential consequences of a lack of masking on outcomes. That study reported altered behavior in one of the three participants when the device was powered off. Indeed, those researchers designed additional experiments to attempt to control for “a perceived lack of participant engagement in tasks when using system off. We observed this situation in [participant] P2, corresponding to a significantly eccentric starting position and minimal head movement (i.e., small response offset), in spite of our instruction to attempt the task with residual vision.”

Our review discovered only three clinical trials that reported comparable vision assessments before and after implantation. All other clinical trials described preimplantation vision in only the qualitative terms of clinical assessment, not directly comparable to postimplantation measures (e.g., “In all 6 participants, this [preimplantation vision] was recorded as vague nonlocalizing perception of light in the eye to be implanted”). This method, too, is problematic; without a comparable outcome assessment before and after implantation, it is not possible to separate the conflated effects of the device and of residual vision on performance. Several studies implied that participant performance with the device powered off provides an estimate of residual vision. However, outcome assessment with the device off is potentially biased as discussed elsewhere in this article, and therefore a poor estimate of residual vision. Furthermore, retinal implantation may have an overall deleterious effect on participants’ vision; the device off versus on comparison is insensitive to any deleterious effects. The best way to quantify the effect of residual vision on an implanted participant’s performance is to make comparable, quantitative measurements before and after implantation. Preimplantation measurements would facilitate an answer to the central question: Is retinal implantation effective as compared with no treatment? At present, this question is open.

The placebo effect—benefit resulting solely from the administration of treatment—may arise owing to a variety of factors, including participant expectation of an effect. The placebo effect is an important consideration in drug, surgical, and device trials alike, especially where subjective outcomes are assessed. None of the clinical trials discovered by our review attempted to control for the placebo effect. Two studies, both of which made questionnaire-based vision assessments before and after implantation, discussed the placebo effect as a potential confound. Both studies argued that, because their participants’ improved vision persisted throughout long follow-up periods, their results were “most likely not the result of placebo, which is known to deteriorate over time.”

In the reviewed studies, visual performance was variable, whether compared across participants within a single study of a particular device or across participants drawn from studies of different devices. An understanding of the sources of this variability could be used to refine the design of devices and the criteria used to select candidates for implantation. At present, there are too few published data, and the outcomes assessed are too varied, to attempt to quantify factors predictive of visual outcomes of implantation. Indeed, there are likely many factors, including “genetic subtype of retinal disease, age at time of implant, age at time of total vision loss, apposition of the array and retinal surface, and many others.” The identification of prognostic factors is an ongoing challenge to retinal implant engineers, scientists, and clinicians.

We used a “target trial” to guide our judgements of risk of bias of the uncontrolled clinical trials discovered by our review. A target trial is a hypothetical trial designed in such a way that it lacks any features that put assessed outcomes at risk of bias. The target trial need not necessarily be ethical, nor feasible. This target trial may guide the design of future studies of retinal implantation—or array implantation at other sites in the visual pathway—and for that reason we describe its main features here. Our target trial enrolled a cohort of patients with RP. Participants were randomly allocated to one of two groups; participants and outcome assessors were masked to this allocation. Before implantation, all patients participated in the forced-choice experiments for the assessment of light perception, spatial vision, and motion perception, as described by Bach et al. The use of multiple preimplantation follow-ups ensured that practice effects were at asymptote at the time of implantation. These assess-
ments were made monocularly in each eye (with the fellow eye patched), as well as binocularly. At implantation, participants in one group received an active device, whereas those in the other group received a sham device. The only difference between the sham device and the active device was that, during the trial, the electrodes comprising the sham device carried no current (the sham device was designed to be activated at the conclusion of the trial so as not to dissuade participation). After implantation, all patients participated in the same (monocular and binocular) forced-choice experiments for the assessment of light perception, spatial vision, and motion perception. The use of multiple postimplantation follow-ups enabled participants in the active group to learn to use the device, thus further emphasizing any between-group differences in outcomes. Combined with appropriate statistical analysis of repeated-measures data, this target trial facilitates two main comparisons: first, the comparison of vision outcomes between active and sham groups and, second, the comparison of vision outcomes before and after implantation separately for each group. Thus, our target trial controls for residual vision, participant engagement, and the placebo effect. A systematic review with a meta-analysis is the gold standard for the evaluation of medical interventions; our target trial articulates the pitfalls of attempting to test for vision with methods designed to measure the limits of a visual system that is assumed to function (normally or partially).59 Therein, Peli identifies problems with several tests currently in use to assess vision restoration, arguing that the field’s progress is contingent on the use of appropriate tests, free of bias, to guide the design of new and better devices.

Acknowledgments

The authors thank Shaun Cloherty and Eli Peli for comments on the manuscript. Thanks also to Eli Peli for raising to our attention his recent presentation at the Annual Meeting of ARVO, “Testing for vision rather than testing vision” (Invest. Ophthalmol. Vis. Sci. 2019;60(9): 2844).

Supported by the University of Auckland Faculty of Engineering Research Development Fund.

Disclosure: L.E. Hallum, None; S.C. Dakin, None

References


Appendix A

Study Characteristics
Author
Year
PMID
Extractor
Key question addressed
Study design
Recruitment dates
Participants enrolled (n)
Participants completed (n)
Funding
Inclusion criteria
Exclusion criteria
Comparator
Duration of follow-up
Primary outcome
Secondary outcome
Concurrent treatment
Intervention

Baseline characteristics of predictors
Duration of vision impairment
Anatomy
Other disabilities
Age
Preimplantation vision
Choice of implanted eye
Site
Implanted device

Summary of general study characteristics
Author/year/country
Study design/follow-up
Population (n)
Mean age
Male (%)
Disease
Preimplantation vision
Duration of vision impairment
Mean time between impairment and implantation
Device coding strategy
Study quality
Eye implanted
### Summary of study results

<table>
<thead>
<tr>
<th>Author/year/country</th>
<th>Outcome category</th>
<th>Specific outcome</th>
<th>Intervention group ($n$, mean, 95% CI)</th>
<th>Control group ($n$, mean, 95% CI)</th>
<th>$P$ value</th>
<th>Descriptive analysis</th>
</tr>
</thead>
</table>

#### Signaling questions

1. Was the study prospective?
2. Were eligibility criteria stated and appropriate?
3. Were participants representative of the target population?
4. Were potential confounders reported, and were these confounders accounted for in either the study design or analysis?
5. Were the outcomes assessed relevant to the research question?
6. Was outcome assessment independent, masked, and objective (e.g., computerized)?
7. Were outcome assessment methods comparable with the device powered off and on?
8. Were outcome assessment methods comparable before and after implantation?
9. Do psychophysical methods appear sound (e.g., were forced-choice methods used where possible)?
10. Was attrition reported?
11. Were all participants and data accounted for?
12. Were protocol violations specified?
13. Do data analysis methods appear sound (including statistical analysis methods)?

Y, yes; N, no; NA, not applicable; NR, not reported