Objective Assessment of Spectacle Wear in Infants and Toddlers Using a Wearable Sensor

Joseph M. Miller¹-³, Leslie K. Dennis², Chiu-Hsieh Hsu², and Erin M. Harvey¹,²

¹ Department of Ophthalmology and Vision Science, The University of Arizona, Tucson, AZ, USA
² Mel and Enid Zuckerman College of Public Health, The University of Arizona, Tucson, AZ, USA
³ James C. Wyant College of Optical Sciences, The University of Arizona, Tucson, AZ, USA

Correspondence: Joseph M. Miller, Department of Ophthalmology and Vision Science, University of Arizona, 655 N. Alvernon Way, Suite 204, Tucson, AZ 85711, USA. e-mail: jmiller@eyes.arizona.edu

Received: May 11, 2021
Accepted: July 27, 2021
Published: August 24, 2021

Keywords: infants; toddlers; spectacles; compliance

Citation: Miller JM, Dennis LK, Hsu CH, Harvey EM. Objective assessment of spectacle wear in infants and toddlers using a wearable sensor. Transl Vis Sci Technol. 2021;10(9):29, https://doi.org/10.1167/tvst.10.9.29

Purpose: To assess the feasibility of using a thermal microsensor to monitor spectacle wear in infants and toddlers, to determine the inter-method reliability of two methods of estimating spectacle wear from sensor data, and to validate sensor estimates of wear.

Methods: Fourteen children, 3 to <48 months of age, and one adult were provided pediatric spectacles containing their spectacle prescription. A thermal microsensor attached to the spectacle headband recorded date, time, and ambient temperature every 15 minutes for 14 days. Parents were asked for daily spectacle wear reports, and the adult recorded wear using a smartphone app. Sensor data were dichotomized (wear/non-wear) using two methods: temperature threshold (TT) and human judgment (HJ). Kappa statistics assessed inter-method reliability (child data) and accuracy (adult data).

Results: Data from two child participants were excluded (one because of corrupted sensor data and the other because of no parent log data). Sensor data were collected more reliably than parent wear reports. The TT and HJ analysis of child data yielded similar reliability. Adult sensor data scored using the HJ method provided more valid estimates of wear than the TT method (κ = 0.94 vs. 0.78).

Conclusions: We have demonstrated that it is feasible to deduce periods of spectacle wear using a thermal data logger and that the sensor is tolerated by children.

Translational Relevance: Results indicate that it is feasible to use a thermal microsensor to measure spectacle wear for use in clinical monitoring or for research on spectacle treatment in children under 4 years of age.

Introduction

In some children with amblyopia or accommodative esotropia, spectacle treatment alone can result in the apparent elimination of the amblyopia or strabismus, whereas in other children the benefits of spectacle treatment (sometimes referred to as spectacle adaptation) plateau after several weeks, leaving residual amblyopia or misalignment.¹,² Accurate and reliable assessment of compliance with spectacle wear would be useful in the clinic for ruling out noncompliance in children who are not responding optimally to spectacle treatment³,⁴ and in research efforts to better understand the time course of improvement and effective dose (hours of wear) of spectacle treatment on visual and developmental outcomes, as well as barriers to treatment adherence. However, compliance with spectacle treatment is often assessed through parent report and is very difficult to monitor, as children are not directly observed at all times and can remove their spectacles at will. Drawbacks of relying on parental report are that the accuracy of reports is unknown, there is likely to be variable consistency in reporting over time, and there may be missing or less reliable data for periods of time when the child is under the care of others (e.g., with a child-care provider or at school).

Objective automated measures of compliance have been used successfully with occlusion treatment for amblyopia. Approaches to objective monitoring of occlusion have included use of electrocardiographic electrodes under the patch to measure patch-skin
resistance\textsuperscript{5} and use of temperature sensors.\textsuperscript{6–11} Studies utilizing objective measures of compliance have provided important data on the relationship between frequency (dose) of occlusion and visual outcome and the optimal regimens for treatment of amblyopia.\textsuperscript{4}

Several recent studies have assessed use of wearable temperature sensors for measuring compliance with spectacle treatment,\textsuperscript{8,9,12–15} similar in principle to the use of occlusion dose monitors in occlusion treatment studies. Recent studies have demonstrated that temperature sensors can provide reliable estimates of spectacle wear in adult participants\textsuperscript{8,12–15} and have been used to assess the relationship between hours of spectacle wear and visual outcome in 3- to 12-year-old children prescribed spectacles for anisometropic or strabismic amblyopia.\textsuperscript{9}

The aims of the present study were (1) to assess the acceptance and feasibility of using a commercially available microsensor/data logger attached to a pediatric spectacle headband to measure the duration of spectacle wear in infants and young children, (2) to determine the inter-method reliability of the two methods of estimating spectacle wear in young children based on raw sensor data compared with wear logs recorded by a parent, and (3) to determine the validity of raw sensor data obtained by the two methods of estimating spectacle wear compared with detailed, accurate, and reliable wear logs recorded by an adult participant.

### Methods

#### Subjects

Child participants were 3 to <48 months of age, had a current eyeglass prescription, and were recruited through the Banner–University Ophthalmology Clinic (Tucson, AZ) and by referral by other participating parents. Prior to participation, written informed consent was obtained from a parent. The adult participant was one of the investigators (EMH). This study complied with the tenets of the Declaration of Helsinki, was approved by the Institutional Review Board of The University of Arizona, and conformed to the requirements of the United States Health Insurance Portability and Privacy Act.

#### Study Materials

We used a commercially available thermosensor and data logger called the TheraMon (MC Technology GmbH, Hargelsberg, Austria). The TheraMon stores ambient temperature readings at set intervals. The TheraMon software allows the user to set sampling frequency, although higher frequency sampling comes at a cost to battery life. In the present study, samples were obtained every 15 minutes. The time and temperature data are uploaded from the sensor via a radio-frequency identification reader to the manufacturer’s cloud-based software (TheraMon Azure, version 1.2.0.11), where data can be viewed in graphical format (time by temperature plots). The basic assumption underlying use of the TheraMon software to assess compliance is that recorded temperatures should approach body temperature when the sensor is being worn by the participant.

The TheraMon sensor was attached to the headband supplied with flexible pediatric eyeglass frames, either Dilli Dalli (ClearVision Optical, Hauppauge, NY) or Miraflex (Miraflex Glasses, Doral, FL). The sensors were fastened near the temple tip attachment hook for the most comfort (Fig. 1). Because a loose sensor could introduce a choking hazard due to its small size or potential for internal injury if ingested due to its battery,\textsuperscript{16} the sensors were secured to the eyeglass headbands using a cut-through and puncture-resistant medical grade heat-shrink tubing (Xtra-Shield HS-714, 3/8 inch; Insultab, Woburn, MA) that encapsulated the sensor and a portion of the headband. The tubing is difficult for an adult to remove without using a sharp tool, and a toddler is unlikely to have the strength or coordination to remove it. Parents were given safety instructions to cease use of the sensor/headband if the material attaching the sensor became damaged and to seek emergency care for their child if the child accessed and ingested a loose sensor.

#### Procedures

Study participation required one visit for the parent/child participants. At the study visit, informed consent procedures were performed, the parent and child selected a suitable pediatric eyeglass frame for the child, and the parent provided a copy of their child’s current spectacle prescription. The parent was also consulted with regard to how and when the study team should contact them (call, text, e-mail) for daily spectacle wear reports when the spectacles had been dispensed. The spectacles were then ordered.

Prior to dispensing the spectacles, a TheraMon sensor was activated and attached to the headband, and several data points were collected and downloaded to verify the functionality of the sensor. Either parents picked up the spectacles or they were mailed to the child’s home. Written and verbal safety instructions regarding use of the sensor and instructions on how...
Figure 1. TheraMon sensor and method of attaching sensor to the spectacle headband. (A) TheraMon sensor size compared to US penny. (B, C) TheraMon attached to eyeglass headband using shrink tubing. B shows the top (outer facing) view, and C shows the bottom (inner facing) view. (D) Pair of spectacles with sensor attached to the headband. The sensor was always attached in the same location (near the attachment hook for the frame temple tip), so the sensor was behind the ear when the glasses were worn.

and when parents would be contacted for daily wear reports were provided upon dispensing or mailing. Parents were given a daily log sheet and examples of the type of information and format requested, such as “wore glasses from 9 to 1, 3 to 7”; “he was sick, did not wear them today”; “forgot in the morning, 3 to bedtime (around 8)”; or “spent the day with grandma, so not sure if/when he wore them.” Parents were asked to make their best estimate of when the glasses were worn (time on, time off), providing as much detail as possible.

After dispensing, a study team member contacted the parent (call, text, or e-mail) either each evening to obtain a report of the child’s eyeglass wear for the day or each morning to obtain a report about the child’s eyeglass wear for the previous day (depending on parent preference). After 14 days of attempting to obtain daily reports, a new headband was sent to the family along with a postage-paid return envelope. Parents were asked to remove the headband with the sensor attached and return it via mail.

One adult participant also wore a pair of pediatric-style spectacles (Dilli Dalli; ClearVision Optical) with the TheraMon sensor attached to the headband in the same location used for the child participants (Fig. 1). In order to obtain several periods of wear and non-wear throughout each day, the spectacles contained only near correction and were worn intermittently as needed for reading over a 14-day period. To obtain precise records of actual wear (to the minute), a smartphone app was used to quickly and easily record “on” or “off” along with the current time with one key press, thus recording the exact timing of wear/non-wear transitions.

Upon return of the sensor, the raw data (which included the unique sensor ID and the date/time and temperature for each observation) was downloaded in spreadsheet (CSV) format for analysis. Each day, 96 observations were generated (once every 15 minutes). The dataset was trimmed to include only the samples obtained during the period that wear was logged by the sensor and adult participant or parent (14-day period starting from the first day after dispensing or receipt of mailed spectacles). Date/time data were recoded into local units to align with the wear logs completed by the parents and adult participant, and the logged wear for each specific time point at which a temperature sample was obtained was added to the dataset.

Temperature Threshold Analysis of Sensor Data

Receiver operating characteristic (ROC) techniques were used to determine the optimal temperature for distinguishing wear from non-wear based on sensor data. The parent and adult wear logs were treated as a dichotomous diagnostic test indicating wear or non-wear for this analysis, although the accuracy of the parent logs is unknown.

Youden’s J (sensitivity + specificity − 1)\(^{17}\) was used to determine the optimal temperature threshold (TT) because it places equal weight on the cost of false positives and false negatives; that is, overestimating spectacle wear time was deemed to be just as detrimental as underestimating spectacle wear time. Youden’s J
Figure 2. Example of time-by-temperature plot as initially presented to human raters (top) and after “wear” data points identified by rater (bottom, filled circles).

statistic was calculated for each candidate temperature threshold within each participant and across all child participants, with temperatures at or above threshold indicating “wear” and temperatures below threshold indicating “non-wear” of the spectacles. Youden’s J ranges from 0 to 1, with 0 indicating that the method cannot reliably distinguish between “wear” and “non-wear” and 1 indicating optimal performance of the method for distinguishing “wear” from “non-wear” (no false positives or false negatives). The optimal temperature thresholds, defined as the temperature at which Youden’s J is maximized, were determined for each participant (individually optimized threshold) and for the child participants as a group (group optimized threshold, determined using a generalized estimating equation model to account for within-subject dependence).

Human Judgment Analysis of Sensor Data

This analysis of sensor data was conducted to determine if human judgment (HJ) and interpretation of time-by-temperature data plots might yield better estimates of wear than the TT analysis. In the training
phase, four human raters (investigators JMM, EMH, LKD, and PCH) reviewed time-by-temperature data plots along with corresponding wear log data to familiarize themselves with the characteristics of the plots during instances of reported “wear” and “non-wear.” In the rating phase, the four raters examined individual daily time-by-temperature plots and categorized each temperature sample as indicating “wear” or “non-wear” as judged within the context of the daily plot (96 samples per day) using a custom scoring program that provided a graphical display of temperature versus time and facilitated scoring to the individual sample level. Final scoring plots did not contain subject identifiers, subject characteristics, date of data collection, or wear log data. Examples of rating phase plots are shown in Figure 2.

The criteria the raters used to dichotomize temperature samples were similar to the heuristics used by human raters in previous studies.8,12–15 Sharp increases or decreases in temperature tended to indicate wear transitions (from wear to non-wear or non-wear to wear); increased noise in the data (small fluctuations across samples) occurring around body temperature tended to indicate wear; and flat or stable temperatures tended to indicate non-wear. Using pairwise kappa statistics, inter-rater agreement was assessed individually for each participant’s data and for the child participants as a group (including controls for within-subject dependence).

**Inter-Method Reliability and Validity**

Inter-method reliability analysis18 of child data assessed agreement between data categorized as wear/non-wear by wear logs and sensor data scored using the TT and HJ methods using Cohen’s kappa, which does not require the assumption that one method is the “gold standard.” Intraclass correlations were examined for the total minutes of wear each day estimated by the human raters.18

The validity of sensor data scored using the TT and HJ methods was assessed through comparisons of the adult data as the gold standard (reliable and accurate wear logs) using Cohen’s kappa. Validity could only be determined for the adult participant, as the accuracy of the parent wear log data wear logs was not known.

### “Waking Hours” Analysis

The data analyses described above were conducted with two different inclusion criteria. In the primary “all hours” analyses, data collected during all daily time intervals were included. In the secondary “waking hours” analyses, data recorded from midnight to 6 AM were excluded (a time when most children are not likely to be wearing spectacles) in order to determine if including the entire time range for each day tended to overestimate agreement between the parent logs and the sensor analyses. Detailed results are reported only for the “all hours” analyses.

**Results**

**Subjects**

Fourteen children and a parent of each, as well as one adult, participated. Data from two children were excluded from analyses, because the parent of one child provided no wear log data and the sensor data from another child was found to be corrupted (temperature values were out of expected local temperature ranges). The final child sample included 12 children (eight females, four males) ranging in age from 4.1 to 39.7 months.

Among the child participants, 14 complete days of sensor data (1344 observations) were collected for 11 of the 12 children (Table 1). One child (participant 2) was missing 6 full days of sensor data; the parent was unable to complete the full 14 days of data collection due to family obligations and returned the sensor after 8 days, but 8 full days of sensor and wear logs were collected. Complete parent wear log data were successfully collected for all 14 days for four children (partici-
Table 2. Summary of ROC Analysis Conducted on “All Hours” (24 hr/d) of Individual Child Data, Grouped Child Data, and Adult Data

<table>
<thead>
<tr>
<th>Participant</th>
<th>Area Under the Curve</th>
<th>Individual Threshold Method</th>
<th>Group Threshold Method (≥28.00°C)</th>
<th>Samples Included in Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Threshold (°C)</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>1</td>
<td>0.92</td>
<td>27.45</td>
<td>0.87</td>
<td>0.90</td>
</tr>
<tr>
<td>2</td>
<td>0.90</td>
<td>23.13</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>3a</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>0.88</td>
<td>23.44</td>
<td>0.84</td>
<td>0.82</td>
</tr>
<tr>
<td>5</td>
<td>0.80</td>
<td>28.70</td>
<td>0.74</td>
<td>0.89</td>
</tr>
<tr>
<td>6</td>
<td>0.95</td>
<td>25.17</td>
<td>0.95</td>
<td>0.93</td>
</tr>
<tr>
<td>7</td>
<td>0.80</td>
<td>26.91</td>
<td>0.72</td>
<td>0.88</td>
</tr>
<tr>
<td>8</td>
<td>0.72</td>
<td>23.98</td>
<td>0.71</td>
<td>0.78</td>
</tr>
<tr>
<td>9</td>
<td>0.71</td>
<td>29.41</td>
<td>0.67</td>
<td>0.99</td>
</tr>
<tr>
<td>10</td>
<td>1.00</td>
<td>27.61</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>11</td>
<td>0.93</td>
<td>26.91</td>
<td>0.88</td>
<td>0.92</td>
</tr>
<tr>
<td>12</td>
<td>0.85</td>
<td>25.08</td>
<td>0.78</td>
<td>0.92</td>
</tr>
<tr>
<td>Groupb</td>
<td>0.88</td>
<td>28.00</td>
<td>0.78</td>
<td>0.92</td>
</tr>
<tr>
<td>Adult</td>
<td>0.93</td>
<td>28.72</td>
<td>0.95</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Summary data include the area under the curve, the temperature threshold that yielded the maximum Youden’s J statistic (individually optimized threshold), sensitivity and specificity at the individual participant’s optimized threshold, and the sensitivity and specificity for each participant using the group threshold of ≥28.00°C.

a Per parent of participant 3, child had no periods of wear.

b Includes all child data based on a generalized estimating equation model to account for within-subject dependence.

Identification of Optimal Temperature Thresholds

Table 2 summarizes results of the ROC analyses for individual participants and for the child participants as a group using data for “all hours.” For individual children, temperatures that optimally dichotomized sensor data to agree with parent wear log data ranged from 23.13°C to 29.41°C. An optimal temperature for participant 3 could not be determined, as the parent reported that the child had no periods of wear. The temperature that optimally dichotomized temperature values across all 12 child participants was ≥28.00°C. In subsequent analyses including the TT method, ≥28.00°C was used as the threshold indicating “wear.” For the adult participant, the individually determined optimal temperature threshold (≥28.72°C) and the child group optimal threshold (≥28.00°C) yielded excellent sensitivity and specificity (0.94–0.95).

Inter-Rater Reliability of Human Judgments

The intraclass correlation among human raters counting the total minutes of wear per day (thus ignoring if start and end times of wear shifted an interval) was 0.71 (95% confidence interval [CI], 0.52–0.85) for child participants, and it was 0.99 (95% CI, 0.97–0.99) for the adult participant.

Inter-Method Reliability of Child Data

Inter-method reliability for grouped child sensor data scored using both the TT method ($\kappa = 0.84$) and HJ method ($\kappa = 0.83$ to 0.88 for the four raters) compared with parent wear logs was moderate (Table 3). For individual children, the kappa values were similar for the TT and HJ methods. Reliability for both methods varied widely across children, ranging from
Table 3. Summary of Pairwise Concordance Analyses (Cohen’s kappa) for “All Hours” (24 hr/d) of Data for Individual Child Data, Grouped Child Data, and Adult Data

<table>
<thead>
<tr>
<th>Participant</th>
<th>Wear Log vs. Temperature Threshold (95% CI)</th>
<th>Human Raters</th>
<th>Wear Log vs. Rater 1</th>
<th>Wear Log vs. Rater 2</th>
<th>Wear Log vs. Rater 3</th>
<th>Wear Log vs. Rater 4</th>
<th>Mean (Min, Max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.77 (0.74–0.81)</td>
<td>0.78</td>
<td>0.76</td>
<td>0.75</td>
<td>0.76</td>
<td>0.76 (0.75, 0.78)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.51 (0.45–0.57)</td>
<td>0.63</td>
<td>0.65</td>
<td>0.60</td>
<td>0.63</td>
<td>0.63 (0.60, 0.65)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.53 (0.46–0.60)</td>
<td>0.53</td>
<td>0.51</td>
<td>0.47</td>
<td>0.50</td>
<td>0.50 (0.47, 0.53)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.58 (0.53–0.62)</td>
<td>0.59</td>
<td>0.63</td>
<td>0.61</td>
<td>0.62</td>
<td>0.61 (0.59, 0.63)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.87 (0.84–0.90)</td>
<td>0.87</td>
<td>0.87</td>
<td>0.86</td>
<td>0.87</td>
<td>0.87 (0.86, 0.87)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.60 (0.53–0.67)</td>
<td>0.57</td>
<td>0.64</td>
<td>0.64</td>
<td>0.53</td>
<td>0.60 (0.53, 0.64)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.45 (0.39–0.50)</td>
<td>0.43</td>
<td>0.45</td>
<td>0.45</td>
<td>0.44</td>
<td>0.44 (0.43, 0.45)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.14 (0.04 to 0.31)</td>
<td>—</td>
<td>0.19</td>
<td>0.13</td>
<td>0.23</td>
<td>0.18 (0.13, 0.23)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.97 (0.95–0.98)</td>
<td>0.97</td>
<td>0.98</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97 (0.96, 0.98)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.76 (0.73–0.80)</td>
<td>0.74</td>
<td>0.79</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76 (0.74, 0.79)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.49 (0.42–0.55)</td>
<td>0.71</td>
<td>0.75</td>
<td>0.71</td>
<td>0.71</td>
<td>0.72 (0.71, 0.75)</td>
<td></td>
</tr>
<tr>
<td>1–12c</td>
<td>0.84 (0.83–0.85)</td>
<td>0.86</td>
<td>0.88</td>
<td>0.83</td>
<td>0.85</td>
<td>0.86 (0.83, 0.88)</td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td>0.78 (0.74–0.82)</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94 (0.93, 0.95)</td>
<td></td>
</tr>
</tbody>
</table>

Analyses assessed agreement between wear logs (adult or parent) and sensor data scored using the TT method and the human raters (four raters). Kappa statistics could not be determined for participant 3, as the parent reported no spectacle wear, or for participant 9/rater 1, as the child had no spectacle wear per rater 1 assessment.

a Using group optimized threshold of ≥28.00°C.
b Per the parent of participant 3, the child had no periods of wear.
c Includes all child data and controls for within-subject dependence.

poor to excellent. For individual children, agreement among each of the four human raters and the parent logs was similar, but agreement varied widely across children (0.18–0.97).

Table 4. Summary of Estimates of Hours of Spectacle Wear (Average Per Day) Based on Wear Logs and Sensor Data Scored Using Temperature Threshold and Human Rater Analysis Methods

<table>
<thead>
<tr>
<th>Participant</th>
<th>Wear Loga</th>
<th>Temperature Thresholdb</th>
<th>Human Raters</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.89</td>
<td>10.34</td>
<td>9.86</td>
<td>10.05</td>
</tr>
<tr>
<td>2</td>
<td>10.13</td>
<td>6.34</td>
<td>7.94</td>
<td>8.19</td>
</tr>
<tr>
<td>3</td>
<td>0.00</td>
<td>2.23</td>
<td>0.07</td>
<td>0.71</td>
</tr>
<tr>
<td>4</td>
<td>6.81</td>
<td>7.57</td>
<td>6.98</td>
<td>7.52</td>
</tr>
<tr>
<td>5</td>
<td>9.42</td>
<td>10.11</td>
<td>7.48</td>
<td>8.09</td>
</tr>
<tr>
<td>6</td>
<td>11.82</td>
<td>12.02</td>
<td>11.98</td>
<td>11.95</td>
</tr>
<tr>
<td>7</td>
<td>8.67</td>
<td>8.59</td>
<td>3.68</td>
<td>3.77</td>
</tr>
<tr>
<td>8</td>
<td>9.15</td>
<td>8.30</td>
<td>7.93</td>
<td>8.36</td>
</tr>
<tr>
<td>9</td>
<td>0.05</td>
<td>0.45</td>
<td>0.00</td>
<td>0.48</td>
</tr>
<tr>
<td>10</td>
<td>10.36</td>
<td>10.21</td>
<td>10.21</td>
<td>10.05</td>
</tr>
<tr>
<td>11</td>
<td>9.96</td>
<td>8.96</td>
<td>9.02</td>
<td>9.91</td>
</tr>
<tr>
<td>12</td>
<td>6.41</td>
<td>4.07</td>
<td>5.50</td>
<td>6.38</td>
</tr>
</tbody>
</table>

Mean (SD) 7.72 (3.89) 7.96 (3.93) 7.43 (3.53) 6.72 (3.80) 7.19 (3.91) 7.13 (3.69) 7.09 (3.80)

Adult 3.70 3.72 4.89 3.54 3.73 3.70 3.57

a Model M1 assumed that observations with missing wear log information indicated no wear. Model M2 excluded observations with missing wear log information and then converted (prorated) it to wear time per day.
b Using group optimized threshold of ≥28.00°C.
spectacles were being worn. Agreement between the sensor data scored using the TT method and the wear logs was moderate (0.78), and agreement between the sensor data scored using the HJ method and the wear logs was excellent (0.94) (Table 3).

### Comparison of Wear Estimates Across Method

Table 4 summarize estimates of average daily wear determined using the wear logs, TT method, and the HJ method for analyses using “all hours” data. Missing wear log data were handled in two different ways in order to assess the potential impact of different assumptions about missing log data on estimates of wear. For one method (M1), missing wear observations (periods of time when parent did not provide a wear log) were considered “non-wear,” possibly underestimating wear. The other method (M2) excluded observations with missing wear log data, and daily wear time was prorated based on available observations. Estimates of mean hours of wear per day were identical using M1 and M2 for nine children (including the four children with no missing log data), <1 hour higher using M2 for two children, and more than 1 hour greater using M2 for one child.

For child subjects, there was good agreement on average between estimates of hours of wear: mean hours of wear for child subjects were 7.72 (M1) and 7.96 (M2) based on the wear log, 7.43 for the TT method, and 6.72 to 7.19 for the HJ method. For the adult subject, hours of wear were similar for the wear log and HJ method of scoring sensor data, but the TT method, using the threshold generated from the grouped child data, yielded longer estimates of wear (>1 hour).

The analyses summarized above were also conducted using only data collected during “waking hours.” Results were similar with the two methods.

### Discussion

The aim of the present study was to determine if a commercially available microsensor and data logger attached to a pediatric spectacle headband could reliably measure the duration of spectacle wear in young children. To our knowledge, this is the first study to report objective measures of duration of spectacle wear using a wearable sensor in children younger than 3 years of age.

Data were collected from an adult participant to assess the validity of the sensor for distinguishing between wear and non-wear compared with precise wear log data. A single subject was sufficient for our purposes because that single subject provided 19 known variable-duration episodes of spectacle wear with exact recording of on/off times and which included known episodes of the sensor being left unworn in both room-temperature environments and elevated temperatures (car in hot weather). We assessed two methods of analyzing sensor data (temperature threshold and human judgment). Results indicated that the HJ method was more consistent with the wear log ($\kappa = 0.93–0.95$ for the four raters) than the TT method ($\kappa = 0.78$). These data suggest that there are characteristics of the plots that could be detected by the human raters as wear, or non-wear, that could not be accurately distinguished by using a simple temperature threshold method.

Data were collected from infants and young children to determine the feasibility of using the sensor to measure the duration of spectacle wear in this age group. Use of the sensor was acceptable to parents and well tolerated by the children. Sensor data were more consistently collected than the parent log data over the 14-day data collection interval (Table 1). Parent wear logs were successfully collected for all 14 days for only four of the 12 children, compared with sensor data, for which 14 days of data were collected for 11 of the 12 children. This difference highlights an important benefit of using the sensor rather than relying on parent reports of wear. However, there are also potential challenges to the use of the sensor in terms of feasibility and minimizing missing data. For example, there is the potential for long periods of missing data if the sensor is damaged, malfunctions, or is not returned. We successfully retrieved all of the sensors in this 14-day study, but lost sensors are likely to occur with longer follow-up intervals. In this study (which utilized 15 sensors), we observed one instance of sensor data becoming corrupted.

Data from infants and young children were also collected to determine the inter-method reliability of the two sensor methods of analysis in comparison with the parent wear logs. These data differ from the adult data in that the accuracy of the parent wear logs is unknown and therefore cannot serve as a “gold standard” reference for assessment of the validity of the sensor with the child participants. For the TT analysis, we first determined the optimal temperature for dichotomizing sensor results based on individual and group child wear log data. Results indicated that optimal thresholds varied widely across children. Thus, one limitation of the TT method is that it assumes that the data used in establishing a temperature threshold (in this study, the child group optimized temperature...
Individual optimal temperature thresholds may be influenced by several factors, including proximity of the sensor to the child’s head (e.g., long hair may yield temperatures closer to ambient temperatures than short hair; loose spectacles may yield temperatures closer to ambient temperatures than tightly worn spectacles) or extreme ambient temperatures (high ambient temperatures at or above threshold may appear to indicate “wear” when the spectacles were not worn, and low ambient temperatures may yield lower temperatures, below threshold, even when the spectacles are worn). Previous studies have attempted to address these issues by using two sensors, arranged so that one is near to or touching the skin and the other is farther from the body so that ambient temperatures can be taken into account when assessing wear, or by using a temperature range to indicate wear rather than a single threshold value.

We included four human raters to assess inter-rater agreement of the HJ analysis method. Results indicated that there was good agreement among raters for child participant data and excellent agreement among raters for the adult participant data, which should simplify implementation and quality control. However, the HJ analysis is more time consuming than the TT analysis, which can be easily determined using the manufacturer’s software. For research applications, the use of human raters offers advantages for measurement validity if rater bias is minimized by masking raters.

Inter-method reliability for grouped child sensor data scored using both the TT and HJ methods compared with parent wear logs was moderate. However, reliability for both methods varied widely across children, ranging from poor to excellent. Given the excellent validity findings from the adult data, much of this variability is likely the result of variability in accuracy and precision of parent logs.

Results of the present study suggest that the HJ method of analyzing sensor data has advantages over the TT method in terms of validity. The principal advantage of the HJ method appeared to be that, in instances where spectacles were not worn but temperatures were recorded above the threshold temperature, judgments as to whether the shape and speed of the temperature curve resembled the step function typically seen for wear starting and stopping were readily made by the human observers, as was detection of wear time when no parent report was available. Our results are consistent with other studies that have found that time-by-temperature data plots could be reliably categorized as indicative of spectacle wear or non-wear by human observers. An example of this is seen in the data from the adult participant. Specifically, when the ambient temperature was above body temperature on a day when the sensor was left in a vehicle, a misclassification can occur (especially in warm regions such as Tucson), resulting in overestimates of duration of wear. Figure 3 shows a time-by-temperature plot from the adult participant on a hot day on which the sensor was left in a vehicle. The raters judged these extreme temperatures as not indicating wear, whereas the TT method categorized them as indicating wear.
that the coin cell battery that is commonly used in miniature electronics presents a tremendous burn hazard to children if swallowed, and we were looking to use a manufactured device that completely encapsulates the battery. The TheraMon sensor was originally developed for intraoral use, as its stated purpose is to document compliance with orthodontic device wear recommendations. Although our application does not embed the sensor in a device, it is attached to a pair of spectacles too large to swallow with US Food and Drug Administration–compliant food-safe polyolefin tubing that is difficult to remove with scissors.

The results of the present study indicate that a wearable temperature microsensor/data logger can provide accurate estimates of spectacle wear duration and that this method is feasible for use with infants and very young children. The method could be used for clinical monitoring of individual patients, as well as for research on the dose–response effects of spectacle wear with regard to visual improvement and ocular alignment, as well as the dose–response relationship between spectacle wear and child development and learning. In addition, use of an objective method of monitoring spectacle wear will be useful in assessing the level of adherence to spectacle wear across a wide age range, beginning in infancy, for use in studies determining barriers to treatment compliance across age. There is currently little information in the literature on spectacle treatment adherence in children younger than 3 years of age.20 Use of wearable technology to obtain objective assessments of compliance with spectacle wear in young children is feasible and could contribute significantly to the design and methodology of future clinical studies.

Acknowledgments

Supported by grants from the National Eye Institute, National Institutes of Health (UG1 EY029657 to JMM and EMH) and the Virginia G. Piper Charitable Trust.

Disclosure: J.M. Miller, Arizona Board of Regents of University of Arizona (P); L.K. Dennis, None; C.-H. Hsu, None; E.M. Harvey, Arizona Board of Regents of University of Arizona (P)

References


