Deriving Objective Probability of Detection for Missing-Person Search: Validating Use of Effective Sweep Width and Associated Mathematical Models

Kenneth B. Chiacchia, PhD

Heather E. Houlahan, AB

Mountaineer Area Rescue Group

Appalachian Search and Rescue Conference

Morgantown, WV

U.S.A.

Email kchiacchia@asrc.net

Abstract

Effective sweep width (*W*) promises objective probability of detection (POD) values for guiding missing-person search efforts. However, methods for measuring *W* produce large uncertainties. Also, models for generating POD from *W* have not been validated for ground-based search. The authors applied least-square fits of POD data collected in the field for air-scent dog teams as well as human searchers using the two most prevalent models to derive *W* values. The method routinely fits the data to an R-square of >.8 with more statistical power than previous methods, and appears to be detector-agnostic. The authors present recommendations for optimizing its use in the field.

KEYWORDS: SAR, effective sweep width, probability of detection, search theory, dogs, smell

Introduction

In ground-based search for missing persons, resource allocation is a major imperative. Available resources — particularly trained searchers and specialized search resources — are virtually never sufficient to look everywhere the search subject could potentially be. Thus, operational search managers employ Bayesian logic to prioritize which "segments" — fractions of a given search area small enough to be searched by a single team/modality — will be searched at a given time via the equation:

1. $POS_o = \sum(POD \times POA)$

where POS_o is overall probability of success for the entire search effort (*i.e.*, finding the subject; Koopman BO, 1980; Charnes A, Cooper WW, 1958). POD, probability of detection, is the probability for a given detector/group of detectors to identify a given search object in a segment under given environmental

conditions, assuming the object is located in that segment. POA, or probability of area, is the probability that the desired search object *is* contained within that segment.

By summing POD X POA (*i.e.*, the POS for each segment) over all segments searched to date, a current estimate for POS_o can be derived, because each unsuccessful search effort (*i.e.*, not generating a find) in a segment reduces its POA. (The new POA is [1 – POD] X the initial POA.) This approach — sometimes called "shifting POAs" — continually reassesses POAs throughout the search area, allowing search managers to reallocate efforts toward the areas with the highest POAs (weighted by time required to search), thus optimizing POS_o (Charnes A, Cooper WW, 1958). Shifting POAs can thus be used to determine whether to continue searching current segments, expand to a larger search area, or indeed to suspend the search pending better information on the subject's possible location.

Much work has been done to amass historical data to estimate initial POAs (Syrotuck WG, Syrotuck JA, 2000; Koester RJ, 2008; Koester RJ, 2020) with some confidence. On the other hand, objectively measuring POD, given the likely effects of different detectors and environmental factors, remains an unfinished project within the search-and-rescue (SAR) community. An expanding body of work has focused on obtaining objective PODs through field measurement of effective sweep width, or *W* (the distance defining an envelope within which a given combination of detector/search object/conditions experiences an equal number of detections and misses; Cooper DC *et al.*, 2003; Koester RJ *et al.*, 2004; Chiacchia KB, Houlahan HE, 2010; Koester RJ *et al.*, 2014; Chiacchia KB, 2020). New mapping/search planning computer applications are beginning to incorporate *W* values. The old practice of determining POD through subjective estimation by search-team leaders, which produces highly inaccurate values (Koester RJ *et al.*, 2004), may still be more common than use of *W*; but the SAR community is making progress.

Issues remain in the use of *W* to derive objective POD values. One study involving computer modeling and field measurements suggested that navigational errors in search patterns may impose errors of ~10-20% in prospective POD determinations of intended search tasks (Perkins D, 2018) — though use of GPS tracks for searchers can document the actual search pattern and so avoid this issue in retrospective assessment of already-completed efforts. In addition, use of *W* almost certainly requires adoption of computerized search planning, since the nonlinear relationship between *W* and POD is unintuitive (Koopman BO, 1980; also, see Results and Discussion).

A more significant problem may derive from the limits of field-measurement of *W*. *W* values can be derived from the canonical crossover method (Koopman BO, 1980; Cooper DC *et al.*, 2003; Koester RJ *et al.*, 2004), which determines the "crossover point" at which the number of detections, cumulated from infinite distance inward, equals the number of misses, cumulated from zero distance outward. The lateral distance of the crossover point is *W*/2, as the detector "sweeps" this distance to its left and its right as it moves through the search area. By making a crossover plot for each individual searcher or dog-and-handler team, we can derive mean *W* values and statistical parameters for those means. However, the small data from each individual determination tends to result in high uncertainties, limiting us to only modest statistical comparisons of detectors and effectors of *W* (Chiacchia KB, Houlahan HE, 2010; Chiacchia KB *et al.*, 2015).

An examination of a typical crossover plot, seen in Figure 1 (adapted from Chiacchia KB *et al.*, 2015), explains part of the problem. Again, the plot cumulates detections from infinity and misses from zero distance, with the crossover point defining *W*/2. In this particular case, generating the crossover plot involved collection of more than 250 detection opportunities (DOs, lateral detections or misses). However, a simple thought experiment shows that adding detections below the crossover point or misses above it does not change the crossover point and thus *W*. The value for *W* rests solely on detections at the crossover and above and misses at the crossover and below — again in this case, only slightly more than 50 out of 250 DOs "count" toward *W*, a significant loss of data.

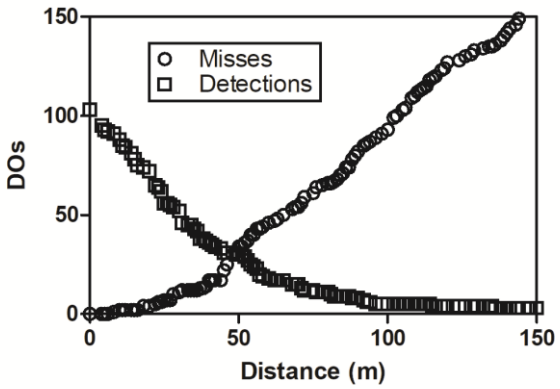


Figure 1: Crossover plot for air-scent dog teams, adapted from the authors' earlier report. Note that, as detections are cumulated from infinity and misses cumulated from zero lateral distance, only detections ≥ the crossover distance or misses ≤ the crossover distance actually contribute to determining W (which is twice the crossover distance). Reprinted from Wilderness & Environmental Medicine*, 26(2), Chiacchia KB* et al.*, Deriving Effective Sweep Width for Air-scent Dog Teams, p. 146, Copyright (2015), with permission from Elsevier.*

An alternative method, the lateral range curve, can be used to determine *W* and makes more complete use of the data (Koopman BO, 1980; Cooper DC *et al.*, 2003; Koester RJ *et al.*, 2004). But this method depends on calculating detections at each given lateral distance from a detector, which either means the huge noise of low-DO (<~30) determinations at each distance (sometimes an impossibility, when there is only a single DO at that distance) or binning of distances, which adds averaging bias. See Figure 2 for an example of this problem in a typical lateral range curve, generated by the data for the low-visibility mannequin in summer in our earlier report (Chiacchia KB, Houlahan HE, 2010). (As POD is by definition the probability of detection within a given area, the detections here, being single-distance probabilities, are instead labeled "% Detected;" Koopman BO, 1980)

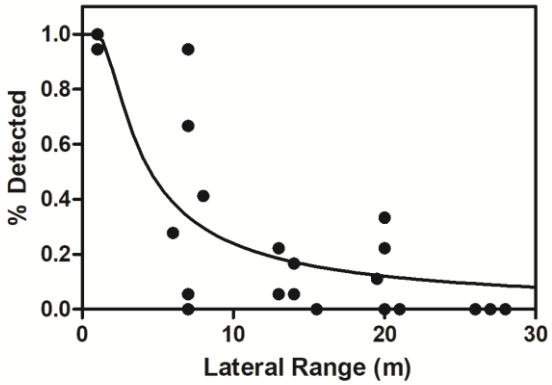


Figure 2: Typical lateral range curve, from the low-visibility mannequin in State Game Lands 203, summer (Chiacchia KB, Houlahan HE, 2010), showing the high degree of noise in individual points based on low number of DOs. The y-axis is labeled "% Detected," as single-distance detections do not fit the area-based definition of POD. The fit line, for illustration of the trend, is arbitrarily based on the inverse-cube model (see below).

Another problem with determining objective PODs from *W* is that the predominantly used "random-search" model (a name referring to random error in the detector's intended path, not a detector searching a segment in a random path) for deriving POD from *W*, while used successfully for some time in the maritime- and aerial-search spheres (Koopman BO, 1980), has never been explicitly validated in ground-based search. Hence, some SAR researchers and practitioners are skeptical that the method produces accurate results in that sphere.

In this report, the authors present a new way of determining *W* that explicitly links the value to field-measured PODs for both air-scent dog teams and human visual searchers. Unlike the crossover method, it employs virtually all the data collected; and without binning it produces much less noise in individual points in the curve than the lateral range method. The more efficient use of data helps to enable better least-square fits and thus more extensive statistical comparison of detectors and effectors of *W*, as well as testing of the appropriateness of specific theoretical models for the relationship between POD and *W*.

Methods

Air-Scent Dog Team Experiments

This report analyzes data collected anew, between 2016 and 2020, as well as data from our previous report (Chiacchia KB *et al.*, 2015). We collected new air-scent dog team data using the methodology described in that report. Briefly, we recorded GPS data on randomized, blinded daylight air-scent tasks of 5 ha or larger (to ensure the possibility of longer-range finds and misses), using a Garmin GPSMap 60csx or 62sc GPS receiver (Garmin International, Olathe, KS, USA), tracking the handler's path and recording the position at which the dog signaled a find to the handler and the position of the subject when the dog

led the handler in to the subject. No finds by the dog without such a "refind" that successfully led the handler in to the subject were scored as "detections;" lateral approaches to the subject without such a successful refind, whether misses or detections without a refind, were scored as misses. As reported, all detection or miss distances were recorded laterally from the handler's track and not directly between the handler's position at the refind and the subject, in some cases requiring extrapolation of the handler's intended path. Detections and misses were recorded manually using a 1:3,000 or 1:4,000-scale printout of the task using either MyTopo Terrain Navigator Pro (various versions, MyTopo, Billings, MT, USA) or the online version of SARTopo (CalTopo LLC, Truckee CA, USA).

For the current report we collected DOs from 178 air-scent dog tasks at 27 locations throughout western PA and northern WV, USA. Of these tasks, 11 were excluded from the study: 1 lacked a defined search area in our records, 2 because noticeably poor performance by the dog was followed by a diagnosis requiring Lyme disease treatment, 2 because the subject was not inside the assigned search area, 3 because the subject inadvertently unblinded the problem, 1 because fencing not reflected on the USGS maps prevented the team from fully accessing the assigned area, 1 because the team searched the wrong area, and 1 because the subject was in motion during the task and so relative position with the handler's track was not possible to obtain. The remaining tasks generated 150 detection and 393 miss distances, a total of 543 DOs. Note that when the team completed the task without a find (which happened in 10 tasks), the problem was unblinded so that the dog could make a find for training purposes and only misses scored before the unblinding were counted. As with the earlier study, and because it is difficult to distinguish olfactory finds by the dogs and (relatively rare) visual finds by the handler, both were scored as "dog team finds," with the difference in *W*s between dog teams and individual human searchers demonstrating what the dog adds to the team's search efficacy (Chiacchia KB *et al.*, 2015).

Locations for tasks were chosen arbitrarily, following the team's training program and goals. Meteorological measures of insolation and wind were taken in the field by each handler using the methods previously described (Chiacchia KB *et al.*, 2015). Vegetation density was derived by scoring subjective descriptions of each area taken at the time of the task, using terms such as "open" or "sparse" as denoting open vegetation, "heavy" or "thick" as denoting heavy vegetation, and "moderate" or combinations of open and heavy to describe parts of the same area as moderate/mixed vegetation. Ecoregions for each task were scored by registering a map of level III and IV ecoregions in U.S. EPA Region 3 against the task locations using SARTopo (See Figure 3). Humidity and temperature data for each task were determined with the online Weather Underground site, which collects data for the nearest weather station to a given location (Online at https://www.wunderground.com/).

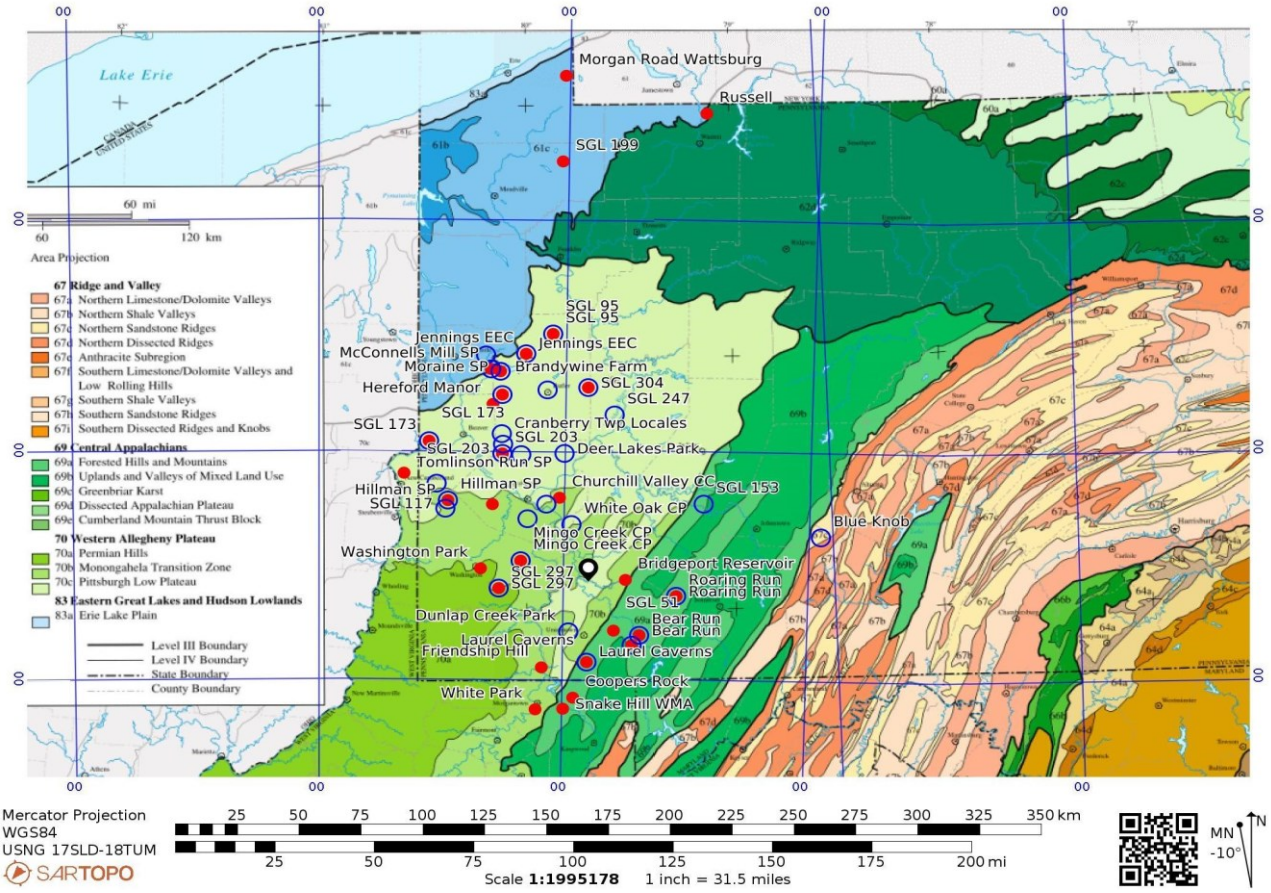


Figure 3: Location of air-scent-dog-team tasks, on a U.S. EPA level III and IV ecoregion map. Sites in the current study are solid red circles; sites in the previous study (Chiacchia KB et al., 2015) are open blue circles.

The authors collected data from 12 air-scent dog teams consisting of 2 handlers (denoted handler 1 and 2) alternating in handling 6 dogs. Note that both handlers began the study with nearly identical training and operational histories, having both become operational in December, 1992. Dog/handler pairs were determined at each training with a rotation, in an attempt to ensure that each pair was maximally represented. Due to staggered entry of dogs into the study and the occasional absence for medical reasons, each team was nevertheless not equally represented. Data collection for the study was also suspended from May through November 2019 due to an author/handler's shoulder injury and subsequent recovery. The dogs were:

- Dog A, Pipistrelle, an intact female English shepherd, born in 1999. Pip died in 2012 and so did not participate in the new data collection, though data from the previous study used in this report includes her.
- Dog B, Sophia, a neutered female German shepherd dog, born in 2005. Sophie entered the current study at its beginning, and retired from SAR and the study due to age-related medical reasons in April, 2017. She was the only dog represented in both data sets.
- Dog C, Brier Rose, a female English shepherd born in 2007. Note that Rosie is Pip's daughter and was neutered in 2019. Rosie entered the current study at its beginning.

- Dog D, Cole, a neutered male English shepherd, born in 2008. Cole entered the current study at its beginning. Unlike the other dogs in this study, which were purpose-bred for either police, farm, or SAR work, as a puppy Cole was a rescue from a felony animal cruelty case.
- Dog E, Charlotte, an intact female English shepherd born in 2013. Charlie is Rosie's daughter, and entered the current study at its beginning.
- Dog F, Verity, an intact female English shepherd born in 2017. V is Charlie's daughter, and entered the current study in December, 2019.

Dog teams, consisting of the dogs above paired with the two individual handlers, were certified via an external testing organization, the Pennsylvania Search and Rescue Council, using their 2010 air-scent testing standard (corresponding to NIMS type III non-discriminating air-scent team; online at https://img1.wsimg.com/blobby/go/8cfbf535-53eb-493c-8c8c-c88150fa0928/downloads/1d2vbidlq_354974.pdf?ver=1611603462319). The exceptions were teams 1E, 2D, and 2F (*i.e.*, handler 1/dog E, etc.), which were not certified during the duration of the current study. The certified dog-and-handler teams also worked together on multiple real SAR missions.

Human Visual Experiments

We conducted the human visual sweep width experiments in the manner of Koester RJ *et al.*, 2004, with modifications previously described (Chiacchia KB, Houlahan HE, 2010), using the IDEA Microsoft Excel worksheet provided by R. Koester and N. Guerra to automatically generate a randomized plan for a sweep-width course. We calibrated the courses either by measuring average maximum detection range (AMDR) values on site, or from previous experiments in the same area and season.

In addition to data from the two effective-sweep-width experiments from our previous report (Chiacchia KB, Houlahan HE, 2010), we used data from two additional experiments. The results from these contributed to another previous report (Koester RJ *et al.*, 2014), but have not previously been published in toto. These experiments took place at Clarence Schock Memorial Park, Mount Gretna, Pennsylvania, U.S.A., in the spring of 2006 and State Game Lands 203 near Wexford, Pa., U.S.A., in the summer of 2010.

The area chosen for the Mount Gretna experiment was the north slope of Governor Dick Hill in Clarence Schock Memorial Park. Starting point for the course was 18TUK7658057276 (U.S. National Grid, Datum NAD83); total length was 3,350m, with the last clue placed at 3,320m. The area (U.S. EPA Ecoregion 64, Northern Piedmont) is largely second-growth forest, dominated by a birch, beech, and maple canopy common to the Allegheny Mountain ridges, with a mid-story of birch saplings and spicebush and a ground cover of light bramble, birch, and spicebush saplings. The vegetation had just begun to leaf, producing perhaps a transitional situation compared with winter or summer sight lines. Another major feature of the area was frequent outcroppings of large (one- to two-meter or larger diameter) boulders of a green-gray hue quite similar to that of the green adult mannequin (see Discussion below).

The days of the experiment (May 4-6, 2006, in coordination with the Pennsylvania Search and Rescue Council's then-annual SAR-EX conference) proved largely clear, with cool mornings and mildly warm days,

and little cloud cover. 26 searchers followed a course consisting of 19 real and 3 virtual high-visibility human targets, and 17 real and 2 virtual low-visibility human targets. ("Virtual" objects occurred when the random placement would have made the object completely invisible from the course. In these cases, the virtual location was counted as an automatic miss, with the real object moved toward the course until just visible to a cued experimenter.) With discounted DOs, this made for a total of 553 DOs on the high-visibility target and 485 on the low-visibility target.

The first three searchers completed the course in the late afternoon/early evening of May 4. Sundown occurred while these searchers were still on the course, making the relevance of subsequent detection opportunities (DOs) to daytime visual search questionable. Therefore, we used local sundown that day (20:03 Eastern Daylight Time) as an endpoint, counting neither detections nor misses recorded after that time. Upon our confirmation run of the course at the end of the experiment, we found one of the high-visibility mannequins to have moved onto the trail, presumably from wind. We discounted any detections or misses on this search object after the last searcher made a positive identification on it in its proper place, which had happened earlier that day. Another issue arose on the morning of May 6, when a park staff member removed one of the high-visibility mannequins near the trail (lateral distance 1m) and subsequent navigational flags. Upon clearing up the matter with the staff member (who had mistakenly thought the mannequins and flags to be unauthorized), we replaced the navigational flags and the dummy. No searchers missed this target, so apparently none had passed by its position in the time between the ranger removing it and our replacing it.

Starting point for the Wexford experiment at State Game Lands 203 was 17TNE7396998772; total length was 2,478m, with the last clue placed at 2,464m. The area (Ecoregion 70, Western Allegheny Plateau) is largely second- and third- growth forest that had been utilized for farming as well as strip mining, gas and oil drilling, and more recently has been managed for game-bird habitat via clear-cutting and resulting succession. The woods are dominated by a mixed maple, tulip poplar, oak, cherry, and hickory canopy common to the Allegheny Plateau, with a mid-story of spicebush and sassafras and a ground cover of ferns, mayapple, and naturalized multiflora rose. Occasional stands of naturalized black raspberry bramble and fox grapevines also dot the area. The vegetation was in full leaf for the experiment, with full summer growth of the brambles. A few leaves of the spice-bush plants had gone yellow, presumably due to stress from a dry spell, but the great majority of the vegetation was still green. General vegetation types included open woods; fields and mowed trail areas; brambles; and sucker stands, the latter being the five- to 10-year succession vegetation after clear-cutting.

The weather for the single day of the experiment (Sept. 11, 2010), was mild, beginning in the low 10s°C in the morning, hitting a high in the low 20s around 15:30, and staying there until sundown. Skies were clear for the duration of the experiment. 24 searchers followed the course, which was laid out intentionally not following extant trails, with search objects consisting of 17 real and 2 virtual low-visibility-mannequins; 16 real and 5 virtual orange gloves, 15 real and 3 virtual blue gloves; and 15 real and no virtual brown gloves.

When confirming the search object placements after the experiment, we found that one orange glove that had randomized to a crossing of our course with an extant trail had been moved, presumably by a passerby

assuming it had been accidentally dropped and intending to place it where it would be easier to see. We discounted all detections and misses on this glove. We also discarded all data from one searcher (number 11) from whom we did not obtain a signed release form. The end result was 435 low-visibility-mannequin, 483 orange-glove, 413 blue-glove, and 345 brown-glove DOs.

The low- and high-visibility mannequins used in both of the new experiments were the same ones used in our previous report (Chiacchia KB, Houlahan HE, 2010). We obtained brown work gloves (Tractor Supply, Brentwood, TN, USA) as low-visibility clues; blue jersey work gloves (Grand Rapids Industrial Products, Grand Rapids, MI, USA), intended to be medium-visibility clues; and orange gloves (Boss Manufacturing Company, Kewanee, IL, USA), as high-visibility clues. Note, however, that the blue of the gloves was quite bright and returned a *W* value similar to that of the orange glove (see below). For the SGL 203 but not the Mount Gretna experiment, as previously reported (Chiacchia KB, Houlahan HE, 2010), we forced the IDEA worksheet to place the low-visibility objects (green mannequins and brown gloves) to a maximum distance of 1 X AMDR rather than the usual 1.5 X AMDR to avoid placing these objects too far from the course to obtain crossover plots (Koester RJ *et al.*, 2014). We also made use of the post-experiment confirmation box in the search object placement worksheet to jot down a short subjective description of the vegetation surrounding each search object.

Deriving Effective Sweep Width

In addition to determining individual teams' crossover *W* values as previously described (Chiacchia KB, Houlahan HE, 2010; Chiacchia KB *et al.*, 2015), we determined PODs for given distances in the following manner. In the random-search model:

2. $POD = 1 - e^{-C}$

where C, or coverage, is defined as:

3. $C = Wd/A$

W being effective sweep width for a particular detector (or group of detectors) and search object under given conditions, d the length of the detector's path within the search segment, and A the area of that segment.

Each lateral approach of a detector past a search object can be considered as an area defined by:

4. $A = 2Ld$

where L is the lateral distance at the detector's closest approach to the search object, and again d is the length of the detector's track. We use 2L because *W* is measured from the detector's right and left as it moves through the search area. When a DO occurs, then, its lateral distance thus defines a coverage for that DO and lateral distance. A POD can be determined for each value of C, then, by cumulating any detections between the detector's path and that value of L and dividing them by the number of detections divided by detections plus misses within that L (See Figure 4, modified from Cooper DC *et al.*, 2003).

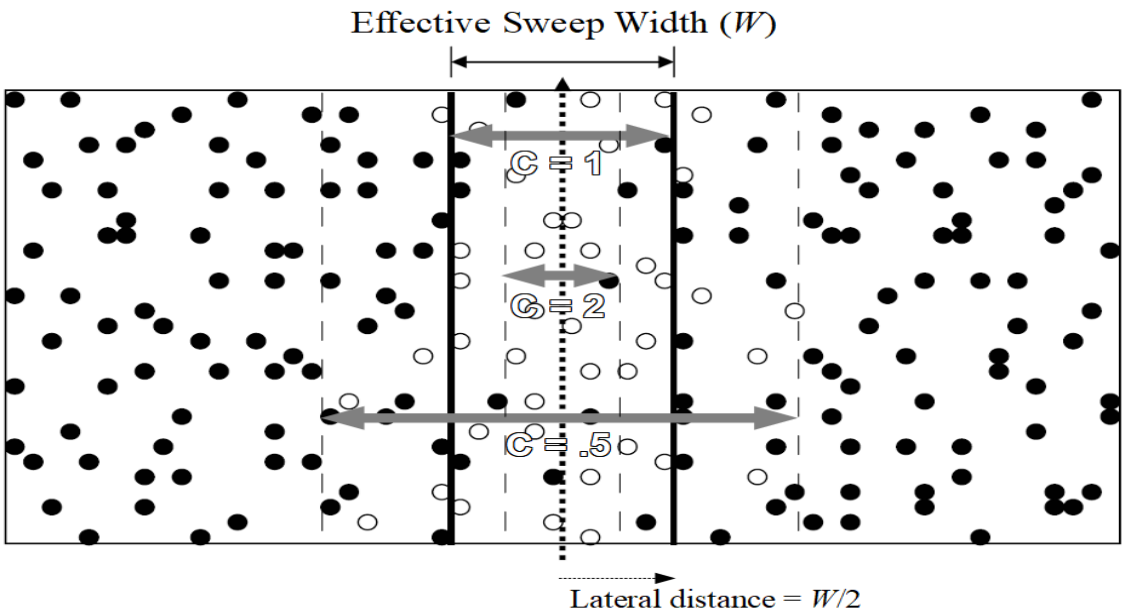


Figure 4: Determining coverage for individual detection opportunities (DOs) using the envelope defined by lateral distance (L). Black circles represent misses, open circles detections. Gray arrows and dashed lines show the envelope that would define coverages of 0.5, 1, and 2, each defined as 2L/W; the dotted line/arrow shows the detector's path.

For each pass, and substituting equation 4 into equation 3 and then 3 into 2, the POD equation, then, becomes:

5. $POD = 1 - e^{-(W/2L)}$

We want to derive *W*, which we do not yet have. Since the coverage for a detector with an arbitrary *W* of 50m at a lateral distance L would be:

6. $C_{50} = 50m/2L$

for a detector of unknown *W* we can then substitute:

7. $POD = 1 - e^{-[(50m/50m)(W/2L)]}$

to get:

8. $POD = 1 - e^{-(C_{50}W/50m)}$

Thus, by plotting POD for a given detector and group of DOs against C_{50} — again, the coverage at the distance of a given DO assuming a *W* of 50m — we can derive the detector's actual *W* as a constant from a least-square fit of the data against equation 8.

W may be derived in the same way using different models for the relationship between POD and *W*, such as the "inverse-cube" model (Koopman BO, 1980):

9. $POD = erf[(\sqrt{\pi})C/2]$

where erf is the statistical error function and C is coverage as above (also see Results and Discussion).

The question of how far from the detector's path to begin counting DOs (*i.e.*, how high a C_{50} to include in the curve) is important, as the small numbers of DOs close to the detector would result in high noise of POD due to low N (the noise in Figure 2 stems from the same source), and thus the points in this part of the graph will have a high degree of variance. At low C_{50} we have a different problem: Because of the finite size of the search areas in the study, longer Ls will become sparse and, because we are cumulating the POD from the detector's path outward, if there is a gap between two DOs/Ls we will have the same decrease in measured POD as if the DOs were close together. We would thus expect the data points to artifactually begin to "lift off" of the curve at low C_{50}. To partially ameliorate both problems, we selected an arbitrary number of DOs both from the detector outward and from the longest DO inward before beginning to count and tally detections and misses for the POD determination.

The issue is how best to select that arbitrary number. In the current study, we leveraged the limitations of the human visual sweep width data. Our second such experiment, in summer 2006, involved 18 human searchers, the smallest number in our visual-search datasets (Chiacchia KB, Houlahan HE, 2010). Because DOs in the human experiments are number of searchers multiplied by number of search objects, setting the minimum number of DOs below 18 would have meant discarding results from two each of the 23-29 search objects of each type placed in that experiment — a major loss of data. Thus, we set the minimum number of DOs to begin calculating POD both from the detector outward and the longest DO inward at 18 for both canine and human data.

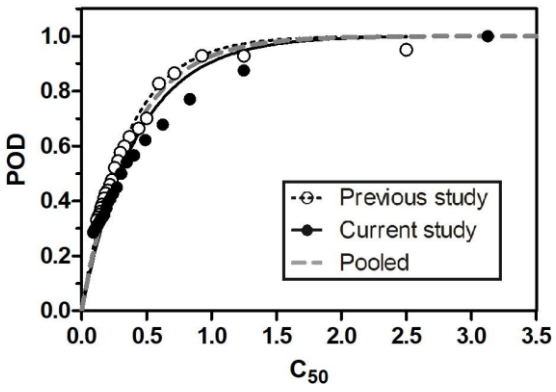
Statistical Methods

The current report employed GraphPad Prism version 5.00 for Windows (GraphPad Software, San Diego, CA, USA, www.graphpad.com) to perform statistical testing. Data are expressed as mean±standard deviation (SD). All tests were two-tailed, with $P<.05$ set as the threshold for statistical significance. We used Microsoft Excel and Apache OpenOffice to create a worksheet that performed simple calculations, including correcting P-values for repeated testing using the step-down Hochberg/Šidák method.

Results

Air-scent Dog Teams

Of the 167 included tasks in the new data, 10 resulted in no find before intentional unblinding, and a further 7 resulted in refind failure (*i.e.*, the dog found but did not tell the handler and so it was counted as a miss). The resulting PODs are plotted against C_{50} in Figure 5a. We also plotted the data from our previous report (Chiacchia KB *et al.*, 2015). For clarity, Figure 5a only shows one point out of every six, though all points were included in this and subsequent analyses.



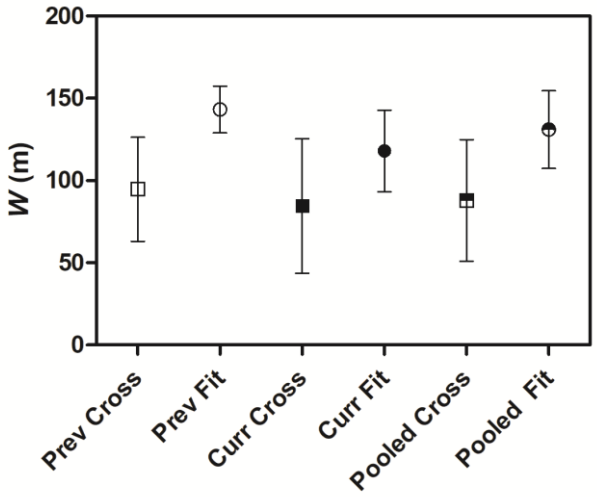


Figure 5a (left): Plots of C_{50} (coverage assuming W *of 50m) versus probability of detection (POD) and the resulting least-square fits of the air-scent data to the random search model. Data shown are from the previous study (Chiacchia KB et al., 2015), open circles and dotted line; the current study, black circles and solid line; and the fit to both sets of points, thick gray dashed line. For clarity, only every sixth point is pictured.*

5b (right): W *values for the previous, new, and combined data. Open symbols are the previous study; black symbols current; and half-black, half open combined data. Squares are* W *determined by the crossover method; circles are from the fits in 5a. Shown are* W±*SD.*

The result of the least-square plot is that *W* for the new data is 118±25m, with an R-square of .895 and an N, after excluding points at the upper and lower part of the graph as described in Methods, of 138. These 138 points in the graph are underlain by 526 included DOs (with the POD at the lowest C_{50} value calculated from the full 526 DOs, and the highest from 18 DOs). The data from the previous report returned a *W* of 143±14m, R-square .978, N=139, and 320 DOs. An extra sum-of-squares F-test produced P<.0001 (P-value corrected for repeated testing <.005) for the null hypothesis that the resulting curves share a common *W* value, despite the very small difference in POD between the two curves at all points (on the order of ≤~5%).

One factor to be kept in mind is that the 50% POD level may be a "tipping point" for shifting search efforts to new areas, when using shifting POAs (see Introduction and Charnes A, Cooper WW, 1958). Thus it is useful to observe the difference in C_{50} at the curves' 50% POD level: as $C=Wd/A$ (Equation 3) and $d=vt$ (the detector's velocity multiplied by time searching in a given area), differences in C_{50} in these curves will be proportional to the time needed to reach a given level of POD.

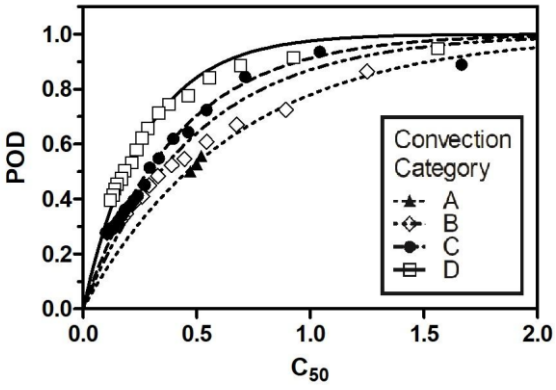
A Wolfowitz runs test for deviation from the random-search model was significant, with P<.0001 (corrected to <.004 for the current data, <.005 for the previous curves; see Discussion for more on this). Because of the small POD difference between the two datasets, we also did a fit of the two sets of points pooled together. The result was *W*=131±24m, R-square .926, N=268, and 846 DOs.

Figure 5b shows the resulting "fit" *W* values compared with those acquired from averaging crossover plots of individual dog/handler teams for the current, previous, and pooled data. The latter were, respectively, 85±41m, 95±32m, and 88±40m, with Ns of 8 (because not all 10 teams in this group had sufficient data for a successful crossover plot), 4, and 12 (because for the purpose of this last comparison dog B's data were combined, instead of divided between the previous and current studies). These values were smaller than those derived from the least-square fits, and the uncertainties of the mean *W* values were higher, phenomena examined below.

The unlikelihood that a POD difference of ~5% will ever be operationally significant (See Discussion) and the excellent R-square value of the pooled POD data against C_{50} (and thus *W*) led the authors to pool the data between the two datasets in subsequent comparisons, to achieve higher statistical power.

Atmospheric convection proved a significant correlate with *W* in the authors' previous report (Chiacchia KB *et al.*, 2015; the rationale being that convection causes turbulence that dilutes scent, see Graham H, 1994). Therefore, we addressed this issue again, plotting POD for the pooled data against C_{50} in the different daytime convection conditions (Figure 6a) — A through D, with A representing the strongest convection and most hypothesized disruption of scent transport (*i.e.*, lower *W*). Again, for clarity, only every sixth point is shown for conditions B through D in Figure 6a, though all were included in the analysis. All points are shown for condition A.

In Figure 6a and 6b, we see a correlation of stronger convection with poorer search performance. *W*s for convection conditions A, B, C, and D were, respectively, 75±3m, 102±21m, 121±12m, and 184±18m, with R-squares of .856, .908, .980, and .971. Note that the low SD for the A value is likely an artifact of the scant data for that curve; Ns for each condition were, respectively, 4, 127, 119, and 100, with 20, 353, 253, and 204 DOs. An ANOVA of the parameters generated by the fit returned P<.0001 (corrected <.002); a Tukey's multiple comparison post-test showed P<.0001 for differences between each convection condition except A vs. B, which was P<.05. The fit for convection conditions B through D displayed a significant runs test (P<.0001, corrected to <.004 for each) but the fit for A did not (P=.667, corrected .963).



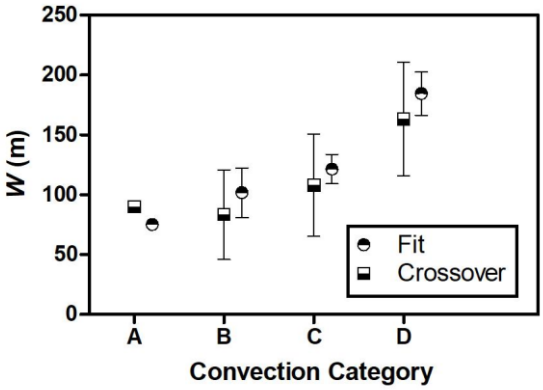


Figure 6a (left): Plots of C_{50} versus POD for the pooled air-scent data at the four measured convection conditions (A, strongest convection, D, weakest). For clarity, only every sixth point is pictured except for A, in which all points are shown.

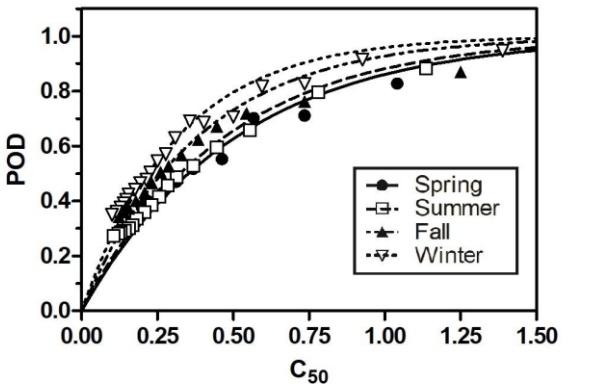
6b (right): W *values at the different convection conditions. Squares are W determined by the crossover method; circles are from the fits in 6a. Shown are* W±*SD.*

Also shown in Figure 6b are the pooled data from these comparisons as calculated via crossover plots. Again, the crossover *W*s were both smaller and had larger errors than the fit values: respectively, 90m, 83±37m, 108±43m, and 163±47m. Ns were 1, 10, 7, and 9, respectively. The A-convection crossover value has no SD because only one of the teams possessed enough data in these conditions to generate a crossover plot; for its fit value, the SD was smaller than the size of the circular symbol. Because of the additional data, the crossover values in this figure differ slightly from those in the otherwise equivalent Figure 6 in our earlier report (Chiacchia KB *et al.*, 2015).

Human visual-searcher *W* values correlate strongly with season (Chiacchia KB, Houlahan HE, 2010). Thus, any relationship between air-scent dog performance and season would be useful for comparing the two search modalities. The authors' previous report did not detect such a relationship for air-scent dog teams (Chiacchia KB *et al.*, 2015), but the smaller SDs of the fit *W* values and the increased Ns of the pooled data encouraged us to make the comparison with the expanded data set.

The result of that comparison can be seen in Figures 7a and b. Again, in 7a only every sixth point is shown. The resulting *W*s for summer, fall, winter, and spring were 106±15m, 131±19m, 158±21m, and 98±11m;

R-square .956, .912, .949, and .961; N=105, 84, 118, and 59; and 236, 172, 225, and 143 DOs. A Tukey's post-test of an ANOVA of the parameters (P<.0001, corrected <.003) returned P<.001 for all pairwise comparisons except for spring versus summer, which was P<.05. The difference between the high and low curves (winter versus spring) approached the ~20% level at C_{50} values of approximately .5. (Note this would relate to an actual C of around 1 or higher using the specific *W* values in each season.) Runs tests were significant for summer and fall, P<.0001 corrected to <.004 for each, but spring lost its significance upon repeated-measures correction (P=.003, corrected .067); neither was the winter runs test significant (P=.361, corrected .932).



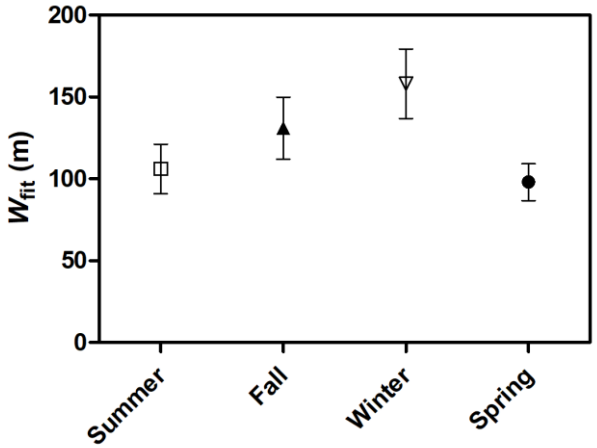
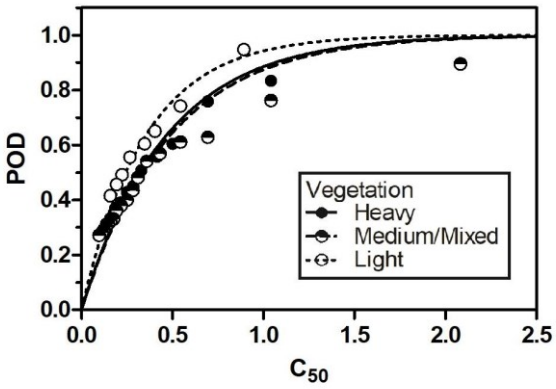


Figure 7a (left): Plots of C_{50} versus POD for the pooled air-scent data in the different seasons. For clarity, only every sixth point is pictured.

7b (right): W *values in the different seasons from the fits in 7a. Shown are* W±*SD.*

The authors previously detected a difference among *W*s in human visual search for search objects randomized to areas subjectively judged as heavy versus light vegetation (Chiacchia KB, Houlahan HE, 2010). For another possible correlate of air-scent performance that allows apples-to-apples comparison with human searchers, we investigated the relationship between POD and C_{50} (and thus *W*) versus subjective vegetation thickness (see Methods), shown in Figures 8a and b (again, only every sixth point shown for 8a). As this information was not collected in our earlier study, this comparison relied only on the new data. For heavy, medium/mixed, and light vegetation, the resulting *W*s were respectively 108±13m,

103±23m, and 142±17m; R-square .939, .866, .918; N=67, 107, and 45; and 127, 78, and 96 DOs. In a Tukey post-test of an ANOVA (P<.0001, corrected <.003), the light-vegetation *W* was significantly larger than each of the others (P<.001), but the heavy and medium/mixed values were not significantly different. The runs tests lost their significance upon correction (heavy vegetation P=.002 corrected .052; medium/mixed P=.005 corrected .088; light P=.031 corrected .331).



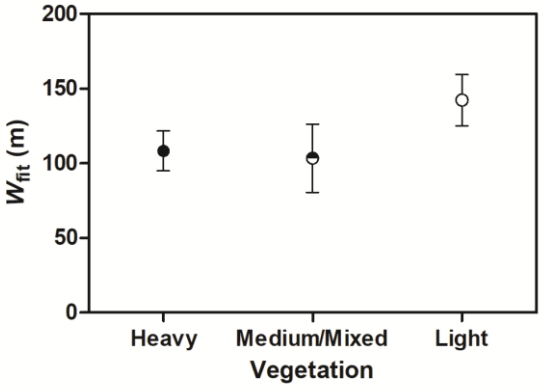


Figure 8a (left): Plots of C_{50} versus POD for the pooled air-scent data in different ground-level vegetation thickness scored subjectively (heavy, light, and either medium or a mix of heavy and light). For clarity, only every sixth point is pictured.

8b (right): W *values in different vegetation thickness from the fits in 8a. Shown are* W±*SD.*

Another variable that seems likely to be relevant to air-scent dog team performance is the U.S. EPA ecoregion in which the task takes place. Ecoregions differ by climate, elevation, vegetation, and terrain in ways that will likely impact both scent transport and handlers' choices of search tactics. Thus we compared POD curves and resulting *W*s for the ecoregions in which our study tasks were undertaken — which again were arbitrarily chosen and not equally distributed. The largest such number of tasks (199) was in Ecoregion 70, the Western Allegheny Plateau. For the next region, we pooled 67 tasks in Ecoregions 67 and 69, respectively Ridge and Valley (specifically 67c, Northern Sandstone Ridges) and Central Appalachians, which are subjectively similar rugged-terrain regions. Finally, a smaller number of tasks (13) took place in Ecoregion 61, Erie/Ontario Drift and Lake Plains. (See Figure 3 for the map of tasks and ecoregions.)

In Figures 9a and b we see the results (every sixth point shown in 9a except for Ecoregion 61, in which all are shown), with *W*s for Ecoregions 70, 67/69, and 61, respectively, of 130±26m, 120±17m, 90±8m; R-square .923, .942, .904; N 199, 67, and 13; and 677, 123, and 35 DOs. A Tukey post-test of an ANOVA (P<.0001, corrected <.003) gave P<.001 for 70 versus 61 and 67/69 versus 61 and P<.01 for 70 versus 67/69. Runs test P-values were <.0001 corrected <.003 for Ecoregion 70 and Ecoregions 67/69, and .003 corrected .054 for Ecoregion 61.

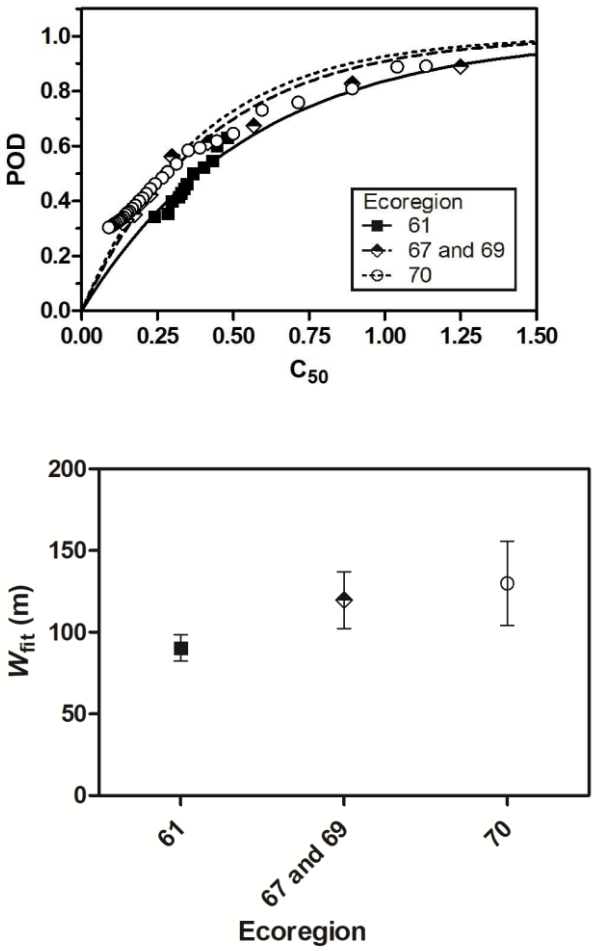


Figure 9a (left): Plots of C_{50} versus POD for the pooled air-scent data in the different U.S. EPA ecoregions. For clarity, only every sixth point is pictured for Ecoregions 70 and 67-and-69; all points are shown for Ecoregion 61.

9b (right): W *values in the Ecoregions from the fits in 9a. Shown are* W±*SD.*

Another variable of interest in the dog-handler community is that of humidity, which arguably could increase scent via bacterial metabolism creating more odorants. As the authors did not collect humidity data in the field, this comparison required the use of historical weather data. As the authors also did not in their initial report retain start and end times for each task once sun angles/convection were calculated, this comparison also was limited to the new data. We separated the data based on three humidity ranges of as similar number of DOs as possible, to obtain the best separation between the different conditions while maximizing the number of DOs for each.

Figures 10a and b show a comparison of the POD curves and the resulting *W* values, respectively. For the 20-49% low-humidity range, *W* was 114±20m; mid-humidity >49-67%, 106±21m; and high-humidity 68-100%, 113±16m. An ANOVA of the difference between these values returned P=.030, corrected .346 (R-square for each fit .905, .851, and .928, respectively; N 72, 68, and 81; DOs 156, 157, and 162). Runs test P-values were <.0001, corrected to <.004 for each curve.

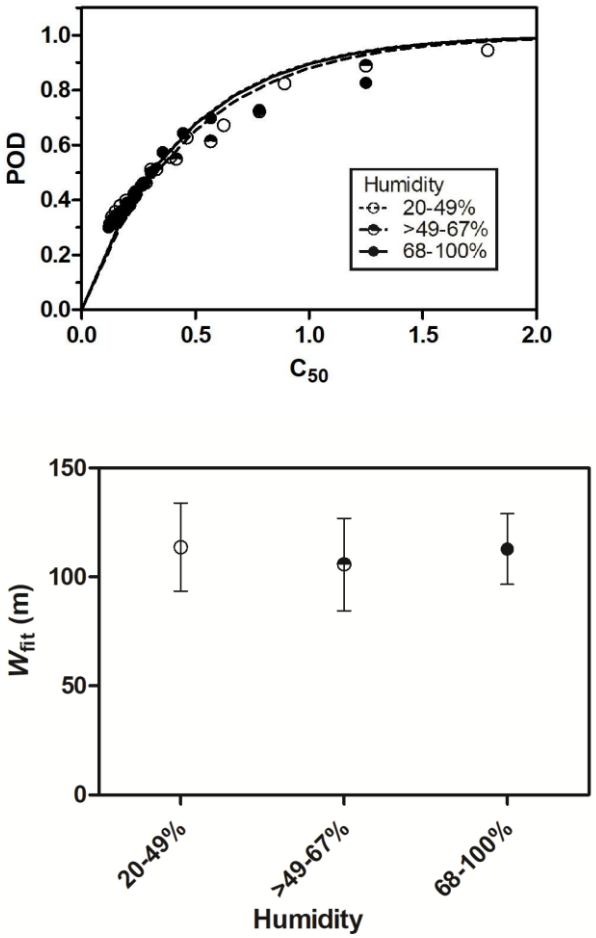
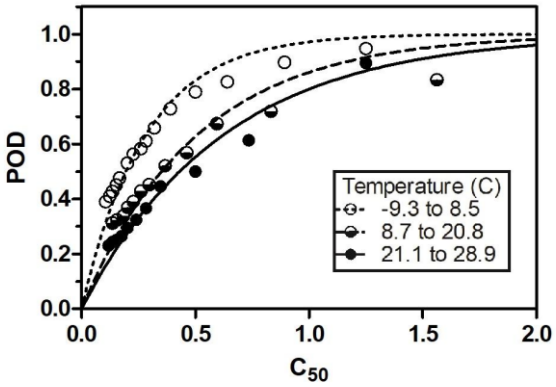


Figure 10a (left): Plots of C_{50} versus POD for the pooled air-scent data at different levels of humidity. For clarity, only every sixth point is pictured.

10b (right): W *values at the humidity levels from the fits in 10a. Shown are* W±*SD.*

Temperature is a matter of great concern to dog handlers, dogs being much more vulnerable to heat illness than humans. Subjectively, handlers note that dogs are less comfortable and may work less effectively at high temperatures. For that reason, temperature is another promising factor for determining *W* for dog teams. As with humidity, temperatures were determined from historical records as described in Methods. The authors compared POD curves and resulting *W* values from the fits in Figures 11a and 11b, respectively, again dividing the tasks into 3 temperature spans of as roughly equal size as possible in DOs (once again only every fifth point pictured in 11a).



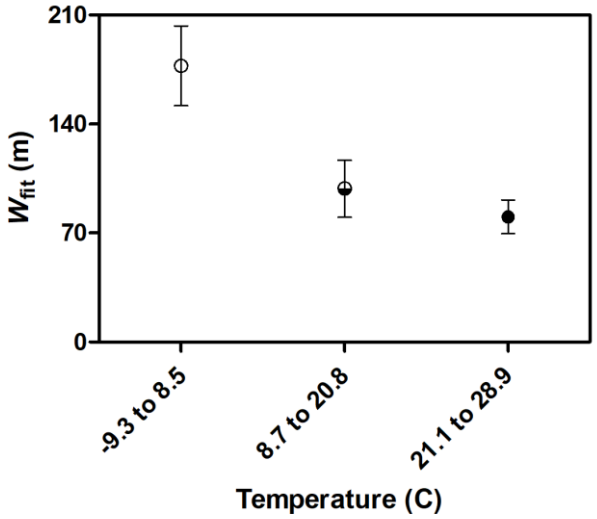


Figure 11a (left): Plots of C_{50} versus POD for the pooled air-scent data at different temperature ranges (C). For clarity, only every sixth point is pictured.

11b (right): W *values at the temperature ranges from the fits in 11a. Shown are* W±*SD.*

As can be seen in Figure 11a, the POD separation between the highest-POD (lowest T) and lowest-POD (highest T) curves is fairly large, on the order of the separation seen for convection in Fig. 4a. For the lowest temperatures, -9.3 to 8.5ºC, *W* was 177±26m (R-square .930; N 86; DOs 159). For the medium range, 8.7 to 20.8ºC, *W* was 99±18m (R-square .868; N 70; DOs 153). For the highest temperatures, 21.1 to 28.9ºC, *W* was 80±11m (R-square .955; N 62; DOs 169). A Tukey post-test of an ANOVA (P<.0001, corrected <.003) revealed significance comparing all of the temperature ranges against each other (P<.001). Runs test for the low-temperature curve was P<.0001 corrected to <.004, the others <.0001 corrected <.003.

Human Visual Searchers and Combined Analysis

The authors applied the curve-fit method for deriving *W* both to data from our previous report for visual human searchers at a single location on the Western Allegheny Plateau (Chiacchia KB, Houlahan HE, 2010) in the summer and winter as well as new data collected at that same location in the summer and a location in the spring in Ecoregion 64, Northern Piedmont, in Mount Gretna, Pennsylvania.

Figures 12a through c plots the POD versus C_{50} data and fits for the human experiments. Figure 12d compares the *W*s derived from these curve fits versus crossover method. *W,* SD, R-square, and runs-test P-values and corrected P-values for each can also be seen in Table 1 (the table also provides the R-squares for an inverse-cube-model fit to the data, see below). Note that for two of the fits, P-value corrections removed the significance of the runs test (the orange glove in the SGL 203 experiment P=.025 corrected to .316; the low-vis mannequin in the Mount Gretna experiment P=.012 corrected to .195); the others were not significant before correction.

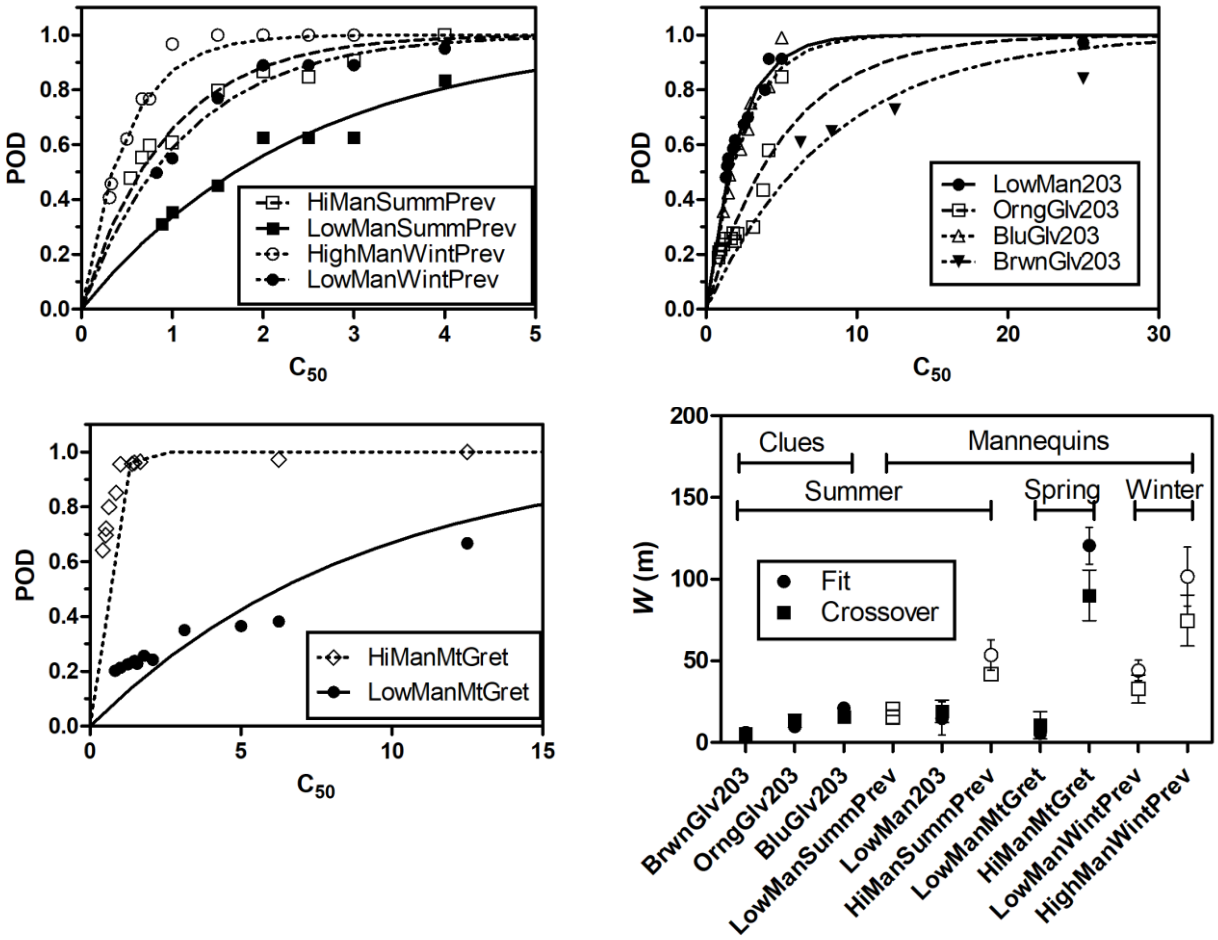


Figure 12a (top left): Plots of C_{50} versus POD for the previously reported human visual search values. HiManSummPrev: the high-visibility (white and blaze orange) mannequin in the previous summer State Game Lands 203 experiment; LowManSummPrev: the low-visibility (olive drab) mannequin in the summer experiment; HighManWintPrev: the high-vis mannequin at the winter 203 experiment; LowManWintPrev: the low-vis mannequin in the winter experiment.

12b:(top right): Plots of C_{50} versus POD for the newly reported summer SGL 203 experiment. LowMan203: the low-vis mannequin in the new 203 experiment; OrngGlv203: the orange glove at that experiment; BluGlv203: the blue glove at that experiment; BrwnGlv203: the brown glove at that experiment.

12c (bottom left): Plots of C_{50} versus POD for the spring Mount Gretna experiment. HiManMtGret: the high-vis mannequin at that experiment; LowManMtGret: the low-vis mannequin at that experiment.

12d (bottom right): W values for detection of different search objects by human visual searchers at different sites, in different seasons. Values from previous studies are open symbols, the current study black; crossover-method values are squares, random-search-fit values circles.

Table 1: Fit (random-search model) versus crossover W *values for human visual searchers and R-square for inverse-cube-model fits (*W *and SD in meters).*

	Low-Vis Mannequin Summer SGL 203			Orange Glove Summer SGL 203			Blue Glove Summer SGL 203			Brown Glove Summer SGL 203			High-Vis Mannequin Spring Mt. Gretna			Low-Vis Mannequin Spring Mt. Gretna			High-Vis Mannequin Summer Previous Study			Low-Vis Mannequin Summer Previous Study			High-Vis Mannequin Winter Previous Study			Low-Vis Mannequin Winter Previous Study		
	W	SD	N	*W*	SD	N	*W*	SD	N	*W*	SD	N	*W*	SD	N	*W*	SD	N	*W*	SD	N	*W*	SD	N	*W*	SD	N	*W*	SD	N
Random-search fit value	14	10	14	10	4	15	21	3	12	6	3	7	120	11	12	6	2	12	54	9	10	21	3	8	101	18	11	44	6	8
Crossover value	19	7	24	13	4	24	16	3	24	5	2	24	90	15	26	11	8	26	42	4	18	15	4	18	75	16	20	33	9	20
R-square of fit	.975			.868			.966			.264			.973			.555			.933			.939			.964			.956		
P for runs test of fit (corrected)	.070 (.581)			.025 (.316)			.643 (.984)			.130 (.752)			.833 (.833)			.012 (.195)			.071 (.555)			.800 (.960)			.191 (.852)			.200 (.832)		
R-square of inverse cube fit	0.913			0.878			0.973			Not converged			0.938			0.290			0.794			0.864			0.991			0.928		

To further investigate the relationship between the fit and crossover values for *W*, we plotted the crossover versus fit *W* values for the new and previous dog data in Figure 5 as well as all the human *W*s in Figure 12. That plot can be seen in Figure 13. The apparent proportionality of the crossover and fit *W*s led us to carry out a linear least-square fit of the data points, with the Y intercept constrained to 0. We obtained a slope of 1.44±.24, R-square .904. Note, however, that while the R-square appears to suggest reasonable predictability in the relationship, the runs test showed P<.0001 (corrected <.002), indicating that the relationship only approximates linearity.

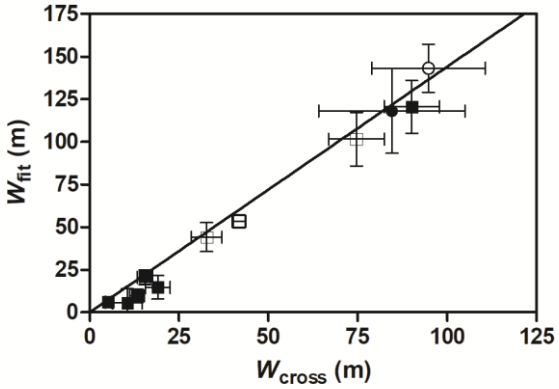


Figure 13: Plot of the crossover values of W *(*W_{cross}*) versus the random-search-fit values (*W_{fit}*) and the resulting least-square-fit line. Previous studies are open symbols; current are black. Human visual searcher values are squares, air-scent dog teams circles.*

While the SDs for the fit *W*s were smaller than those for the crossover values in the dog data, this was no longer apparent for the human data. In an attempt to clarify this issue, we conducted a paired, nonparametric Wilcoxon signed rank test of the fit versus the crossover SD values for the new and previous dog *W*s in Figure 5 as well as all the human *W*s in Figure 12, a comparison that can be seen in Figure 14. The result was not significant (P=.339, corrected .945), though again the dog SDs (circles in Figure 14) do subjectively appear to be higher for the crossover values.

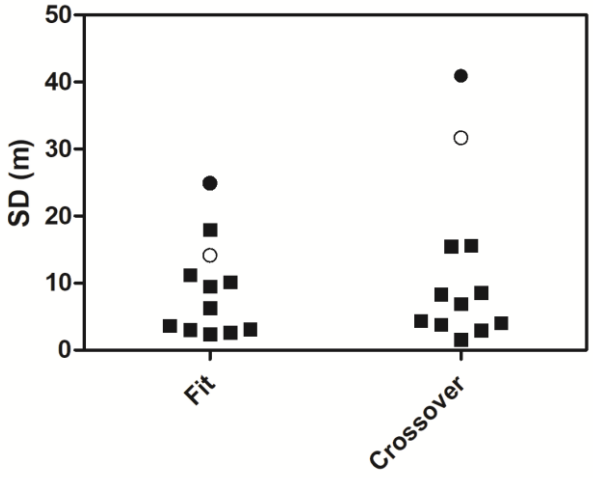


Figure 14: SD values for the random-search-fit W*s versus the crossover* W*s. Human values from all experiments are black squares; air-scent values are circles, with the previous study open and the current study black.*

Appropriateness of Random-Search Model for Ground-Based Search Efforts

The data for the air-scent dog teams as well as for the human searchers fit well to the random-search model in terms of R-square values, showing R-squares mostly >.8. Nonetheless, judging by runs tests, the model often fails (see above). In addition, fits to this model return a *W* value that is larger than the crossover value (by 1.44-fold, Figure 13 and above). This is curious given that the latter method defines *W* (Koopman BO, 1980; Cooper DC *et al.*, 2003; Koester RJ *et al.*, 2004), suggesting that the model may systematically underestimate POD and thus the fit must generate an inflated *W* to fit the data. For this reason, we compared the fit provided by the random-search model with that of another, the inverse-cube model (Koopman BO, 1980).

Figure 15 shows a plot of the new and previous dog POD data, each plotted against C_{50} and separately fit against each model. As reported above, the values for *W* using the random-search model were, respectively, 118±25m, R-square .895; and 143±14m, R-square .978. Applying the inverse-cube model, we find corresponding *W* values of 90±26m, R-square .741; and 107±19m, R-square .907. As with the random-search model, in both cases the inverse-cube model failed a runs test for deviation from the model (each P<.0001, corrected <.003). By comparison, the crossover values for each air-scent data set were 85±41m and 95±32m.

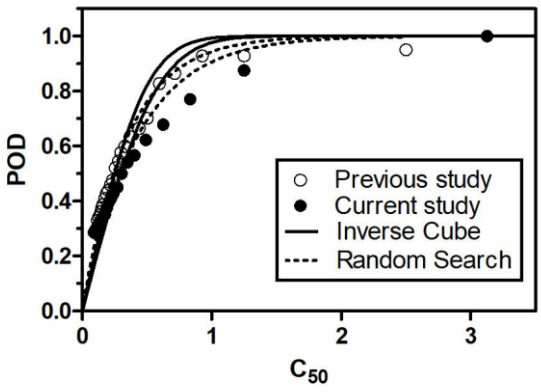


Figure 15: Plot of C_{50} versus POD for the previous (open circles) and current (black circles) air-scent data, fitting each separately to either the inverse-cube (solid line) or random-search (dotted line) models. For clarity, only every sixth point is shown.

Applying an Akaike's information criteria (AICc) test to select the model most likely to have generated the data, we find the results shown in Table 2. (The inverse-cube fit against the brown glove at the SGL 203 experiment did not converge and so that search object was not included.) The mean value for the probability that the random-search model generated the data is >75±37%, compared with the inverse-cube model.

Table 2: AICc test results for the random-search versus inverse-cube models against each dataset

Search object/experiment	*Probability of model generating data*	
	Random Search	*Inverse Cube*
Low-Vis Mannequin SGL 203	>99.9%	<.1%
Orange Glove SGL 203	35.8%	64.3%
Blue Glove SGL 203	22.1%	77.9%
Brown Glove SGL 203	n/a	Not converged
High-Vis Mannequin Mt. Gretna	98.9%	1.1%
Low-Vis Mannequin Mt. Gretna	92.9%	7.1%
High-Vis Mannequin Summer SGL 203 Previous Study	99.4%	0.6%
Low-Vis Mannequin Summer SGL 203 Previous Study	94.2%	5.8%
High-Vis Mannequin Winter SGL 203 Previous Study	0.1%	99.9%
Low-Vis Mannequin Winter SGL 203 Previous Study	84.5%	15.5%
Dog Teams Previous Study	>99.99%	<.01%
Dog Teams Current Study	>99.99%	<.01%
Mean	**>75±37%**	**<25±37%**

Discussion

The authors' intent in this report was to establish a new means of deriving *W* that offers an explicit link to field-derived POD values, smaller uncertainties, stronger statistical inference enabling comparison of more effectors of search efficacy, and the ability to distinguish between competing models for deriving POD from *W*. While our analysis accomplished most of these goals, the instances in which it did not are interesting and worthy of further study (see below).

The reader will note that, in the dog data above especially, the data points tend to sag below the fit lines at high C_{50} values. This phenomenon may indeed reflect a shortcoming of both models when applied to air-scent dog teams. On the other hand, the points here derive from a relatively small number of DOs (<~50), and thus have a noise level similar to that of the points in a lateral range curve (Figure 2). The reader should also bear in mind that there are many more points toward the low-C_{50} side of the graph, and so these low-C_{50} values (which themselves are based on high DOs and so are likely to be less affected by low-N variance) dominate the fit statistically. Another factor to be considered is that this difference may not be operationally significant as it lies in the ≤~10% range throughout the curves (see below for more on this). Ultimately, when using the shifting POA method for resource allocation, the critical point is at a POD of ~50% — that is when we would mathematically expect the method to begin strongly shifting future efforts from current high-POA areas to new areas (Charnes A, Cooper WW, 1958). Importantly, the data at this POD do not deviate widely from the curves in terms of absolute percentages in any of the fits above.

One important lesson offered by the POD curves comes from comparing the curves at the 50% POD mark, as the time required to reach that level of POD for a given detector, search object, and set of conditions is proportional to C_{50} (C being a function of d which in turn is a function of time spent searching, see Equation 3 and Results above). Thus, comparing the C_{50} of each curve allows us to see the concrete impact of that difference on time necessary to reach the arguable "tipping point" for shifting POAs.

For air-scent dog teams, the previously reported data (Chiacchia KB *et al.*, 2015) and those from the current study are in general agreement, as judged by the small (≤~5%) difference in PODs along the C_{50}/POD curve. An F-test comparing the two datasets did return a significant difference between the resulting *W*s at the chosen $P<0.05$ level. The reader can find a discussion of the difference between statistical and operational significance below; but the small POD difference convinces the authors that the operational significance in this case would likely be nil (*i.e.*, we suspect that search managers would seldom if ever find a POD difference of less than 5% to be actionable). More broadly, the POD-curve method for deriving *W* appears to provide good predictive agreement with the data and possibly smaller SD values, at least for the dog data (though see below re. the latter).

The authors used ANOVA and other tests associated with the POD-curve fits to gauge the significance of various effectors of search efficacy either reported previously (Koester RJ *et al.*, 2004; Chiacchia KB,

Houlahan HE, 2010; Chiacchia KB *et al.*, 2015) or subjectively believed to be important among operational dog handlers. Of these, and for the dog teams, the three most effective comparisons (*i.e.*, providing the largest significant difference between the compared groups) were, in decreasing magnitude of difference, atmospheric convection, temperature, and season.

In a perfect world, the authors would recommend using convection to track search efficacy operationally for air-scent dog teams precisely because it offers the largest POD separation between the comparison groups. In addition, these measurements seem more likely to apply to disparate climates and terrains than, for example, the seasonal measurements. However, few operational dog handlers formally track cloud, wind, and sun angle differences through a task. While our own experience shows that such measurements are indeed possible using a minimum of extra equipment (thanks largely to the efforts of meteorologists in the wildland firefighting service — Lavdas LG, 1976; Lavdas LG, 1997) and the ubiquity of smartphone apps for determining sun-altitude angle, expanding this practice to the larger community may be a heavy lift among dog handlers already trying to formulate search tactics, read moment-by-moment wind direction, interpret dog behavior, and navigate through an unfamiliar area.

Perhaps a bigger limitation of the use of convection is it can only be done retrospectively; weather forecasts are neither accurate nor granular enough to capture the wind and cloud conditions likely to be encountered on a given search task. While dog handlers can report convection conditions on debriefing at the end of a task — information that is helpful for retrospective analysis of search efforts and thus planning of future tasks — they are not useful for prospective planning (*i.e.*, search managers telling dog handlers what level of POD they would like to achieve in a given task and thus selection of tactics to meet that level).

Because of this, temperature may be an attractive alternative to convection. Indeed, convection drives daytime air temperature (either locally or in generating the temperature of winds accompanying weather fronts) and thus the two are interrelated. Temperature data are much more simply collected. Unlike convection conditions, temperature is more stable over time and likely to be comparable between the command post and the usually close-by area of a given task, making it unnecessary for field teams to collect the data. In addition, temperature implicitly collects another factor subjectively believed by handlers to be important: The dog's comfort level, as dogs are clearly more animated and comfortable when the temperature is not high. However, to some extent the inaccuracy of weather forecasts (particularly with respect to arrival of fronts that change the temperature quickly) limit temperature as a prospective planning tool just as it does with convection.

For prospective planning, the seasonal differences are thus likely to be the most useful among those reported here. They provide an easily derived value for *W* that search planners and dog handlers alike can use to make prospective decisions (albeit possibly only relevant to the U.S. Northeast/Appalachian region in which they were derived and areas climatically similar — see the limitations section below).

Just as temperature and convection are tightly related, so are convection, temperature, and season. Indeed, the likely drivers of seasonal differences in *W* are numerous, including convection conditions (especially

driven by daytime sun angles), temperature, vegetation density differences due to seasonal leaf shedding/die back, *etc.* Just as importantly, as seasonal differences have proved to be a major determinant for human visual searcher efficacy (Chiacchia KB, Houlahan HE, 2010, as well as the current human data), use of this factor for deriving air-scent *W* values provides a single, apples-to-apples comparison with human search efforts, simplifying search managers' assessments.

The true significance of the other effectors is harder to judge. Most of them returned significant P-values. However, the smaller POD differences (≤~10% through the curves) seen with vegetation thickness and ecoregion are of indeterminate *operational,* as opposed to statistical, significance. Of course, the ecoregions sampled in this study are fairly similar subjectively; repeating the measurements in more disparate environments (*e.g.*, high-altitude mountains, warmer marshy areas, desert areas) might identify more clearly significant differences. The humidity conditions, on the other hand, correlated with no significant difference in *W* upon repeated-measures correction. This was surprising, at least among the dog-handler community, in which humidity is subjectively considered to be important for scent generation (though see the Limitations section below).

As noted in the Introduction, one study suggested that navigational errors impose errors of ~10-20% in prospective POD determinations (Perkins D, 2018). Thus, it may be that, for prospective planning at least, differences of ≥~20% may be necessary to ensure operational significance. On the other hand, acquisition of GPS tracks of teams — possible, through smartphone apps, for every searcher in a human team even if sufficient numbers of dedicated GPS units are not available — allows accurate determination of retrospective PODs despite this limitation. Note that this issue will not affect the results reported here, as the human searchers were walking a set course and the dog handlers' tracks were determined by GPS.

One could argue, with some justification, that the statistical tests used here are over-sensitive for determining operational significance. This stems, perhaps, from the nonlinear relationship between POD — which is the quantity needed to manage search efforts — and *W* — a tool for deriving that quantity. Indeed, as can be seen in these results, an intuitively large absolute difference in *W*s can produce a surprisingly small difference along the resulting POD curves. Defining operational significance, then, will likely stem from practical imperative rather than statistical significance of *W* differences.

The results in applying the POD-curve method to human visual searcher data were largely in consonance with those for the dog data, with the exception that the POD-curve-derived *W*s did not have smaller SDs than those of the crossover values for the humans. One possible explanation for this is that, while obtaining a crossover *W* for each detector suffers from the large variance in *W* inherent in the small number of DOs for individual dog/handler teams or human searchers, the latter are represented by larger Ns — *i.e.*, similarly small DOs for each individual but more individuals contributing to the mean crossover *W* value. The previous and current dog data reported here rest on 4 and 10 teams, respectively, while the smallest of the human sweep-width experiments collected data from 18 searchers, and the largest 26 searchers. More

work will be necessary to determine if this is the reason for the difference, but at the very least the POD-curve method does not return *larger* SDs than the crossover method.

The curious result that the POD-curve fits to the random-search model provide larger *W* values than the (definitive) crossover method (as defined by slope>1 in the least-square plot in Figure 13), the authors believe, is tightly linked to the limitations of that model. Indeed, that model has been preferred for operational use because of its theoretical insensitivity to uneven search patterns as much as its fit to the data (Koopman BO, 1980; Charnes A, Cooper WW, 1958; Cooper DC *et al.*, 2003; Koester RJ *et al.*, 2004). In addition, though the inverse-cube model (derived from a hypothetical relationship between POD and the size of an object's image on a human retina) was primarily intended as a theoretical tool, it has been cited as nevertheless fitting well to field data in some cases (Koopman BO, 1980; Koester RJ, 2020)

The problem is that the inverse-cube model generates *W* values arguably closer to the definitive crossover values at the cost of a somewhat poorer fit to the data (reflected in its <25% score on the AICc test — smaller than that for the random-search model, though not small enough to rule it out). In particular it can overestimate POD values compared with the random-search model starting around a POD of ~60% (Figure 15). This is not far from the ~50% point at which, again, we would expect the shifting POA method to begin directing search efforts to new segments. Operationally this could be a problem when using inverse-cube, causing search efforts to be redeployed when further search in an area may be warranted.

Of course, the most likely explanation for these results is that neither model completely captures the phenomenon. The often-significant P-values for the runs tests above, for either model, further support that interpretation. Both models provide a close physical fit to the data, however, with R-square values >.8 in a majority of cases.

Smaller random-search R-square values of .264 and .555 obtained, respectively, from the fit for the brown glove at the second summer State Game Lands 203 experiment and the low-vis mannequin at the spring experiment in Mount Gretna (Table 1). This is interesting because these search objects both had extremely low *W* values (the latter is the smallest measured so far that the authors are aware of for an adult-sized mannequin). Relevant to the former, *W* values have proved difficult to obtain for low-vis clues in some cases (Chiacchia KB, Houlahan HE, 2010). Relevant to the latter, the green mannequins in that experiment had an unusually low value of *W* compared with their average maximum detection radius (AMDR, the mean distance at which a searcher is capable of seeing an item whose location the searcher is cued to, as opposed to a searcher who does not already know the location of an item, as in the *W* determinations; Koester RJ *et al.*, 2014). This apparent discrepancy between ability to *see* the search object and ability to *notice* it may be because, as reported by data loggers at that experiment, the large green rocks in that area appeared to cause the searchers to progressively ignore objects of that approximate size and color, with many close-range misses that the (cued) loggers thought should have been obvious detections. The phenomenon of attention affecting detections over and above simple visibility of an object has been documented in the literature (Koester RJ *et al.*, 2014; McClanahan S, 2021). Determining whether these

observations offer any kind of window into explaining the poorer fit for these two search objects will require more work.

Taking the results reported here *in toto*, the authors would recommend the larger *W* values produced by the POD-curve fit in concert with the random-search model for operational use over the canonical crossover *W* values with the inverse-cube model (or the fit-derived *W*s using that model) because of the former's marginally greater agreement with the data. The approximately linear relationship between the random-search-fit and crossover values for *W* is reassuring on this count. Indeed, that simple relationship could allow quick conversion of previously measured crossover *W*s for use with the random-search model. Alternatively, the IDEA spreadsheet for planning sweep-width experiments contains a "Data Summary Object" sheet that, with minimal reworking, can be converted into a calculator to derive C_{50} versus POD, allowing easy re-analysis of previous sweep width experiments to provide a curve to derive *W* using any statistical package allowing nonlinear least-square fits. On the other hand, either of these approaches (larger, fit-derived *W*s with random search versus canonical or fit *W*s with inverse cube) appears to provide high-R-square agreement with the data and thus could be used operationally.

The one combination the authors would *not* recommend is using crossover *W*s with the random-search model (arguably the most common practice currently), as this may systematically underestimate POD. While intuitively this might seem attractive — a deflated POD being more "conservative" by one way of thinking because it reduces, using shifting POAs, the chance of missing a subject in areas being searched — it would also tend to cause search efforts to linger in empty areas rather than expand to new areas that might contain the search subject (Charnes A, Cooper WW, 1958). The most conservative POD values are accurate ones, the operational derivation of which the authors hope to contribute to with this report.

One objection sometimes raised to determination of objective POD via effective sweep width is that, occasionally, individual search segments may vary significantly in character from the average surrounding environment and so would not provide comparable PODs. This is likely true; but just as outlier responses to pharmaceuticals do not make their use impossible, outlier *W* values will not prohibit operational use of sweep width. Granted, it is important to understand such outliers and when and how they occur to maximize accuracy of objective POD determination. Yet this argument, if taken to its logical conclusion, given the known unreliability of obtaining POD from subjective assessments (Koester RJ, 2004), is not against use of sweep width, but any use of POD at all. Ultimately, and if the PODs reported here were not reasonably consistent — and they do appear to be — this line of reasoning calls for a completely different approach to search management, which is yet to be elaborated.

Pulling back to a more theoretical level, one would not expect the same model to capture the relationship between visual acquisition of a search object and its cognitive detection by a human searcher, *as well as* the relationships between transport of odorants in a turbulent medium, their detection by a dog's olfactory epithelium, cognitive recognition of the signal, localization of the source by the dog in that turbulent medium, and the complex psychological factors in the dog/human relationship that produce a clear signal by the dog

that compels the handler to follow her to the search subject. In retrospect, the apparently detector-agnostic nature of the relationship between *W* and POD suggests to the authors that factors common to these detection methods may be rate-limiting to the search process, rather than the many differences. Future progress will be necessary to tease apart this very large question.

Limitations of the Current Study

The current study has several limitations that must be borne in mind. First, despite the authors' attempts to expand the datasets, pool data from multiple experiments, and leverage as much of the collected data as possible to derive the results above, the effort remains in the realm of "small data," and thus is vulnerable to outlier effects. Also, some of the P-values that lost their significance to repeated-measures correction may in fact represent real if weaker effects. A means for greatly expanding the data — perhaps via a smartphone application that collects data while generating a log entry for users at training events or deployments — might enable collection on a vastly larger scale. Application of the results to searchers and dog teams with different techniques or training protocols, at greater statistical robustness, and above all in areas with different climate and terrain may depend on such expanded data. The authors would be interested in working with groups or individuals capable of such software development.

Another limitation is that the relationships between the external factors measured and search efficacy are by necessity correlative. While these results help document robust correlations of weather, season, terrain, or vegetation with search results, they cannot establish causality and so must be interpreted with some caution. Uncontrolled factors and correlations between effectors that were not considered in the present analysis could bias results. In addition, the complex interrelationships between season, weather, and vegetation means that many of the factors investigated above are not independent and cannot be confidently aggregated — for example, what would *W* be for an air-scent team searching in convection condition C, in the spring, with an air temperature of 15ºC, in heavy vegetation, in ecoregion 70? The factors will almost certainly not combine in a linear fashion and so for the time being must be considered separately. One possible means past this limitation might be a Bayesian causal inference analysis (Gelman A, Meng X-L, 2004) — which, again, would require expanded data.

The limitations of measuring humidity via historical records at the nearest weather station may have played a role in obscuring any effect associated with this factor. Also, the cutoff points in the humidity categories, chosen to provide groups of similar DO counts, may not have captured transitions meaningful to air-scent detection. Along the same lines, while the freezing point of water would have been an attractive temperature cutoff for that comparison, the authors did not collect enough data below 0ºC to divide the data at that point. Also, the values derived from the POD curve for convection condition A and for ecoregion 61 rested on a very small number of DOs (20 and 35, respectively), which may not be sufficient for a reliable result (Koester RJ *et al.*, 2004).

The mannequins used for the human visual experiments are similar in visual cross-section to a supine or prone human figure and are generally considered a workable stand-in for a human subject (Koester RJ *et*

al., 2004). However, the possibility remains that obligate use of live humans as search subjects in the air-scent tasks resulted in an unknown bias compared with the human-searcher results. As subjects for the dog tasks wore medium- to low-visibility colors (with the handlers lending camouflage when their clothes were high-vis), the dog data may systematically under-estimate PODs for dog teams searching for subjects in high-vis colors.

Finally, while the authors have endeavored to measure effectors of search efficacy as objectively as possible, we have also confined ourselves to equipment that can reasonably be carried and managed by a searcher whose primary mission is searching and not environmental measurement. The tools for measuring convection in the field, for example, may not be precise but they are accurate — and are precise enough to determine convection category (Lavdas LG, 1976; Lavdas LG, 1997; Graham H, 1994). Such factors as season, ecoregion, and sun angle are easily obtained; others, such as humidity and temperature, were derived from historical records that, as stated above, reflected these factors measured at the nearest weather station to each search task, and thus may diverge from the true values in an unpredictable way. The vegetation determination, on the other hand, was entirely subjective and thus would profit greatly from an objective method. The authors have not yet encountered a GIS data layer that documents the character of vegetation under the canopy with good-enough granularity for this application; but such a layer, if it exists, would be invaluable. Last, the assignment of which areas to train in on a given date and which dog and handler would perform each task were arbitrary; while this was necessary to meet training goals, randomization of tasks and teams would have been more objective.

Conclusions

In this report, the authors have presented a means of deriving POD — the critical number necessary to judge search efforts and plan future efforts — objectively via the relationship between effective sweep width (*W*), a distance-quantified value specific to a given detector or group of detectors, search object, and environmental conditions, and POD as measured in the field. Making maximal use of actual detection and miss distances, this method provides the first explicit link between previously measured search efforts (via coverage, or C) and POD for prospective ground-search efforts with uncertainties at least as good as and possibly better than the definitive crossover method, and with a level of statistical sensitivity likely to be greater than that needed to measure operationally significant differences between detectors, search objects, and environmental factors. In addition, the method appears to be detector-agnostic, providing good agreement between predictions and field-measured POD for both air-scent dog teams and human visual searchers using both of the most common theoretical models. The authors recommend operational use of the slightly larger-than-canonical *W* values obtained with this method in concert with the most popular, random-search model, as providing the best fit to the data by the criteria described.

Acknowledgements

Thanks are due to Keith Conover, Allegheny Mountain Rescue Group (AMRG); Stephen McClanahan, Search and Rescue Ohio (SAR-OH) and Mountaineer Area Rescue Group (MARG); and Don Scelza, AMRG and MARG, for invaluable comments on this manuscript. Crucial discussions on SAR theory came from Robert Koester and Charles Twardy.

Thanks go as well to teammates, colleagues, and volunteers in MARG, Fayette County Sheriff's Tactical Search and Rescue, Northwest Pennsylvania K9 Search and Rescue, SAR-OH, AMRG, and elsewhere for help in setting up and hiding for air-scent dog tasks. The most profound thanks go to the dogs whose work, directly or indirectly, contributed to this study: Charlie, Cinders, Cole, Friday, Mel, Moe, Rosalyn, Rosie, Pip, Sophie, Verity, and particularly Lilly, who made us look good long before we had any possible claim to it.

We would like to thank the staff and leadership of the Pennsylvania Search and Rescue Council and the Pennsylvania SAR-EX conference for their organizational assistance and for expenses related to the study site permit at the Mount Gretna experiment. Thanks go to Jack Frost and Quincy Robe for their assistance in conducting the experiment and tearing down the course. We also thank the individual searchers at SAR-EX who participated in the experiment and their teams. This experiment was one of five supported partially or in whole by contract DTCG32-02-D-R00010 from the U.S. Coast Guard. The final development of IDEA was supported by contract HSCG32-04-DR00005 from the U.S. Coast Guard. GraphPad Software donated its Prism software.

Thanks go to the members of AMRG who participated in the new State Game Lands 203 experiment, in particular Paul Jung, whose hard work helped set up the course; AMRG, for financial support from the team's general budget in obtaining some of the disposable supplies used; as well as members of the Youth Outdoors program of Cleveland Metroparks and 4H, and the individual volunteers who came out to participate as searchers and data loggers. Particular thanks go to the Commonwealth of Pennsylvania Game Commission, for permission to conduct the experiment and for special access to the Game Lands.

About the Authors

Kenneth B. Chiacchia has been an operational SAR dog handler since he first certified with Search Dogs Northeast (SDNE) of Antrim, NH, in 1992. He is currently a dog handler and search manager with Mountaineer Area Rescue Group (MARG), Morgantown, WV, of the Appalachian Search and Rescue

Conference, as well as a firefighter/EMT with Harmony Fire District in Butler County, Pa. He is also a lead evaluator and instructor in the National Association of Search and Rescue (NASAR) GSAR program and an evaluator with the Pennsylvania SAR Council air-scent dog testing program. Ken has responded to searches in Pennsylvania, West Virginia, Ohio, and elsewhere. A professional nonfiction and fiction writer, Ken trained originally as a biochemist. He is currently the senior science writer at the Pittsburgh Supercomputing Center.

Heather E. Houlahan is a professional dog trainer, writer, and farmer who lives on a small farm near Pittsburgh. A SAR dog handler and wilderness EMT with MARG, she first certified as operational with SDNE in 1992. Heather has deployed as a dog handler in Pennsylvania, West Virginia, Ohio, New England, and elsewhere, including a 2005 deployment to Mississippi after Hurricane Katrina. She is also a lead evaluator and instructor in the NASAR GSAR program, as well as the head behaviorist for National English Shepherd Rescue of the U.S. Heather writes the "Raised by Wolves" blog, at http://cynography.blogspot.com/

They are married.

Abbreviations

AMDR: Average maximum detection radius, the mean distance at which an observer cued to the location of an object can see it.

C: Coverage, the ratio of effective sweep width times the path length of a detector within an area to its size.

C_{50}: Coverage assuming an effective sweep width of 50m.

DO: Detection opportunity, a detection or miss of a search object at a right angle (lateral) to the detector's path.

POA: Probability of area, the estimated probability that a search object is contained within a given area.

POD: Probability of detection, the probability that a given detector will detect a given search object within an area under certain environmental conditions, assuming that object is in the area.

POS: Probability of success, the probability that a given search effort will detect a search object within a given area being searched (a "segment"). Equal to POA X POD.

POS_o: Overall probability of success, the sum of POSs over all segments being searched.

W: Effective sweep width, a distance-denominated term defining the envelope within which a given detector's number of misses on a search object equals the number of detections outside, under specific environmental conditions.

References

Charnes A, Cooper WW. The theory of search: optimum distribution of search effort. *Management Sci.* 1958;5:44-50.

Chiacchia KB, Houlahan HE. Effectors of Visual Search Efficacy on the Allegheny Plateau. *Wilderness Environ Med.* 2010;21(4):188-201.

Chiacchia KB, Houlahan HE, Hostetter RS. Deriving Effective Sweep Width for Air-scent Dog Teams. *Wilderness Environ Med.* 2015;26(2):142-149.

Chiacchia KB. Critical Separation Versus Effective Sweep Width: Bridging the Old and New Search-and-Rescue Worlds. *Wilderness Environ Med.* 2020;31(1)44-49.

Cooper DC, Frost JR, Robe RQ. *Compatibility of Land SAR Procedures with Search Theory.* Alexandria, VA: Potomac Management Group Inc; 2003.

Gelman A, Meng X-L (eds). *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. West Sussex, England: John Wiley & Sons Ltd; 2004.

Graham H. Probability of detection for search dogs or how long is your shadow? *Response Magazine*. Winter 1994;1-7.

Koester RJ, Cooper DC, Frost JR, Robe RQ. *Sweep Width Estimation for Ground Search and Rescue.* Alexandria, VA: Potomac Management Group Inc; 2004.

Koester RJ. *Lost Person Behavior.* Charlottesville, VA: dbS Productions; 2008.

Koester RJ, Chiacchia KB, Twardy CR, Cooper DC, Frost JR, Robe RQ. Use of the visual range of detection to estimate effective sweep width for land search and rescue based on 10 detection experiments in North America. *Wilderness Environ Med.* 2014;25(2),132-142.

Koester RJ. Enhancements to Statistical Probability of Area Models based upon updated ISRID data collection for Autistic Spectrum Disorders and Typically Developing Children. *J Search Rescue.* April 2020;4(1):35-63.

Koopman BO. *Search and Screening: General Principles with Historical Applications.* Elmsford, NY: Pergamon Press; 1980.

Lavdas LG. A groundhog's approach to estimating insolation. *J Air Pollution Control Assn.* 1976;26:794.

Lavdas LG. *Research NoteSRS-4: Estimating Stability Class in the Field.* Juliette, GA: U.S. Department of Agriculture; 1997.

McClanahan, S. Object recognition and detection: Potential implications from vision science for wilderness searching. *J Search Rescue.* July 2021;5(1):1-17.

Perkins D. *The Variable Nature of Probability of Detection for Ground Search Teams.* Ashington, Northumberland, UK: The Centre for Search Research; April 2018.

Syrotuck WG, Syrotuck JA. *Analysis of Lost-Person Behavior.* Mechanicsburg, PA: Barkleigh Productions; 2000.