Missing Person Behaviour
An Australian Study

Final Report to the Australian National SAR Council

Prepared by:
Charles R. Twardy
Robert Koester
Sgt. Rob Gatt

sarbayes.org

June 3, 2006

Executive Summary

Background: This is the final report to the Australian National SAR Council on the SAR*Bayes* data collection begun in late 2000. There were two earlier versions, both of limited release. The preliminary report (November 2003) had 288 cases (245 non-vehicle), but was missing 3 out of the 8 states and territories. The interim report (May 2005) had 526 cases (445 non-vehicle), filled most gaps, and was revised for clarity.

This version: This report has about 30 extra cases, but more importantly, the data has been thoroughly reviewed, cleaned up, and consolidated. We were able to determine many values that were previously unknown. A new field, 'Incident Type' allows us to select just the Missing Person cases for analysis (excluding straight rescues, water searches, etc.) We believe this is the first report to analyse cases by retrospective Scenario (another new field). We have tested categories for significant differences (from the remainder).

Findings: The key findings, relative to earlier discussions:

- Distributions are largely consistent with previous reports, and with expectations. Despondents, however, were not clearly bimodal.
- The 25%, 50%, and 75% zones for Groups are about *double* those for Singles, even within a single category (Hikers). The 95% zone is about the same. Overall fatality rates are lower in groups, but the difference is weaker within a particular category (Hikers).
- Rural injuries and fatalities (28% and 14%) are much higher than for Wilderness (18% and 6%), probably reflecting the different composition of case types.
- Our median Distance for Alzheimer's patients has come down to about 1.3 km, which roughly matches that of the Virginia data.
- Our form asked *too many questions*. Most questions were not answered. Many gathered more detail than is useful for predicting lost person behavior. Future efforts should *seriously* limit the variables measured, and use automated map methods for fields like Last Known Point, Find Location, Distance Travelled, etc.

We have decided to present a relatively straightforward statistical summary. Predictive models and comparisons should follow in due course. The data will be made available on the SAR*Bayes* website, `sarbayes.org`. It has also been incorporated into the International SAR Incident Database, ISRID.

Acknowledgements

This SAR*Bayes* project is possible only because of the cooperation of many individuals and organisations, not all of whom we will succeed in naming.

Special thanks to Cheryl Morahan for tireless work collecting, reviewing, and entering the data. Thanks also to all the police officers who took the time to fill out yet another form.

Cheryl Morahan, Rik Head, David Albrecht, and Jim Donovan reviewed drafts. David provided very helpful statistical advice, and Rik helped greatly with clarity. Any remaining murkiness or inaccuracy is solely the fault of the authors.

The National Science Foundation (U.S.) supported Charles Twardy from 2000 through 2002. The Monash Data Mining Centre supported Charles' work from 2002 through early 2005, and provided webservers and other computer resources through early 2006. Thanks to all at Monash for encouragement and advice, especially Lloyd Allison, Trevor Dix, Adam Golding, Luke Hope, Kevin Korb, and Ann Nicholson.

Fronterra Search & Rescue Software (U.S.) has supported the efforts of Robert J. Koester and Charles Twardy since 2005.

Contents

Chapter 1

Background

To begin viewing the data, skip straight to Chapter 2. This chapter provides background, motivation, and method.

The SAR*Bayes* project collected data on Australian missing person searches from late 2000 through mid-2005, including some cases from before 2000. The ultimate goal is to improve the chance of finding a lost person quickly. The idea has been around since at least the 1970s, though most of the data has been from North America. Since 2001 the Centre for Search Research has published annual lost person reports for the United Kingdom. We have, with permission, adopted some of their conventions and methods. While it is interesting to have more data, we believe further improvements in search performance are unlikely until this information is incorporated into "live" predictive models in search planning software.

SAR*Bayes* was conceived by Charles Twardy in order to provide such models. The project began in 2000 as a collaboration between the Monash Data Mining Centre at Monash University (where Charles was working on Bayesian networks), Victorian Police Search & Rescue, and VicWalk's Bushwalkers' Search & Rescue, to collect and analyze the data presented here. Charles returned to the U.S. in 2005, but Monash continues to support SAR*Bayes*via affiliation and continued collaboration and computer access.

Further copies of the report are available from the authors or the SAR*Bayes* website (`sarbayes.org`), as is the database.

1.1 Relation to Previous Work

We believe our database is comparable to other well-known databases such as those compiled by Mitchell (1985) for the USA, Koester and Stooksbury (1999); Koester (2003) for Virginia, USA, Hill (1991, 1999a) for Nova Scotia (Canada), Perkins et al. (2001, 2002, 2004, 2005) for the United Kingdom, and of course Syrotuck (2000) for New York and Washington, USA.[1] Of these, Mitchell (1985) is probably the largest and most thorough, but also the least

[1]Syrotuck had a few cases from other states as well.

well-known.[2] Like these databases, our data was collected from specialised Search & Rescue units.

> **NOTE**
>
> Our data consists only of those cases which notified **specialised SAR resources**. If the case was resolved quickly by local resources, we do not have it. Therefore our database probably represents only the longer searches, so it would *not* be appropriate to use it to evaluate average SAR response time or outcome, although it *would* still be appropriate to investigate the relationship between times and outcomes within this database.

There are other important studies which are not directly comparable. Kelley (1973) thoroughly investigated other factors like the reasons and timings for searches, and factors influencing survival. Rik Head of Emergency Systems Technology Pty. Ltd. has extended and implemented some of Kelley's work in a computer program used by the Victoria Police Search & Rescue Squad. Kelley also deserves credit for being possibly the only person in land SAR prior to Frost (1999) to examine mathematical search theory (see Kelley's Appendix III). However, he does not really break up his data by category.

There is also an excellent report out of the University of Toronto that makes use of GIS information (Csillag et al., 2000). That report is more in line with Kelley, preparing for responses by profiling the most frequent SAR cases by area, age, etc. However, it is a rich report making use of decision trees and other forms of nonlinear regression that we also hope to use in the future.

Alone among all the studies we have seen, Heth and Cornell (1998) fit parametric models to their data, found natural clusters based on groups with similar parameters (except possibly for scale), and compared those clusters. This approach is statistically more powerful, and potentially more robust than the usual "straight" data approach. Their paper is well worth reading.

Twardy and Hope (2004) performed a cluster analysis on the 2001 Virginia dataset Koester (2001), which had 242 relevant cases. We found 5 clusters: Children, Alzheimers with quick response (20% fatalities), Alzheimers with slow response (65% fatalities), Medical (Despondent, Retarded, Psychotic), and Miscellaneous adults. However, differences in Distance were not predictive, in part because so many values were unknown.

Using the same dataset, Allison (2005, 2006) at Monash created *hybrid* Bayesian network models of the same data. Even though his models explicitly handled unknown values, and did not need to discretize the continuous variables, they still found no link between Type and Distance.

[2] Mitchell has 3,511 cases. Although he does not analyze them all, it does give him over 600 hikers, which he analyses in three groups of about 200 each.

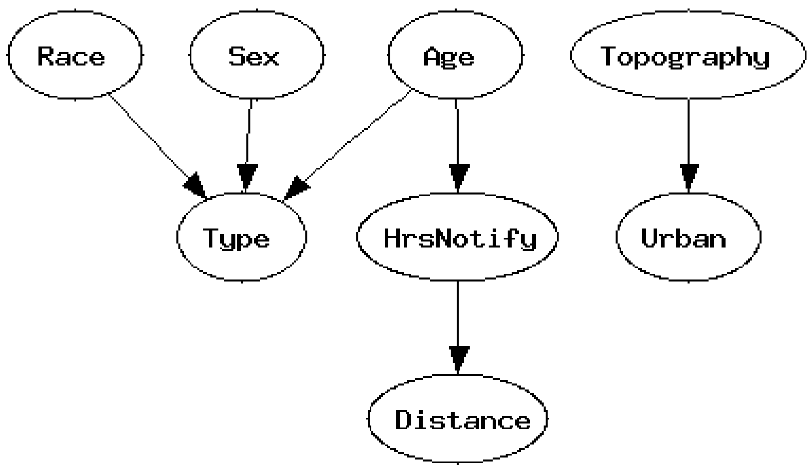

Figure 1.1: Bayesian network found by Lloyd Allison for 8 variables in the Virginia 2001 dataset. Distance From LKP was predicted entirely by HrsNotify, the elapsed time SAR rescources were notified. Given the dataset size (242 cases) and the large number of unknown values (see text), no other links to Distance were warranted.

The structure of the 8-variable model is shown in Figure 1.1. The variables are:

- **Age:** in years (continuous, near enough) [10% unknown]
- **Race:** White, Black[3] [80% unknown]
- **Sex:** Male, Female [80% unknown]
- **Type:** Alzheimers, Child, Despondent, Hiker, Other, Psychotic, Retarded [6% unknown]
- **Topography:** Mountains, Piedmont, Tidewater [80% unknown]
- **Urban:** Rural, Suburban, Urban [80% unknown]
- **HrsNotify:** Hours until SAR notified (continuous) [74% unknown]
- **Distance:** Distance in km from LKP (continuous) [43% unknown]

The model suggests that Type and Distance will be *correlated*, because they both descend from Age, but that this correlation will disappear once we know Age. Futhermore, they say that the correlation between Age and Distance disappears once we know HrsNotify (the elapsed time until SAR is called).

It makes sense that HrsNotify is a good predictor of Distance, since we expect our subjects to keep moving for at least 4 hours.[4] A fast response means less time to wander. But it was surprising that nothing else mattered. So surprising, in fact, that we attribute it mostly to sparse data.

When Allison included 7 other post-find variables, there was a link from Type to Find Location, which is obviously useful to search management. There was a link from Distance to Find Location, and other suggestive links that were less clearly useful. (For example, the link from Type to Find Resource may reflect mostly what kinds of resources get *used* on a search.) We look forward to letting Allison work on the larger ISRID.

1.2 Methods

1.2.1 Forms

In November 2000 all eight states and territories in Australia agreed to submit land SAR data for the SAR*Bayes* project (National SAR Council meeting, Canberra, Ref. WP24/4/1&2). At that meeting they saw and commented on a preliminary form. Based on their comments, we revised the form and

[3]Only collected on Alzheimers patients, so not generalizable. But it was there as a predictor, so it got included in Allison's model.

[4](Mitchell, 1985, p.18, Figs. 33–35) found about that 75% of Hikers in the Western U.S. kept moving for at least 4 hours, though only about 10% kept moving longer than 24 hours.

released it in February 2001. This version asked specifically for data on land search in rural or wilderness areas for people traveling under their own power.

We found that respondents were not consistent about what they included and excluded. After a few inquiries in May 2002 we instructed them to send *all* their searches, with a view to filtering them consistently at our end. To help in filtering, at the end of 2002 we added fields specifying whether the subject used a vehicle after the Last Known Point, and whether the search was urban, rural, or wilderness.[5] In addition we made some fields easier to use. That form is (**dataform2003**): `sarbayes.org/dataform2003.pdf` or `sarbayes.org/dataform2003.doc`.

We asked too many questions. Many fields were too variable or too infrequently answered to be useful. For example, respondents rarely listed (or knew) specifics about gear and clothing. In retrospect, we don't care either. All we want to know is whether the gear and clothing were *adequate to the situation*. We could have saved a dozen questions. At the end of our data collection, we streamlined the form. The new form is **dataform2005**: `sarbayes.org/dataform2005.pdf` or `sarbayes.org/dataform2005.doc`. We will streamline it still further for the International SAR Incident Database, ISRID.

The forms are quite similar to the one later implemented on the NASAR website, and a version has been available on dbS Productions SARDISK since 2001. See for example, dbS Productions (2003).

1.2.2 Data

Almost all the data was submitted on paper forms and then added to a Microsoft Access™ database by Cheryl Morahan. Using a single person improved consistency. Nevertheless, some cases were submitted online by the reporting agencies (Tasmania and Western Australia) and merged into the Access database. Many cases were sent on regional police forms, which did not have all of our fields. Where possible we followed up on these, but in many cases key fields like Distance remain unkown.

For the final report, the authors reviewed the cases using Microsoft Access and cleaned the data with an eye towards merging with the International SAR Incident Database.

New fields. We added two new fields: IncidentType and Scenario. **IncidentType** is used to select only "Missing Person" incidents for analysis, leaving off straight rescues, water incidents, most criminal cases, and the like. **Scenario** gives the type or cause of the incident, separating "just plain lost" from overdue, evading, despondent, investigative, medical mishaps, and trauma victims. Detailed definitions are given in Appendix A.

[5] For data collected on earlier forms, we went back and entered appropriate values for these new fields.

Unknown Values. Wherever possible, we filled in unknown (a.k.a. missing) values by hand, using the free-text Notes field, entries in other fields, and where possible, review of original documents or requests to the officer reporting the case. We were able to assign almost all cases a TotalTimeLost this way, based on start and end dates.

⚠ **CAUTION** ⚠

The resulting TotalTimeLost values are in general only precise to within 6 or 12 hours, depending on the length of the search. However, that is good enough for our purposes.

Values consolidated. We reviewed several key fields and combined similar entries to reduce the number of types. We also created some new values based on repeated entries in some "Other" fields.

- **Find Location** had the following name changes and new fields:
  - *Water:* In or next to water. Includes stream, river, lake, riverbank, dam.
  - *Flat:* includes park, beach, similar
  - *Cliff:* added because several mentioned 'cliff'.

  ⚠ GENERAL NOTE: this category is flawed. The values are not exclusive: 'stream' is also 'valley', 'track' is often on a 'ridge', etc.

- **Resource Types and Find Technique** cleaned up similarly and made to match where appropriate.
  - *Motorbike:* includes motorcycle, trail, quadbike
  - *Mounted*: includes all mounted: Vehicle, Horse, Motorbike, etc. Probably includes some that should be 'Car'.
  - *Car*: when it was clear that it was a patrol car doing road patrol.
  - *Self*: Any case of self-recovery. Previously entered variously as "walk-out", "self recovery", or similar.
  - *Investigative*: solved by normal police investigation.
  - *Civilian*: found by anyone not formally part of the search, including family, friend, etc. Should *not* include park rangers and State Emergency Services (SES) responders, though we caught several cases where that had happened.
  - *Communication*: found by contact with MP; overlaps with Investigative.

- **TradCateg** expanded based on answers found in 'Other':

  - *Autistic*: we had 8 cases, and it will be a category in ISRID.
  - *Motorist*: 4WD, Motorbike, Car. Specific type available under Activity.

- **Activity** expanded based on text in ‘Other’:

  - *Walking*: for example, child going home from school, etc.
  - *4WDriving*: usual definition
  - *Motorbiking*: includes motorcycle, trailbike, quadbike
  - *Driving*: in a car, on a paved road

Other changes. In addition to a great deal of general correction and cleaning, removing duplicates, etc., we made the following changes.

- Many fields were consolidated and dropped, including **TimeOnScene**, **DateFrom**, **DateTo** (in favor of other dates), some absolute times, some “other” fields, and all “units” fields, which offered a pernicious freedom of choice.

- **TotalTimeLost** filled in. It was often unknown. However, start and end dates provided significant constraints. The free text also helped. We were able to fill in values for almost all cases, to within 6 or 12 hours, which is adequate for our purposes. *On the whole, TotalTimeLost should not be considered more accurate than 6 hours.*

- Some distances were filled in from coordinates. Some coordinates were looked up from placenames. Unfortunately, we did not get as many new values as we wanted.

Much of the data is unused here. For example, we have not touched the weather data, which is bound to help predict survival. See variable names in Appendix C.

1.2.3 Reporting the Results

We follow Perkins et al. and report full statistics only for categories with enough cases. In tables, when there are fewer than 15 cases, we just show “−−” in place of the statistic (like median distance). In figures, we begin the report with overall distributions for variables of interest (like Status). Then, when investigating the effect of another variable (like Traditional Category), we make subplots *only for those subsets (like Despondents) whose distribution is likely to be reliably different from the remainder (here, non-Despondents).*

For more details, please see Appendix B.

1.2.4 Programs

We wrote a set of Python programs to automate the analysis. Here we describe a few details, for interested researchers.

The programs let us easily subset the data and for any subset, generate the histograms, pie charts, tables, and quantiles you see in later sections. The original programs used in the preliminary and interim reports were rewritten to be object-oriented, and provided with a testing suite based on a sample dataset. The programs are open-source, and available from `sarbayes.org`. They make use of the SciPy (Jones et al., 01) and Matplotlib (Hunter et al., 04) libraries. They can be used interactively from within Python, or run from the command line to generate the whole report (also requires LaTeX). They should be useful with minimal editing for any similar *flat-file* database. We exported our Access™ data to a flat file in CSV format (comma-separated value), and used that directly.

Groups are counted only once, to keep large groups (up to 22 people!) from being over-represented.[6] Distances for groups are calculated by taking the distance for a random member of the group.[7] Group outcomes are determined by the worst-case member, with "No Trace" counted worse than "Fatality", on the grounds that it is probably a fatality, and also a search failure. In theory, that overstates the risk to individuals in a group, since if 1 member of a 20-person group dies, we count that group as a fatality. In practice, it didn't make much of a difference, but we also look at individual risks in Chapter 3.

NOTE

Most of Chapter 2 is generated automatically by the Python programs. Consequently, it is largely charts and tables until the end. There is very little flavour text.

[6]The original Aussie data used one record per person. The ISRID data lists them all in one line but may separate data by slashes. For example, Age might be "18/25/15" for a group of three. Different branches of the programs have routines for both formats. We are converting everything to ISRID format.

[7]Namely, the first. We saw no difference using mean, median, farthest, or nearest.

Chapter 2

Overall Summaries

2.1 Data Summary as of June 3, 2006

We have 688 records comprising 550 cases (lost groups contribute many records, but only one case.)

We have excluded 74 cases where the subject traveled in a motor vehicle after the Last Known Point (LKP),[1] leaving 476 cases where the subject is traveling under his or her own power. Of those cases, 458 were Missing Persons. (The Incident Types were: Recovery, ELT/EPIRB, Evidence, Water, Other, Missing Person, Rescue.)

Of the Missing Persons cases, 97 were groups, (comprising 339 people), and 361 were single persons. In 3.2 we investigate possible differences between groups and singles.

2.2 Representation by State

Figure 2.1 and Table 2.1 show the breakdown by state or territory, and for comparison, the Australian population distribution.

[1]The LKP is sometimes called the Initial Planning Point, IPP.

State	Cases	%	% Pop'n
ACT	21	4	2
NSW	154	28	34
NT	14	3	1
QLD	111	20	19
SA	95	17	8
TAS	7	1	2
VIC	103	19	24
WA	45	8	10
	550		

Table 2.1: Number of cases by state or territory, compared to population.

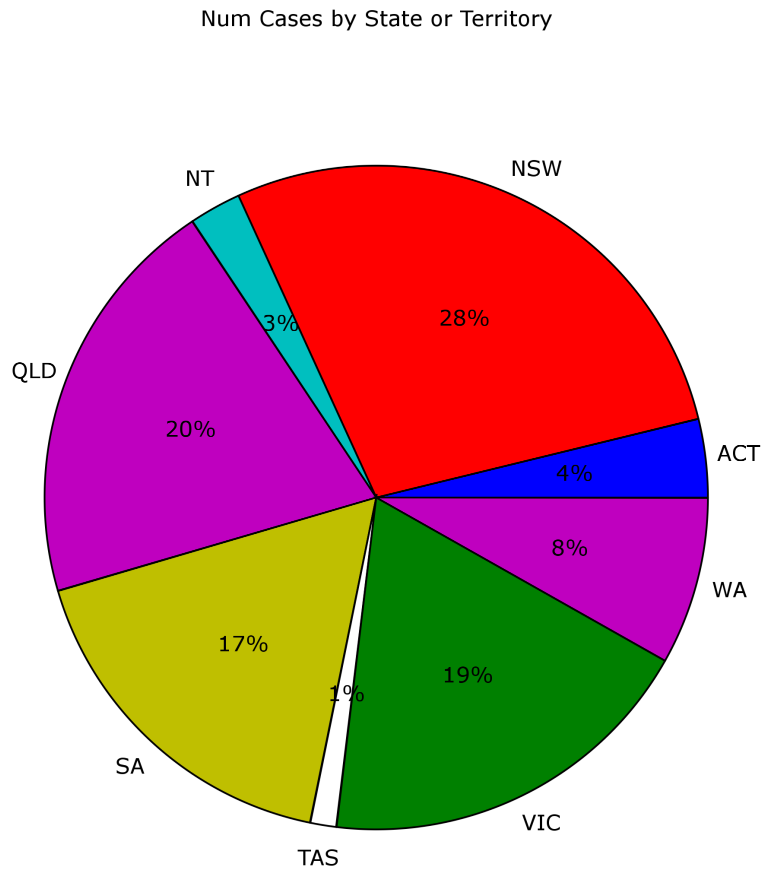

Figure 2.1: Breakdown by state

2.3 MP-All

The dataset *Missing Persons* is the "whole" dataset: all the cases which fit our criteria. As noted previously, we excluded Water searches, straight Rescues, cases where the MP hopped on a bus after the LKP, etc. Groups were condensed to a single entry by taking the record for the first member, but substituting the worst-case for Status.

2.3.1 Overall

Here we summarize the key variables in the dataset, in tables and figures. Because they have so many values, Table **??** omits Activity, Find Location and Traditional Category. However, the distributions for Find Location and Traditional Category are available in figures in this section, and that of Activity is available in Section **??**.

			Percentiles			
	N	N_r	25%	50%	75%	95%
DistFrLKP	458	238	1.0	2.0	6.0	20.0
Age	458	429	20.0	36.0	59.0	83.0
TotalTimeLost	458	435	6.5	15.2	25.0	69.1

Table 2.2: Summary table for key **numeric** variables in MP-All. N is the total number of cases in this dataset, and N_r is the number reporting that category. (50% is the median.)

	N	N_s	Distribution of Values				
AreaKnowledge	458	299	Poor	Fair	Average	Good	Excellent
			125 (41%)	37 (12%)	53 (17%)	64 (21%)	20 (6%)
Dehydrated	458	195	No	Yes			
			133 (68%)	62 (31%)			
Experienced	458	287	V.Inexperi	Inexperien	Average	Experience	V.Experien
			57 (19%)	77 (26%)	88 (30%)	51 (17%)	14 (4%)
FindLoc3	458	140	Downhill f	Neither	Uphill fro		
			50 (35%)	63 (45%)	27 (19%)		
Fitness	458	238	V.unfit	Unfit	Average	Fit	V.fit
			14 (5%)	31 (13%)	97 (40%)	76 (31%)	20 (8%)
Hiding	458	228	No	Yes			
			172 (75%)	56 (24%)			
Hypertherm	458	178	No	Yes			
			172 (96%)	6 (3%)			
Hypotherm	458	186	No	Yes			
			157 (84%)	29 (15%)			
Seeking	458	233	No	Yes			
			104 (44%)	129 (55%)			
Setting	458	426	Wilderness	Rural	Urban	Other	Unknown
			216 (50%)	84 (19%)	84 (19%)	9 (2%)	33 (7%)
Sex	458	446	Female	Male			
			125 (28%)	321 (71%)			
Status	458	456	Unhurt	Injured	Fatality	No Trace	
			300 (65%)	100 (21%)	41 (8%)	15 (3%)	
Visibility	458	249	Concealed	Easily Vis			
			65 (26%)	184 (73%)			

Table 2.3: Summary of key **categorical** variables in MP-All. N is the total number of cases in this dataset, and N_r is the number reporting each variable. (Percents rounded to nearest whole number.)

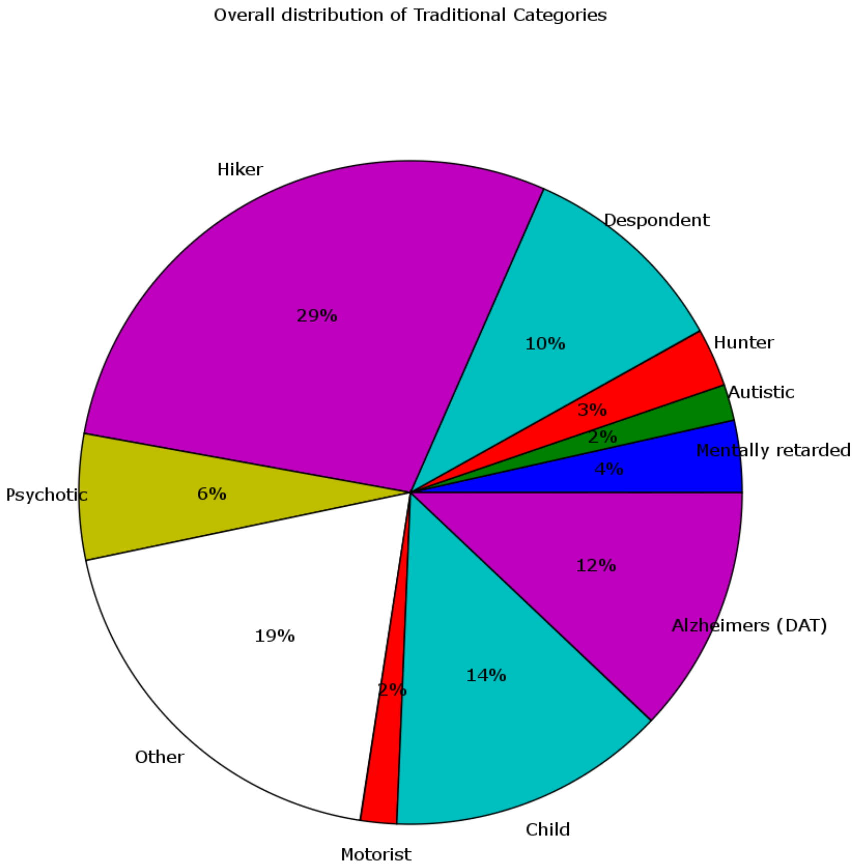

Figure 2.2: Overall Category distribution in MP-All.

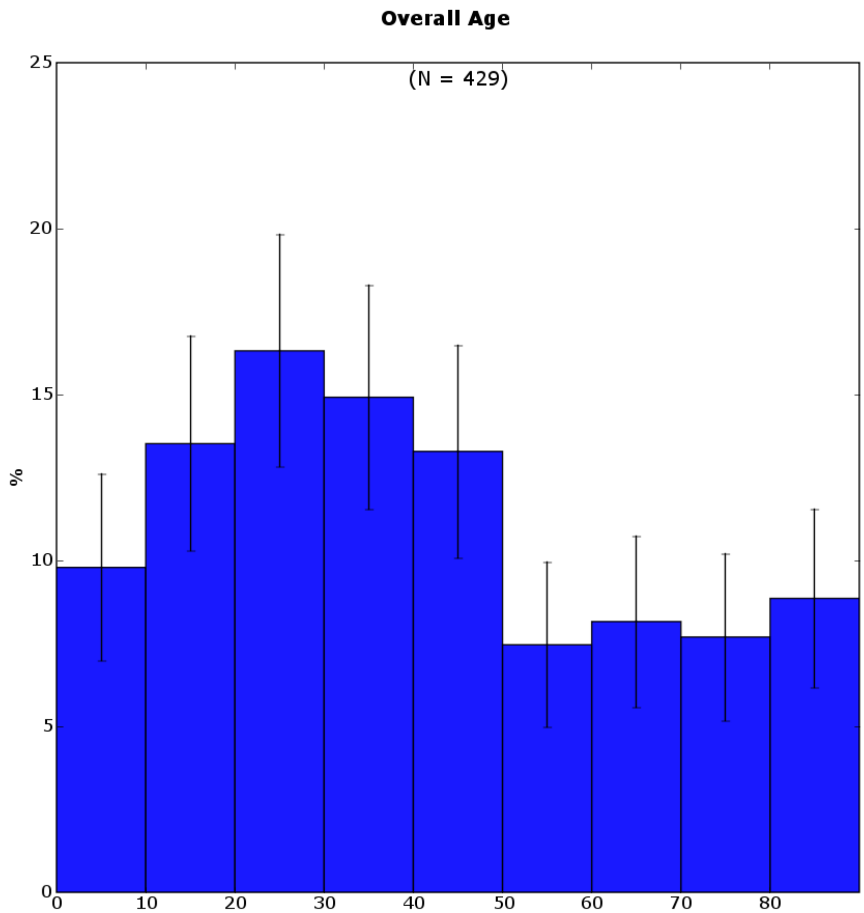

Figure 2.3: Overall Age (yrs) in MP-All. Bars show 95% confidence intervals.

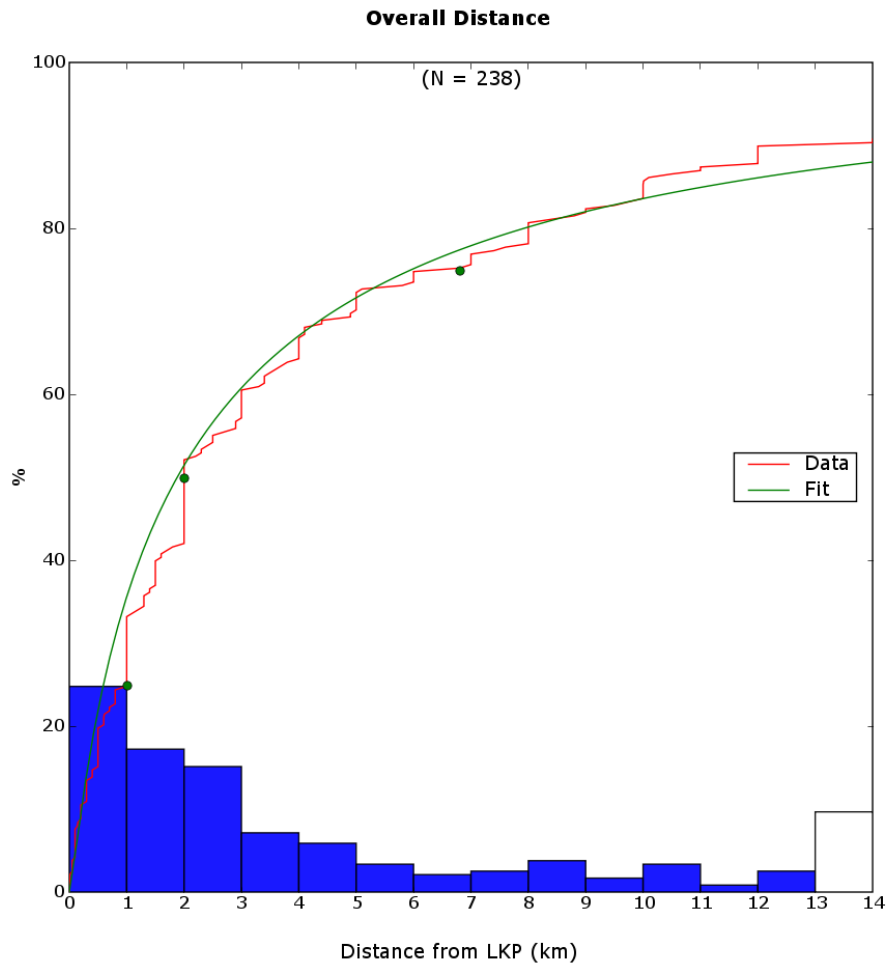

Figure 2.4: Overall Distance From LKP (km) in MP-All. Lines show cumulative percent. Dots mark 25%, 50%, 75%, and 95% points (some of which may be off the graph).

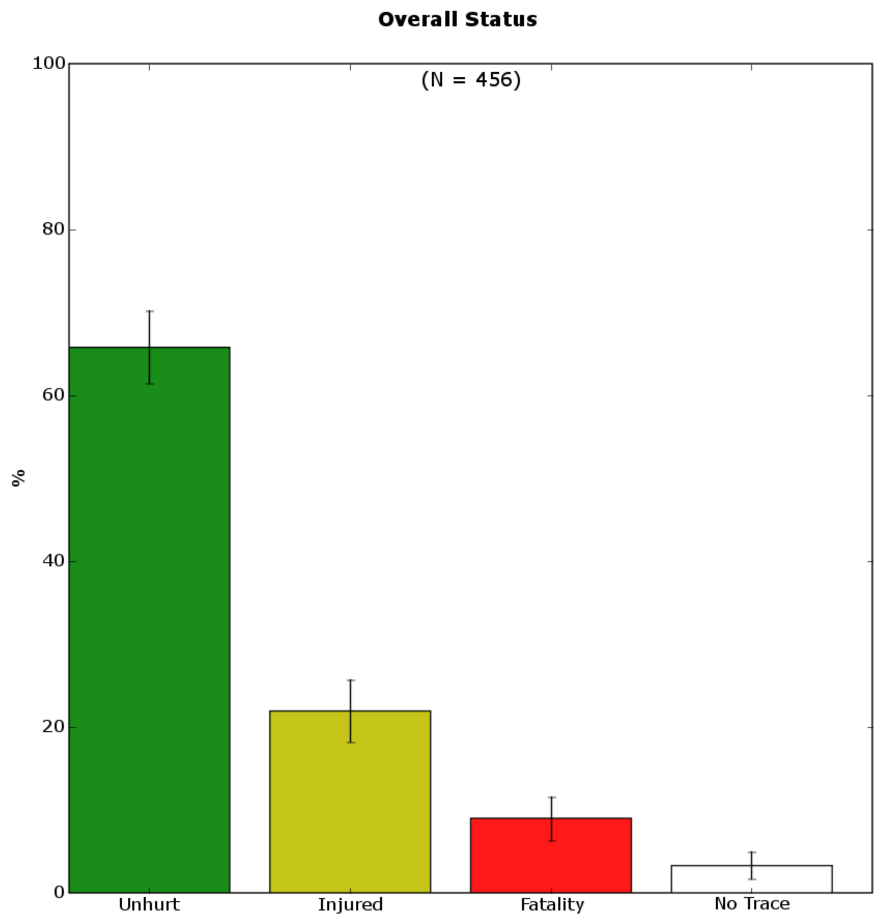

Figure 2.5: Overall Status in MP-All. Bars show 95% confidence intervals.

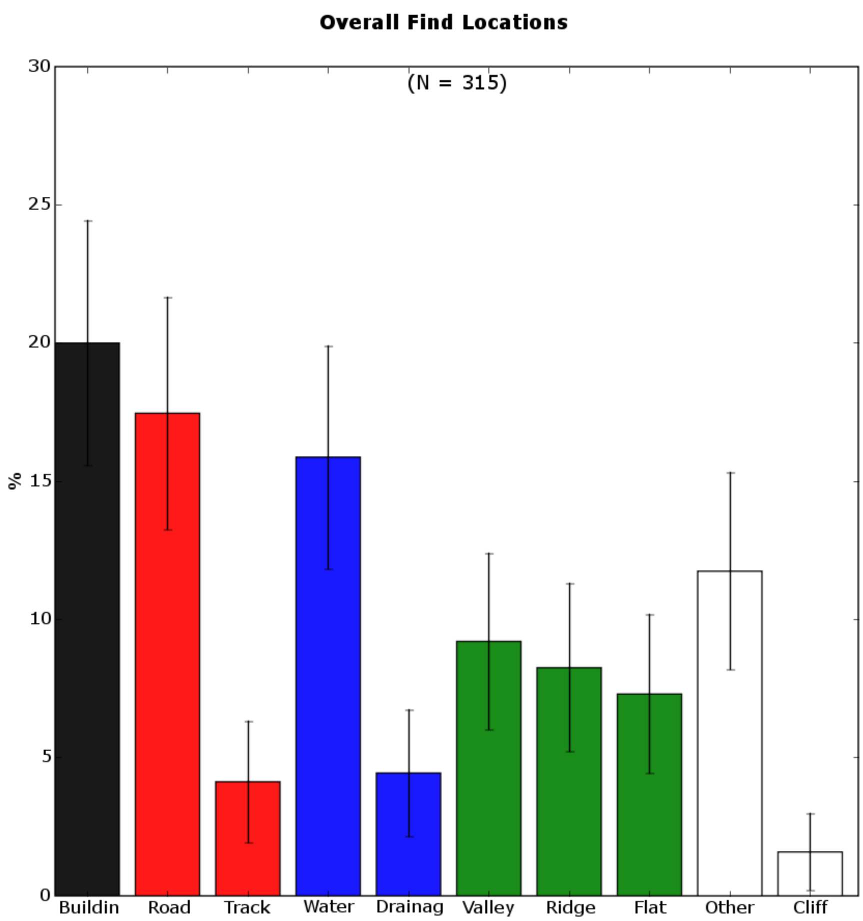

Figure 2.6: Overall Find Location in MP-All. Bars show 95% confidence intervals.

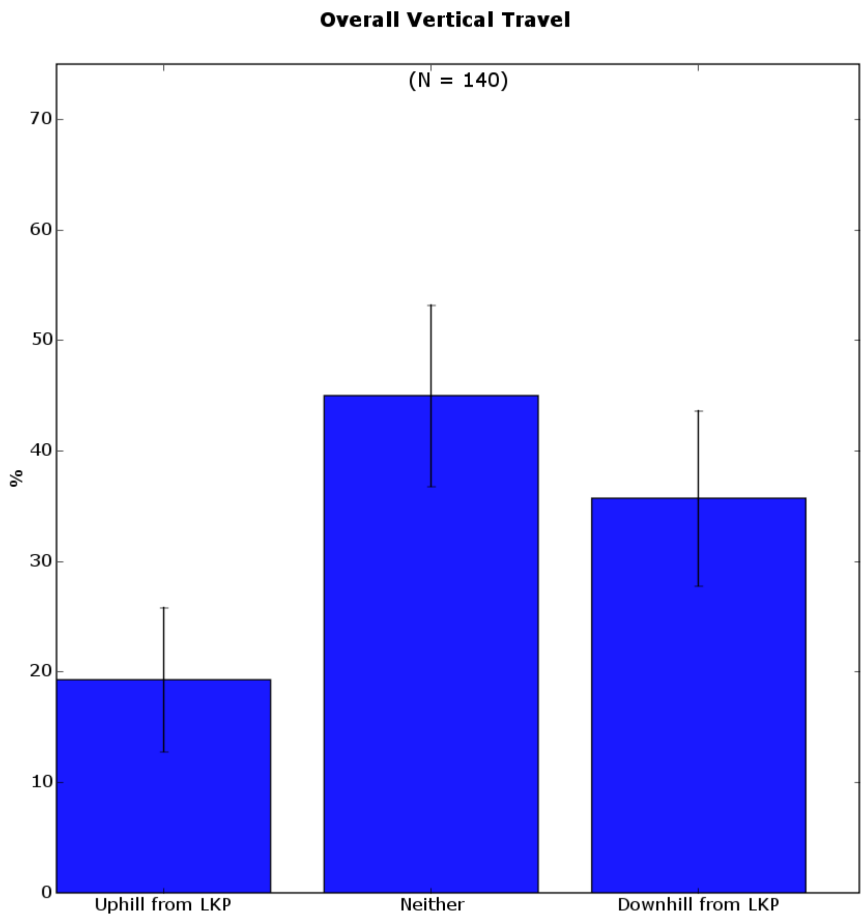

Figure 2.7: Overall Vertical Travel in MP-All. Bars show 95% confidence intervals.

Scenario

2.3.2 By Scenario

			Percentiles			
	N	N_d	25%	50%	75%	95%
Criminal	1	0	- -	- -	- -	- -
Despondent	41	20	0.5	1.0	2.0	3.3
Evading	22	15	0.1	1.0	5.9	27.2
Investigative	29	12	- -	- -	- -	- -
Lost	304	164	1.0	2.5	6.0	18.7
Medical	4	2	- -	- -	- -	- -
Overdue	27	14	- -	- -	- -	- -
Trauma	11	4	- -	- -	- -	- -

Table 2.4: Distances (km) from LKP, by Scenario in MP-All. N is the total number of cases in this dataset, and N_d is the number reporting d. (50% is the median.)

	N	N_s	Unhurt	Injured	Fatality	No Trace
Criminal	1	1	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Despondent	41	41	11 (26%)	14 (34%)	12 (29%)	4 (9%)
Evading	22	22	16 (72%)	5 (22%)	0 (0%)	1 (4%)
Investigative	29	29	26 (89%)	2 (6%)	0 (0%)	1 (3%)
Lost	304	303	212 (69%)	74 (24%)	14 (4%)	3 (0%)
Medical	4	4	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Overdue	27	27	26 (96%)	1 (3%)	0 (0%)	0 (0%)
Trauma	11	11	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)

Table 2.5: Status by Scenario in MP-All. N is the total number of cases in this dataset, and N_s is the number reporting status. (Percents rounded to nearest whole number.)

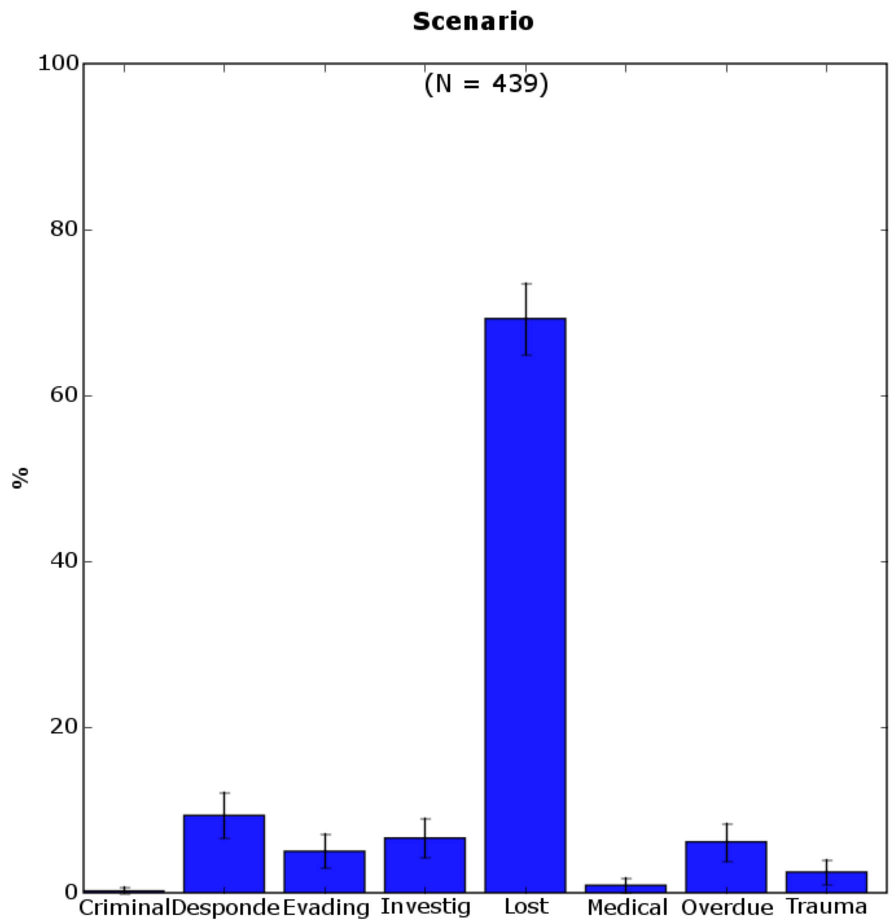

Figure 2.8: Distribution of Scenario in MP-All.

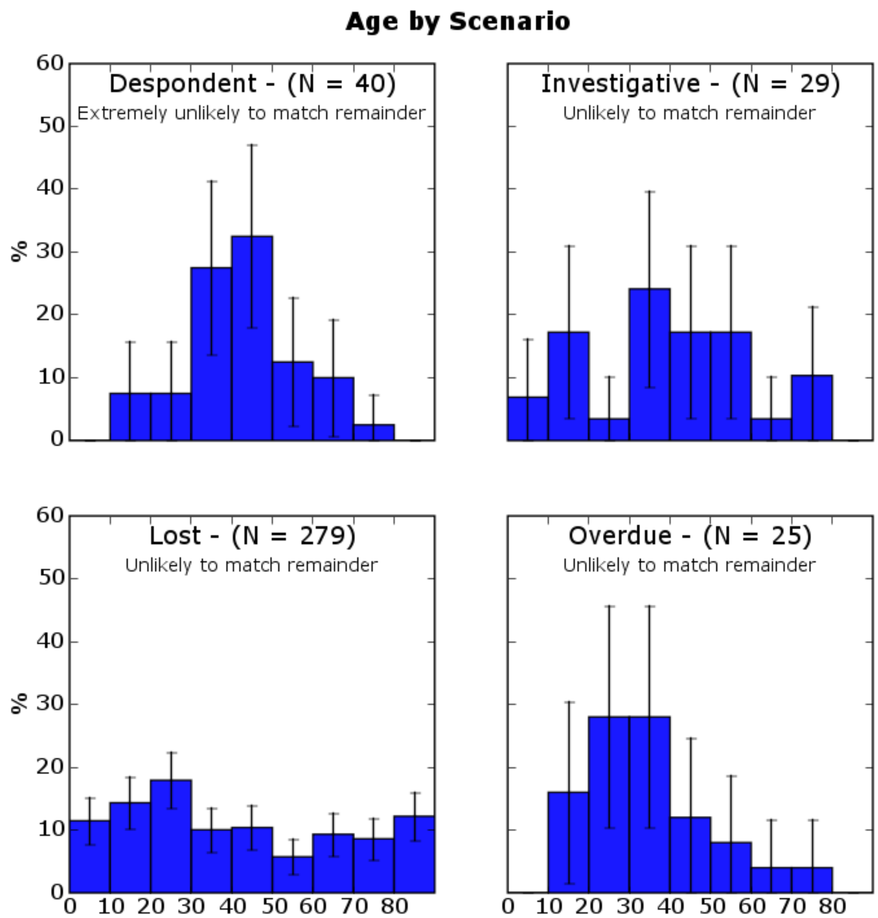

Figure 2.9: Age (yrs) by Scenario in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

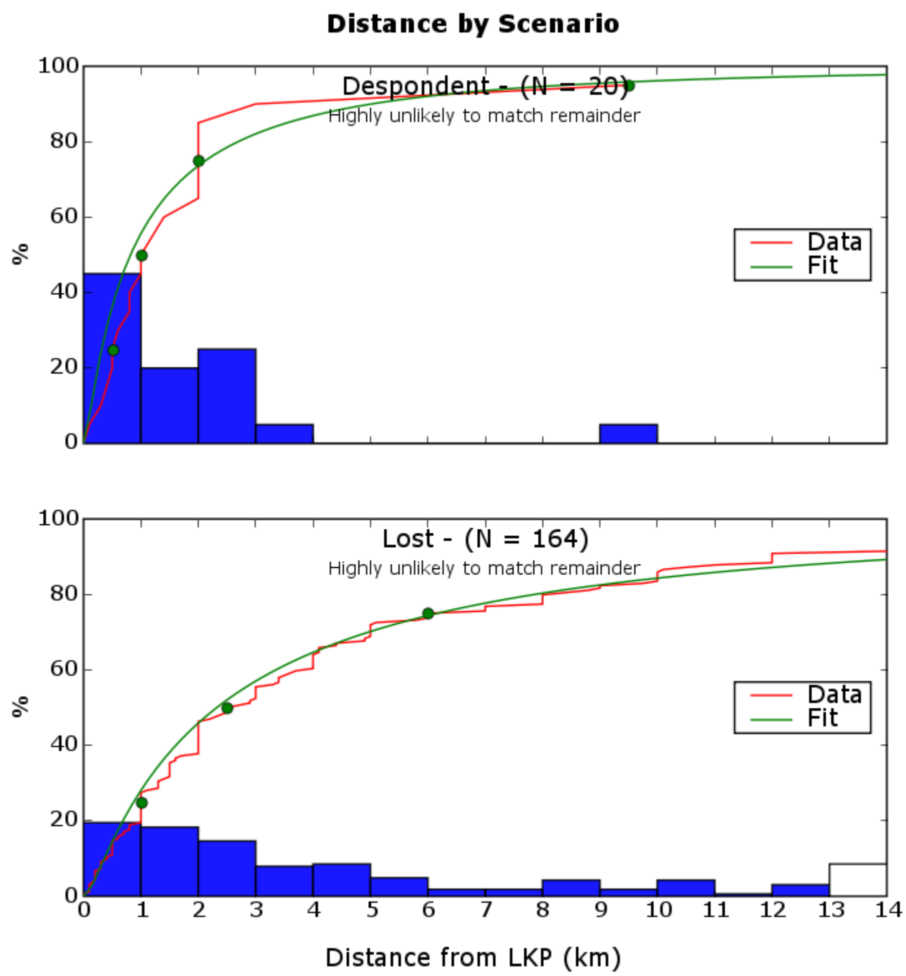

Figure 2.10: Distance From LKP (km) by Scenario in MP-All. Lines show cumulative percent. Dots mark 25%, 50%, 75%, and 95% points (some of which may be off the graph). Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

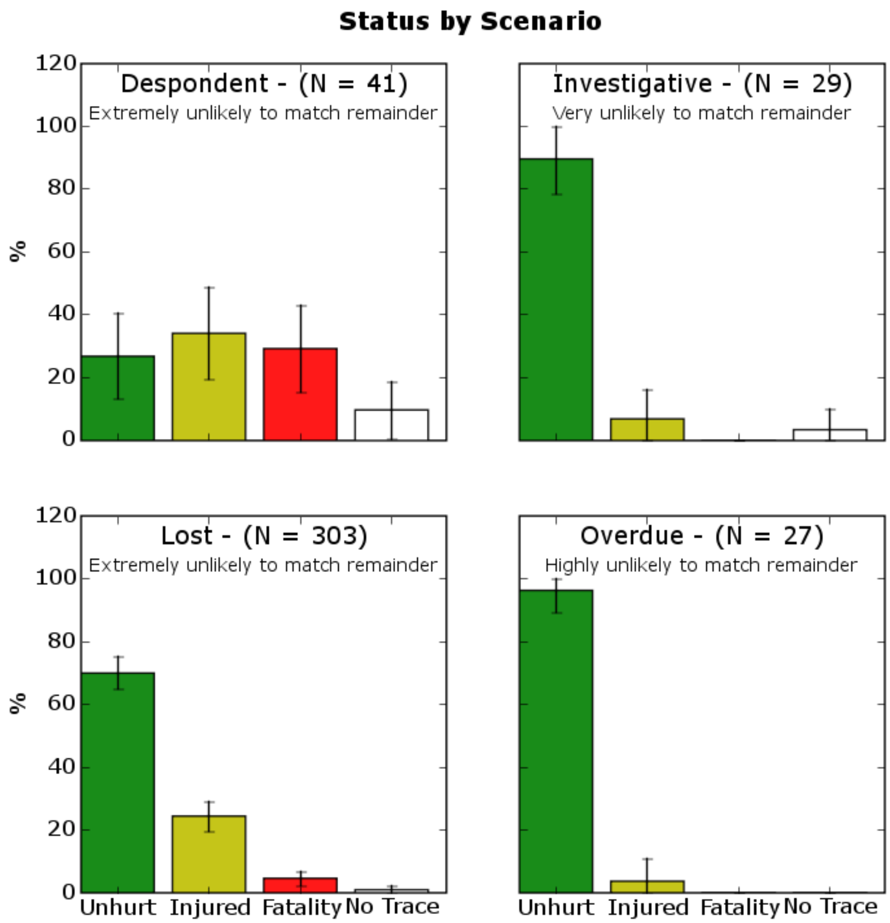

Figure 2.11: Status by Scenario in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

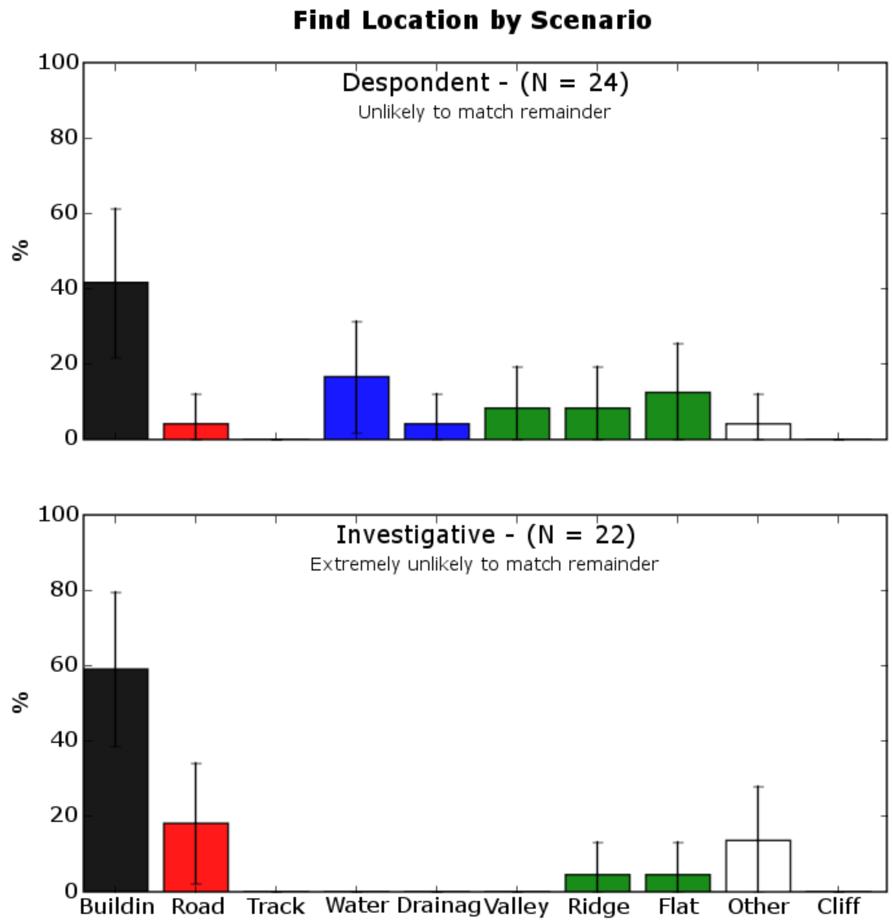

Figure 2.12: Find Location by Scenario in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

Figure 2.13: Vertical Travel by Scenario in MP-All *No categories were reliably different from the remainder.* See Section 2.3.1.

Setting

2.3.3 By Setting

	N	N_d	Percentiles 25%	50%	75%	95%
Other	9	4	- -	- -	- -	- -
Rural	84	54	0.5	1.6	4.0	15.4
Unknown	33	17	0.7	1.7	2.9	6.4
Urban	84	40	0.9	1.5	3.2	20.2
Wilderness	216	113	1.1	3.3	9.5	22.0

Table 2.6: Distances (km) from LKP, by Setting in MP-All. N is the total number of cases in this dataset, and N_d is the number reporting d. (50% is the median.)

	N	N_s	Unhurt	Injured	Fatality	No Trace
Other	9	9	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Rural	84	84	48 (57%)	24 (28%)	12 (14%)	0 (0%)
Unknown	33	33	20 (60%)	9 (27%)	1 (3%)	3 (9%)
Urban	84	84	55 (65%)	21 (25%)	6 (7%)	2 (2%)
Wilderness	216	216	157 (72%)	39 (18%)	13 (6%)	7 (3%)

Table 2.7: Status by Setting in MP-All. N is the total number of cases in this dataset, and N_s is the number reporting status. (Percents rounded to nearest whole number.)

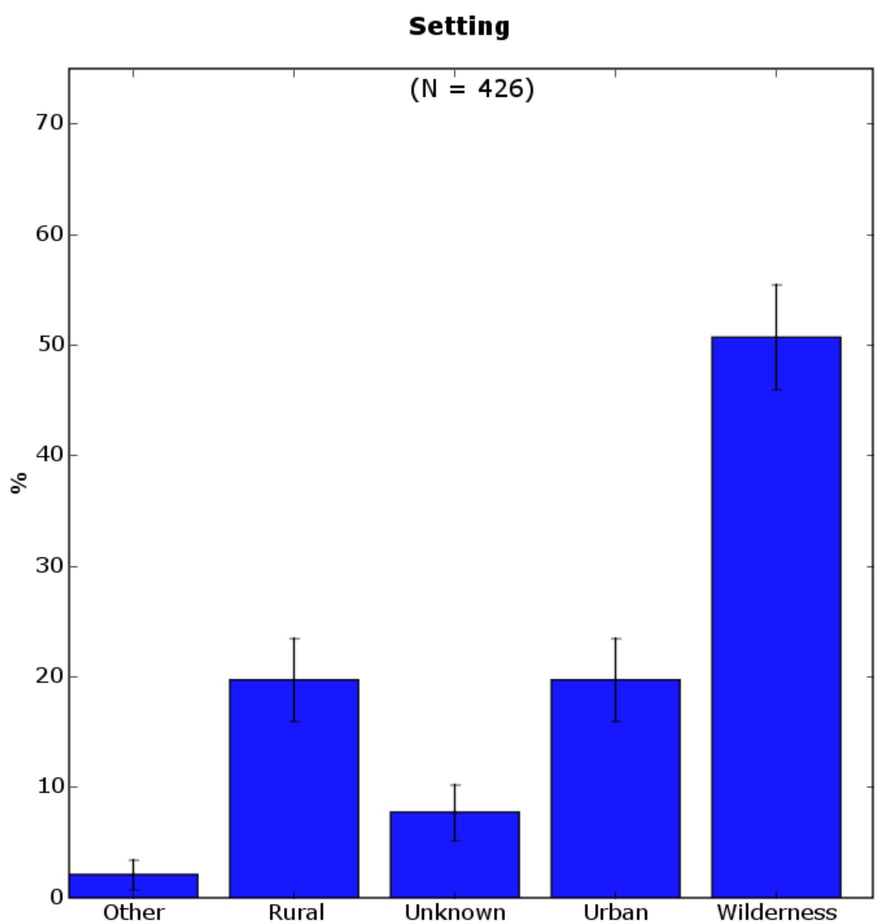

Figure 2.14: Distribution of Setting in MP-All.

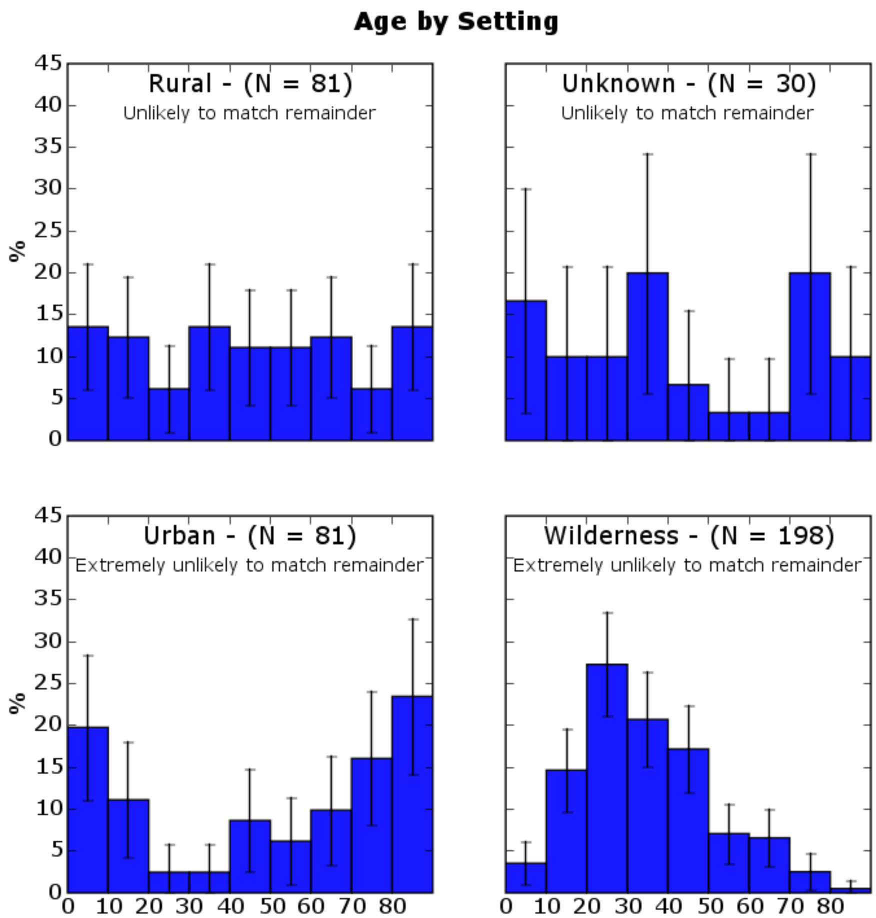

Figure 2.15: Age (yrs) by Setting in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

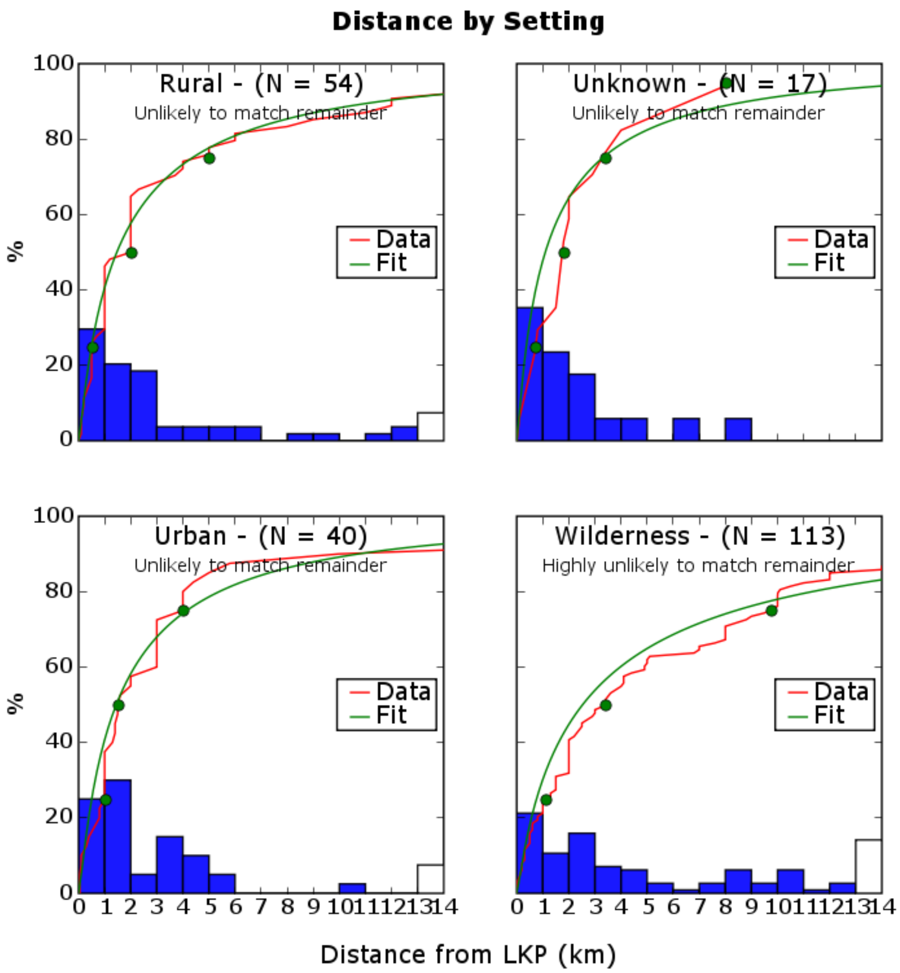

Figure 2.16: Distance From LKP (km) by Setting in MP-All. Lines show cumulative percent. Dots mark 25%, 50%, 75%, and 95% points (some of which may be off the graph). Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

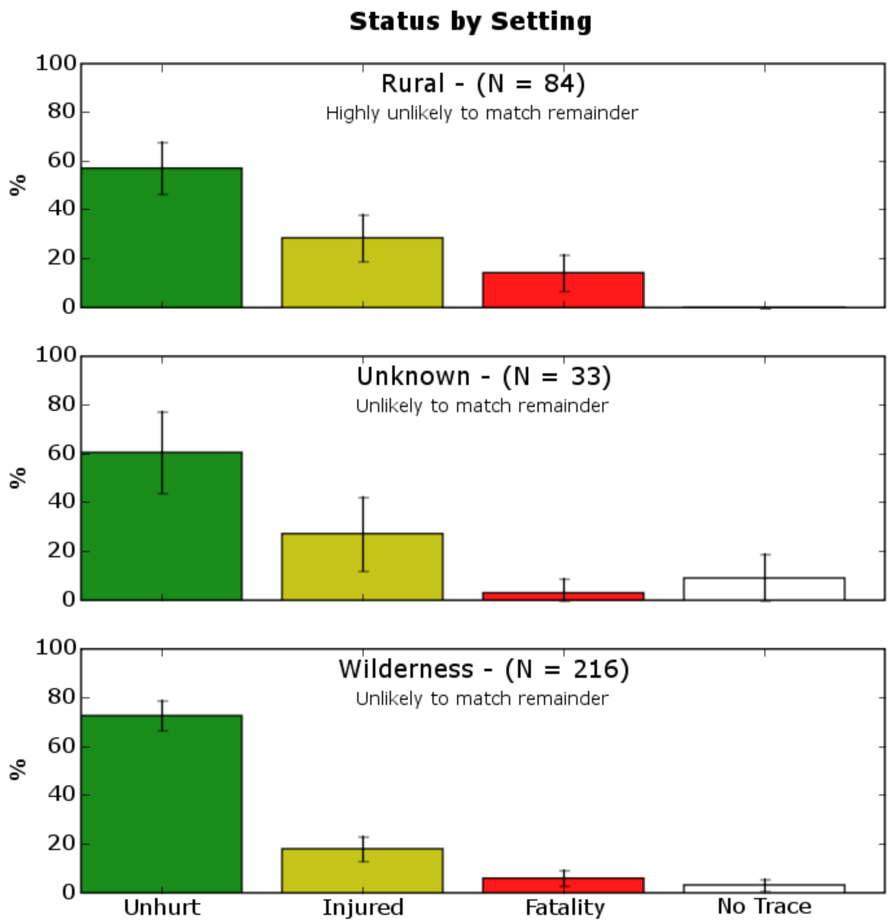

Figure 2.17: Status by Setting in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

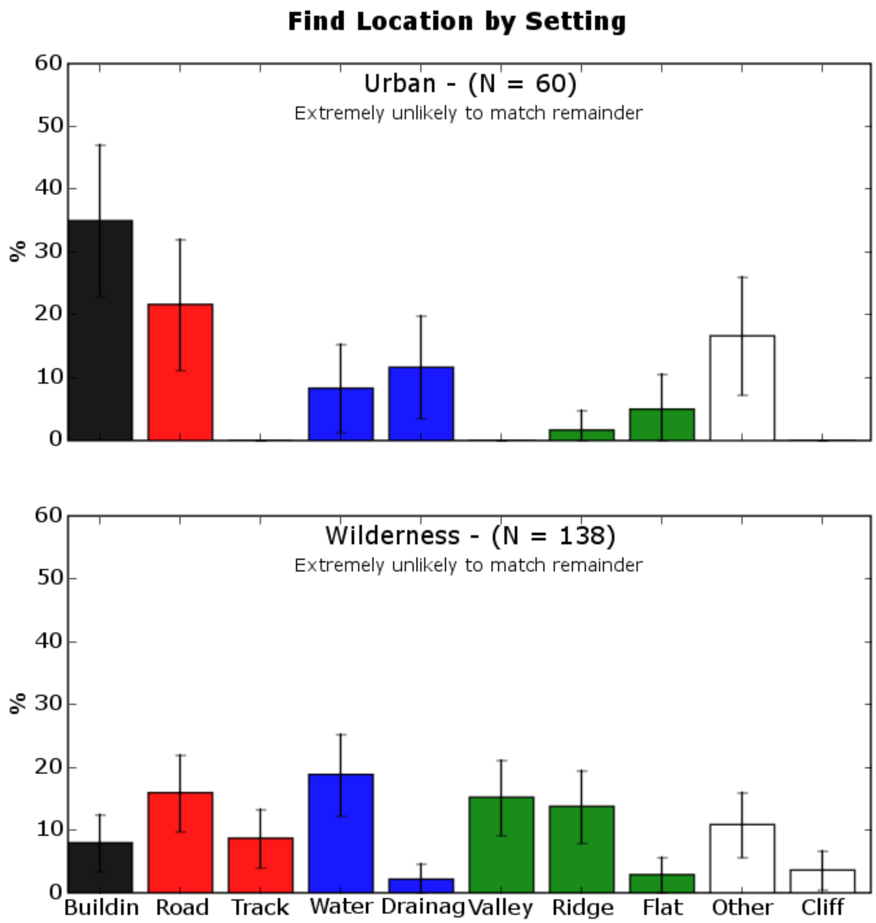

Figure 2.18: Find Location by Setting in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

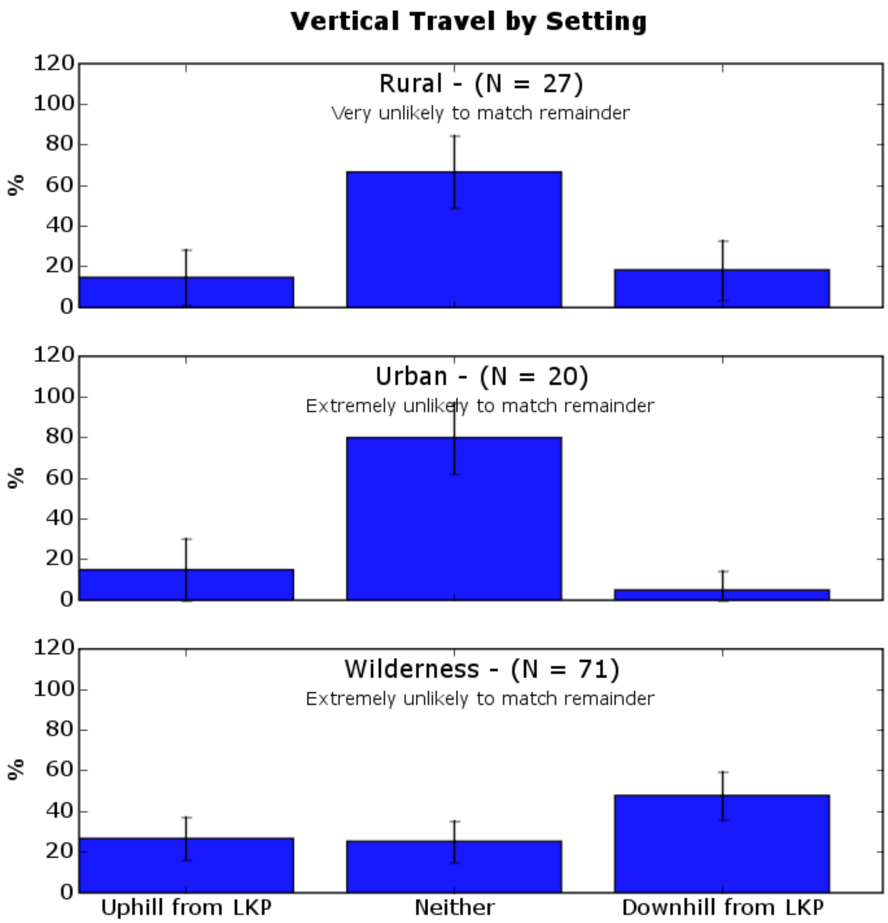

Figure 2.19: Vertical Travel by Setting in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

TradCateg

2.3.4 By TradCateg

			Percentiles			
	N	N_d	25%	50%	75%	95%
Alzheimers (DAT)	55	30	0.5	1.3	4.0	28.2
Autistic	8	5	- -	- -	- -	- -
Child	62	34	0.6	1.1	2.0	5.0
Despondent	47	23	0.6	1.4	2.0	23.5
Hiker	131	72	1.5	3.2	8.1	17.4
Hunter	13	10	- -	- -	- -	- -
Mentally retarded	16	7	- -	- -	- -	- -
Motorist	8	6	- -	- -	- -	- -
Other	88	33	1.0	2.2	8.0	26.0
Psychotic	28	17	0.5	1.0	3.8	10.2

Table 2.8: Distances (km) from LKP, by TradCateg in MP-All. N is the total number of cases in this dataset, and N_d is the number reporting d. (50% is the median.)

	N	N_s	Unhurt	Injured	Fatality	No Trace
Alzheimers (DAT)	55	55	25 (45%)	25 (45%)	5 (9%)	0 (0%)
Autistic	8	8	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Child	62	61	55 (90%)	4 (6%)	2 (3%)	0 (0%)
Despondent	47	47	18 (38%)	15 (31%)	10 (21%)	4 (8%)
Hiker	131	130	106 (81%)	20 (15%)	3 (2%)	1 (0%)
Hunter	13	13	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Mentally retarded	16	16	11 (68%)	5 (31%)	0 (0%)	0 (0%)
Motorist	8	8	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Other	88	88	51 (57%)	18 (20%)	14 (15%)	5 (5%)
Psychotic	28	28	10 (35%)	8 (28%)	5 (17%)	5 (17%)

Table 2.9: Status by TradCateg in MP-All. N is the total number of cases in this dataset, and N_s is the number reporting status. (Percents rounded to nearest whole number.)

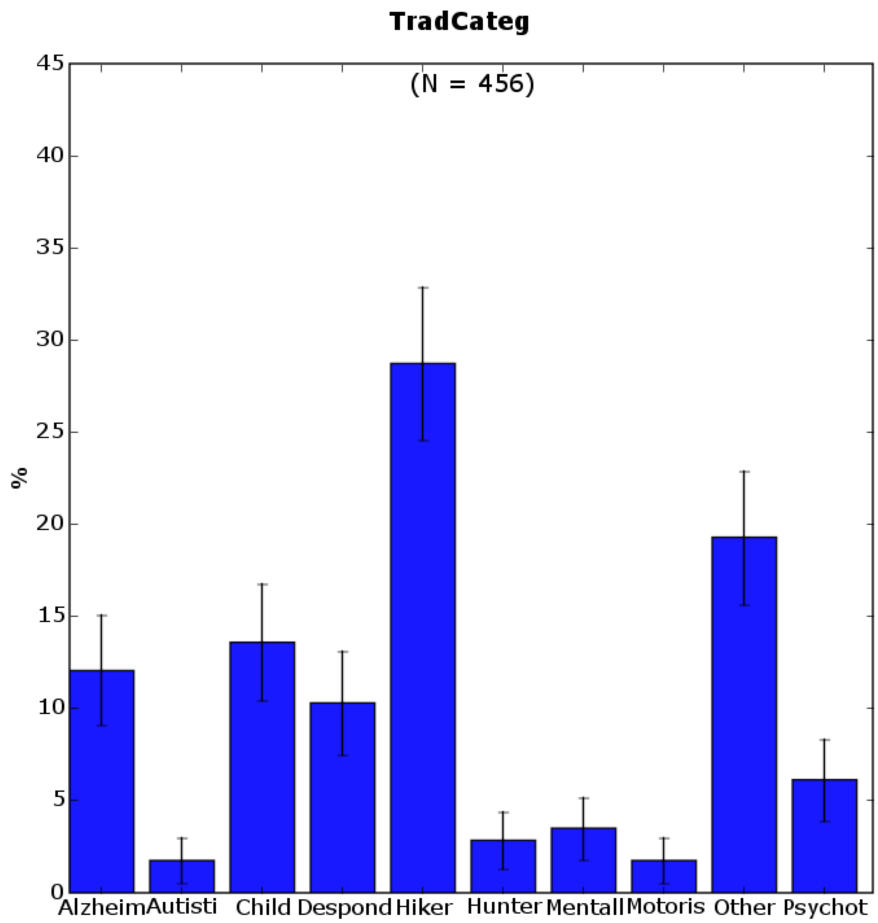

Figure 2.20: Distribution of TradCateg in MP-All.

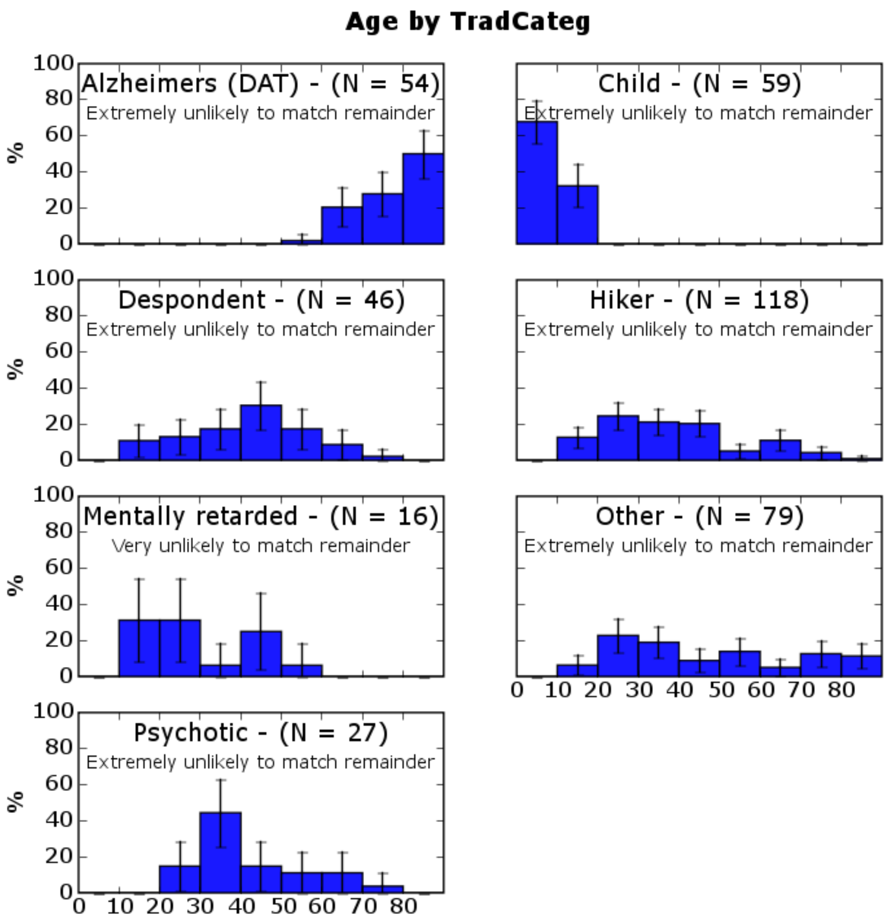

Figure 2.21: Age (yrs) by TradCateg in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

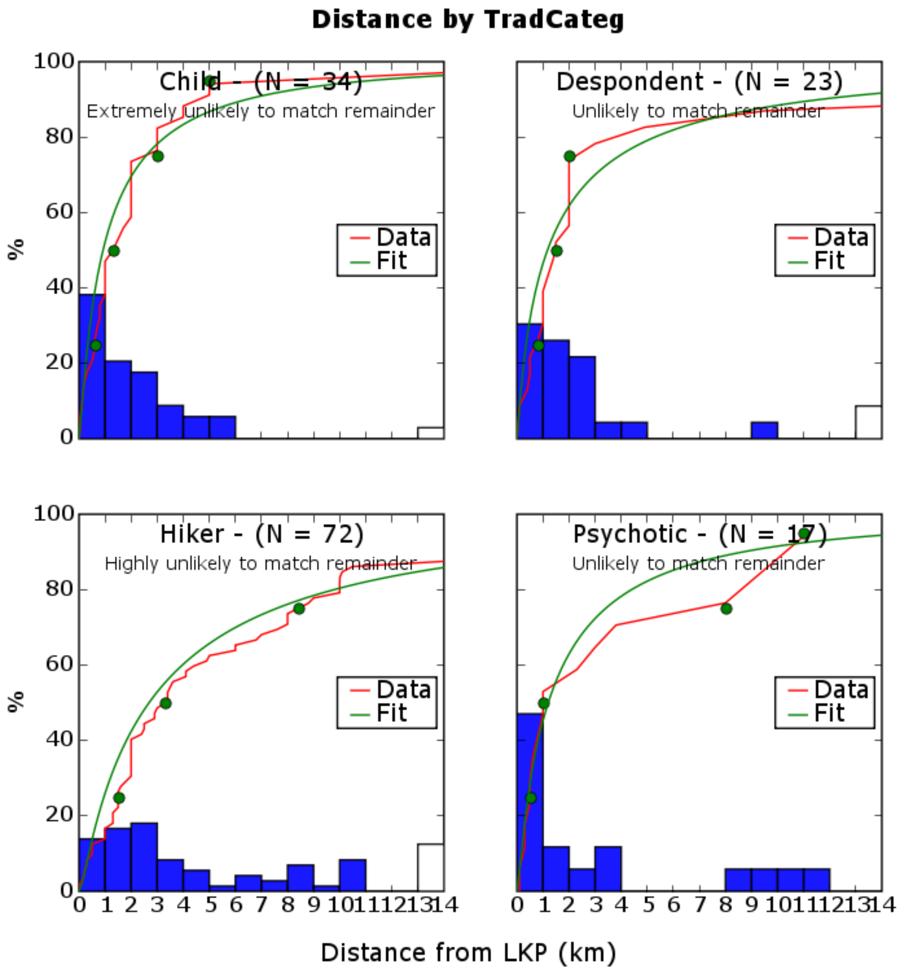

Figure 2.22: Distance From LKP (km) by TradCateg in MP-All. Lines show cumulative percent. Dots mark 25%, 50%, 75%, and 95% points (some of which may be off the graph). Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

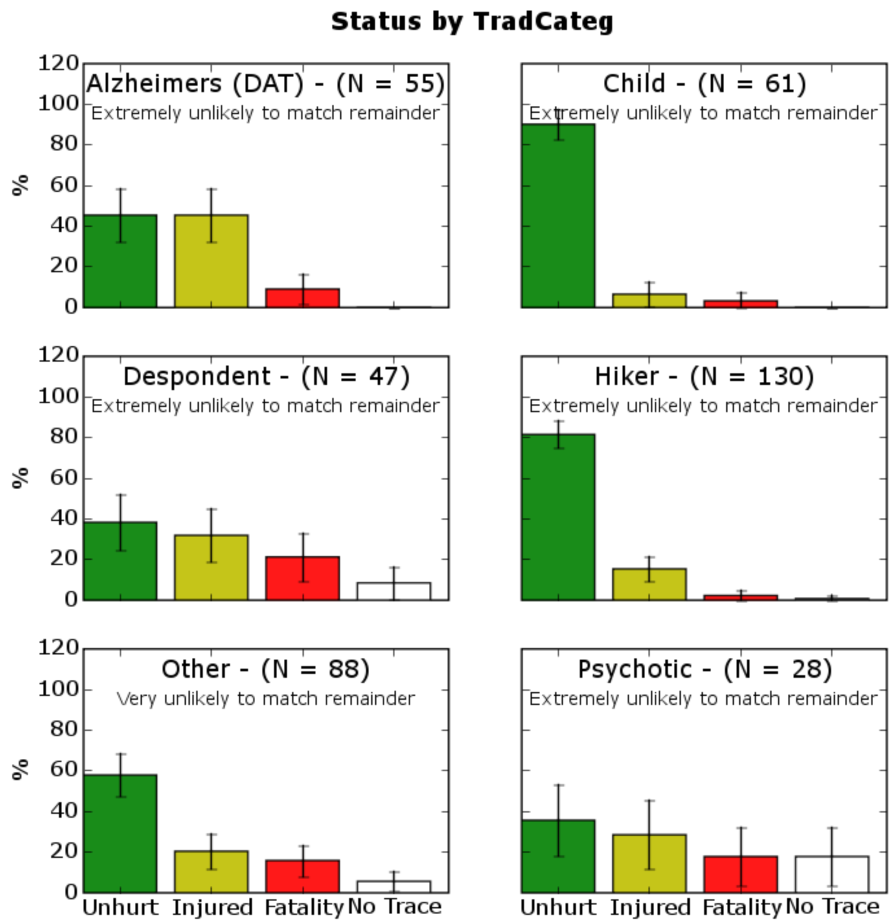

Figure 2.23: Status by TradCateg in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

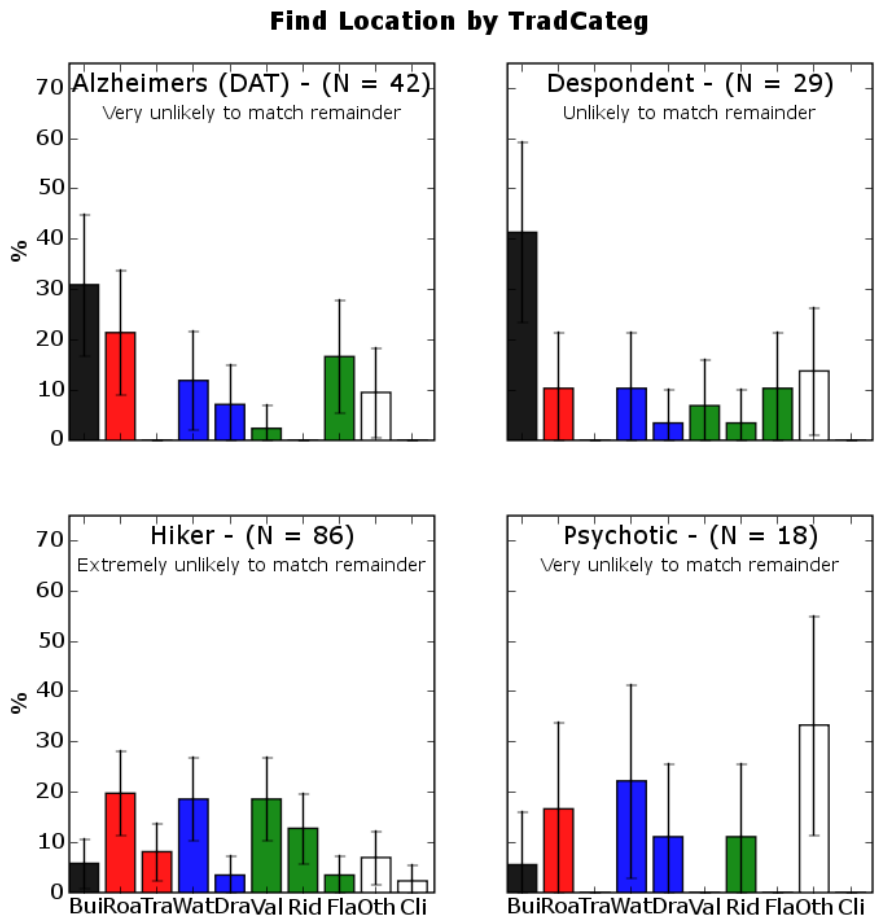

Figure 2.24: Find Location by TradCateg in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

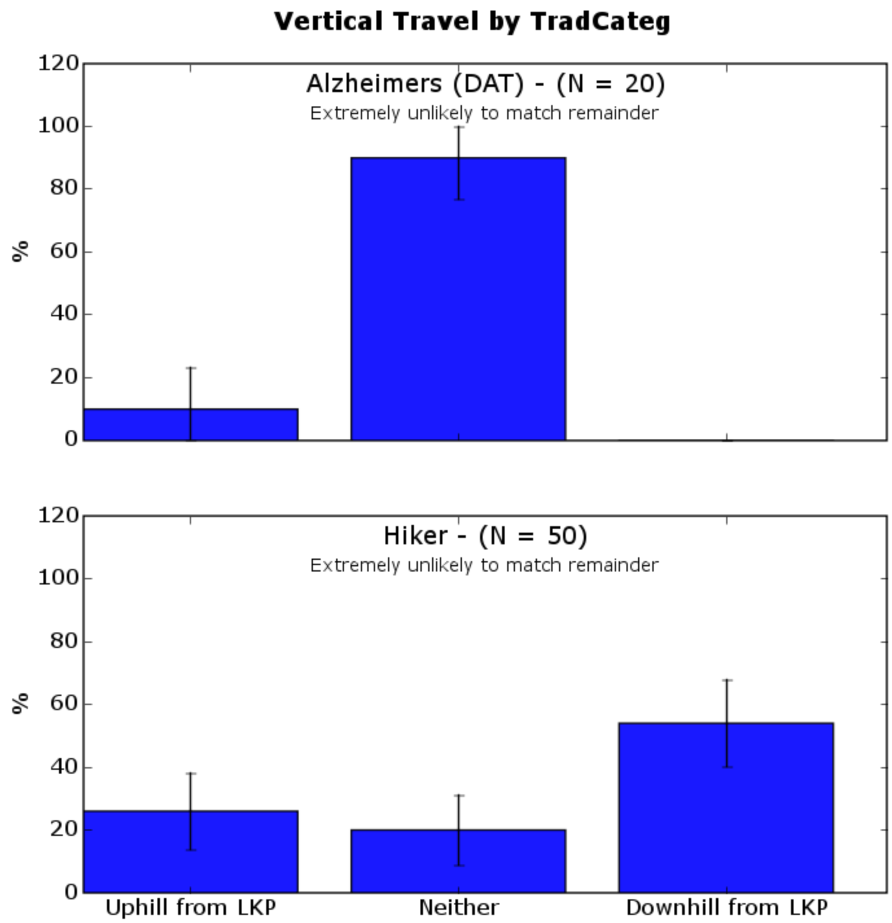

Figure 2.25: Vertical Travel by TradCateg in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

Activity

2.3.5 By Activity

			Percentiles			
	N	N_d	25%	50%	75%	95%
4WDriving	2	1	- -	- -	- -	- -
Alpine skiing	7	3	- -	- -	- -	- -
Backpacking	28	20	1.5	3.0	7.7	19.2
Canoeing etc.	6	0	- -	- -	- -	- -
Climbing	7	1	- -	- -	- -	- -
Cycling	7	1	- -	- -	- -	- -
Dayhiking	93	45	1.3	2.4	7.0	13.2
Driving	5	5	- -	- -	- -	- -
Fishing	2	1	- -	- -	- -	- -
Hunting	10	9	- -	- -	- -	- -
Motorbiking	1	1	- -	- -	- -	- -
Nordic skiing	2	2	- -	- -	- -	- -
Other	48	27	1.5	6.0	10.5	25.0
Runaway	36	21	0.2	1.4	3.0	14.0
Suicide	20	8	- -	- -	- -	- -
Walking	16	8	- -	- -	- -	- -
Wandering	152	75	0.6	1.5	3.8	13.4

Table 2.10: Distances (km) from LKP, by Activity in MP-All. N is the total number of cases in this dataset, and N_d is the number reporting d. (50% is the median.)

	N	N_s	Unhurt	Injured	Fatality	No Trace
4WDriving	2	2	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Alpine skiing	7	7	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Backpacking	28	28	24 (85%)	4 (14%)	0 (0%)	0 (0%)
Canoeing etc.	6	6	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Climbing	7	7	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Cycling	7	7	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Dayhiking	93	93	77 (82%)	14 (15%)	2 (2%)	0 (0%)
Driving	5	5	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Fishing	2	2	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Hunting	10	10	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Motorbiking	1	1	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Nordic skiing	2	1	- - (- -%)	- - (- -%)	- - (- -%)	- - (- -%)
Other	48	48	26 (54%)	11 (22%)	8 (16%)	3 (6%)
Runaway	36	35	21 (60%)	12 (34%)	0 (0%)	2 (5%)
Suicide	20	20	5 (25%)	4 (20%)	8 (40%)	3 (15%)
Walking	16	16	8 (50%)	3 (18%)	4 (25%)	1 (6%)
Wandering	152	152	90 (59%)	48 (31%)	10 (6%)	4 (2%)

Table 2.11: Status by Activity in MP-All. N is the total number of cases in this dataset, and N_s is the number reporting status. (Percents rounded to nearest whole number.)

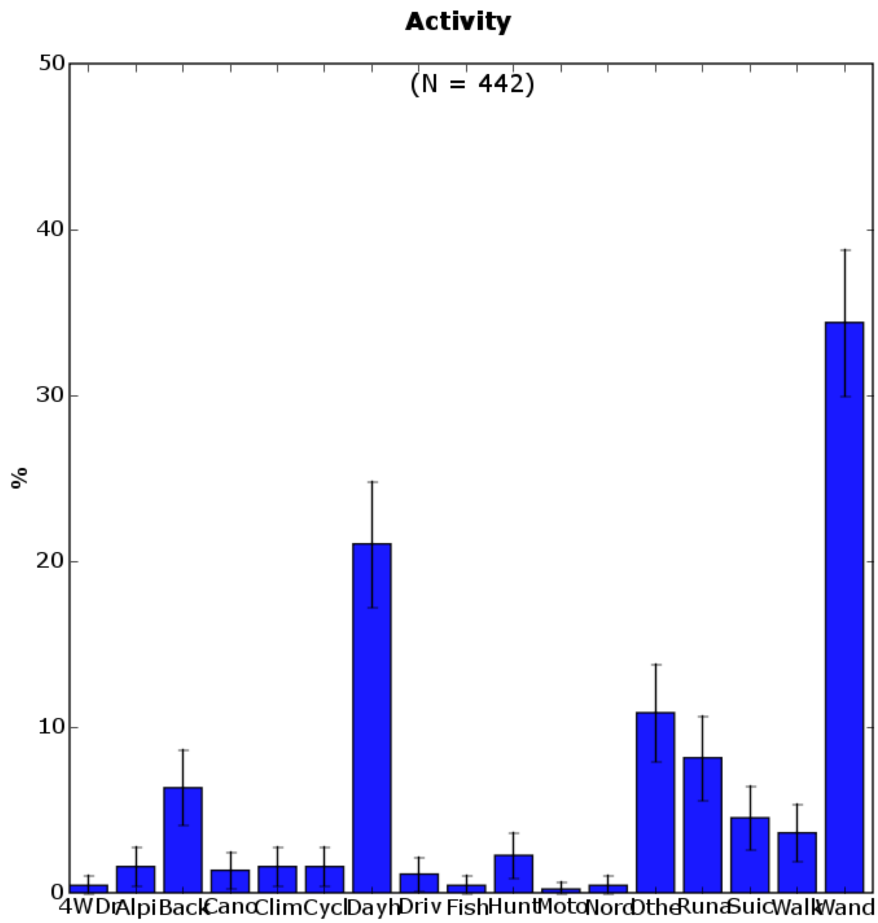

Figure 2.26: Distribution of Activity in MP-All.

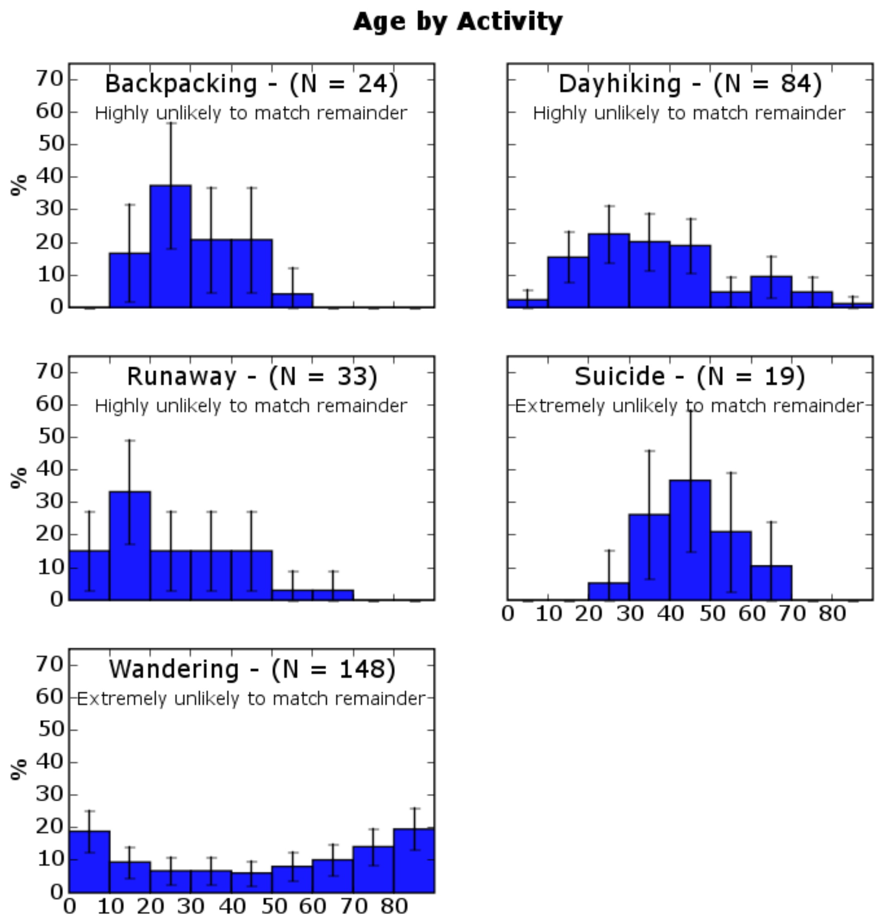

Figure 2.27: Age (yrs) by Activity in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

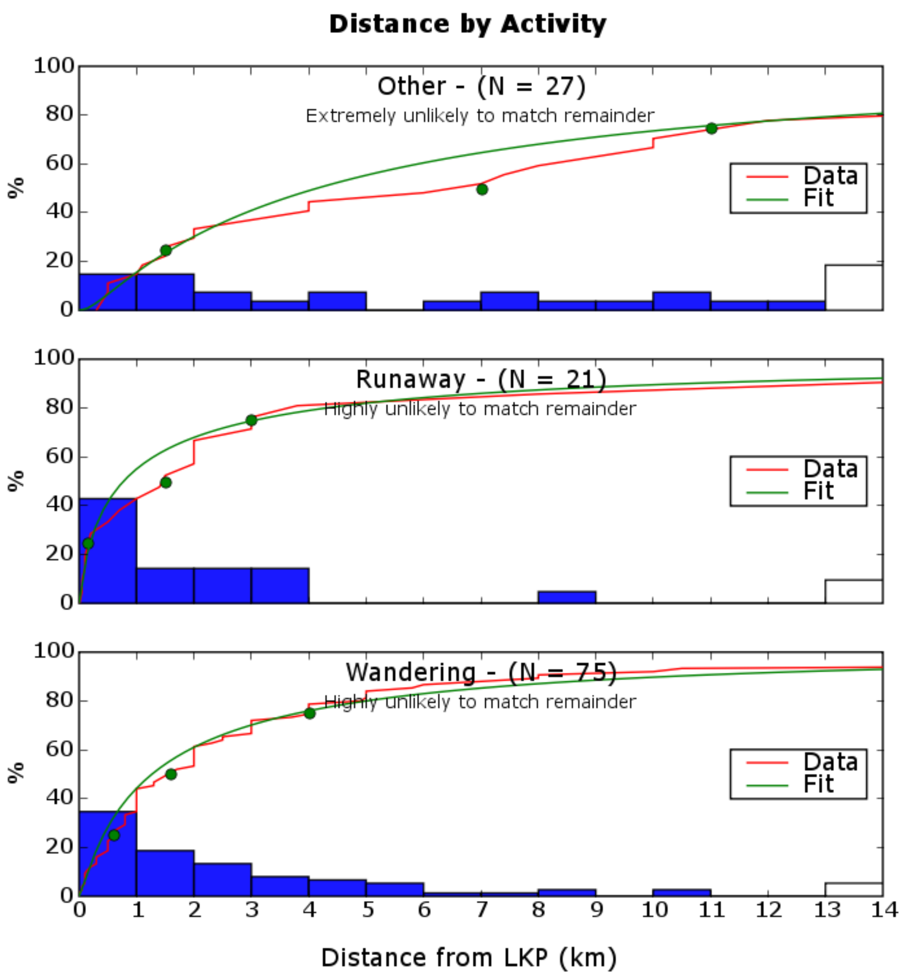

Figure 2.28: Distance From LKP (km) by Activity in MP-All. Lines show cumulative percent. Dots mark 25%, 50%, 75%, and 95% points (some of which may be off the graph). Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

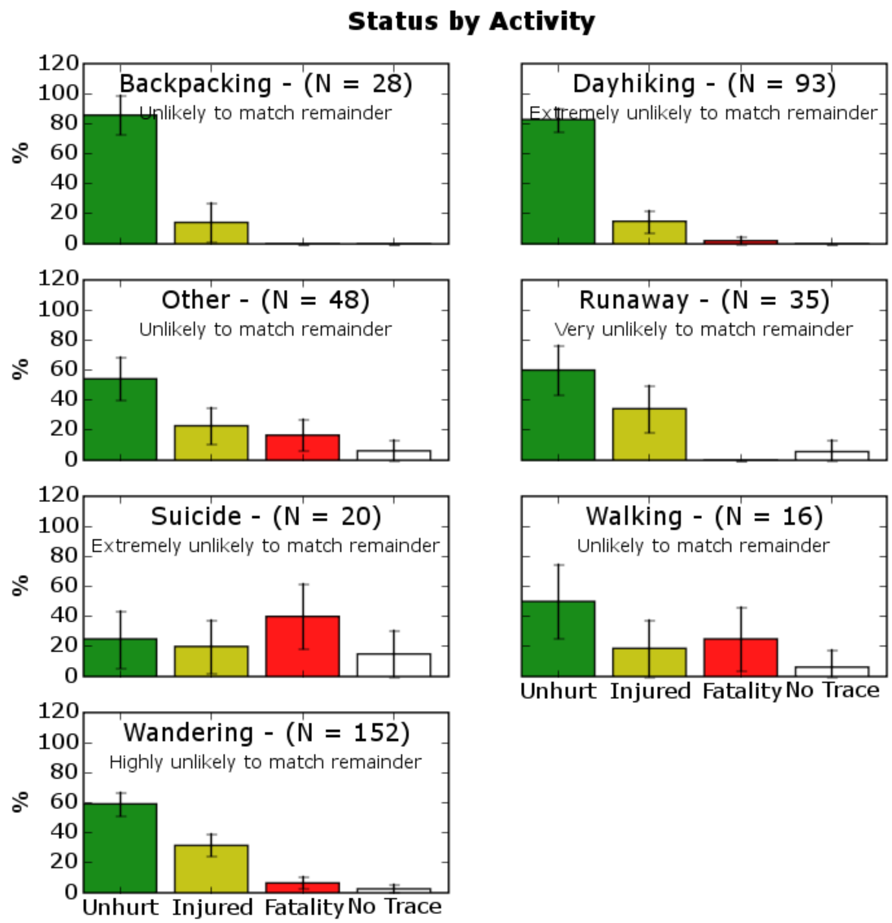

Figure 2.29: Status by Activity in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

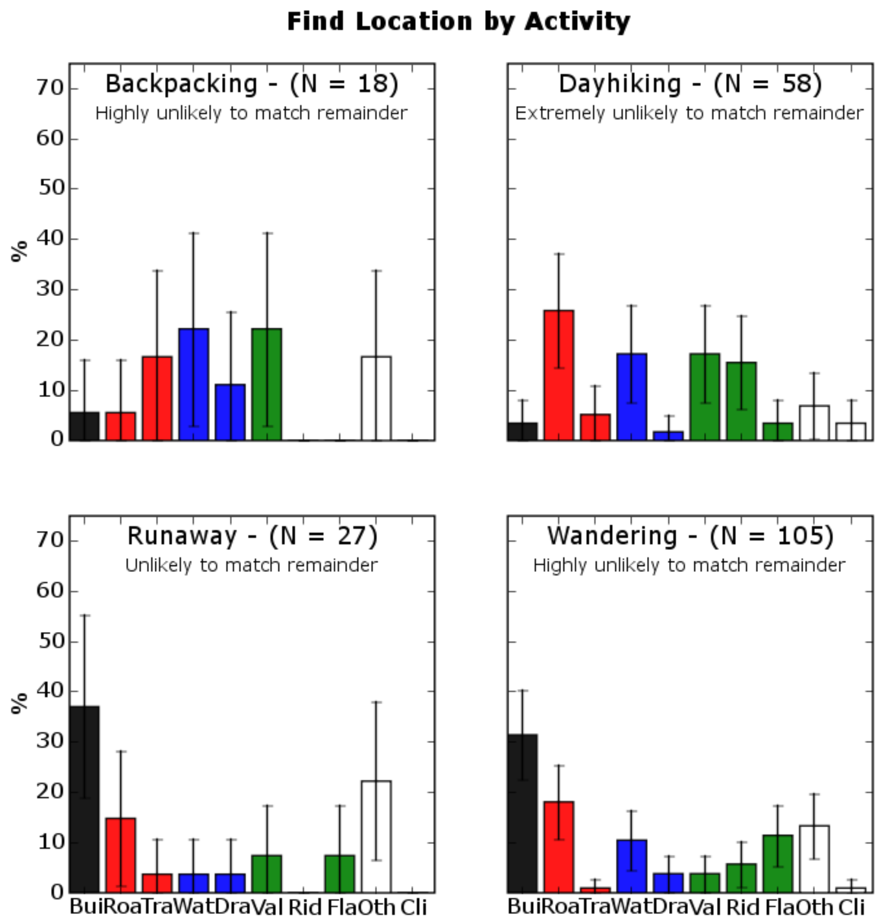

Figure 2.30: Find Location by Activity in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

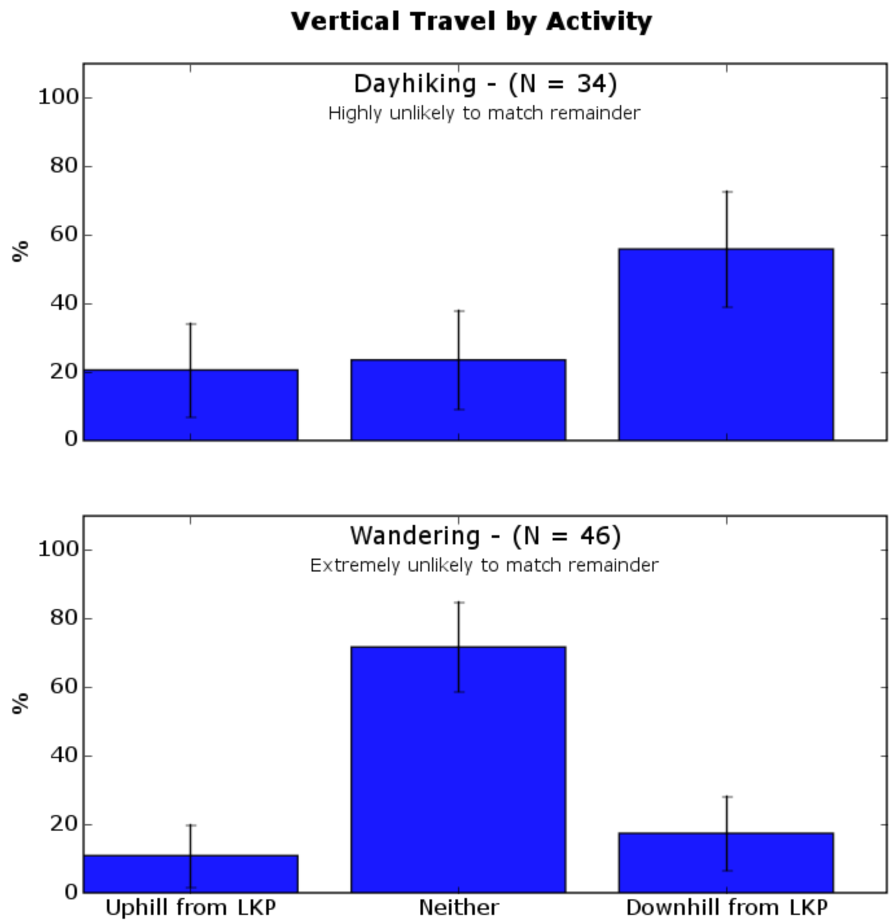

Figure 2.31: Vertical Travel by Activity in MP-All. Bars show 95% confidence intervals. Only categories *differing* from the remaining data are shown. For others, see Section 2.3.1. Categories with fewer than 15 cases not shown.

Final comments?

Chapter 3

Groups vs. Singles

This chapter compares groups to singles on several variables, especially Status and Distance. We begin with an overall summary, in keeping with the draft report. However, since Despondents and Alzheimers rarely travel in groups, we then ask whether there is still a reliable difference on a matched category: Hikers.

Basic psychology and subject reports certainly suggest that groups should be less likely to panic, so for many reasons we would expect groups to do better. We might also hope that with 2 or more people, there is a better chance of keeping oriented, etc. But group dynamics can play out in many ways. It is possible that each person in the group suspects they are going the wrong way, but says nothing because they defer to what they think everyone else knows. Or perhaps worse, if 2 or more people make the same error, it can lead to overconfidence (groupthink). Group makeup may encourage "showing off" by some members. In the future, we hope to examine a few basic kinds of group compositions separately.

Figure 3.1 shows the distribution of group sizes.

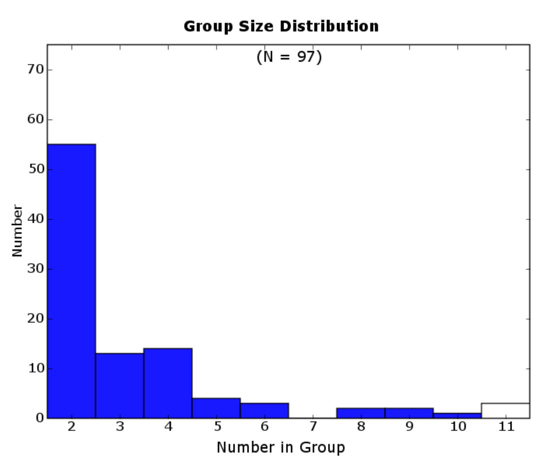

Figure 3.1: Distribution of Number Lost, in Groups. Maximum group size was 22

	Rate	N_{tot}
Fatality rate for singles	11%	361
Fatality rate for individuals in a group	1%	339
Percentage of groups with at least 1 fatality	2%	97
Percentage of groups with at least 1 **survivor**	99%	97

Table 3.1: Comparative fatality and survival rates: groups vs. singles.

Variable	Overall	Hikers only
Status	Likely to be the same	Likely to be the same
Distance	Extremely unlikely to be the same	Unlikely to be the same
Find Location	Extremely unlikely to be the same	Extremely unlikely to be the same
Vertical Travel	Extremely unlikely to be the same	Unlikely to be the same
Seeking Help	Extremely unlikely to be the same	Unlikely to be the same

Table 3.2: Do Groups and Singles differ on the listed variables? All groups represented by their First case. The second column controls for Traditional Category by selecting Hikers, the only category for which Groups have a substantial number of cases.

3.1 Overall Fatality Rates

We begin with *overall* fatality and survival rates for Groups and Singles. Table 3.1 compares fatality and survival rates for Groups and Singles. The fatality rate for Groups assumes that all fatalities have been recorded in the database. (We rarely have separate records for all members of a group, but we presume we have the records for the fatalities.)

⚠ **CAUTION** ⚠

These may not be meaningful: Despondents has the highest fatality rate, yet how many are represented as groups?

3.2 Groups vs Singles

Table 3.2 reports on the differences between distributions (for Groups vs. Singles) on several variables. (See Appendix B for details). For this test we use the Status of the *first* member of the group, rather than the worst-case.

3.3 Comparative Tables

We provide some distance and status summary tables showing the comparative distributions for Singles and Groups, in Hikers. Group status is represented by the first case.

			Percentiles			
	N	N_d	25%	50%	75%	95%
Group	68	34	2.0	4.0	10.0	17.4
Single	63	38	1.1	2.7	6.0	15.0

Table 3.3: Distances (km) from LKP, by Group in Hikers (Firstcase). N is the total number of cases in this dataset, and N_d is the number reporting d. (50% is the median.)

	N	N_s	Unhurt	Injured	Fatality	No Trace
Group	68	68	59 (86%)	9 (13%)	0 (0%)	0 (0%)
Single	63	62	47 (75%)	12 (19%)	2 (3%)	1 (1%)

Table 3.4: Status by Group in Hikers (Firstcase). N is the total number of cases in this dataset, and N_s is the number reporting status. (Percents rounded to nearest whole number.)

Appendix A

Definitions

A.1 Incident Types

The new field IncidentType answers the question, "What kind of response was this?" We use it here to select only Missing Person cases for analysis. We defined the following incident types, in accordance with the ISRID.

- **Missing Person**: our target case: people missing on land, usual notification
- **Water**: anything that started AND ends in water. Ex: canoeing w/o lifejacket, drowned
- **Rescue**: we know where they are, we go help. Ex: injured climbers
- **Recovery**: we know where they are, they're dead, we go get. Ex: climber who fell
- **ELT/EPIRB**: incident started by an ELT or EPIRB signal. The base does not know if it is a MP or Rescue, etc.
- **Criminal**: usually abduction
- **PLB:** incident started by a PLB signal. As with ELT/EPIRB, but because PLB is new, some databases may track them separately to measure increased usage.
- **Training:** a planned training event.
- **Disaster:** natural or anthropogenic disaster, mass casualty, etc.
- **Evidence:** SAR units called to find physical evidence for investigation, criminal prosecution or similar. Often looking for small fragments.

The Australian database has no PLB, Training, Disaster, or Evidence cases. PLB cases are normally handled by AusSAR, not the police SAR units. The other three were screened out before data entry.

A.2 Scenario

The new field Scenario asks, "given all else, what do we think happened?" It is a retrospective assessment. It expands and replaces our earlier Lost/NotLost field, which was inadequate for distinguishing "genuine" cases from false alarms. Values are:

- **Criminal**: person missing against their will. Example: abduction/murder.
- **Despondent**: *actively* Suicidal. Trumps 'Evading' and most other things. Many, but not all those with TradCateg of 'Despondent' get this scenario. Some are depressed, but just out for a walk and get lost. But any with 'suicidal' or 'suicide' in some field were put in this scenario.
- **Evading**: *deliberately* missing, *hiding* from at least some searchers, and *not* suicidal (else Despondent). Lots of psychotics here, some children, plus those decamping the scene of an auto accident to avoid DUI/DWI charges. Many of those with Hiding=Yes get this scenario.
- **Investigative**: false alarms, often called "bastard searches." MP is often unaware of the search, and happy elsewhere. Often MP failed to notify, there was a miscommunication, etc. Solved by investigative techniques. Example: husband drives off to Ballarat for the night, or aunt forgets that father was picking child up today.
- **Lost**: just plain lost — disorientation is main or major reason they're missing. Alzheimers are here because we presume they were at some stage confused.
- **Medical**: the reason they're missing is a heart attack or such.
- **Overdue**: *never* lost, doing OK, just taking longer than expected. May or may not need assistance. Example: experienced backpackers meet a swollen river and wait the night for it to diminish. Also, most bogged vehicles go here. They may have a long walk back, but they know where they are. Usually they are either waiting or walking back along road.
- **Trauma**: missing because of injury or major mishap like capsize/drown. Does *not* include bogged!

A.3 Traditional Categories

Alzheimers (DAT): The category "Alzheimer's (DAT)" denotes those with Alzheimer's disease and Dementia of the Alzheimer's Type. Although technically we should always use "DAT," it is common to refer to the broad group as "Alzheimer's" even though Alzheimer's can only be diagnosed by autopsy (so far).

Autistic: The category "Autistic" denotes those diagnosed with autism, Asperger's syndrome, or possibly related traits. These cases are usually children, but "Autistic" is the more salient category. Children who are known to be autistic are classified as "Autistic."

Child: The category "Child" covers most children, usually taken to be 12 years old or younger. However, medical categories like "Autistic" and "Mentally Retarded" usually trump "Child."

Despondent: The category "Despondent" covers depressed subjects and includes those known to be suicidal, but not exclusively those. People being treated for depression sometimes get lost even when not trying to kill themselves.

Hiker: The traditional category "Hiker" includes all those on some kind of directed walk. In Australia, the proper term would be "Bushwalker". "Hiker" includes those whose Activity is either dayhiking or backpacking.

Hunter: "Hunter" denotes anyone hunting any kind of game. However, for cases described as "collecting roos from the roadside" we replaced it with "Motorist". Typically hunters travel off-track, and are prone to different sorts of mishaps, so are worth considering separately. We have very few Hunters in the dataset.

Mentally Retarded: "Mentally Retarded" covers patients with many disorders that slow mental development, resulting in a "mental age" notably below the subject's physical age. One of the better-known of these disorders is Down's syndrome, but that is only one of many. Note that we do *not* currently measure the *level* of retardation (for example by recording "mental age"). Were we to do so, no doubt we could refine the profile. Note: a 30-year-old with a "mental age" of 10 nevertheless has 20 years' extra life experience.

Motorist: "Motorist" includes anyone who was in a motor vehicle when they became lost, stranded, injured, etc. Many of these are "Vehicle" cases (such as the elderly gentleman who kept driving past his destination, until he ran out of fuel), which get screened out. However, if the vehicle serves as the LKP for subjects travelling on foot (or waiting), these are legitimate cases for our purposes. Such MPs typically are not lost, but also typically are not expecting to be travelling on foot. "Motorist" includes Activities "4WD," "Motorbike," and "Car."

Other: Unsurprisingly, the category "Other" covers everything not listed in any other category, such as birdwatchers, berry-pickers, station hands returning to their stations, or geologists out prospecting. Their profile should not be considered that of a particular class of people, but rather a best prediction in a state of ignorance given that the subject is *definitely not* one of the known categories.

Psychotic: "Psychotic" includes patients who have clinical psychosis and also those with temporary psychosis such as that induced by drugs or alcohol, particularly when the psychosis is considered to be a major factor in the incident.

A.4 Activity

Most of these are self-explanatory. Some, like Wandering and Runaway are hard to tease apart, and we try to *emphasize* any non-obvious decisions. We did not set criteria in advance, so we report here how we came to divide the reported data, especially the many cases which did not fit our initial category divisions.

- **4WDriving:** Out in a 4WD vehicle on 4WD tracks.
- **Alpine Skiing:** Downhill skiing. *Includes snowboarding.*
- **Backpacking:** Bushwalking with a pack, intending to stay overnight and prepared for it.
- **Canoeing etc.:** Canoeing, Rafting, Kayaking, etc. We did not have enough to split them up.
- **Climbing:** Climbing or abseiling (rappelling): ascending or descending cliffs, usually with technical gear.
- **Cycling:** Travelling by bicycle, with no or little motor assist. Includes street and mountain bikes.
- **Dayhiking:** Bushwalking intending short duration. Not carrying overnight gear.
- **Driving:** Driving a car, not intending other activity, not a dedicated 4WD trip. Lots of stranded/bogged scenarios, including (perhaps wrongly?) the 6 people collecting roos by the roadside. The idea is people who were expecting to just be out for a regular drive, and had a mishap.
- **Fishing:** Any sort. The trip's intent is to catch fish.
- **Hunting:** Any sort. Excludes collecting roadkill, etc.
- **Motorbiking:** *Motorcycle, Quad bike, ATV.* Not quite a car, usually not on paved roads.

- **Nordic Skiing:** Cross-country skiing.

- **Other:** Borderline cases, unknown or unclear activities, or unusual or diverse things like: chasing camels, visiting friends, at casino, some psychotic episodes, cutting or gathering wood, at a rave, beach, bird-watching, hangliding, paragliding, crossing river, rogaining, firefighting, prospecting, ...

- **Runaway:** Typically children, but includes *any deliberate attempt to escape, in people 65 years or less.* Runaways over 65 reclassified as Wandering. *This category overlaps with Wandering.*

- **Suicide:** Known or *strongly* suspected suicide attempt. Suicide "trumps": other activities (like Hiking, or 4WDriving) become secondary to the suicide attempt.

- **Walking:** MP was out for a regular walk. Includes children walking home from school, people walking their dogs, etc. *Could* be reclassified as Wandering or such, but these just didn't seem quite right there.

- **Wandering:** Any aimless wandering (children on up), disorientation, or confusion. Most Alzheimer's searches, even if the MP seemed to know what was going on, and any escape behavior in those over 65, even if it otherwise would be Runaway.

A.5 Find Location

⚠ GENERAL NOTE: this category is flawed. The values are not exclusive: 'stream' is also 'valley', 'track' is often on a 'ridge', etc. Neither were respondents given a list of definitions.

- **Building:** Pre-made structure, from a hut or shed to a hospital. Including one houseboat.

- **Road:** Usually paved, but may include some 4WD tracks.

- **Track:** Usually walking track or trail, but also 4WD track and desert track.

- **Water:** In *or next to* water. Includes stream, river, lake, riverbank, dam. Consolidates many entries like "near dam".

- **Drainage:** Stream, river, ditch, culvert. Anything that at least occasionally drains water away. Not consolidated to "Water" because these are often dry, and we can't tell from the data entry.

- **Valley:** User chose Valley over Drainage. Presumably therefore more likely to be dry, or broad.

- **Ridge:** Including peaks, etc. May also be Track, but user chose Ridge as the best fit.
- **Flat:** Includes park, beach, similar.
- **Cliff:** Added because several cases mentioned 'cliff' in their text. Found on or at the base (or edge) of a cliff.
- **Other:** User entered "Other" and any additional info did not enable us to choose a category, or make a new one. Including: steep sloping ground, dunes, CBD, Bush, National Park.

A.6 Comparison to UK Categories

A.6.1 Traditional Category

There is a rough, but imperfect mapping between categories in the UK report and our categories. The following table may help for comparison.

UK 2004 Category	**SAR*Bayes* Category**	**SAR*Bayes* Notes**
Child (1 to 6 year) Child (7 to 12 year)	Child	
Despondent	Despondent	
Climber Fellrunner Mountain Biker Skier	*Not used*	Some of these will show up in our "activities" field.
Youth (13 to 16 yr.)	*Not used*	*Might* be child, if no other.
Vulnerable	Alzheimer's Mentally Retarded Psychotic	We split "Vulnerable" into 3 classes to cope with the very different mental processes of the groups.
Hiker/Walker	Hiker	We allow under 17, in theory.
Miscellaneous	Other	But Other also has Skiers etc.
Organised Party	Group	We don't require "recognised leader or purpose".
Not used	Autistic	In UK, probably Child or Vulnerable.
Not used	Hunter	

A.6.2 Conditions

We use the same categories for subject condition as the UK report, and so we have used the UK labels throughout the report. Here is how they describe the terms:

Fatality	dead when found
Injured	required significant medical treatment when found
Unhurt	not Injured
No Trace	not located, outcome not known

Appendix B

Statistics

B.1 Reporting likelihood: U.K. conventions

We wish to highlight when a subpopulation (like Despondents) differs in a reliable way from the overall population (of Missing Persons). Often, but not always, one can trust intuition, if given the sample size. However, it is customary and perhaps even beneficial to include some measure of the confidence we have in apparent differences. We do so by putting a note to this effect in the figures. (In fact, we only *show* figures for subpopulations that seem to be reliably different from the remainder.)

So, for example, when looking at Status of Despondents, we see that they have many more fatalities than other people. We calculate the chance of getting such an outcome *if* the real probabilities were actually the same as for non-Despondents. The smaller that chance (which is usually called p), the more confidence we can have that the apparent difference is real.

Rather than reporting p directly, we have adopted the plain-language phrases suggested in Perkins et al. (2005). They are:

Probability (p)	**Chance**	**Phrase**
$p < .01$	$< 1\%$	Extremely unlikely to match...
$p < .05$	$< 5\%$	Highly unlikely to match...
$p < .1$	$< 10\%$	Very unlikely to match...
$p < .25$	$< 25\%$	Unlikely to match...

Where the U.K. reports also say "could possibly have occurred by chance" when $p > .25$, we simply say nothing. Given the very relaxed standards, it's worth taking a "no difference" result seriously.[1]

[1]The early U.K. reports (Perkins et al., 2002, 2001) used much more demanding p values: "extremely unlikely to" was reserved for $p < 0.001$ while everything with $p > 0.10$ became "could possibly". From a straight statistical standpoint, we would be happier with the more stringent requirements. However, we can regard the relaxed criteria as reflecting the strong prior beliefs that categories (etc.) do indeed matter.

> **NOTE**
>
> We chose to compare a category (like Despondents) with the *remainder* (like non-Despondents) rather than the overall population. The reason is that some subpopulations (like Hikers) will not vary much from the overall population because largely, they *are* the overall population!
> Also note: a subpopulation that does not differ from the remainder may still differ from other subpopulations! We would have to make a separate comparison, as we undertake in the chapter on Groups vs. Singles.

B.1.1 Agonizing details

The numerically inclined may wish to know that except for Distance, we use a "chi-squared" (χ^2) test – a standard test for comparing two discrete (i.e. "binned") distributions. We use the test in the statistical package "R", called from Python via the "rpy" interface because it can use simulation when there are too few observations in a bin to meet the assumptions of χ^2.

For Distance, we found that the log of Distance closely approximates a Normal distribution, which is quite sensible, so:

1. We could fit lognormal curves to the data, allowing fairly robust estimates as these need only two parameters.

2. We could perform a t-test on the difference of means (of log distance), which is a better test for continuous data (as long as the distribution fits).

We note that the log-transform and t-test was also recommended by Heth and Cornell (1998). Heth & Cornell first clustered groups that might have somewhat differently-shaped curves, before performing t-tests to detect differences *within* the two clusters they found.

We were able to use their technique manually, but could not easily automate it. We did perform a cluster analysis that we think is at least as informative, but decided not to use it for this report, as people might be confused as to why Hikers, Hunters, and Autistics were in the same cluster.[2]

The lognormal is a good choice for many reasons, but we do not seek to defend it as uniquely appropriate. The Weibull and Gamma (and no doubt others) have similar shapes and properties, and would probably fit about as well. It would take more data to reliably distinguish between them.

[2] In this case, because we were comparing only on Distance, and Hikers, Hunters, and Autistics in our dataset tended to travel further than average, but close enough to each other that no further splitting was warranted by our distance data. (Including other data would almost certainly have split them.) The second cluster (Alzheimers, Psychotic, Despondents, and Children) tended to go less far, and the 6 motorists form their own weird little group that travelled very far.

For determining curve shape, Heth & Cornell used the Wakeby because it can be expressed nicely in quantile form and is quite flexible. Given the range of shapes that lognormal, Weibull, and Gamma can already take, we would argue that 5 parameters is somewhat too flexible, and prefer these two-parameter curves.[3] In practice there was little difference, because their software ended up fixing two parameters to 0, and estimated the lower bound to be 0 for one cluster, and nearly 0 for the other, leaving essentially a two-parameter Pareto curve.

B.2 Error Bars:

In addition to reporting a sense of reliability, we included error bars to provide an immediate visual sense variability. We feel they are worth the additional clutter. The less they overlap, the more likely the two estimates are to really be different. They also help focus the reader on the real task, which is estimating probabilities, rather than on significance tests.

Our error bars show standard 95% “confidence intervals”. These are known to be conservative: that is, somewhat too wide. But correction techniques (GraphPad Software, 1999) would make it harder to compare our graphs with other reports.

We also considered showing absolute numbers rather than proportions. However, both Mitchell (1985) and Perkins et al. (2005) show proportions, and doing so makes it easier to compare figures. We provide raw numbers in the tables, and we show the sample size in every figure.

[3]Well, 3-parameter, but fixing the lower bound to 0, since some people *are* found at the PLS.

Appendix C

Variable Names

Here are the field (variable) names and the percentage of cases where that field is unknown, as of 2006-06-03. Percentage unknown is given for *all* cases, with groups represented as a single case. Overall, there were **550 cases**, and there were **108 fields**. Most of these fields have not been used in the analysis. They are listed here for the benefit of other investigators, who may wish to know what is available. (The file was generated on 2006-06-03 by running listfields.py on data.csv.)

Fields are: **#, Variable Name, % Blank**.

#	VarName	%Blank
0	KeyID	0
1	UserID	0
2	IncidentNum	0
3	DataEntryDate	0
4	PreparedBy	3
5	Organisation	0
6	Contact	6
7	Locality	6
8	City	7
9	State	0
10	PostCode	53
11	Local_Coord	76
12	IncidentType	0
13	Scenario	5
14	NumLost	0
15	Notes	4
16	SubjNum	0
17	Age	8
18	Sex	3
19	TradCateg	0
20	CategOther	83
21	Activity	3
22	ActivOther	82
23	NotFound	0
24	Status	0
25	Setting	7
26	DateLost	2
27	TimeLost	11
28	DateFound	5
29	TimeFound	15
30	TotalTimeLost	7
31	LKP	7
32	LKP_Coord	73
33	Find_Coord	81
34	DistFrLKP	49
35	FindLoc1	32
36	FindLoc1Other	84
37	FindLoc2	61
38	FindLoc2Other	98
39	FindLoc3	72
40	FindTechniques	20
41	DateCalled	7
42	TimeCalled	20
43	DateBaseClosed	1
44	TimeBaseClosed	15
45	MinPersonHrs	46
46	MaxPersonHrs	46
47	Openness	25
48	Steepness	26
49	Hazards	39
50	HazardOther	88
51	WxMinTemp	55
52	WxMaxTemp	55
53	WxWind	64
54	WxDesc	35
55	Mobility	17
56	Alertness	20
57	Consciousness	21
58	Visibility	48
59	Wet	52
60	Sheltered	54
61	Weight	63
62	Height	52
63	Build	42
64	Fitness	50
65	Impediment	53
66	Precondition	53
67	Experienced	39
68	AreaKnowledge	36
69	Personality	66
70	TraitsOther	94

71	Plans	51
72	PlansOther	89
73	TimeEvac	71
74	GearOther	85
75	Vehicle	8
76	Sighting	15
77	Confinement	54
78	Hasty	32
79	Efficient	40
80	Thorough	53
81	Grid	52
82	Mantracking	55
83	Dog	45
84	Mounted	40
85	Aerial	24
86	Attraction	55
87	Night	46
88	Radiobeacon	51
89	Car	64
90	DogType	91
91	MountType	65
92	AerialType	48
93	OtherMethods	84
94	Hiding	52
95	Seeking	51
96	Hypotherm	61
97	Hypertherm	63
98	Dehydrated	60
99	Injured	58
100	Waterproof	37
101	Windproof	38
102	Warm	43
103	ShelterGear	32
104	SleepingBag	33
105	Water	32
106	Food	32
107	Fire	39

Bibliography

Allison, L. (2005). Inductive programming. Website. http://www.csse.monash.edu.au/ lloyd/Seminars/2005-II/.

Allison, L. (2006). A programming paradigm for machine learning, with a case study of bayesian networks. In *ACSC2006*. Australian Computer Science Conference.

Csillag, F., Menes, P., Marosz-Wantuch, M., and Burchfield, M. (2000). Ground search and rescue (GSAR) baseline study. Report submitted to Emergency Measures Ontario, Department of Geography, University of Toronto, Ontario, Canada. `http://eratos.erin.utoronto.ca/fcs/PRES/gsar_rep.pdf`.

dbS Productions (2003). SARDISK 2003. CDROM. Published yearly.

Frost, J. (1999). Principles of search theory. *Response*, 17(2,3). `http://www.sarinfo.bc.ca/Library/Planning/PrincSrchThry_S.pdf`.

GraphPad Software, I. (1999). The confidence interval of a proportion. Website.

Heth, C. D. and Cornell, E. H. (1998). Characteristics of travel by persons lost in Albertan wilderness areas. *Journal of Environmental Psychology*, 18:223–235.

Hill, K. (1999a). Categories of lost persons. In Hill (1999b), chapter 5, pages 45–60.

Hill, K., editor (1999b). *Lost Person Behaviour*. Number 2 in Canadian Ground Search and Rescue Study. National Search and Rescue Secretariat, Canada, Ottawa, Ontario K1A 0K2.

Hill, K. A. (1991). Predicting the behaviour of lost persons. In *Proceedings of the Annual Meeting of the National Association for Search and Rescue*, volume 20, pages 159–175. NASAR.

Hunter, J. D. et al. (2004–). Matplotlib.

Jones, E., Oliphant, T., Peterson, P., et al. (2001–). SciPy: Open source scientific tools for Python.

Kelley, D. E. (1973). *Mountain Search for the Lost Victim*. Dennis E. Kelley, P.O. Box 153, Montrose, CA 91020, USA.

Koester, R. (2003). SAR research at dbS-SAR. Website. http://www.dbs-sar.com/Research.htm.

Koester, R. J. (2001). Virginia dataset on lost-person behavior. Excel file. Contact author at: http://www.dbs-sar.com.

Koester, R. J. and Stooksbury, D. E. (1999). Behavioral profile of wandering Alzheimer's patients. In Hill (1999b), chapter 6, pages 68–78.

Mitchell, B. L. (1985). A summary of the National Association for Search and Rescue data collection and analysis program for 1980–1985. Report, National Association for Search and Rescue, Inc., Washington, DC, USA.

Perkins, D., Roberts, P., and Feeney, G. (2001). Missing person behaviour: A U.K. study (interim report). Report, Centre for Search Research and Mountain Rescue Council, United Kingdom. http://www.mountain.rescue.org.uk/publications/MissingPersonBehaviourInterimReport2001.pdf.

Perkins, D., Roberts, P., and Feeney, G. (2002). Missing person behaviour: A U.K. study (full report). Report, Centre for Search Research and Mountain Rescue Council, United Kingdom. http://www.mountain.rescue.org.uk/publications/MPBReport_2002.pdf.

Perkins, D., Roberts, P., and Feeney, G. (2004). The U.K. missing person behaviour study. Report, Centre for Search Research and Mountain Rescue Council, United Kingdom. http://www.mountain.rescue.org.uk/publications/MPBReport2004.pdf.

Perkins, D., Roberts, P., and Feeney, G. (2005). The U.K. missing person behaviour study. Report, Centre for Search Research and Mountain Rescue Council, United Kingdom. http://www.mountain.rescue.org.uk/publications/MPBReport2005.pdf.

Syrotuck, W. G. (1976, 1977, 2000). *Analysis of Lost Person Behavior*. Barkleigh Productions, Inc., Mechanicsburg, PA. NASAR version, 2000.

Twardy, C. R. and Hope, L. R. (2004). Missing data on missing persons. Unpublished manuscript prepared for ECML04.