Estimating attendance at non-ticketed non-gated events

C.J. Schwarz*

Abstract
Methodology is developed to estimate attendance at day-long, non-ticketed, non-gated events. It consists of two complementary surveys. First, instantaneous counts of individuals are recorded at one or more times during the event. Second, either an access-survey, where parties who are leaving the event are intercepted, or a roving-survey where parties are intercepted during attendance at the event, record information on the start and end of the attendance for that party. The proportion of parties that were active during the instantaneous counts is used to expand the instantaneous counts. The estimators account for the cluster sampling of individuals within parties and the length-biased sampling in the roving survey. The methodology is applied to estimating attendance at a community festival in Vancouver, Canada.

Key Words: length-biased sampling, combining information, attendance

1. Introduction

There are many day-long community events that are non-ticketed (i.e. people do not purchase a ticket to attend) and non-gated (i.e. there is no control on admissions). People arrive at will, and enjoy the event. How can the total number of people attending the event during the entire day be estimated?

In large settings, it may be difficult to assess the number of people that attend a day-long event. This can be due to one or more of the following factors:

- There are numerous entrances with no "main" entrance,
- The event lasts for a substantial period of time (i.e. a day).
- Attendees are not required to register or to purchase tickets,
- Attendees may arrive and leave at various times.

There are two types of crowd estimation. First, is when the crowd arrives for a single event (e.g. a free concert or a demonstration at a particular time). Here the peak attendance is of primary interest. Watson and Yip (2011) calls this static crowd estimation and suggest several methods. For example, aerial photographs can be taken and processed manually or automatically such as outlined by Jiang et al. (2014). Mamei and Colonna (2016) showed how cellular network data could be used. Unfortunately, these methods may not be feasible for smaller festivals run by volunteers who do not have access to these technologies. Yip et al. (2010) suggested a method for when participants join a moving crowd moving down a linear route. Two spots are chosen near the start and the end of the route. Counts are made a both spots and added together. In addition, a random sample of people at the end of the route are asked if they also passed the start of the route. This gives an estimate of the fraction of people that are double counted and the total from the two counts is adjusted accordingly. Capture-recapture methods can also be used. A sample of people that enter the crowd are "tagged" (e.g. a sticker on their clothing; a free trinket). Later, a second sample

*Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, B.C., Canada V5A 1S6, cschwarz@stat.sfu.ca

of people is selected and the proportion that were “tagged” is determined. The standard closed-population Lincoln-Petersen estimator could be used. Here a key assumption is that sample of “tagged” people mixes thoroughly with the crowd which is difficult to obtain in practice. Similarly, for smaller events, point counts on the ground can also provide estimates of peak attendance. However, for many events, attendees do not stay for the entire event, and so peak counts may miss substantial number of attendees.

The above methods are not easily extended to counting when people may arrive and leave through the day, such as at a community festival. Multiple peak counts can be taken (e.g. at 12:00 and 15:00) and the instantaneous crowd numbers estimated as above, but unless information is available on the average length of stay, it is unknown if the people counted at 12:00 are also counted at 15:00 or what fraction of attendees have left and how many new attendees arrived between 12:00 and 15:00. As well, people who arrive and leave before the first count, or arrive and leave after the last count, would not be enumerated in the instantaneous counts. If capture-recapture methods are used, more complex methods must be used because the population is no longer closed with attendees departing (which look like “deaths” in natural populations) and arriving (which looks like '“recruitment” in natural populations) throughout the day. Open-population methods such as the Jolly-Seber method can be used but need to be modified to estimate what is known the super-population (total number of animals ever exposed to the sampling process, Schwarz and Arnason, 1996). Surveys would be scheduled say every 3 hours. On each survey, people would be selected (‘captured’) at random. Captured people would have their name recorded. This would then allow you construct a capture history for each person surveyed showing in which occasions each person was “captured”. For example, if there were 5 surveys, then a capture history of 10110 would indicate a person who was “captured” at surveys 1, 3 and 4, but not “captured” at surveys 2 and 5. These capture-histories would then be analyzed using the POPAN formulation of the Jolly-Seber model outlined by Schwarz and Arnason (1996). Open-population capture-recapture methods would be too complex to implement for most volunteer-run festivals.

For these types of events, the methods above are usually not suitable because of the longer time scale over which the crowd must be monitored. Streich et al. (2003) considered the problem of estimating attendance at large outdoor events that take place over an entire day. Their methodology consists of two complementary surveys. In the first survey, a count is made of some attribute (e.g. total number of drinks sold) at the concession. This count is assumed to be without error (i.e. a census). In the second survey, people leaving the event are surveyed at a convenient access point and asked if they attended the event and if yes, how drinks they purchased. The total number of drinks sold divided by the average number of drinks purchased per person is then an estimate of the total attendance. Streich et al. (2003) implicitly assumed that each individual leaving the event at that access point was a simple random sample from all attendees when finding the variance of the the estimate. However, this is often untrue with people arriving and leaving together in a party (e.g. a family) so that the second survey is a cluster sample. This affects how the estimator is computed and how the variance is found.

Rather surprisingly, estimating event attendance for day long events is analogous to estimating the total catch of fish from a lake (Dauk ad Schwarz, 2001). There are two complementary components in these surveys. First, instantaneous counts of the number of active fishing parties are obtained at one or more times during the day using, for example, aerial overflights. Second, interviews of fishing parties are obtained using one of two methods. In an aerial-access survey, the overflight information is augmented by interviews of fishing parties at the completion of their fishing trip at an access point where the fishing party is asked about their catch and (importantly) when the trip started and when the trip ended. In an aerial-roving survey, the overflight information is augmented by interviews of fishing parties by a surveyor who wanders among the fishing parties and samples fishing parties. In this case, the start of the fishing trip, the catch up to the point of interview, and the expected end of the fishing trip are recorded. Here the individual anglers naturally form clusters (fishing parties) which are the sampling units.

The questions about the start and end of the fishing trip provide information on what fraction of the fishing parties surveyed (using either the roving or access method) were active during the overflights and is used to estimate an expansion factor to estimate the total catch and other attributes.

This paper provides a "low-tech" method that can be used by festivals with a limited budget and without needing sophisticated technology such a face recognition or accessing cellular data. This paper shows how to modify these creel survey methods for use in estimating attendance at day-long, non-gated, non-ticketed events. Our method has two components. First, an instantaneous count of individuals at selected times of the day is taken. This is augmented with an access or roving survey. The cluster nature of the interview (e.g. parties are intercepted rather than individuals) is accounted for when developing the methodology for both cases. A simulation study investigates the performance of the estimators. The methodology is applied to a community festival from Vancouver, British Columbia, Canada. Finally we discuss how this methodology can be extended to using other attributes such as the total sales at the concessions.

2. Survey design and analysis

2.1 Notation

We let upper case letters refer to population values and lower case letters to refer to sample values.

P	Number of parties attending the event .
S_i	The size of party i, i.e. the number of people in the party.
S	Total attendance at the event. $S = \sum\limits_{i=1,\ldots,P} S_i$. **This is the parameter of interest**.
T	Total time the event is open during the day.
A_i	Arrival time of a party. $0 \leq A_i \leq T$.
D_i	Departure time of the party. $0 \leq D_i \leq T$.
L_i	Length of stay of the party at the event. $0 \leq L_i = D_i - A_i \leq T$.
C	Number of instantaneous counts conducted.
n	Number of parties interviewed (either access or roving). Each party is measured for the number of people in the party (s_i) and the arrival (a_i) and departure times (d_i). The party would stay at the event for length ℓ_i.
t_c	Time of the instantaneous count, $c = 1, \ldots, C$.
O_{t_i}	Instantaneous count at time t_i.
δ_{ic}	An indicator variable that is 1 if party i is active during instantaneous count c, i.e. $\delta_{ic} = 1$ if $A_i \leq t_c \leq D_i$ and 0 otherwise.
$\delta_{i\cdot}$	For each party counts the number of instantaneous counts at which the party was active, i.e. $\delta_{i\cdot} = \sum\limits_c \delta_{ic}$.

2.2 Design

The context for this paper are community events that are often of interest to families and groups of people. We define a party as a group of people that arrive and depart together (e.g. a family or a group of friends).

We will assume for simplicity, that the event takes place over one day. For multi-day events, the design and analysis proceeds by first stratifying by day, performing a separate survey each day, estimating the attendance for each day, and then combining the independent estimates of attendance (and their standard errors) in the usual way.

An attendee is defined as a person who visits the events at some point during the day. This paper assumes that each attendee only visits the event once in a day and multiple visits are treated as separate attendees. This may lead to an overestimate of the total attendance in a day.

Attendees arrive and depart in a party of 1 or more. Often these parties are families, but the design and analysis does not depend on a strict definition of a party and will still work if two families arrive/depart together but are treated as a single party.

Both variants of the methodology assume that a headcount at one or more times during the event can be obtained. For example, volunteers can be assigned to perform a headcount at small portions of the event grounds and the counts combined together. These headcounts are typically performed during peak attendance, and if several headcount are done, they should be spaced so that some parties would have a chance to arrive and depart between the counts. Notice that the counts are of individuals – it is not necessary to try and estimate the number of parties attending. This part of the survey design is denoted as the *aerial* component.

In the aerial-access method, a station is set up where (some) departing parties may leave the event. For example, there may be several streets by which parties could leave to access parking, transit, etc. At this access point, a sample of departing parties (e.g. every 5th party) is intercepted and asked about their party size (number of people in the party) and their arrival and departure times from the event. We assume that parties randomly choose their method of access to the event so the parties that are intercepted are a simple random sample of all parties.

In the aerial-roving method, a surveyor wanders around the event grounds and samples individuals (e.g. every 50th individual). Notice that using a systematic sample automatically spreads the sample "evenly" among attendees and does not over/undersample during the day. For each selected individual, the size of the party to which they belong, the arrival time of the party, and the expected departure time of the party are recorded.

Notice that in both cases, the sample of parties/individual should be spread throughout the day and the sample size in any hour during the day should be roughly proportional to the attendance at the event. Serious bias could arise if, for example, an equal number of parties were surveyed in all hours of the event regardless of the pattern of attendance. A systematic sample as suggested above would automatically allocate the sample size proportionally to attendance during the day. We are relying on "self-randomization" of the individuals within the event where we assume that a systematic sample is equivalent to a simple random sample.

2.3 Analysis: aerial-access design

Following Dauk and Schwarz (2001), the estimator for the total attendance is a ratio-of-means estimator

$$\widehat{S}^{access} = \frac{\sum\limits_{i=1,\ldots,n} s_i}{\sum\limits_{i=1,\ldots,n} s_i \sum\limits_{c=1,\ldots,C} \delta_{ic}} \times \sum_c O_{t_c} = \frac{\sum\limits_{i=1,\ldots,n} s_i}{\sum\limits_{i=1,\ldots,n} s_i \delta_{i\cdot}} \times \sum_c O_{t_c}$$

The instantaneous counts are inflated by fraction of the people in the interviewed parties that were active during the instantaneous counts. An alternate estimator could be formed using the fraction of the parties (rather than individuals) that were active during the instantaneous counts by replacing all s_i by the value of 1 in the above equation.

In the aerial-access design, parties are selected with equal probability regardless of their length of time they attended the event and we assume that due to self-randomization of the parties, that all selections are independent of each other. Consequently, standard results for the bias, variance, consistency of a ratio-of-mean estimator in a simple random sample is presented in many books on survey sampling such as Cochran (1977). Detailed results for this estimator are also provided in Dsuk and Schwarz (2001). For example, the estimated standard error of the estimator is found by defining $r = \frac{\sum\limits_{i=1,\ldots,n} s_i}{\sum\limits_{i=1,\ldots,n} s_i\delta_{i\cdot}}$ and $\epsilon_i = s_i - rs_i\delta_{i\cdot}$.

Then

$$\widehat{se}(\widehat{S}^{access}) = \sqrt{\frac{1}{\overline{s\delta}^2}\frac{s_\epsilon^2}{n}} \times \sum_c O_{t_c}$$

where $\overline{s\delta} = \frac{\sum\limits_{i=1,\ldots,n} s_i\delta_{i\cdot}}{n}$ is the mean of the denominator in the estimator, and s_ϵ^2 is the usual sample variance of the ϵ values.

For simplicity we have ignored any finite-population correction factor, but Dauk and Schwarz (2001) provide details on how this may be computed.

2.4 Analysis: aerial-roving design

In the aerial-roving design, parties do not have equal probability of being selected. Parties with longer lengths of stays at the event or parties with more people has a larger probability of being selected. This leads to the well known problem of length-biased sampling (Nowell and Stanley, 1991). Again following Dauk and Schwarz (2001) and Nowell and Stanley (1991), we modify the estimator to account for this length-biased sampling:

$$\widehat{S}^{roving} = \frac{\sum\limits_{i=1,\ldots,n} \frac{s_i}{s_i\ell_i}}{\sum\limits_{i=1,\ldots,n} \frac{s_i\delta_{i\cdot}}{s_i\ell_i}} \times \sum_c O_{t_c} = \frac{\sum\limits_{i=1,\ldots,n} \frac{1}{\ell_i}}{\sum\limits_{i=1,\ldots,n} \frac{\delta_{i\cdot}}{\ell_i}} \times \sum_c O_{t_c}$$

Deriving an estimator for the standard error is difficult. While the above estimator takes the form of a Horvitz-Thompson estimator, the actual probabilities of selection are unknowable because they depend on the (unknown) population size. Wolter (2007, Section 5.2.4) provides information on the bias, variance and consistency of this estimator. He also indicates that a good approximation for the variance is found by treating the sample as if selected with replacement, and using a simple bootstrap (i.e. select the units with replacement and equal probability) to estimate the standard error. It is not necessary to select the bootstrap sample with unequal probabilities because the observed sample is already weighted with the higher selection probability items.

Because we intercept parties in the middle of their visit to the event, we implicitly assume that their reported time of departure (d_i) are correct. On average, parties should be intercepted half way through their visit and so a comparison of the planned length of stay vs. twice the time from the interview back to the arrival time provides some information to assess this assumption.

A key problem with length-biased sampling is that the estimator is sensitive to parties that have small lengths of stay. For example, a party that reports they plan to stay for only one minute gets expanded by a large amount and inflates the overall estimated attendance significantly. In practice, we discard any interviews where the length of stay is very small (e.g. less than 30 minutes) during a 12 hour event.

3. Simulation Study

Because the aerial-access estimator is a variant of a ratio-estimator in a simple random sample, its properties are well known (e.g. Cochran, 1977). Similarly, the roving-access estimator is a variant of a Horvitz-Thompson estimator from sampling with unequal probability and its properties are also well known (e.g. Wolter, 2007) Consequently, it is not necessary to use simulation to verify these known results. Rather, a small simulation study was performed to evaluate the performance of the estimators (see Supplemental material) when there are problems in the data collection that were encountered during our application of these methods.

When all of the assumptions are satisfied, the estimators have little bias, estimated standard errors estimate the actual variability of the estimators, and the actual coverage of confidence intervals is close to nominal levels.

A well know problem with self-reported times is heaping where respondents ‘round’ their actual times of arrival and departure (e.g. to the nearest hour). The simulation study showed that this could lead to bias in the estimates, but this was related to an artefact of also having the instantaneous counts take place at the ‘heaping’ times, e.g. on the hour. By taking the instantaneous counts off the heaping time (e.g. at 10 past the hour rather than on the hour), this bias can be reduced.

Errors in the reported time of departure (especially for roving surveys) can introduce biases in many directions. If, for example, intercepted parties tend to underreport/overreport their time of departure, then the number of active parties during the head counts is under estimated/over estimated, leading to a positive/negative bias in the estimated attendance.

4. Example

The Powell Street Festival (`http://www.powellstreetfestival.com`) is a community event celebrating Japanese-Canadian arts and culture since 1977 in Vancouver, British Columbia, Canada that takes place on the first weekend on August. There is no admission fee (except for special events) and the event takes place in a local park with many entrances. The festival has many exhibits and booths and open air concerts.

In 2014, an aerial-roving survey was conducted on each of two days to estimate attendance during the event. Instantaneous head-counts were obtained at 13:00, 15:00, and 17:00 by dividing the park into small sections and having volunteers count the number of people at the festival (Table 1).

Volunteers circulated through the festival and interviewed approximately every 50th person encountered (Table 1). A plot of the time of arrival and (planned) time of departure is shown in Figure 1. The times of arrival and departure overall tend to be symmetrical about the point of interview, but of course parties intercepted earlier in the day are usually intercepted near the start of their visit, and parties intercepted at the end of the day, are usually intercepted near the end of their visit. There were a total of 6 parties intercepted who indicated that they planned to stay less than 30 minutes in total. These were discarded because their inclusion leads to estimates that were unreasonable.

The estimated attendance on each day was about 9000 and 7000 respectively, for a total estimated attendance of just under 16,000 (SE 1100) people.

Table 1: Summary statistics about the Powell Street Festival, 2014

Statistic	Day 1	Day 2	Total
Instantaneous counts			
13:00	2442	3123	
15:00	3252	3119	
17:00	2448	1819	
Interviews			
Number of interviews	95	80	
Average party size[a]	3.3	3.2	
Average planned length of stay (hours)[a]	3.0	3.3	
Interviews with planned length of stay $<$ 30 minutes[b]	5	1	
Parties active at 13:00	36	39	
Parties active at 15:00	43	51	
Parties active at 17:00	42	39	
Estimates using aerial-roving design			
Estimated attendance	9093	6599	15692
Estimated SE	877	616	1073

[a]Not adjusted for length-biased sampling.
[b]Discarded when estimates of attendance obtained.

5. Discussion

This estimator has a number of advantages over other methods to estimate crowd size. It only requires a low level of technical expertise and is easily done by volunteers. There is no need to access sophisticated technologies such as face recognition or cellular data. It is also much easier to implement than capture recapture methods.

A key advance of this paper over Streich et al. (2003), is the recognition that for many events, individuals do not act completely independently, but often arrive and depart in groups (e.g. families). By ignoring the cluster nature of the sampling design, estimators are unaffected, but estimates of precision are typically biased downwards because of the lack of independence.

This methodology has wide applicability for estimation of attendance at undated, non-ticketed events. If an access method is used, any convenient access point can be used and parties that have completed their visit are intercepted. Here the key assumption is that each party has the same chance of being intercepted regardless of when they are leaving, e.g. using a systematic sample of every 5th departing party would accomplish this. This may be difficult to do if people leave en masse (e.g. after a concert) and so sufficient resources need to be available to keep the probability of selection constant over time.

If a roving method is used, a similar concern arises, i.e. a constant fraction of the population must be surveyed over time. Again, a systematic sample is an easy way to accomplish this, but the same potential problem could arise if attendance spikes at certain times during the event. Furthermore, the roving component suffers from potential length-of-stay bias and this must be accounted for in the estimator process. Some post-processing of the roving data is required to exclude participants with very small length-of-stay to avoid inflating the estimate.

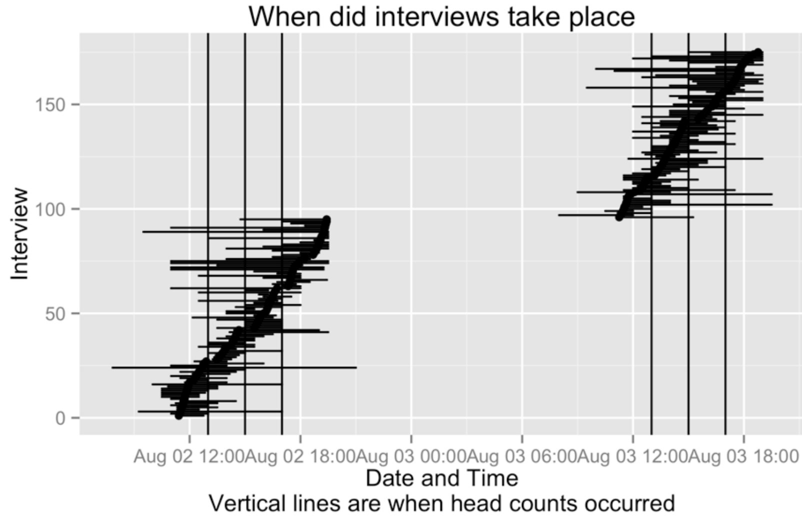

Figure 1: Plot of the reported time of arrival and proposed time of departure for interview in the roving survey. Vertical lines are the time of the instantaneous head counts.

Because these estimators are design-based and variants of standard survey estimators, they perform well when all assumptions are satisfied and collected data is perfect. However, some care is needed to deal with a common problem of heaping where parties round their time of arrival and actual or forecasted time of departure to common values (such as hourly). To avoid biases introduced when these rounded values coincide with the time of the instantaneous count, it is advisable to schedule the instantaneous count at a "non-heaped" time, such as 10 past the hour.

Another concern of using roving surveys is that the parties must predict correctly the time of departure. This can introduce potential sources of errors into the procedures if there is a consistent direction of error in reporting. For this reason, roving surveys are less recommended than access surveys.

A capture-recapture method is an alternative, but as shown in the simulation study would require substantial effort – almost 1/3 of parties would have to be sampled at each of the three sampling occasions. And because of the non-identifiability concerns at the first and last sampling occasions (Schwarz and Arnason, 1996), a potentially unrealistic model assuming equal probability of capture at each sampling occasion must be used. Our methods are much simpler and require much less effort.

We developed estimators using instantaneous counts. However, as noted by Streich et al. (2003), an alternate method for use with an access-component could use any activity that can be readily ascertained by the intercepted parties such as the number of drinks purchased at the concession. The modifications to the aerial-access estimators are trivial. Roving surveys are more problematic as now the respondents are required to predict their purchases in advance. We expect this to be poorly estimated and so do not recommend using roving surveys for this modification.

5.1 Acknowledgements

We thank the volunteers at the Powell Street Festival who gathered the data.

5.2 Funding

This research was funded by the Natural Science and Engineering Research Council (NSERC) of Canada with grant RGPIN/165058-2011.

5.3 Notes

All computations were done using R (R Core Team 2018) version 3.5.1 using the survey (Lumley, 2017) and boot (Canty and Ripley, 2017) packages.

REFERENCES

A. Canty, A., and B. Ripley, B. (2017). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-20.

Cochran, W.G. (1977), *Sampling* 3rd Edition, Wiley, New York.

Dauk, P., and Schwarz, C. J. (2001), "Catch estimation with restricted randomization in the effort survey", *Biometrics*, 57, 461–468.

Jiang, M., Huang, J., Wang, X., Tang, L., and Wu, C. (2014), "An approach for crowd density and crowd size estimation", *Journal of Software*, 9, 757–762.

Lumley, T. (2017), survey: analysis of complex survey samples. R package version 3.32.

Mamei, M., and Colonna, M. (2016), "Estimating attendance from cellular network data," *International Journal of Geographic Information Science*, 30, 1281–1301.

Nowell, C., and L.R. Stanley, L. R. (1991), "Length-biased sampling in mall intercept surveys," *Journal of Marketing Research*, 28, 475–479.

Schwarz, C. J., and Arnason. A. N. (1996), "A general methodology for the analysis of capture-recapture experiments in open populations," *Biometrics*, 52, 860–873.

Streich, A. M., Marx, D. B., Stafford, J. M., Rodie, S. N., and Todd, K. W. (2003), "Estimation of attendance at a large outdoor event," Journal of Extension, 41, `http://www.joe.org/joe/2003october/tt5.php`. Accessed 2018-09-11.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Watson, R. and Yip, P. (2011), "How many were there when it mattered," *Significance*, 8, 104–107.

Wolter, K. M. (2007), *Introduction to variance estimation,* 2nd edition. Springer, New York..

Yip, P. S. F., Watson, R., Chan, K. S., Lau, E. H. Y., Chen, F., Xu, Y., Xi, L., Cheung, D. Y. T., Ip, B. Y. T., and Liu, D. (2010), "Estimation of the number of people in a demonstration," *Australian & New Zealand Journal of Statistics*, 52, 17–26.

A. Details of the simulation study

A simulated population of 4000 parties attending an event was created. Each party size was drawn from a Poisson distribution with a mean of 3.2 (close to the observed party size in the example) except that generated party sizes of 0 were set to 1. This gave a total attendance of 13,344 people.

The length of time for the event was arbitrarily set a 100 units of time. Parties arrived to the event following a uniform distribution between a time of 0 and 80. The length of stay followed a log-normal distribution with a mean and standard deviation (on the logarithmic scale) of $\log 20$ and 5 respectively. Any departure time greater than 100 was truncated to 100. Any length of stay < 20 units was arbitrarily increased to 20 units. This gave a mean length of stay of 37 units of time (about 1/3 of the length of the event).

Three head counts were done at times 20, 40, and 60 with counts of 3164, 4962, and 6153 respectively.

The access-survey was simulated using a sampling fraction of 5% leading to an average of 200 interviews in each of 1000 simulations. The roving-survey was simulated using a sampling fraction of 5% but with parties selected with a probability proportional to the product of the number of people in the party and their length of stay at the event leading to an average of 200 initial interviews in each of 1000 simulation. Because all parties stayed a minimum of 20 units of time, no cutoff to eliminate short stays was needed.

We examined the performance of the estimators under the following scenarios

- When all assumptions are satisfied. The estimators are expected to be unbiased; the estimated standard errors are expected to reflect the actual variability in the estimators; and the coverage should be close to nominal levels.
- When attendees "round" their time of arrival and departure to convenient values – this is known as heaping. If rounding takes place at random (i.e. equally likely to be rounded up as down), bias is not expected to be introduced to the estimators, and this extra noise is expected to increase the estimated standard error relative to that when no rounding took place. Coverage is not expected to be affected.
- When attendees misreport their departure time and leave earlier than reported. A negative bias is expected because now more attendees appear to be active during the head counts than actually are present. Standard errors are also expected to be negatively biased. The combination of negative bias in both the estimator and standard error is expected to lead to a drop in coverage – refer to Cochran (1977). Similar effects in the opposite direction are expected when attendees misreport their departure time and leave later than reported.

For each simulation, the estimated total attendance, its standard error, a 95% confidence interval for the total attendance and a coverage indicator if the 95% confidence interval contained the true attendance were recorded. The performance of the estimators was assessed in the usual ways. Bias is assessed by comparing the average of the estimates over the simulations with the known attendance of 13,344. The performance of the estimate standard error was assessed by comparing the average estimated standard error with the actual standard deviation of the estimates over the simulation. The coverage was assessed by computing the proportion of estimated confidence intervals that included the known attendance over all simulations.

Results for all scenarios are shown in Table 2.

As expected, when all assumptions were satisfied, the estimators have little bias and the estimated standard errors are close to the actual variability of the estimator for both survey methods. Observed coverage was close to the nominal 95% coverage.

Heaping, where people round their time of arrival and departure to convenient values (such as the nearest hour) was simulated by rounding the arrivals to the nearest 10 value (corresponding to about 1/10 of the event time). Of course in an access component, the actual time of interview is known, so any heaping for the time of departure is expected to be small. Rather surprisingly, this can lead to extensive (negative) bias in the estimated attendance using both methods (Table 2) which reduces coverage considerably below the nominal value. This was an unexpected finding and is caused by an interaction with the time of the instantaneous count which also were taken at times that are at the 10's. Consider a party which arrives before the headcount. If the time of arrival is rounded downwards, it does not affect the 'active at headcount' status. If the time of the arrival is rounded upwards, the worst than can happen is that it now matches the time of the headcount and so is included. So heaping has no impact on the activity status of these parties. But consider a party that arrives after a headcount. Heaping may round downwards the time of arrival and party is now counted. So the estimated number of active parties at the head counts can only increase, which increases the denominator in the ratio, lowers the expansion factor, and leads to a negative bias in the estimates, as observed. This bias can be counter acted by scheduling the head counts at ”odd” times (such as 10 after the hour) rather than the hour to avoid heaping issues. This was confirmed in a simulation study (Table 2).

In the last scenario parties over reported their time of time departure (i.e. they actually left earlier than reported) by 30 units of time. Not unexpectedly, a substantial negative bias was introduced (Table 2) because now the estimated number of active parties during the head counts will tend to be over-estimated leading to a negative bias in the expansion factor and estimate of the total count. A similar finding (not shown) occurs when parties tend to under report their time of departure (i.e. they actually leave later than reported). It is more difficult to make errors in reporting the departure time for access-surveys because the intercepted parties are typically interviewed shortly after leaving the event.

We also considered estimation using a capture-recapture method. Rather than performing an instantaneous count at times 20, 40, and 60, parties were selected at random with equal probability. This is actually difficult to do in practice, but those practical difficulties will be ignored. If a party is selected, then the party size and the name of the party is recorded. The lists of names of parties sampled at the three sample times are collated and a capture history (see introduction) is create for each unique group names. These histories were then used to estimate the superpopulation number of parties from a Jolly-Seber model assuming equal probability of capture at each survey occasion but the fraction of parties leaving or the number of new parties arriving was allowed to vary between sampling occasions. It is assumed that the mean party size is known with trivial error (collected when parties are interviewed) so that an estimate of the total number attendees can be obtained from the estimate of the number of parties. Through trial and error, it was found that a sampling rate of about 35% was required at each sampling event (i.e. about 1/3 parties must be sampled) in order to get estimates with approximately the same precision as from our estimators. Results are reported at the top of Table **??**. Because parties are interviewed in person, heaping and errors in arrival and departure time are not relevant.

The capture-recapture method required sampling an average of almost 1300 parties at each of the three sampling occasions. The capture-recapture estimators showed some bias because some parties either arrived prior to and departed prior to the first survey, or arrived after the third survey. The estimated standard errors reflected the actual variability in the estimate, but because of the bias, the coverage probability was less than the nominal level.

B. R code and Notebook

Available from the Author.

Table 2: Summary of simulation study of aerial-access, aerial-roving, and capture-recapture estimators. Actual population size was 13,344. Results based on 1000 simulation replicates. Bootstrap estimates of the se for the aerial-roving survey based on 100 replicates. Bottom three tables not relevant for capture-recapture model.

Statistic	Aerial-access	Aerial-roving	Capture-recapture
All assumptions satisfied			
Average of $\widehat{S}$	13,384	13,544	12,517
Average of estimated se	883	732	845
Actual SD($\widehat{S}$)	886	717	856
Coverage of 95% c.i.	0.95	0.95	0.87
Heaping for time of arrival			
Average of $\widehat{S}$	11,417	11,767	
Average of estimated se	669	628	
Actual SD($\widehat{S}$)	649	599	
Coverage of 95% c.i.	0.21	0.21	
Heaping for time of arrival but time of headcount shifted			
Average of $\widehat{S}$	13,377	13,699	
Average of estimated se	847	754	
Actual SD($\widehat{S}$)	782	910	
Coverage of 95% c.i.	0.96	0.95	
Departure time under estimated by 30 minutes			
Average of $\widehat{S}$	9,779	10,406	
Average of estimated se	515	624	
Actual SD($\widehat{S}$)	504	590	
Coverage of 95% c.i.	0.00	0.12	