Abstract Title Page
Not included in page count.

Title:

Operationally comparable effect sizes for quantifying changes in behavior, with application to meta-analysis of single-case studies

Authors and Affiliations:

James E. Pustejovsky
Northwestern University
Department of Statistics

Abstract Body

Background / Context

Single-case designs (SCDs) are a class of research methods for evaluating intervention effects by taking repeated measurements of an outcome over time on a single case, both before and after the deliberate introduction of a treatment. SCDs are used heavily in fields such as special education, school psychology, social work, and applied behavior analysis (Busse, Kratochwill, & Elliott, 1995; Horner et al., 2005; Kazdin, 2011; Kennedy, 2004; Odom et al., 2005), frequently in combination with behavioral observations. Indeed, the focus on outcome measurements based on direct observation is considered a hallmark of single-case methodology, in that treatment impacts on behavioral outcomes often have immediate and recognizable social implications for individual participants and the broader populations that they represent (Hartmann & Wood, 1990; Horner et al., 2005). Given the prominence of behavioral observation data in single-case research, primary investigators and meta-analysts need effect size measures that are appropriate and interpretable when applied to such data.

Several different operational procedures are commonly used to record direct observations of human behavior, ranging from continuous duration recording to interval recording methods (Altmann, 1974; Barlow & Hersen, 1984; Kazdin, 2011). I describe these in more detail below. The variations in recording procedures have important implications for meta-analysis. In a collection of studies, some studies may have used one recording procedure while others used another. In order for a meta-analysis of such a heterogeneous collection of studies to be scientifically interpretable, the results of each study must be expressed on a common scale. If the basic input units into the meta-analysis—effect sizes—are not comparable across recording procedures (or what I term "operationally comparable"), average across and contrasts between results based on different recording procedures will be confounded by differences of scale.

Purpose / Objective / Research Question / Focus of Study

This methodological research will describe a model for behavioral observation data that allows definition of an intuitively interpretable, operationally comparable effect size, the *prevalence odds ratio* (POR). After defining the POR, I describe basic estimators based on data from several different recording methods.

Significance / Novelty of study

Many different effect sizes have been proposed for meta-analysis of single-case studies, but nearly all are subject to serious criticisms (Allison & Gorman, 1993; Beretvas & Chung, 2008; Shadish, Rindskopf, & Hedges, 2008; Wolery, Busick, Reichow, & Barton, 2010). Current proposals for single-case effect sizes can be classified into three broad categories: parametric models for single cases (Busk & Serlin, 1992; Center, Skiba, & Casey, 1985; Swaminathan et al., 2008), hierarchical models for groups of cases (Hedges, Pustejovsky, & Shadish, 2012; Van den Noortgate & Onghena, 2003a, 2003b, 2007, 2008), or non-overlap statistics (Parker, Vannest, & Davis, 2011). Both types of parametric approaches have focused largely on standardized mean differences, which are appropriate for continuous, interval scale data but less useful when measurements are discrete or have bounded ranges. Non-overlap statistics, many of which are inspired by non-parametric test statistics, have been criticized for being un-interpretable as measures of effect magnitude, as well as for lacking known sampling distributions and for being sensitive to design features (such as number of repeated measurements in a phase) that are not of

scientific interest (Beretvas & Chung, 2008; Shadish & Rindskopf, 2007; Wolery et al., 2010). Crucially, none of the current proposals for single-case study effect sizes are designed to address the implications of different recording procedures.

The properties of the recording procedures have long been subject to scrutiny and debate. Much of the debate has centered on the theoretical interpretation and practical utility of interval recording methods (Altmann, 1974; Harrop, Daniels, & Foulkes, 1990; Mann, Ten Have, Plunkett, Meisels, & Have, 1991). The sensitivity of results to variation in recording methods has been studied through simulations (e.g., Rapp, Colby-dirksen, Michalski, Carroll, & Lindenberg, 2008) and through empirical examples (e.g., Rapp et al., 2007), but rarely through explicit statistical modeling. The most relevant exception is Rogosa and Ghandour (1991), who used an alternating renewal process model (like the one described below) to study the psychometrics of behavioral observations; however, these authors focus mostly on behavior frequency measures, rather than prevalence. Other authors have used an alternating poisson process formulation to study the properties of momentary time sampling (Brown, Solomon, & Stephens, 1977; Griffin & Adams, 1983).

Statistical, Measurement, or Econometric Model:

In order to define an operationally comparable effect size, I posit a model for the sequence of behavioral events that occur over a single observation session; this sequence of events in time is sometimes called the *behavior stream* (Hartmann & Wood, 1990; Rogosa & Ghandour, 1991). Based on the behavior stream, I describe the properties of several different recording methods. I then define the prevalence odds ratio (POR) and, based on a simple between-session model, consider how to estimate the POR from data generated by different observation recording procedures.

Behavior stream data. During a single observation session, the behavior stream can be described as follows. Assume that, within session *t*, events occur sequentially and can be numbered $u = 1,2,3,\ldots$. Let D_{tu} denote the duration of event *u*; let E_{tu} denote the length of time between the end of event *u* and the beginning of event $u + 1$, sometimes called the *inter-event time* (IET); let E_{t0} denote the length of time until the first event, with E_{t0} if event 1 is occurring at the beginning of the observation period. The quantities $\{E_{t0},D_{t1},E_{t1},D_{t2},E_{t2},D_{t3},E_{t3},\ldots\}$ are the underlying data that describe the behavior stream during session *t*. Figure 1 depicts the behavior stream.

Recorded data. I now describe several different recording procedures, denoting the recorded datum from session *t* and recording method *m* as Y_t^m, $m \in \{C, M, E, P, W\}$. Assume that the session is of length *T*.

- In continuous duration recording, the recorded datum Y_t^C measures the proportion of session time during which the behavior occurs.
- In momentary time sampling, the observer records the presence or absence of a behavior at each of *K* time-points during a session (typically, time-points are equally spaced). The reported datum Y_t^M measures the proportion of time-points at which the behavior was occurring (see Figure 2a).
- In event counting, the recorded datum Y_t^E measures the number of times that an event begins during the course of the session.

- In partial-interval recording, the observer divides the session into *K* intervals, each of length *L*. The recorded datum Y_t^E measures the proportion of intervals during which the behavior occurred for any length of time (see Figure 2b).
- Whole-interval recording is structured identically to partial-interval recording, except that each interval is scored only if the behavior occurs for the entire interval (see Figure 2c). The recorded datum Y_t^W therefore measures the proportion of intervals in which the behavior occurred for the duration.

Within-session model. I will assume that, within each session, the behavior stream can be modeled by an equilibrium alternating renewal process (EARP). The EARP and is a broad classes of models for describing a behavior that is either present or absent, and has been used to study the psychometric properties of behavioral observation data (Rogosa & Ghandour, 1991). In an equilibrium alternating renewal process, it is assumed that IETs are identically distributed random quantities, that event durations are also identically distributed, and that all IETs and event durations are mutually independent. For session *t* of length *T*, the main parameters of the model are then the average event duration $\mu_t = \mathrm{E}(D_{t1})$ and the average IET $\lambda_t = \mathrm{E}(E_{t1})$. Table 1 reports the expectations of each type of recorded data, under the assumptions of the EARP. Derivations are omitted due to space constraints.

Target effect size. If one has to choose a single parameter to describe session-to-session changes in the behavior stream, the best candidate may be the prevalence odds ratio, which measures proportional change in the ratio of μ_t to λ_t. For comparing session *a* to session *b*, define:

$$\Omega = \frac{\mu_b / \lambda_b}{\mu_a / \lambda_a}. \tag{1}$$

There are several reasons for making the POR the target of inference. First, the POR compares the prevalence across sessions, and from the point of view of an interventionist, prevalence is presumably the most substantively important aspect of behavior. This is because reducing the incidence of an undesirable behavior without changing its prevalence (i.e., fewer incidents of longer duration) is not a clear improvement. Second, if the average event duration μ_t is constant from session to session, the POR has the simple interpretation of a proportionate change in inter-event rates. For example, if $\mu_a = \mu_b$, then a POR of $\Omega = 1/2$ means that from session *a* to session *b* the average inter-event rate has halved and, equivalently, that the average IET has doubled. Similarly, if the average inter-event time is constant across sessions, the POR represents a proportionate change in the event duration. The POR therefore provides an intuitive means of equating reductions in event duration with increases in IET. This final property is particularly desirable in a meta-analysis context in which one may wish to compare some interventions that reduce the duration of an undesirable behavior with others that reduce the frequency of the behavior. For purposes of meta-analytic modeling, it is often helpful to use scales that that have no upper or lower limit. Because, the POR ranges from 0 to positive infinity, it is therefore useful to use the log of the prevalence odds ratio rather than the ratio itself. Thus, the target of estimation and inference is $\omega = \log(\Omega) = \log \mu_b - \log \lambda_b - \log \mu_a + \log \lambda_a$.

Between-session model. I consider a very basic model for a set of sessions. Suppose that the first n_0 observation sessions occur in a baseline phase, which is immediately followed by a treatment phase consisting of n_1 observation sessions. Further assume that within a phase, observation sessions are independent and identically distributed (i.i.d.), so that

$\mu_1 = \cdots = \mu_{n_0} = \mu_a$, $\lambda_1 = \cdots = \lambda_{n_0} = \lambda_a$, $\mu_{n_0+1} = \cdots = \mu_{n_0+n_1} = \mu_b$ and $\lambda_{n_0+1} = \cdots = \lambda_{n_0+n_1} = \lambda_b$. Admittedly, the assumption that repeated measurements are i.i.d. is probably unrealistically strong, in that it does not allow for between-session trends or serial dependence among repeated measurements. I maintain it here in order to illustrate the relationship between recording procedures and estimators.

I now describe some very simple moment estimators of the log-POR comparing the baseline with the treatment phase. The estimators are summarized in Table 2, along with approximate delta-method variance estimators. Due to space constraints, I do not comment on all recording methods. To begin, note that continuous duration recording and momentary time sampling provide direct estimates of prevalence, so that ω can be estimated by taking the log-odds ratio of the phase means. Event counting estimates incidence rather than prevalence; to estimate prevalence, it may sometimes be reasonable to assume that the average duration is known and constant across phases: $\mu_a = \mu_b = \mu$. Under this assumption, the POR is equivalent to the ratio of λ_a / λ_b, which can be estimated as described in Table 2. Because in partial interval recording, the expectation of the recorded datum depends on the full distribution of the IETs, further parametric assumptions are needed. Assuming that $E_{t1} \sim Exp\left(1/\lambda_t\right)$, it follows that $\mathrm{E}\left(Y_t^P\right) = 1 - \lambda_t e^{-L/\lambda_t} / \left(\mu_t + \lambda_t\right)$. Further assuming a known value for $\mu_a = \mu_b = \mu$ leads to a moment estimator for $\log\left(\lambda_a / \lambda_b\right)$.

Usefulness / Applicability of Method:

A recent meta-analysis of single-case studies (Shogren, Faggella-luby, Bae, & Wehmeyer, 2004) evaluated the effects of allowing children greater autonomy to make choices. Of the 13 studies in the original meta-analysis, 9 studies (including 27 participants) used one or more of the recording methods described above to measure participants' levels of problem behavior or task dis-engagement (see Table 3). Table 4 details the observation recording procedures used for each case, as well as the assumed value of μ for event count and partial interval data. Figure 3 presents a forest plot of the log-prevalence odds ratios for each case, along with the estimated overall average effect size based on a random effects meta-analysis. In this preliminary analysis, the average effect of allowing choice-making is estimated to be -1.51 with a 95% confidence interval of [-2.01, -1.02]. This overall average effect corresponds to a reduction of between 64% and 87% in the prevalence odds from the no-choice baseline condition. The between-case variance in the true effects is estimated to be 1.43 [$Q(26) = 378$, $p < .0001$, $I^2 = 94\%$].

Conclusions

The model described here provides a basis for defining an operationally comparable effect size, the prevalence odds ratio, that allows comparisons and meta-analytic summaries of treatment effects measured using different behavioral observation recording procedures. The model highlights the strong assumptions needed to estimate the effect size based on event counting or partial interval recording data. Future work will evaluate the sampling distribution of the moment estimators proposed here, as well as considering alternative estimators such as those based on making second-moment or full distribution assumptions regarding the event durations and inter-event times. I will also consider three-level meta-analytic models to account for study-level effects.

Appendices

Not included in page count.

Appendix A. References

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: the case of the single case. *Behaviour research and therapy*, *31*(6), 621–31.

Altmann, J. (1974). Observational Study of Behavior: Sampling Methods. *Behavior*, *49*(3/4), 227–267.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: strategies for studying behavior change*. Pergamon Press, Inc.

Beretvas, S. N., & Chung, H. (2008). A review of meta-analyses of single-subject experimental designs: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention*, *2*(3), 129–141. doi:10.1080/17489530802446302

Brown, M., Solomon, H., & Stephens, M. A. (1977). Estimation of Parameters of Zero-One Processes by Interval Sampling. *Operations Research*, *25*(3), 493–505.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-Case Research Design and Analysis: New Directions for Psychology and Education* (pp. 187–212). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Busse, R. T., Kratochwill, T. R., & Elliott, S. N. (1995). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology*, *33*(4), 269–285.

Center, B. A., Skiba, R. J., & Casey, A. (1985). A methodology for the quantitative synthesis of intra-subject design research. *The Journal of Special Education*, *19*(4), 387.

Cox, D. R. (1962). *Renewal Theory*. Great Britain: Metheun & Co. Ltd.

Dunlap, G., DePerczel, M., Clarke, S., Wilson, D., Wright, S., White, R., & Gomez, A. (1994). Choice making to promote adaptive behavior for students with emotional and behavioral challenges. *Journal of Applied Behavior Analysis*, *27*(3), 505–518.

Dyer, K., Dunlap, G., & Winterling, V. (1990). Effects of choice making on the serious problem behaviors of students with severe handicaps. *Journal of Applied Behavior Analysis*, *23*(4), 515–524.

Frea, W. D., Arnold, C. L., & Vittimberga, G. L. (2001). A demonstration of the effects of augmentative communication on the extreme aggressive behavior of a child with autism

within an integrated preschool setting. *Journal of Positive Behavior Interventions*, *3*(4), 194.

Griffin, B., & Adams, R. (1983). A parametric model for estimating prevalence, incidence, and mean bout duration from point sampling. *American Journal of Primatology*, *4*(3), 261–271. doi:10.1002/ajp.1350040305

Harrop, A., Daniels, M., & Foulkes, C. (1990). The Use of Momentary Time Sampling and Partial Interval Recording in Behavioural Research. *Behavioural Psychotherapy*, *18*(2), 121–127. doi:10.1017/S0141347300018231

Hartmann, D. P., & Wood, D. D. (1990). Observational methods. In A. S. Bellack, M. Hersen, & A. E. Kazdin (Eds.), *International Handbook of Behavior Modification and Therapy* (2nd ed., pp. 107–138). New York, NY: Plenum Press.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*. doi:10.1002/jrsm.1052

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S. L., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165–179.

Jolivette, K., Wehby, J. H., Canale, J., & Massey, N. G. (2001). Effects of choice-making opportunities on the behavior of students with emotional and behavioral disorders. *Behavioral Disorders*, *26*(2), 131–145.

Kazdin, A. E. (2011). *Single-Case Research Designs: Methods for Clinical and Applied Settings*. New York, NY: Oxford University Press.

Kennedy, C. H. (2004). *Single-Case Designs for Educational Research*. Boston, MA: Allyn & Bacon.

Kern, L., Mantegna, M. E., Vorndran, C. M., Bailin, D., & Hilt, A. (2001). Choice of task sequence to reduce problem behaviors. *Journal of Positive Behavior Interventions*, *3*(1), 3–10.

Mann, J., Ten Have, T. R., Plunkett, J. W., Meisels, S. J., & Have, T. T. (1991). Time sampling: A methodological critique. *Child development*, *62*(2), 227–241.

Moes, D. R. (1998). Integrating choice-making opportunities within teacher-assigned academic tasks to facilitate the performance of children with autism. *Research and Practice for Persons with Severe Disabilities*, *23*(4), 319–328.

Odom, S. S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, *71*(2), 137–148.

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, *35*(4), 303–322. doi:10.1177/0145445511399147

Powell, S., & Nelson, B. (1997). Effects of choosing academic assignments on a student with attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis*, *30*(1), 181. doi:10.1901/jaba.1997.30-181

Rapp, J. T., Colby, A. M., Vollmer, T. R., Roane, H. S., Lomas, J., & Britton, L. N. (2007). Interval recording for duration events: a re- • evaluation. *Behavioral Interventions*, *22*(June), 319–345.

Rapp, J. T., Colby-dirksen, A. M., Michalski, D. N., Carroll, R. A., & Lindenberg, A. M. (2008). Detecting changes in simulated events using partial-interval recording and momentary time sampling. *Behavioral Interventions*, *23*, 237–269.

Rogosa, D., & Ghandour, G. (1991). Statistical Models for Behavioral Observations. *Journal of Educational Statistics*, *16*(3), 157–252.

Romaniuk, C., Miltenberger, R., Conyers, C., Jenner, N., Jurgens, M., & Ringenberg, C. (2002). The influence of activity choice on problem behaviors maintained by escape versus attention. *Journal of applied behavior analysis*, *35*(4), 349–62. doi:10.1901/jaba.2002.35-349

Seybert, S., Dunlap, G., & Ferro, J. (1996). The effects of choice-making on the problem behaviors of high school students with intellectual disabilities. *Journal of Behavioral Education*, *6*(1), 49–65.

Shadish, W. R., & Rindskopf, D. M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. *New Directions for Evaluation*, *2007*(113), 95–109.

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, *2*(3), 188–196. doi:10.1080/17489530802581603

Shogren, K. A., Faggella-luby, M. N., Bae, S. J., & Wehmeyer, M. L. (2004). The effect of choice-making as an intervention for problem behavior. *Journal of Positive Behavior Interventions*, *6*(4), 228–237.

Swaminathan, H., Horner, R. H., Sugai, G., Smolkowski, K., Spaulding, S. A., & Hedges, L. V. (2008). Application of generalized least squares regression to measure effect size in single-case research: A technical report.

Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly*, *18*(3), 325–346.

Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior research methods, instruments, & computers*, *35*(1), 1–10.

Van den Noortgate, W., & Onghena, P. (2007). The aggregation of single-case results using hierarchical linear models. *Behavior Analyst Today*, *8*(2), 196–209.

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence based Communication Assessment and Intervention*, *2*(3), 142–151. doi:10.1080/17489530802505362

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education*, *44*(1), 18–28. doi:10.1177/0022466908328009

Appendix B. Tables and Figures

Not included in page count.

Table 1
Expectations of recorded datum produced by various recording methods

Recording method	Expectation	Source
Continuous duration recording	$\mathrm{E}\left(Y_t^C\right) = \frac{\mu_t}{\mu_t+\lambda_t}$	Cox (1962, p. 101; see also Rogosa & Ghandour, 1991, p. 226)
Momentary time sampling	$\mathrm{E}\left(Y_t^M\right) = \frac{\mu_t}{\mu_t+\lambda_t}$	Cox (1962, p. 87)
Event counting	$\mathrm{E}\left(Y_t^E\right) = \frac{T}{\mu_t+\lambda_t}$	Cox (1962, p. 46)
Partial interval recording	$E\left(Y_t^P\right) = \frac{\mu_t + \int_0^L \tilde{F}_E(v)dv}{\mu_t + \lambda_t}$	Author derivation, based on Cox (1962, p. 85)
Whole interval recording	$E\left(Y_t^W\right) = \frac{\mu_t - \int_0^L \tilde{F}_D(v)dv}{\mu_t + \lambda_t}$,	Author derivation, based on Cox (1962, p. 85)

Notes:

T denotes the total length of the session**.**

$\tilde{F}_E(v) = \Pr\left(E_1 > v\right)$ is the complement of the cumulative distribution function of the IETs **.**

$\tilde{F}_D(v) = \Pr\left(D_1 > v\right)$ is the complement of the cumulative distribution function of the event durations.

Table 2
Effect size estimators and variance estimates

Recording method	Estimate of POR	Variance estimate
Continuous duration recording[1]	$\hat{\omega}_C = \text{logit}\left(\bar{y}_1^C\right) - \text{logit}\left(\bar{y}_0^C\right)$	$Var\left(\hat{\omega}_C\right) \approx \frac{s_{C1}^2}{n_1\left(\bar{y}_1^C\right)^2\left(1-\bar{y}_1^C\right)^2} + \frac{s_{C0}^2}{n_0\left(\bar{y}_0^C\right)^2\left(1-\bar{y}_0^C\right)^2}$
Momentary time sampling[1]	$\hat{\omega}_M = \text{logit}\left(\bar{y}_1^M\right) - \text{logit}\left(\bar{y}_0^M\right)$	$Var\left(\hat{\omega}_M\right) \approx \frac{s_{M1}^2}{n_1\left(\bar{y}_1^M\right)^2\left(1-\bar{y}_1^M\right)^2} + \frac{s_{M0}^2}{n_0\left(\bar{y}_0^M\right)^2\left(1-\bar{y}_0^M\right)^2}$
Event counting[2]	$\hat{\omega}_E^{\mu} = \log\left(\frac{T}{\bar{y}_0^E} - \mu\right) - \log\left(\frac{T}{\bar{y}_1^E} - \mu\right)$	$Var\left(\hat{\omega}_E\right) \approx \frac{T^2 s_{E1}^2}{n_1\left(\mu\left(\bar{y}_1^E\right)^2 - T\bar{y}_1^E\right)^2} + \frac{T^2 s_{E0}^2}{n_0\left(\mu\left(\bar{y}_0^E\right)^2 - T\bar{y}_0^E\right)^2}$
Partial interval recording[3]	$\hat{\omega}_P^{\mu} = \log\hat{\lambda}_0 - \log\hat{\lambda}_1$	$Var\left(\hat{\omega}_E\right) \approx \sum_{i=0}^{1} \frac{\hat{\lambda}_i^2\left(\mu+\hat{\lambda}_i\right)^2 s_{E1}^2}{n_1\left(1-\bar{y}_1^P\right)^2\left(\mu\hat{\lambda}_i + L\mu + L\hat{\lambda}_i\right)^2}$

Notes:

For recording method $m \in \{C, M, E, P, W\}$, the within-phase means are $\bar{y}_0^m = \frac{1}{n_0}\sum_{t=1}^{n_0} Y_t^m$ and $\bar{y}_1^m = \frac{1}{n_1}\sum_{t=1}^{n_1} Y_t^m$;

the within-phase sample variances are $s_{m0}^2 = \frac{1}{n_0 - 1}\sum_{t=1}^{n_0}\left(Y_t^m - \bar{y}_0^m\right)^2$, $s_{m1}^2 = \frac{1}{n_1 - 1}\sum_{t=1}^{n_1}\left(Y_t^m - \bar{y}_1^m\right)^2$.

[1] $\text{logit}(p) = \log(p) - \log(1-p)$.

[2] Assuming a known value for $\mu = \mu_{n_0} = \mu_{n_0+n_1}$.

[3] Assuming a known value for $\mu = \mu_{n_0} = \mu_{n_0+n_1}$ and that $E_{t1} \sim Exp\left(1/\lambda_t\right)$. The estimator $\hat{\lambda}_i$ is defined implicitly as the solution to $\bar{y}_i^P = 1 - \lambda e^{-L/\lambda} / \left(\mu + \lambda\right)$, and must be evaluated numerically except in certain special cases.

Table 3
Studies in meta-analysis by Shogren et al. (2004)

Study	Citation	Design	Number of participants by recording method			
			Continuous duration recording	Momentary time sampling	Event counting	Partial interval recording
Dun1994	(Dunlap et al., 1994)	ABAB				3
Dye1990	(Dyer, Dunlap, & Winterling, 1990)	ABAB				3
Fre2001	(Frea, Arnold, & Vittimberga, 2001)	Multiple baseline			1	
Jol2001	(Jolivette, Wehby, Canale, & Massey, 2001)	Multiple baseline				3
Ker2001	(Kern, Mantegna, Vorndran, Bailin, & Hilt, 2001)	ABAB			1	2
Moe1998	(Moes, 1998)	AB/BA				4
Pow1997	(Powell & Nelson, 1997)	ABAB		1		
Rom2002	(Romaniuk et al., 2002)	ABAB	5		1	
Sey1996	(Seybert, Dunlap, & Ferro, 1996)	ABAB				3

Table 4
Cases in meta-analysis by Shogren et al. (2004)

Study	Case	Recording method	Session length (minutes)	Interval length (seconds)	Assumed μ (seconds)
Dun1994	Ahmad	Partial interval recording	15	10	10
	Sven	Partial interval recording	15	10	10
	Wendall	Partial interval recording	15	15	10
Dye1990	George	Partial interval recording	15	30	10
	Lori	Partial interval recording	15	30	10
	Mary	Partial interval recording	15	30	10
Fre2001	Tim	Event counting	10		0.0001
Jol2001	Bruce	Partial interval recording	15	10	10
	John	Partial interval recording	15	10	10
	Nicky	Partial interval recording	15	10	10
Ker2001	Danny	Event counting	15		0.0001
	Kelly	Partial interval recording	Not reported	10	10
	Shannon	Partial interval recording	30	10	10
Moe1998	Carl	Partial interval recording	20	10	10
	Charles	Partial interval recording	20	10	10
	Chuck	Partial interval recording	20	10	10
	James	Partial interval recording	20	10	10
Pow1997	Evan	Momentary time sampling	30		
Rom2002	Brooke	Continuous duration recording	5		
	Christy	Continuous duration recording	5		
	Gary	Continuous duration recording	5		
	Maggie	Continuous duration recording	5		
	Rick	Continuous duration recording	5		
	Riley	Event counting	5		0.0001
Sey1996	Bob	Partial interval recording	Not reported	10	10
	Maria	Partial interval recording	Not reported	10	10
	Scott	Partial interval recording	Not reported	10	10

Figure 1
The behavior stream

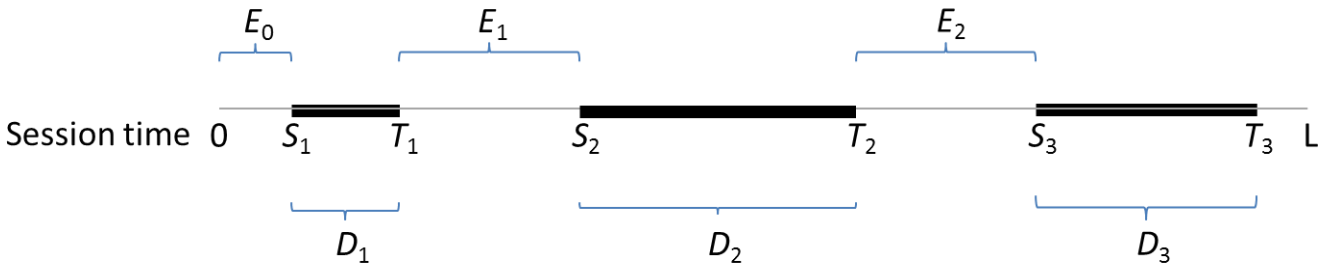
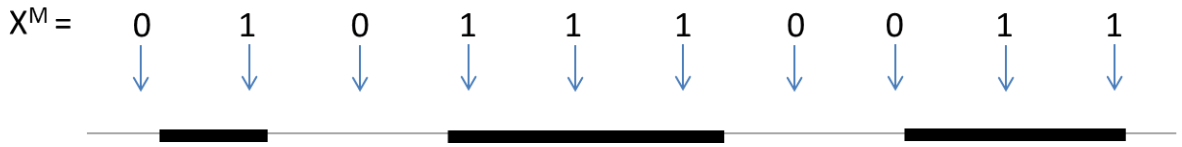
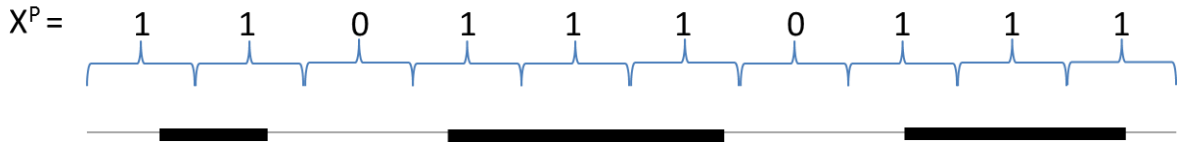


Figure 2
Observation recording methods

(a) Momentary time sampling



(b) Partial interval recording



(c) Whole interval recording

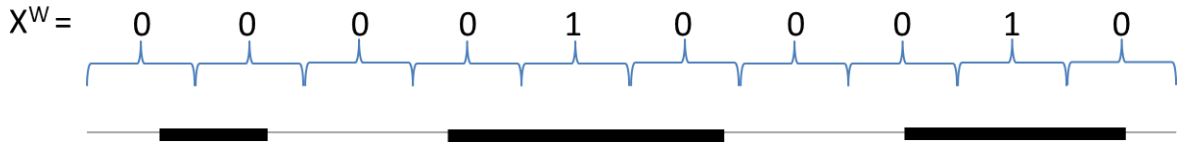


Figure 3
Forest plot of log-prevalence odds ratio effect sizes