Received: 7 July 2023 | Accepted: 5 May 2024

DOI: 10.1002/jeab.925

TECHNICAL REPORT

Journal of the Experimental Analysis of Behavior

SEAB

Treatments for undefined log ratios in matching analyses

Pier-Olivier Caron

Université TÉLUQ, Montréal, Canada

Correspondence
Pier-Olivier Caron, Département des Sciences humaines, Lettres et Communication, Université TÉLUQ, 5800, Rue Saint-Denis, Bureau 1105, Montréal, QC, Canada H2S 3L5.
Email: pocaron19@gmail.com

Editor-in-Chief: Suzanne H. Mitchell
Handling Editor: Sarah Cowie

Abstract

A challenge in carrying out matching analyses is to deal with undefined log ratios. If any reinforcer or response rate equals zero, the logarithm of the ratio is undefined: data are unsuitable for analyses. There have been some tentative solutions, but they had not been thoroughly investigated. The purpose of this article is to assess the adequacy of five treatments: omit undefined ratios, use full information maximum likelihood, replace undefined ratios by the mean divided by 100, replace them by a constant 1/10, and add the constant .50 to ratios. Based on simulations, the treatments are compared on their estimations of variance accounted for, sensitivity, and bias. The results show that full information maximum likelihood and omiting undefined ratios had the best overall performance, with negligibly biased and more accurate estimates than mean divided by 100, constant 1/10, and constant .50. The study suggests that mean divided by 100, constant 1/10, and constant .50 should be avoided and recommends full information maximum likelihood to deal with undefined log ratios in matching analyses.

KEYWORDS

full information maximum likelihood, matching law, missing data, undefined ratios, zero data

Choice behavior has been a long-standing topic of research in both applied behavior analysis and the experimental analysis of behavior. The generalized matching law (GML) is a quantitative model describing an organism's response allocation as a function of reinforcer ratio in a concurrent schedule of reinforcement (Baum, 1974). The relation is represented in the following equation:

$$\log\left(\frac{B_1}{B_2}\right) = a\log\left(\frac{R_1}{R_2}\right) + \log c, \tag{1}$$

which describes the allocation of responses (B) across alternatives 1 and 2 in terms of the distribution of reinforcers (R) across those alternatives—hereafter, behavior and reinforcers in Equation (1) are referred to as the components of the GML. The parameter a refers to the sensitivity to reinforcement—that is, the changes in response allocation, $\log\left(\frac{B_1}{B_2}\right)$, relative to changes in reinforcer allocation, $\log\left(\frac{R_1}{R_2}\right)$—and $\log c$ is the bias—that is, the preference between choice-affecting variables that are kept constant but are different between response alternatives. The GML has been shown to account for a wide variety of species, behavior, reinforcers, and experimental and applied scenarios (Davison & McCarthy, 1988).

A challenge in carrying out matching analyses is to deal with undefined log ratios. If any component equals zero, the logarithm of the ratio is undefined, being either positive infinity, $\log\frac{1}{0} = \infty$, or negative infinity, $\log 0 = -\infty$, depending on whether the numerator or the denominator is null. Some software may produce error instead. This issue occurs in experimental settings when either reinforcer or response rates are extreme—that is, when a single instance of either is rare. In applied settings, undefined log ratios occur because sessions last for too short a duration, reinforcer rates are out of the experimenter's control, subjects may not have had the occasion to respond, or reinforcers were not delivered. Without any treatment, undefined values are ultimately dropped from analyses, becoming missing values despite their meaningfulness as rare but still probable events.

There have been some tentative remedies for undefined values in matching analysis, which are (a) to drop any session that had a single undefined ratio (known as listwise deletion), (b) to replace any undefined values by an arbitrary *perinull* (values close to 0), or (c) to manipulate data or the experimental design to avoid undefined values. For instance, some studies have simply dropped the undefined values (Caron et al., 2017; Rivard et al., 2014), whereas others have used the average response or reinforcer rates divided by 100 (Ecott & Critchfield, 2004; Seniuk et al., 2015, 2020). Although replacing values appears to be a convenient solution, its adequacy depends on the order of magnitude and variability of the components. It should be used with caution rather than systematically. A potential problem is, for instance, that the mean divided by 100 could be greater than 1 (the lowest nonzero value) if the average is greater than 100, breaking ordinality by replacing an undefined value with a value higher than the lowest one. A more systematic approach could be to use a value of $1/2$, $1/10$, or $1/100$ instead. A similar treatment was recommended by Hautus (1995) who suggested adding a constant of .50 to all values in signal-detection experiments, a task similar to matching. It remains arbitrary, but at least it is independent of the sample. The third option is to manipulate data (such as pooling) to avoid undefined ratios, for example, by summing responses and reinforcers over a large number of sessions. This tends to eliminate infinite ratios; however, pooling data across sessions can wash out within-subject variance and lose considerable amounts of information, ultimately distorting results (Caron, 2019). Finally, there is the procedure of continuing sessions until the infinite ratio vanishes—that is, for at least one instance of each component in a session—which strongly favors the nonrare reinforcer and response rates and distorts the results.

A statistical method to consider is full information maximum likelihood (FIML), a well-known method to account for missing data in the methodological literature, even being qualified as the *state of the art* for handling missing data (Enders & Bandalos, 2001; Schafer & Graham, 2002). This treatment is different from the previous ones notably because it handles data at the log-ratios' level rather than at the component level. Because undefined ratios are a type of missing data (data unsuitable for statistical analysis), FIML could have some potential as a treatment. In FIML, missing values are not replaced or imputed but are handled in the analysis. Full information maximum likelihood assumes that data are (a) at least *missing at random*—that is, missingness is related to the observed data but not to some other unobserved phenomenon that undefined values correspond to—and (b) have a multivariate normal distribution. Its goal is to minimize the -2 log likelihood function with all paired data for the covariance matrix's elements and all the available data for the mean vector.[1] All available subjects' data are used even if a single element is missing, which is in sharp contrast with listwise deletion, where the subject is dropped completely if a single element is missing.

Objective

Given the existing treatments to handle undefined values in matching analyses, it is surprising that there has been no systematic comparison. Such study may lead to meaningful recommendations for behavior analysts when they perform matching analyses with extreme ratios or very rare behavioral occurrences. Thus, the purpose of the current article is to investigate the adequacy of five treatments: omit undefined ratios (OMIT), use FIML, replace values by the mean of the component divided by 100 (M/100), replace values by the constant 1/10 (1/10), and add a constant .50 to all components (CONS). These treatments will be compared on estimating variance accounted for (VAF, R^2), sensitivity (slope, a), and bias (intercept, $\log c$).

METHOD

This simulation was carried out in R (R Core Team, 2023). The package `lavaan` (Rosseel, 2012) was used to carry the FIML treatment with the `sem()` function.

Generating data

Generating random data based on the matching law poses three main challenges. First, undefined values cannot be simply added to a data set because they are not produced from mere random processes (e.g., missing completely at random data) but due to their low probability of occurrence. Although their probabilities can be fixed, their exact quantity cannot. Second, there is no consensual algorithm to generate pseudorandom numbers following the model implied by the matching law. The model in itself is complex (Davison, 2021). It implies four discrete correlated yet unknown distributions. Both the reinforcer and response rates, as well as their respective log ratios, must be strongly correlated. Finally, parametrizing these distributions is not straightforward because there are infinitely many possible configurations. There are 14 free parameters to estimate (four means, four variances, and six covariances between components) and four statistical distributions to choose, and the component level must yield an adequate molar level—that is, matching behavior. Worst, there is no analytically derived solution to obtain VAF, a, and $\log c$ directly from the reinforcer and response rates, only numerical estimations, even though there is an approximation for the VAF (Caron, 2017).

To generate matching analysis data at the molar and component levels, we chose to generate, at first, log ratios of response and reinforcer rates from a bivariate normal distribution with known slope, intercept, and

[1] For additional information on the topic of full information maximum likelihood to deal with missing values, see https://real-statistics.com/handling-missing-data/.

VAF. There is some evidence that log ratios are indeed Gaussian distributed (Tustin & Davison, 1978). Due to this, parameters were known a priori, which are useful to compute statistical bias, relative bias, and root mean square error (RMSE). We can also derive the expected values given the doubly truncated multivariate normal (Manjunath & Wilhelm, 2021; Wilhelm & Manjunath, 2010). Then, we derived response and reinforcer rates from their log ratios. To do so, we generated log ratios as a Gaussian variable x for $\log\frac{R_1}{R_2}$, so that

$$\log\frac{R_1}{R_2} = x. \quad (2)$$

To solve the two unknowns (R_1 and R_2) against the one known value (x), we can postulate a maximum number of reinforcers (or responses) by session, R_t, an assumption inspired by Caron (2015), so that $R_t = R_1 + R_2$, which yields

$$\log\frac{R_1}{R_t - R_2} = x. \quad (3)$$

The parameter R_t can now be manipulated. The last step is to solve the equation for R_1, which yields with some algebraic manipulation:

$$R_1 = \frac{R_t e^x}{e^x + 1}. \quad (4)$$

These manipulations can be easily carried to R_2 and B_1 and B_2 by fixing a maximum response rate B_t. To simulate real data, which are discrete values, the obtained values were finally rounded.

To manipulate undefined value proportions, we used thresholds, referred to herein as α, based on the Gaussian distribution. As both undefined log ratios and the tails of the Gaussian distribution correspond to rare events, they can be readily translated. In other terms, a Gaussian value over a certain threshold is deemed too rare or too extreme and was set to have a rate of zero. For instance, the proportion for 10% of undefined values has a bilateral threshold of 1.645. Log ratios over 1.645 yield a denominator (B_2 and R_2) of zero, whereas log ratios below −1.645 have a numerator (B_1 and R_1) of zero. This threshold is reminiscent of the Type I error rate threshold α in null hypothesis testing.

Fitting models

Five treatments of undefined values were assessed: (a) OMIT, which computes the matching analyses by removing undefined values by listwise deletion (removing a session with at least one undefined value); (b) FIML estimation, which uses all available data to estimate the model; (c) M/100, which replaces undefined values with the average response or reinforcer rate divided by 100 before computing the matching analyses; similarly, (d) 1/10 replaces undefined values with the value 1/10 before carrying the analyses; and finally, (e) CONS adds a constant .50 to all components.

Simulations

Four scenarios were investigated: The main one uses $R^2 = .81$, $a = .90$, $\log_c = 0$ (close to matching behavior); the second uses $R^2 = .49$, $a = .70$, $\log_c = 0$; the third uses $R^2 = .64$, $a = 1$, $\log_c = .1$; and the last uses $R^2 = .0$, $a = 0$, $\log_c = 0$ (Type I error rate condition). Four factors were investigated: four levels of sample size (i.e., number of sessions), $n = 25, 50, 75, 100$; eleven levels of undefined values proportions, $\alpha = 0\%$ to 25% by increments of 2.5%; and three levels of maximum response and reinforcer rates, $B_t = 50, 100, 150$, which yield 396 conditions. Each was replicated 5,000 times. For each iteration, an artificial data set was generated. The log ratios were sampled from a bivariate normal distribution with the VAF (R^2), slope (sensitivity, a), and intercept (bias, $\log c$) fixed. Response and reinforcer rates were then computed from the procedure shown in previous sections, and `NA` values (a missing value indicator) were added according to the threshold. Log ratios were computed again based on the obtained rates. On each artificial data set, OMIT, FIML, M/100, 1/10, and CONS were assessed. Estimates of R^2, a, and $\log_c$ were recorded as well as the proportion of undefined values. The Appendix shows a pseudo-R code example for a given scenario.

RESULTS

For the sake of simplicity, only the main scenario, $R^2 = .81$, $a = .90$, $\log_c = 0$, and the conditions where the maximum reinforcer rate equals the maximum response rate are discussed. All scenarios showed similar findings, and there was no visual and statistical difference when the maximum reinforcer rate and the maximum response rate were unequal. All scenarios and conditions are available in the Supplementary Material.

Figures 1, 2, and 3 depict the performance of the five treatments (colored lines) to handle undefined values when estimating VAF (R^2), sensitivity (a), and bias ($\log_c$), respectively, shown as the y-axis in their corresponding figure. The black dotted lines are the parameters at the population level.

Figure 1 shows that FIML had the best performance overall to estimate VAF. All methods underestimated the true VAF. The underestimation increased slightly as the proportions of undefined values increased, an expected result because there is an increasing loss of information on the distribution's tails of the log ratios. The FIML treatment was slightly better than OMIT at all tested

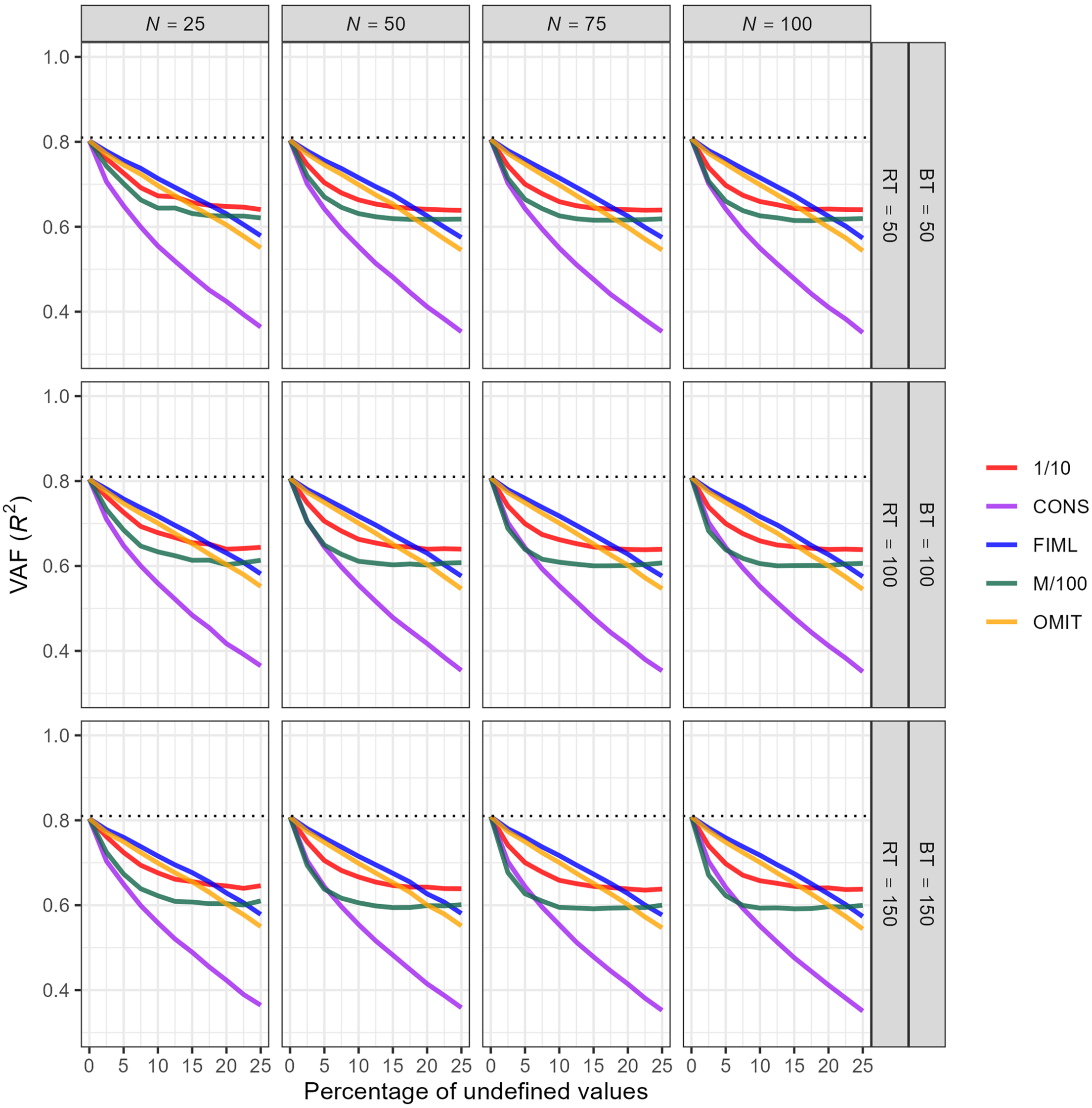

FIGURE 1 The effect of the percentage of undefined values on variance accounted for. Rows depict different levels of maximum responses (BT) and maximum reinforcers (RT), and columns depict different sample sizes (*N*). The black dotted lines represent the expected parameter at the population level. Each colored line represents a treatment to deal with undefined ratios.

levels. The treatments 1/10 and M/100 showed a curvilinear trend. Their performances in estimation decreased rapidly until they reached a plateau at 10%. Treatments 1/10 and M/100 remained constant afterward and even surpassed OMIT and FIML at 20% of undefined values. Although, at this percentage of undefined values, data would be dubious. Both the 1/10 and M/100 treatment had poor performances at low percentages of undefined values, but the 1/10 treatment was slightly better than M/100. The CONS treatment had the worst performance.

Figure 2 depicts performance of the five treatments to handle undefined values when estimating sensitivity (*a*). The FIML and OMIT treatments had the best overall performance among the treatments, both having the exact same performance. The underestimation increased slightly as the proportions of undefined values increased. Treatments M/100 and 1/10 had poor performances at a low percentage of undefined values and reached a plateau at 10% undefined values. They reached and even surpassed the performance of OMIT and FIML at very high percentages (<15%) of undefined values. Again, the CONS treatment had the worst performance. The treatments' performance of sensitivity was similar to that for VAF (Figure 1). There were two main differences: the treatments' estimation was slightly less statistically biased for sensitivity relative to VAF, and M/100 was better to estimate sensitivity than 1/10.

Regarding bias, Figure 3 shows that all methods were equal to recover the true parameter value of 0, except for CONS, which strongly overestimated the parameter in an increasing fashion with the proportion of undefined

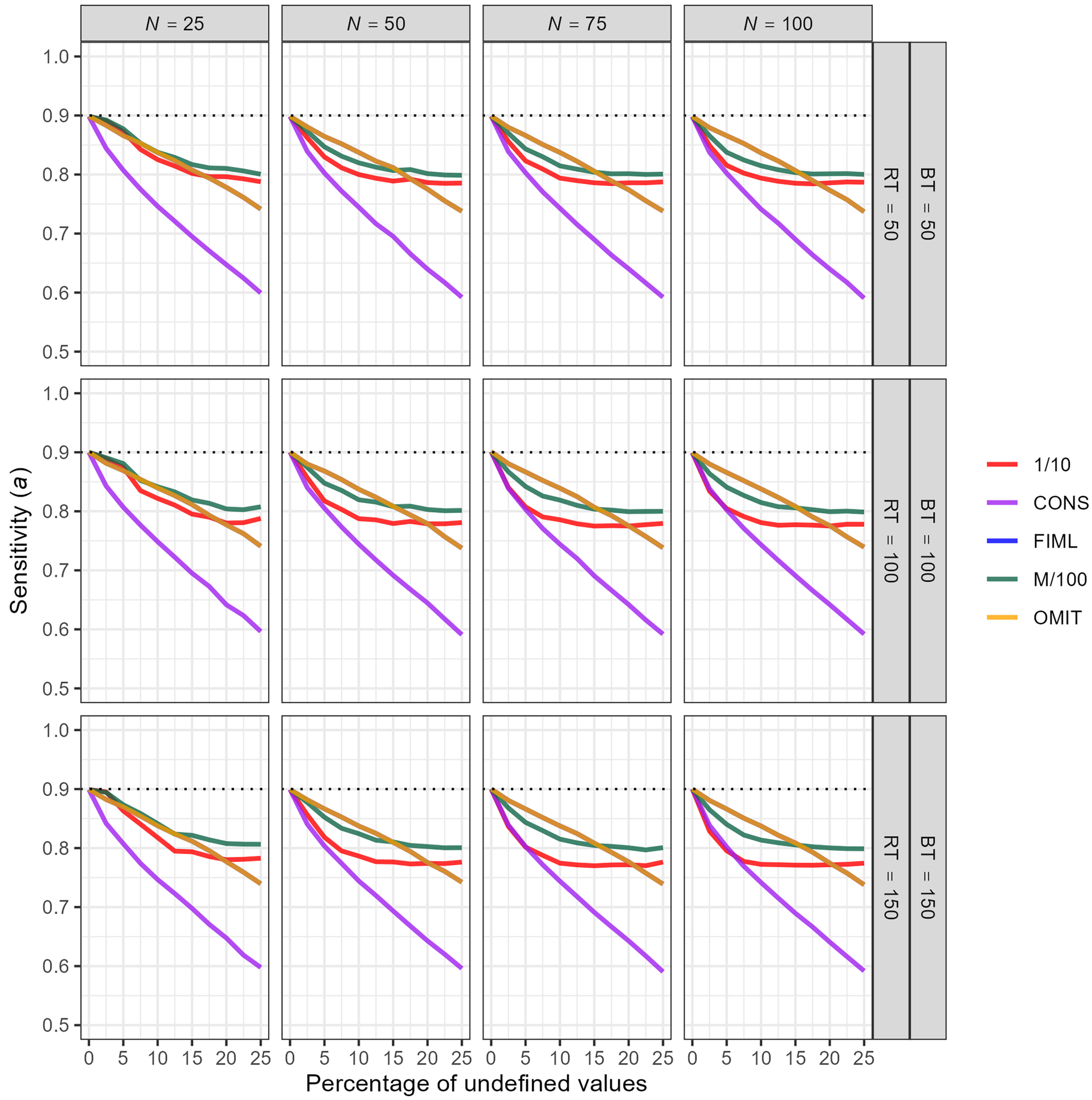

FIGURE 2 The effect of the percentage of undefined values on sensitivity (slope). Rows depict different levels of maximum responses (BT) and maximum reinforcers (RT), and columns depict different sample sizes (N). The black dotted lines represent the expected parameter at the population level. Each colored line represents a treatment to deal with undefined ratios. FIML and OMIT had the exact same performance and cannot be distinguished. Only OMIT is visible.

values. This result is expected because CONS adds a constant to every component, the influence of which is preserved at the molar level. Scenarios with nonzero bias (shown in the Supplementary Material), OMIT and FIML, were better to estimate the parameter than 1/10 and M/100, which showed strong bias in these cases.

As complementary information from previous figures, Table 1 shows the statistical bias, relative bias (RB), and RMSE of the three parameters R^2, a, and $\log_c$ according to proportions of undefined values (α) and treatments. Based on previous results, it is of no surprise that FIML had the lowest bias, relative bias, and RMSE in most conditions, except at very high percentages of undefined values. Regarding bias, even though all treatments, except CONS, reached the same bias, FIML had the lowest RMSE, which further incentivizes its usage.

To get a better understanding of the relation between the maximum number of responses, maximum number of reinforcers, sample size, proportion of undefined values, and the treatment, a regression analysis was carried out with the top-performing approaches (FIML, OMIT). Treatments 1/10, M/100, and CONS were not analyzed because of their consistently poor performance. As the number of replications was high, a significance level of $p < .001$ was used. The complete results are presented in the Supplementary Material. We found four significant results regarding VAF: a double interaction between Undefined × Treatment and its two simple effects, undefined and treatment, as well as a double interaction Undefined × Sample Size. Regarding sensitivity, we found only the effect of treatment. Given results presented in Figures 1 and 2, the results are straightforward: FIML gets

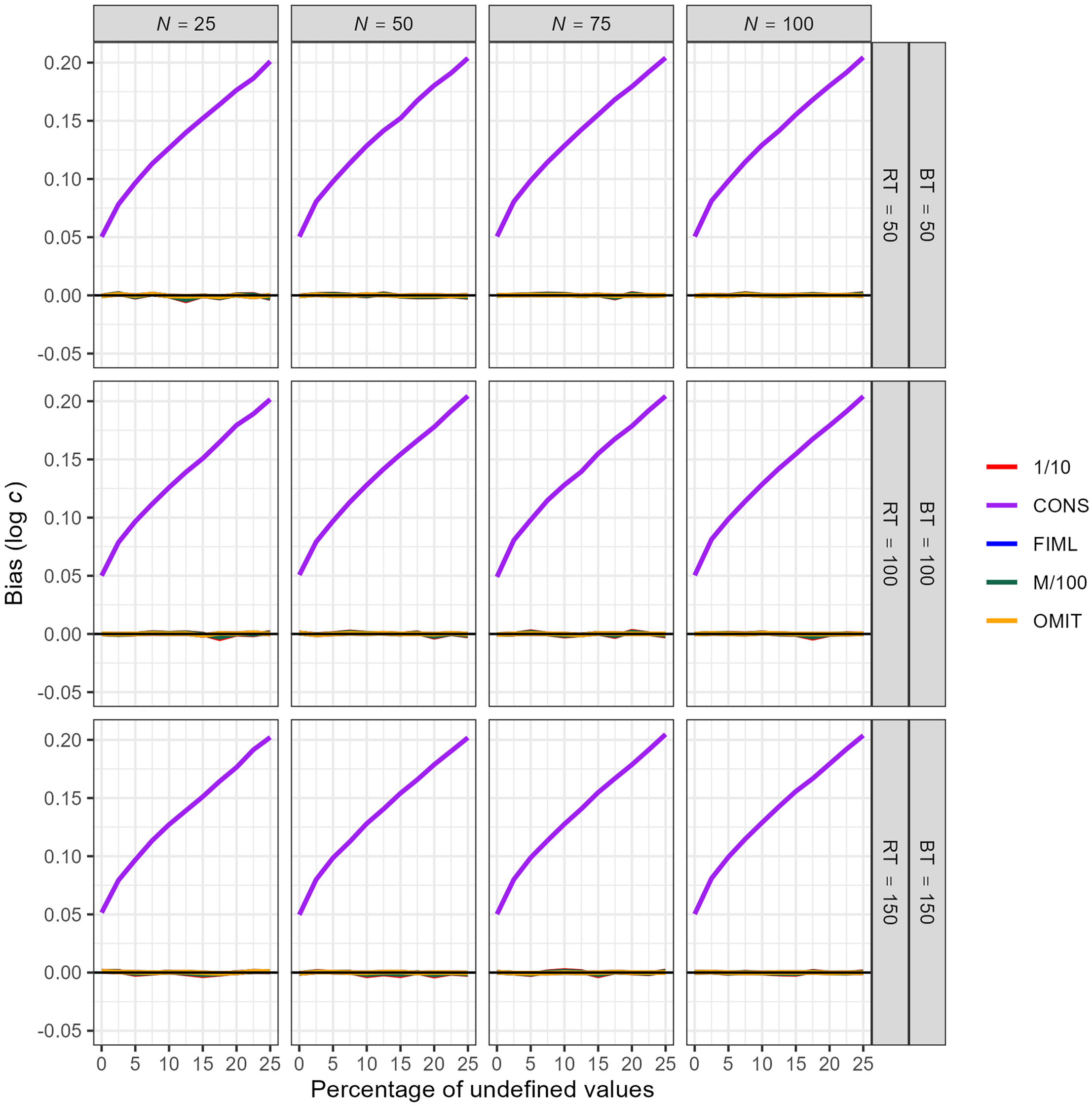

FIGURE 3 The effect of the percentage of undefined values on bias (intercept). Rows depict different levels of maximum responses (BT) and maximum reinforcers (RT), and columns depict different sample sizes (*N*). The black dotted lines represent the expected parameter at the population level. Each colored line represents a treatment to deal with undefined ratios. All treatments but CONS had approximately the same performance and can hardly be distinguished from the expected dotted line.

significantly better than OMIT as the proportion of undefined values increases for VAF, and their performance was the same for sensitivity.

DISCUSSION

The purpose of the current article was to compare five treatments to handle undefined log ratios in matching analyses. These treatments were OMIT, FIML, 1/10, M/100, and CONS. To our knowledge, it is the first study to investigate and compare treatments to deal with undefined log ratios for matching analyses.

Among the five treatments, FIML had the best performance overall, closely followed by OMIT. The FIML treatment was slightly more accurate than OMIT when estimating VAF. They had the same performance when estimating sensitivity. It is worth noting that OMIT's performances matched perfectly the expected values derived from the doubly truncated multivariate normal (Manjunath & Wilhelm, 2021; Wilhelm & Manjunath, 2010), further promoting FIML's effectiveness. The OMIT and FIML treatments were slightly biased for low percentages of undefined values regarding VAF and slope but to a lesser extent than M/100 (for sensitivity) and 1/10 (for VAF), which were both strongly biased even at low percentages. Moreover, the M/100 and 1/10 treatments were biased for the intercept (bias) when the parameters were nonzero (shown in the Supplementary Material). The CONS treatment had the worst performance, producing strongly biased estimates for each parameter, especially when estimating the intercept ($\log_c$). Adding a constant to

TABLE 1 Bias, relative bias (RB), and root mean square error (RMSE) of the estimands by treatments and proportions of undefined values α.

	VAF			Sensitivity			Bias	
α ∼ Treatment	**Bias**	**RB**	**RMSE**	**Bias**	**RB**	**RMSE**	**Bias**	**RMSE**
α = .000								
1/10	0.00	0.00	0.05	0.00	0.00	0.07	0.00	0.03
CONS	0.00	0.00	0.05	0.00	0.00	0.07	0.05	0.07
FIML	0.00	0.00	0.05	0.00	0.00	0.07	0.00	0.03
M/100	0.00	0.00	0.05	0.00	0.00	0.07	0.00	0.03
OMIT	0.00	0.00	0.05	0.00	0.00	0.07	0.00	0.03
α = .025								
1/10	−0.06	−0.08	0.11	−0.04	−0.05	0.22	0.00	0.05
CONS	−0.11	−0.13	0.14	−0.06	−0.07	0.11	0.08	0.10
FIML	−0.03	−0.04	0.06	−0.02	−0.02	0.07	0.00	0.03
M/100	−0.11	−0.13	0.17	−0.02	−0.03	0.16	0.00	0.04
OMIT	−0.04	−0.05	0.07	−0.02	−0.02	0.07	0.00	0.03
α = .050								
1/10	−0.10	−0.13	0.16	−0.07	−0.08	0.24	0.00	0.06
CONS	−0.17	−0.20	0.19	−0.10	−0.11	0.13	0.10	0.11
FIML	−0.05	−0.06	0.08	−0.03	−0.04	0.08	0.00	0.03
M/100	−0.16	−0.19	0.22	−0.05	−0.05	0.18	0.00	0.05
OMIT	−0.06	−0.08	0.09	−0.03	−0.04	0.08	0.00	0.03
α = .075								
1/10	−0.13	−0.16	0.18	−0.09	−0.10	0.24	0.00	0.08
CONS	−0.21	−0.26	0.24	−0.13	−0.14	0.16	0.11	0.13
FIML	−0.07	−0.09	0.10	−0.05	−0.05	0.09	0.00	0.03
M/100	−0.18	−0.22	0.24	−0.07	−0.07	0.18	0.00	0.06
OMIT	−0.09	−0.11	0.11	−0.05	−0.05	0.09	0.00	0.03
α = .100								
1/10	−0.15	−0.18	0.19	−0.10	−0.11	0.23	0.00	0.09
CONS	−0.26	−0.32	0.28	−0.16	−0.17	0.19	0.13	0.14
FIML	−0.09	−0.12	0.12	−0.06	−0.07	0.10	0.00	0.03
M/100	−0.19	−0.24	0.25	−0.08	−0.09	0.18	0.00	0.06
OMIT	−0.11	−0.14	0.13	−0.06	−0.07	0.10	0.00	0.03
α = .125								
1/10	−0.16	−0.19	0.20	−0.11	−0.12	0.22	0.00	0.10
CONS	−0.29	−0.36	0.31	−0.18	−0.20	0.21	0.14	0.15
FIML	−0.12	−0.14	0.14	−0.08	−0.08	0.12	0.00	0.03
M/100	−0.20	−0.25	0.25	−0.08	−0.09	0.17	0.00	0.07
OMIT	−0.13	−0.17	0.16	−0.08	−0.08	0.12	0.00	0.03
α = .150								
1/10	−0.16	−0.20	0.20	−0.11	−0.13	0.21	0.00	0.10
CONS	−0.33	−0.41	0.35	−0.21	−0.23	0.23	0.15	0.17
FIML	−0.14	−0.17	0.16	−0.09	−0.10	0.13	0.00	0.03
M/100	−0.20	−0.25	0.25	−0.09	−0.10	0.17	0.00	0.08
OMIT	−0.16	−0.20	0.18	−0.09	−0.10	0.13	0.00	0.03
α = .175								
1/10	−0.17	−0.20	0.21	−0.12	−0.13	0.20	0.00	0.11
CONS	−0.36	−0.45	0.38	−0.23	−0.26	0.26	0.17	0.18
FIML	−0.16	−0.20	0.18	−0.11	−0.12	0.15	0.00	0.03

(Continues)

TABLE 1 (Continued)

α ∼ Treatment	VAF			Sensitivity			Bias	
	Bias	RB	RMSE	Bias	RB	RMSE	Bias	RMSE
M/100	−0.20	−0.25	0.25	−0.09	−0.10	0.16	0.00	0.08
OMIT	−0.18	−0.23	0.21	−0.11	−0.12	0.15	0.00	0.03
α = .200								
1/10	−0.17	−0.21	0.21	−0.12	−0.13	0.20	0.00	0.12
CONS	−0.40	−0.49	0.41	−0.26	−0.29	0.28	0.18	0.19
FIML	−0.18	−0.23	0.21	−0.12	−0.14	0.16	0.00	0.03
M/100	−0.20	−0.25	0.24	−0.10	−0.11	0.16	0.00	0.09
OMIT	−0.21	−0.26	0.23	−0.12	−0.14	0.16	0.00	0.03
α = .225								
1/10	−0.17	−0.21	0.21	−0.12	−0.13	0.19	0.00	0.13
CONS	−0.42	−0.52	0.44	−0.28	−0.31	0.30	0.19	0.20
FIML	−0.21	−0.26	0.23	−0.14	−0.16	0.18	0.00	0.03
M/100	−0.20	−0.25	0.24	−0.10	−0.11	0.15	0.00	0.09
OMIT	−0.23	−0.29	0.26	−0.14	−0.16	0.18	0.00	0.03
α = .250								
1/10	−0.17	−0.21	0.20	−0.11	−0.13	0.18	0.00	0.13
CONS	−0.45	−0.56	0.47	−0.31	−0.34	0.33	0.20	0.21
FIML	−0.23	−0.29	0.26	−0.16	−0.18	0.20	0.00	0.03
M/100	−0.20	−0.25	0.23	−0.10	−0.11	0.15	0.00	0.10
OMIT	−0.26	−0.32	0.29	−0.16	−0.18	0.20	0.00	0.03

every component is thus problematic because the intercept ($\log_c$) has a meaningful interpretation. The OMIT and FIML did not produce biased estimates of the intercept ($\log_c$), but FIML had the lowest RMSE in all conditions. We thus recommend FIML when dealing with a reasonable proportion of undefined values.

The current simulations have some limitations. Data were simulated on the assumption of a maximal number of responses and reinforcers within a session. It also produced the log ratios before the actual reinforcer and response rates, although the opposite is true with real organisms. This method, however, simplified the computation beforehand by stipulating matching behavior a priori (with known VAF, sensitivity, and bias) and ensuring that the molar level was in accordance with the component level. Alternative computational modeling could be used in future simulations, like different data generation techniques and a wider range of parameters. New treatments and improvements to deal with undefined ratios should be considered.

In conclusion, FIML had the best overall performance, being the most accurate on VAF and sensitivity when undefined values were present. This first study on the treatment of undefined log ratios in matching analysis suggests that treatments like CONS, M/100, and 1/10 should be avoided. Treatments like FIML or to a lesser extent OMIT are recommended to deal with undefined log ratios. We hope this current study will stimulate the development of new methods to deal with statistical issues in behavior analysis as well as to promote good data-analysis practices.

ACKNOWLEDGMENTS

The author would like to thank Sarah Cowie, Michael Davison, and an anonymous reviewer for their helpful comments on a previous draft of this manuscript.

CONFLICT OF INTEREST STATEMENT

The authors report no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the Supplementary Material of this article.

ETHICS APPROVAL

No human or animal subjects were used in the production of this article.

ORCID

Pier-Olivier Caron https://orcid.org/0000-0001-6346-5583

REFERENCES

Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, *22*(1), 231–242. https://doi.org/10.1901/jeab.1974.22-231

Caron, P.-O. (2015). Matching without learning. *Adaptive Behavior*, *23*(4), 227–233. https://doi.org/10.1177/1059712315590484

Caron, P.-O. (2017). On the statistical properties of operant settings and their contribution to the evaluation of sensitivity to reinforcement. *Adaptive Behavior*, *25*(4), 185–193. https://doi.org/10.1177/1059712317719967

Caron, P.-O. (2019). Multilevel modeling of matching behavior. *Journal of the Experimental Analysis of Behavior*, *111*(2), 183–191. https://doi.org/10.1002/JEAB.510

Caron, P.-O., Forget, J., & Rivard, M. (2017). The allocation of playing behavior of children in mainstream day care according to social reinforcer ratios. *European Journal of Behavior Analysis*, *18*(2), 242–250. https://doi.org/10.1080/15021149.2017.1309956

Davison, M. (2021). Concurrent schedules: How responses per reinforcer affects estimates of sensitivity to reinforcement. *Journal of the Experimental Analysis of Behavior*, *116*(1), 114–123. https://doi.org/10.1002/jeab.694

Davison, M., & McCarthy, D. (1988). *The matching law: A research review*. Erlbaum.

Ecott, C. L., & Critchfield, T. S. (2004). Noncontingent reinforcement, alternative reinforcement, and the matching law: A laboratory demonstration. *Journal of Applied Behavior Analysis*, *37*(3), 249–265. https://doi.org/10.1901/jaba.2004.37-249

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Manjunath, B. G., & Wilhelm, S. (2021). Moments calculation for the doubly truncated multivariate normal density. *Journal of Behavioral Data Science*, *1*(1), 17–33. https://doi.org/10.35566/jbds/v1n1/p2

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rivard, M., Forget, J., Kerr, K., & Bégin, J. (2014). Matching law and sensitivity to therapist's attention in children with autism spectrum disorders. *The Psychological Record*, *65*(1), 79–88. https://doi.org/10.1007/s40732-014-0015-1

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://www.jstatsoft.org/v48/i02/

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Seniuk, H. A., Vu, J. P., & Nosik, M. R. (2020). Application of the matching law to mixed martial arts. *Journal of Applied Behavior Analysis*, *53*(2), 846–856. https://doi.org/10.1002/jaba.653

Seniuk, H. A., Williams, W. L., Reed, D. D., & Wright, J. W. (2015). An examination of matching with multiple response alternatives in professional hockey. *Behavior Analysis: Research and Practice*, *15*(3–4), 152–160. https://doi.org/10.1037/bar0000019

Tustin, R. D., & Davison, M. C. (1978). Distribution of response ratios in concurrent variable-interval performance. *Journal of the Experimental Analysis of Behavior*, *29*(3), 561–564. https://doi.org/10.1901/jeab.1978.29-561

Wilhelm, S., & Manjunath, B. G. (2010). tmvtnorm: A package for the truncated multivariate normal distribution. *The R Journal*, *2*(1), 25–29. https://doi.org/10.32614/RJ-2010-005

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Caron, P.-O. (2024). Treatments for undefined log ratios in matching analyses. *Journal of the Experimental Analysis of Behavior*, 1–10. https://doi.org/10.1002/jeab.925

APPENDIX

A pseudo-R code example for a given scenario.

```
# The scenario

N # Sample size, number of sessions

Rt # Maximum reinforcer rate

Bt # Maximum response rate

nreps # Number of iterations

crit # Threshold to which the log ratio is
deemed rare


# The loop

for (i in 1:nreps){

 # Generate two Gaussian variables with

 # VAF = .81, slope = .9, intercept = 0

 logR <- rnorm(n)

 logB <- .9 * x + sqrt(1 - .81) * rnorm(n)


 # Compute the response and reinforcer rates

 D <- data.frame(B1 = round(Bt * exp(logB) /
 (exp(logB) + 1)),
           B2 = round(Bt * exp(-logB) / (exp
           (-logB) + 1)),

           R1 = round(Rt * exp(logR) / (exp
           (logR) + 1)),

           R2 = round(Rt * exp(-logR) / (exp
           (-logR) + 1)))


 # Transform zeros to NA


 D$B1 = ifelse(logB < -crit, NA, D$B1)

 D$B2 = ifelse(logB > crit, NA, D$B2)

 D$R1 = ifelse(logR < -crit, NA, D$R1)

 D$R2 = ifelse(logR > crit, NA, D$R2)

 D[D == 0] <- NA
```

```
# Compute new log ratios from response and
reinforcer rates

D <- data.frame(D,

          B = log(D[,"B1"]/D[,"B2"],
          base = 10),

          R = log(D[,"R1"]/D[,"R2"],
          base = 10))


# The number of undefined log ratios

nmiss <- sum(!(!is.na(D$R) || !is.na(D$B)))

# Carry the five analyses and record VAF,
slope and intercept

res.omit <- omit(D)
res.fiml <- fiml(D)
res.1.10 <- M1.10(D)
res.M.100 <- M.100(D)
res.cons <- cons(D)

}
```