

# Prevalence of Site-Level Missing Data in a National Evaluation of Programs for At-Risk Families

Jaime Ballard, PhD, Adeya Richmond, PhD, Suzanne van den Hoogenhof, PhD, Lynne M. Borden, PhD, & Daniel Perkins, PhD

## Background

- Data can be missing at the individual level or at a higher level, such as family, classroom, or program site.
- Although the impact of missing data has been studied at the individual level and statistical methods have been developed to account for it, missing data at the *site* level is still under-studied.
- Common best practices for reporting missing data currently do not include recommendations about reporting site-level missing data (e.g., Schlomer, Bauman, & Card, 2009).
- Increased knowledge of higher-level missing data is necessary to develop evaluation design and statistical methods to address it.

## The Present Study

We aim to identify the prevalence of site-level missing data in a national dataset of programs serving at-risk families.

Data were drawn from the Children, Youth, and Families at Risk (CYFAR) Grant Program, a national initiative that provides funding to after-school programs supporting at-risk youth and their families.

## Methods

- Sample includes 1,420 adult participants from 17 programs for at-risk families. Participants were 68% female, with an average age of 35.2 (SD = 17.4). 39% were Hispanic or Latino. 52% were white, 15.92% were Black, and 7.9% were another racial minority. 19% had less than a high school education, 23% had a high school degree, and 51% had post-secondary education.
- Participants completed self-report measures of demographics and participation. Each item was coded as present or missing at the individual and site level.

## Results

Table 1. Prevalence of individual and site-level missing data

	Individual missing (includes site-level missing)	Site-level missing	Percentage of missing data due to site-level	Total Percentage of Missing Data
Demographics				
Gender	66	0	0%	5%
Age	242	0	0%	17%
Ethnicity	244	0	0%	17%
Employment	124	0	0%	9%
Education	118	0	0%	8%
Race	19	0	0%	1%
Military	191	54	28%	13%
Active Duty	198	59	30%	14%
Military Branch	186	54	29%	13%
Participation				
Number Sessions	465	227	49%	33%
Hours/ week	588	227	39%	41%
Length of Participation in 4-H	538	227	42%	38%
Other Activity involvement	434	187	43%	31%
Number of other activities	1025	187	18%	72%

## Results (continued)

- Nearly one third of all programs had site-level missing data, including 10 sites from 5 programs.
- Site- or program-level missing data accounted for 27.53% of all missing data.

## Discussion

- The high prevalence of site-level missing data we found in our study suggests an urgent need to refine statistical methods to account for site-level missing data.
- One strategy for minimizing site-level missing data is reducing participant and site burden associated with evaluations, either by selecting only the central variables or by using approaches such as multiform questionnaire protocols, two-method measurement models, or the wave-missing longitudinal design (Little, Jorgenson, Lang, & Moore, 2014).
- Researchers can invest time and support resources in generating buy-in not only at a program level, but at a site level, to ensure that all questions are administered.

## Acknowledgements

The authors gratefully acknowledge funding provided by the United States Department of Agriculture's National Institute of Food and Agriculture (USDA-NIFA) through a cooperative agreement with the University of Minnesota and Pennsylvania State University under Grant 2018-41520-28908.