

# MLDS Center Suppression Guidance for Researchers

Before any material can be released from the secure environment, it must be submitted for “suppression review” and approved by the Executive Director. This document provides guidance for MLDS researchers on how to prepare your materials for this review.

The MLDS Center suppression policy places certain restrictions on the information that can be used from the MLDS system. Essentially, descriptive statistics with any cells with *ns* less than 10 cannot be released. “Descriptive statistics” means any data representing counts or percentages of individuals (e.g., students, workers). Counts and percentages of other units that are not individual people do not constitute descriptive statistics subject to the suppression policy but should be clearly explained for the review process. Occasionally, the Center may allow release of materials with cell sizes less than 10, but only if the Executive Director, in consultation with the Center’s Disclosure Review Advisory Board, determines that the information has little or no risk of revealing identities and the rigid suppression standard would limit the utility of the information being reported.

If the material contains no descriptive statistics (i.e. contains no counts of observations) then suppression is not required. For example, regression analysis results (coefficients, *p* values) do not need to be suppressed, even if they are based on a group with fewer than 10 members, as long as the number of members is not included. However, the researcher still needs to submit this material for the suppression review process in order to remove it from the virtual machine.

Researchers should always involve their principal investigator in the suppression process. Principal investigators are responsible for reviewing and approving the materials prior to their submission for review by the MLDS Center. It is typically helpful to have multiple people review the materials prior to submission. You should make both the suppressed and unsuppressed versions of the materials available for review by copying them to a dedicated folder in your project folder on the O drive. For details on the Center’s suppression policy and how to submit a suppression review request, see <https://mldscenter.maryland.gov/ResearcherSuppressionRequest.html>. For questions about the process, you can contact Bess Rose, MLDS Center Research Branch Statistician, by email at [bess.rose@maryland.gov](mailto:bess.rose@maryland.gov).

This document walks through some examples of data tables, how to assess whether suppression is needed, and how to correctly apply suppression. Problematic rows with *ns* less than 10 are highlighted in yellow. Green highlighting indicates rows where suppression has been applied or rows have been combined to eliminate small *ns*.

# Example 1: Descriptive table with percentages

Figure 1a shows a table of percentages by subgroup and a total N without subgroup n's.

**Figure 1a. Descriptive Table With Percentages That Cannot Be Reviewed Because It Omits Subgroup Ns.**

Race	Percent
American Indian or Alaska Native	3.00%
Asian	11.00%
Black	48.00%
Native Hawaiian or Other Pacific Islander	1.00%
Two or More Races	9.00%
White	28.00%
Total (N=500)	100.00%

First, the lack of subgroup *ns* in Figure 1a makes it difficult for the reviewer to assess the need for suppression in the table, so please always include the *ns* whenever (as in Figure 1a) you are providing the total *N* and subgroup percentages - even if you do not plan to report them in the final product. See Figure 1b that correctly includes an *N* column.

**Figure 1b. Descriptive Table With Ns and Percentages That Cannot Be Released.**

Race	<i>N</i>	Percent
American Indian or Alaska Native	15	3.00%
Asian	55	11.00%
Black	240	48.00%
Native Hawaiian or Other Pacific Islander	5	1.00%
Two or More Races	45	9.00%
White	140	28.00%
Total	500	100.00%

In Figure 1b, because the total *N* is reported (500), it is possible that a reader could calculate the 1% for Native Hawaiian or Other Pacific Islander to *n*=5, which is less than 10, even if you do not plan to report the subgroup *ns*.

One solution would be to remove the total *N* so that the *ns* represented by the percentages cannot be derived (see Table 1c). You would still need to show the *N* column for the reviewer,

but make a note that that column would not be reported. However, it would be rare to have a situation where you are not reporting the total *N*.

**Figure 1c. Descriptive Table With *N*s and Percentages That Can Be Released If *N* Column Is Removed and Total *N* Is Not Reported.**

Race	<i>N</i> (note for reviewer: this column will be removed and the total <i>N</i> will not be reported)	Percent
American Indian or Alaska Native	15	3.00%
Asian	55	11.00%
Black	240	48.00%
Native Hawaiian or Other Pacific Islander	5	1.00%
Two or More Races	45	9.00%
White	140	28.00%

Another solution is to apply **primary suppression** to the small subgroup *ns* and percentages (i.e., replace the numbers with asterisks) and apply **complementary suppression** to another row (even though that row has an *n* greater than 10), because a reader could calculate the *n*=5 (see Table 1d).

**Figure 1d. Descriptive Table With *N*s and Percentages That Can Be Released With Suppression.**

Race	<i>N</i>	Percent
American Indian or Alaska Native	*	*
Asian	55	11.00%
Black	240	48.00%
Native Hawaiian or Other Pacific Islander	*	*
Two or More Races	45	9.00%
White	140	28.00%
Total	500	100.00%

Note. \* indicates value was suppressed (n under 10 or complementary suppression)

Another option is to **combine the category** with the small *n* with another category (see Figure 1e). This would avoid the need for suppression altogether.

**Figure 1e. Descriptive Table With *N*s and Percentages That Can Be Released With Combined Rows.**

<b>Race</b>	<b><i>N</i></b>	<b>Percent</b>
American Indian, Alaska Native, Native Hawaiian or Other Pacific Islander	20	4.00%
Asian	55	11.00%
Black	240	48.00%
Two or More Races	45	9.00%
White	140	28.00%
Total	500	100.00%

## Example 2: Means and standard deviations with sample size

Figure 2a shows a table with school-level means, standard deviations, and sample sizes.

**Figure 2a. Descriptive Table With *N*s Less Than 10 That Cannot Be Released.**

School	Mean	Std. Dev.	<i>n</i>
School 1	41	NA	1
School 2	79.00463	29.22252	108
School 3	35.01333	35.96714	75
School 4	30.14063	34.72151	64
School 5	2	1.732051	3
School 6	28.95	30.05149	160
School 7	36.34375	37.69006	112
School 8	37.83158	34.07927	95
School 9	21.44595	25.76508	74
School 10	36.57489	35.15511	503

The table in Figure 2a cannot be released, because it includes *n*s that are less than 10.

**Figure 2b. Descriptive Table With *N*'s Less Than 10 That Cannot Be Released.**

School	Mean	Std. Dev.	<i>n</i>
School 1	41	NA	LT10
School 2	79.00463	29.22252	108
School 3	35.01333	35.96714	75
School 4	30.14063	34.72151	64
School 5	2	1.732051	LT10
School 6	28.95	30.05149	160
School 7	36.34375	37.69006	112
School 8	37.83158	34.07927	95
School 9	21.44595	25.76508	74

**Figure 2b. Descriptive Table With N's Less Than 10 That Cannot Be Released.**

School	Mean	Std. Dev.	n
School 10	36.57489	35.15511	503

Note. LT10 indicates value was suppressed (n under 10)

The suppression in Figure 2b partially solves the problem by suppressing the small *ns* using **primary suppression**. However, row 1 is clearly an *n*=1 because std. dev. is NA. This will not be accepted. Instead, all 3 statistics for School 1 should be suppressed, as seen in Figure 2c.

**Figure 2c. Descriptive Table With Ns Less Than 10 That Can Be Released.**

School	Mean	Std. Dev.	n
School 1	*	*	*
School 2	79.00463	29.22252	108
School 3	35.01333	35.96714	75
School 4	30.14063	34.72151	64
School 5	2	1.732051	LT10
School 6	28.95	30.05149	160
School 7	36.34375	37.69006	112
School 8	37.83158	34.07927	95
School 9	21.44595	25.76508	74
School 10	36.57489	35.15511	503

Note. LT10 indicates value was suppressed (n under 10). \* indicates values were suppressed (n=1)

An alternative to suppressing the 3 statistics for School 1 would be to delete the row for School 1 altogether, as seen in Figure 2d.

**Figure 2d. Descriptive Table With Ns Less Than 10 That Can Be Released.**

School	Mean	Std. Dev.	n
School 2	79.00463	29.22252	108
School 3	35.01333	35.96714	75
School 4	30.14063	34.72151	64
School 5	2	1.732051	LT10
School 6	28.95	30.05149	160
School 7	36.34375	37.69006	112

**Figure 2d. Descriptive Table With Ns Less Than 10 That Can Be Released.**

School	Mean	Std. Dev.	<i>n</i>
School 8	37.83158	34.07927	95
School 9	21.44595	25.76508	74
School 10	36.57489	35.15511	503

Note. LT10 indicates value was suppressed (*n* under 10)

In Figure 2d, the row with *n*=1 has been deleted. The row for School 5 with sample size less than 10 but greater than 1 has been suppressed. This is allowable.

Another alternative would be to delete the row for School 5 as well, as seen in Figure 2e. This is also allowable.

**Figure 2e. Descriptive Table With No Ns Less Than 10 That Can Be Released.**

School	Mean	Std. Dev.	<i>n</i>
School 2	79.00463	29.22252	108
School 3	35.01333	35.96714	75
School 4	30.14063	34.72151	64
School 6	28.95	30.05149	160
School 7	36.34375	37.69006	112
School 8	37.83158	34.07927	95
School 9	21.44595	25.76508	74
School 10	36.57489	35.15511	503

## Example 3: Counts and percentages by categories

Figure 3a shows a descriptive table with *Ns* by county and percentages by county and category.

**Figure 3a. Descriptive Table With Percentages That Cannot Be Reviewed Because It Omits Subgroup *Ns*.**

	County <i>N</i>	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5
County 10	656	0.425	0.047	0.003	0.477	0.047
County 11	1121	0.362	0.037	0.018	0.55	0.034
County 12	3207	0.355	0.058	0.031	0.509	0.047
County 13	502	0.231	0.048	0.008	0.677	0.036

The decimals in Figure 3a sum to 1 across categories for each county so presumably they are percentages. Because the total *N* for each row (county) is provided, a reader could calculate the underlying subgroup *ns*. The column header needs to clearly indicate that the column is a percentage. In addition, a column including the category *ns* needs to be added so the reviewer can easily see them, as seen in Figure 3b.

**Figure 3b. Descriptive Table With *Ns* and Percentages That Cannot Be Released Because It Reveals Subgroup *Ns* Less Than 10.**

		Cat 1		Cat 2		Cat 3		Cat 4		Cat 5	
	County <i>N</i>	Pct	<i>n</i>								
County 10	656	0.425	279	0.047	31	0.003	2	0.477	313	0.047	31
County 11	1121	0.362	406	0.037	41	0.018	20	0.55	617	0.034	38
County 12	3207	0.355	1138	0.058	186	0.031	99	0.509	1632	0.047	151
County 13	502	0.231	116	0.048	24	0.008	4	0.677	340	0.036	18

As seen in Figure 3b, adding the *ns* reveals two cells where the *n* is less than 10. These need to be suppressed using **primary suppression**, as shown in Figure 3c. **Complementary suppression** also needs to be applied to another percentage and *n* in the same row so that the percentage and *n* for the small cell cannot be derived. Figure 3c shows these suppression techniques applied correctly.

**Figure 3c. Descriptive Table With Percentages That Correctly Applies Primary and Complementary Suppression And Can Be Released.**

		Cat 1		Cat 2		Cat 3		Cat 4		Cat 5	
	County <i>N</i>	Pct	<i>n</i>								
County 10	656	0.425	279	0.047	31	*	*	0.477	313	*	*
County 11	1121	0.362	406	0.037	41	0.018	20	0.55	617	0.034	38
County 12	3207	0.355	1138	0.058	186	0.031	99	0.509	1632	0.047	151
County 13	502	0.231	116	0.048	24	*	*	0.677	340	*	*

Note. \* indicates value was suppressed (*n* under 10 or complementary suppression)

## Example 4: Multilevel model results

Figure 4a shows multilevel model output for 4 models.

**Figure 4a. Multilevel Model Results Including the Number Of Observations That Can Be Released.**

	Model 0		Model 1		Model 2		Model 3	
	OR	SE	OR	SE	OR	SE	OR	SE
<b>FP1</b>								
cons	0.11***	0.01	0.09***	0.01	0.07***	0.01	0.06***	0.01
hf0_cwc			0.24***	0.01	0.28***	0.01	0.29***	0.02
hf2_cwc			2.02***	0.11	1.97***	0.12	2.10***	0.13
r2bl_cwc					0.41***	0.02	0.41***	0.02
r2ot_cwc					0.55***	0.03	0.55***	0.04
msa6rdg_sco_cwc					1.00**	0.00	1.00**	0.00
msa6mth_sco_cwc					0.98***	0.00	0.98***	0.00
hf0pct10_cgm							0.72***	0.04
hf2pct10_cgm							0.00***	0.00
blpct10_cgm							1.17***	0.04
ot3pct10_cgm							0.70***	0.05
msa6schmean_cgm							0.99***	0.00
<b>RP2</b>								
var(cons)	537.50***	243.34	276.31***	111.93	172.60***	63.41	1,961.85***	1,142.91
<b>OD</b>								
dic	26,947.08		25,798.37		24,617.66		23,888.22	
Number of observations	52,610		52,610		52,610		52,610	

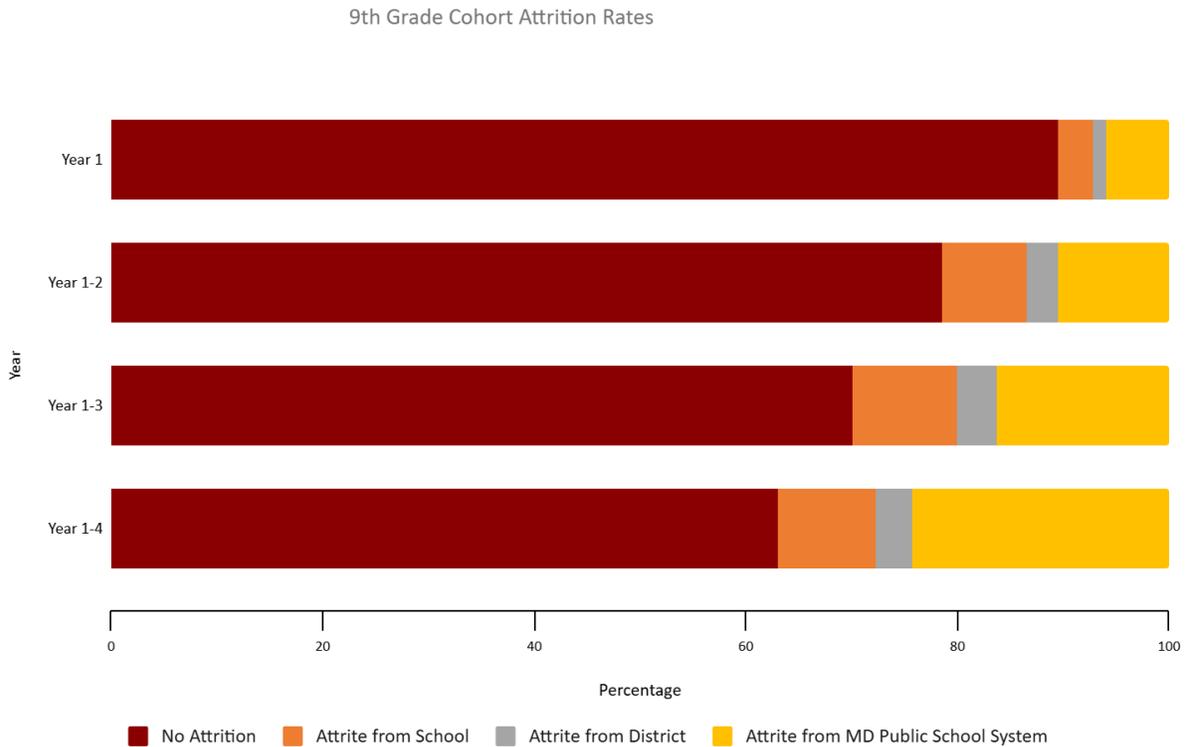
Note. \*  $p < .05$  \*\*  $p < .01$  \*\*\*  $p < .001$ . OR=odds ratio. SE=standard error.

In Figure 4a, the number of observations for each model in the bottom row of the table is considered a descriptive statistic because these are counts of individual people. In this example, these counts do not require suppression because they are all greater than 10. However, the table still must go through the suppression review process to be removed from the virtual machine. It would be helpful to provide the following description to the reviewer: "Figure 4a is a table of multilevel model results predicting the likelihood of dropout. Odds ratios and standard errors are provided for 4 cumulative models. At the bottom of the table there are variance estimates, model fit statistics, and numbers of observations. The numbers of observations reflect counts of individual people. All are greater than 10."

## Example 5: Graph

Figure 5a shows a 100% stacked bar graph of attrition rates.

**Figure 5a. 100% Stacked Bar Graph With Percentages.**



In order for the graph to be reviewed, the researcher must also submit the accompanying data table including percentages and *ns*, as shown in Figure 5b.

**Figure 5b. Data Table to Accompany the Graph in Figure 5a That Can Be Reviewed.**

Year	No Attrition		Attrite from School		Attrite from District		Attrite from MD Public School System		Total
	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	
Year 1-4	63.05	45113	9.28	6642	3.38	2417	24.29	17383	71555
Year 1-3	70.03	50107	9.97	7136	3.72	2664	16.28	11648	71555
Year 1-2	78.61	56248	7.97	5706	2.88	2058	10.54	7543	71555
Year 1	89.58	64097	3.21	2297	1.22	873	5.99	4288	71555

Now the reviewer can see that the underlying  $ns$  for the percentages represented in the graph in Figure 5a are all greater than 10. The graph and table can be released.