



**MLDC Research Areas**

- Definition of Diversity
- Legal Implications
- Outreach & Recruiting
- Leadership & Training
- Branching & Assignments
- Promotion
- Retention
- Implementation & Accountability
- Metrics
- National Guard & Reserve

This issue paper aims to aid in the deliberations of the MLDC. It does not contain the recommendations of the MLDC.

Military Leadership Diversity Commission  
1851 South Bell Street  
Arlington, VA 22202  
(703) 602-0818

<http://mldc.whs.mil/>

## Key Considerations in Interpreting Promotion and Retention Rates

### Abstract

In its tasking from Congress, the MLDC has been asked to examine how the Services’ current policies and practices affect the retention and promotion rates of male and female servicemembers and of servicemembers from different race/ethnicity groups. As part of this tasking, other issue papers (IPs) present raw rates of promotion and retention for key demographic groups to show whether differences exist. To help the commissioners interpret the data presented in those IPs, this IP walks through an example based on raw promotion rates for male and female Air Force colonels. The example shows that simple differences in raw rates seldom provide enough information for firm policy conclusions because of two key limitations. First, it is often difficult to determine whether a difference in raw rates is a good signal of policy importance. Second, comparisons of raw rates by gender or race/ethnicity alone do not account for other factors that affect promotion and retention outcomes for these groups. Although analytical techniques are available to deal with these issues, applying them is not possible within the MLDC time line. Thus, the onus falls on the commissioners to understand the limits of these simple comparisons as they formulate their recommendations.

**T**he congressional charter for the MLDC identifies 16 key taskings for the commissioners. Two of these are to examine “the establishment and maintenance of fair promotion and command opportunities for ethnic- and gender-specific members of the Armed Forces” and “the ability of current activities to increase continuation rates for ethnic- and gender-specific members of the Armed Forces” (10 U.S.C. 596, 2008).

A series of issue papers (IPs) addresses these taskings. In particular, several IPs compare raw retention and promotion rates by demographic group to show what differences exist, if any, and to highlight those differences that merit additional attention to understand what is causing them. To aid the commissioners in properly interpreting the data presented in the other IPs, this IP highlights the key limitations inherent in comparisons of raw rates and explains what conclusions can be drawn from them.

### Interpreting Differences in Raw Rates:

#### A Promotion Example

To illustrate the limitations of comparing raw rates on a single observable characteristic, consider the March 2009 Air Force Colonel Selection results and one demographic variable identified in the charter—gender. Table 1 shows the promotion results separately for male and female line officers in the primary zone.<sup>1</sup>

The promotion data in the table are raw rates, calculated simply by dividing the number selected by the number eligible. For example, 340 men were selected for promotion out of the 784 who were eligible (340/784 equals 43 percent). If we compare the raw promotion-rate percentages alone, it seems that women have been promoted at a greater rate than men. Yet, raw rates such as these raise issues. The first issue is whether the observed difference between the rates is significant. The second is whether the two groups differ across only the dimension in question (in this case, gender) or whether they differ along other dimensions as well.

**Table 1. 2009 Air Force Colonel Selection Results, by Gender**

	Eligible	Selected	Promotion Rate
Men	784	340	43 %
Women	62	32	52 %

### Is the Difference Between the Promotion Rates in the Example Significant?

The first issue in interpreting raw differences in promotion rates is how to determine whether a particular difference is significant. Ordinarily, the term *significant* means “important,” but, in statistical contexts, significance means that any observed difference (such as the difference in promotion rates in the example) is not likely the result of chance alone. This is not to imply that there is randomness in the rates themselves (as if promotion boards were selecting officers by flipping a coin) but rather that the promotion rate is an estimate of the *probability* of promotion.

Given two promotion rates, the test for statistical significance asks, “Do these two groups have the same probability of promotion?” If these two groups differ significantly from one another, the answer to that question would be, “The two groups likely do not have the same probability of being promoted because it would be very unlikely to observe such a large difference if the two groups had the same promotion probability.” If the difference is insignificant, the answer is, “I cannot rule out the possibility that the promotion probabilities between the two groups are equal because such a difference would often arise even if they were.”

There are two vital considerations to remember about statistical significance when interpreting raw data such as the data seen in the example of promotion rates:

- *Statistical significance is mainly driven by sample size.* Statistical significance is a criterion that rules out differences that are easily attributable to random variation. Larger samples yield more-precise estimates, so the bigger the sample, the less likely a given difference is attributable to chance. This means that very small differences can be statistically significant if the sample size is large enough. In contrast, relatively large differences in rates are normal for small samples, even if the underlying probabilities of promotion (or retention) are similar. This means that apparently large differences can be statistically insignificant if the sample size is small enough. Small-sample concerns are especially relevant when interpreting differences in promotion and retention rates among senior officers because so few members reach these ranks.
- *Statistical significance may not be an accurate indicator of importance.* Whether a difference is important enough to require policy intervention is a matter of judgment and context. Statistical significance does not always equate to policy importance; similarly, statistical insignificance does not always equate to policy irrelevance. An analogous concept is clinical significance, which answers the question, “Is the difference between groups large enough to be worth achieving?” The results can be statistically significant yet clinically insignificant.

Going back to the example in Table 1, the difference in male and female rates is not statistically significant. Nor does it seem large from a policy perspective: The 9-percentage-point difference translates into only about five additional

Female colonels than would be the case if male and female rates were the same (27/62 is approximately equal to 43 percent, the male promotion rate), not a large enough gap to be a cause for policy concern.

In general, relatively large gaps (those that are statistically significant) should get the most attention.

### Are the Two Groups Comparable?

The second issue to consider when interpreting raw differences in promotion or retention rates is whether members of the two groups of interest differ in additional ways that also affect promotion or retention outcomes. If they do, then comparisons of the raw rates may confound the effects of the first characteristic with those of the additional factors.

Thus, the important policy question is not whether members of one group are more or less likely to be promoted or retained than members of the other group but whether similarly situated members of each group are equally likely to advance or remain in service. In other words, people with the same qualifications and background should have the same probability of advancement in a “fair” system.

To see this more clearly, return to the example from Table 1. We said that the observed male-female difference in promotion is neither important nor statistically significant. But were the men and women in these data similarly situated in terms of promotion? It turns out that, if we dig a bit deeper, we find a difference in the occupational distribution of male and female Air Force lieutenant colonels at the time of the 2009 promotion board. Notably, male lieutenant colonels were concentrated in the pilot and navigator occupations whereas female lieutenant colonels were concentrated in support occupations. It is possible that promotion rates differ across these occupations, regardless of gender. If that is the case, the fact that men and women were also employed in these occupations at different rates clouds the interpretation of the gender-specific promotion rates because they do not account for the effect of occupation. In other words, the raw promotion rates may be confounding the effects of gender and occupation.

The reality is that the raw gender-specific promotion rates give very little detail about how fair the system is to different demographic groups. Rather, they can be used to identify areas that require further investigation.

### A More-Informative Approach

Suppose the promotion rates were calculated separately for each occupation—that is, that only within-occupation gender differences in promotion rates were examined. This comparison would be an improvement over the simple gender comparison shown in Table 1, but it would still be subject to the same criticism about confounding factors because occupation may be only one of many factors that differentiate men from women. For example, among pilots, men and women may tend to fly different airframes. All such characteristics must be accounted for before drawing conclusions about the

fairness of the system or the appropriateness of certain policies.

Even if one could establish the characteristics on which men and women differ in this example, there would still be problems. One problem is that splitting the sample by considering every characteristic available would eventually make the sample too small for meaningful results. Another problem is that, even if the sample sizes remain large enough as more and more characteristics divide the sample, it may be difficult to assess which characteristics are the most important in explaining the probability of promotion. For example, it could be that both occupation and gender affect the results but that one effect is greater than the other.

The most common way to work around these problems is to assume that the probability of promotion (or retention) depends on *all* the relevant characteristics in addition to race, ethnicity, or gender. Then, researchers typically estimate the separate effect of each variable on the estimated probability of promotion, holding constant the effect of the other characteristics. This technique is known as *regression analysis*. When researchers claim to have “controlled for” a characteristic, they typically mean that that the characteristic was included in the regression model they used to conduct the analysis.

If gender has a significant effect (both statistically and in terms of magnitude) on the probability of promotion in a model that has controlled for a number of other characteristics, then the researcher has evidence that members of one gender are more likely to be promoted than similarly competitive members of the other gender. If no significant gender effect remains after controlling for other characteristics, then the researcher has evidence that members of one gender are no more likely to be promoted than similarly situated members of the other gender (at least in a way that is distinguishable from mere chance).

Furthermore, this methodology allows the researcher to determine which characteristic (or set of characteristics) helps account for the differences in promotion rates, which, in turn, has implications for policy. For example, if the regression results show that occupation has a large, significant effect on the likelihood of promotion and one gender is concentrated in the lower-promotion occupations, then the Services could improve the promotion prospects of that gender by formulating policy that alters the gender composition of occupations.

Thus, more-thorough analyses of promotion and retention outcomes, including regression analysis, can lead to better assessments of the fairness of the personnel system and to more-appropriate policy action. Such analyses are time-consuming to conduct, however, and they require very detailed data. Indeed, regression models are only as good as the data on which they are based. Regression results based on incomplete or inaccurate data can be even more misleading than comparisons of raw rates because they *appear* to be more reliable. It is because of time and data limitations that the MLDC IPs present raw retention and promotion rates only.

A separate IP will, however, summarize the results from more-detailed studies of career progression for both officers and enlisted members.

## Conclusions

Motivated by the simple promotion example from Table 1, this IP demonstrates several important points about how the commissioners should interpret comparisons of promotion and retention rates along a single observable characteristic:

- It is rarely possible to formulate an appropriate policy based simply on raw demographic-specific promotion or retention rates because there may be underlying differences in characteristics other than race, ethnicity, or gender that could produce the differential rates. Instead, comparisons of raw rates highlight areas for further investigation. Specifically, analytical techniques are available to assess the relative impact of race, ethnicity, or gender in addition to the other characteristics that play a role in the promotion and retention processes.
- Demographic differences in promotion or retention rates that are both statistically significant and large in magnitude should receive the most-immediate attention.
- It is important to be mindful of sample size: Large samples can yield significant differences that are small in magnitude, and small samples can produce large differences by chance even if the probabilities are equal. Ultimately, determining whether a gap is large enough to be policy relevant is a matter of judgment.

## Notes

<sup>1</sup>The Air Force classifies officers who are eligible for promotion into “promotion zones” based on time in grade. Most promotions occur “in the primary zone”, while a small percentage of extremely competitive officers are promoted early—“below the primary zone.” Officers who were not promoted in the primary zone still have an opportunity to be promoted late, when they are “above the primary zone.”

## References

Freedman, D., Pisani, R., Purves, R., & Adhikari, A. (1991). *Statistics*, 2nd ed. New York: WW Norton & Co.

Duncan Hunter National Defense Authorization Act for Fiscal Year 2009, Pub L. 110-417, 122 Stat. 4477, October 14, 2008, codified at 10 U.S. Code §596 (2008).