IDENTIFYING INTERNAL MOBILITY PATTERN DIFFERENCES WITHIN THE CANADIAN ARMED FORCES

Michel G. GAUTHIER

Defence Research and Development Canada, Ottawa, Canada

During their career, members of the Canadian Armed Forces (CAF) are frequently assigned to a new position and can even change profession. This high, but normal, level of mobility within the CAF can potentially hide indicators of mobility pattern differences among women and men. This article presents a method to study the mobility of the CAF members over the last decade. We propose a series of statistical tests to detect trends and identify a list of conditions for which the mobility indicators differ from one subpopulation to another. Our analysis uses Fisher's exact test to compare various populations for which sizes can vary from very small (less than 50) to large (tens of thousands). Since our approach allows us to do tests of statistical significance on small samples, we were able to perform a detailed breakdown of the CAF population (down to the occupational level).

Key words: personnel management and training, career analysis, workforce mobility, categorical data, statistical testing, Fisher's exact test.

1. INTRODUCTION

Throughout their careers, members of military workforces around the world are frequently changing position or rank. In addition to that, some military personnel also occasionally change profession during their career. The Canadian Armed Forces (CAF), with its 68,000 Regular Force (RegF) members, is a good example of this high level of mobility among the military workforce. Such high mobility is intrinsic and essential to the military structure in order to maintain its readiness at all times, while also accommodating members' wishes to the extent possible. However, high frequency changes within such a large organization can also prevent managers from detecting potential issues such as mobility pattern differences among subpopulations that can be defined based on gender, age or ethnicity, for example.

The aim of this paper is to present a method to analyze military occupations at the individual level in order to identify differences in workforce mobility patterns between men and women. We will illustrate our method with examples showing the career trajectory differences among women and men from the CAF RegF; however, the focus of this article is the analytical scheme, which could apply to any other dichotomous characterization of a population.

The approach we propose is based on a statistical comparison of a series of mobility indicators between two subpopulations that can be very small (approx. 10) to very large (approx. 10^5). This characteristic makes our technique extremely relevant to the study of military environments since some subpopulations (e.g. specialized occupations) can be very small and thus challenging to statistically compare with other groups.

2. METHODOLOGY

This section presents the steps we followed to obtain our results, from the data collection process to the statistical analysis.

2.1. Data

The data source used to conduct this study is derived from the Human Resources Management System (HRMS) of the CAF. It consists of end of fiscal year (FY) snapshots with demographic and occupational information about all CAF Regular Force (RegF) members. For this paper, we used the HRMS employment records of the last 10 FYs, i.e. for FY 2005-2006 to 2014-2015. More precisely, we extracted the information regarding the occupation and the gender of all RegF members. For each individual, we identified a series of yearly mobility events. A mobility event can either be:

- staying in the same occupation;
- changed occupation;
- or releasing from the RegF the following year.

At this point, we should point out that a few occupations were grouped together during the pre-processing of the data because they are feeder-receptor groups. Transferring occupation within such groups does not represent a fundamental occupation change and we did not want to capture these "false" occupational transfers during the analysis.

We broke down the dataset in a series of 2x2 contingency tables like the one presented in **Figure 1**.

In this study, the first key used

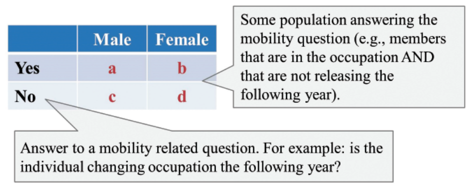

Fig. no. 1. Example of contingency table used for this study. The first dichotomous key (columns) was always based on gender defined as male or female. The second sorting key was always based on a yes/no mobility question such the one presented.

to classify mobility events was always the gender defined as male or female (other or unknown were not considered). The second sorting argument was always the yes/no answer to a precise mobility question (e.g., Was the member released from the RegF the next year?). The actual content of the contingency table presents the aggregate of all the mobility answers collected over the ten-year period.

2.2. List of populations and mobility questions studied

We use 3 case studies to illustrate the utility of our analysis technique. Each case study is defined by a subpopulation of the CAF and a mobility question to be answered (presented in Table 1). Each of these case studies led to many separate analyses since the population was broken down into 83 different occupations (or occupation groups). Our approach thus led to the analysis of hundreds of contingency cases. This illustrates how one can detect statistically significant mobility pattern differences at that small population level. We used our technique to identify occupations where there is difference in mobility patterns between men and women.

2.3. Statistical analysis

Each of the contingency cases discussed in **Table 1** gives a direct measurement of the mobility difference between two (male and female) subpopulations. For example, one can report that men from a given occupation are leaving their occupation in a greater proportion than women from the same occupation, but is this information significant? That depends on several factors such as the size of both populations, the proportions of the positive answers and the level of confidence one is trying to reach.

Table 1. Populations and mobility questions studied.

	Population	Mobility Question
Case Study 1	Trained members that are still in the RegF the following year.	Did the members leave their occupation for another one?
Case Study 2	All members.	Did the members leave the RegF the following year?
Case Study 3	Untrained members that are still in the RegF the following year.	Did the members leave their occupation for another one?

The approach we propose is to use Fisher's exact test for statistical significance [1,2] in order to accept or reject the following null hypothesis (based on 95% confidence intervals):

Null hypothesis: *there is no significant difference between the proportions of men and women answering yes to the mobility question.*

Fisher's test is an exact inferential method that can be used to determine the statistical significance of the differences between populations based on two classification criteria [3]. Such a contingency test can be performed using other approaches (e.g., chi-squared testing) but these methods usually imply approximations that make them valid only for large samples. Fisher's test, on the other hand, calculates the exact probability of observing a given difference between proportions, regardless of the population sizes.

The disadvantage of the Fisher's test is, given its exact nature, that the computational cost increases rapidly with the sample size (due to the usage of factorials). Despite this limitation, we were able to analyze our data in all of the considered scenarios within seconds using built-in routines in R [4]. Consequently, all results from all study cases were obtained using the same calculation methods. Therefore, decisions to reject or accept the null hypothesis are consistent for all case studies presented in this paper.

2.4. Example of detailed results

Table 2 presents an example of the data and results used for Case Study 1 (see Table 1). The first column presents the occupation. The actual occupations are not provided in this paper for confidentiality reasons [5]. The next three columns show the contingency table values for female members as well as the transfer rate (percentage of individuals that are leaving the occupation). The next three columns present the same information for males.

The last columns summarize Fisher's test results:

• The *p*-value represents the probability for the null hypothesis to be true for a given occupation. For example, we estimate to 69.1% the chances that men and women are equally likely to transfer out of Occ. A and that the observed difference between the two transfer rates is simply due to randomness.

Table 2. Example of the data and results obtained for the Fisher's test of Case Study 1 (see Table 1) based on 95% confidence intervals. The letters at the bottom refer to Fig. 1 to indicate the contingency table values.

	Female			Male			Fisher Test			
Occupation	**Transferred**	**Not Transferred**	**Transfer Rate (%)**	**Transferred**	**Not Transferred**	**Transfer Rate (%)**	**p-value**	**Ratio**	**Statistically Different?**	**More Likely to Leave**
Occ. A	16	1054	1.5	75	4257	1.7	0.691	0.9	NO	-
Occ. B	13	827	1.5	165	11745	1.4	0.648	1.1	NO	-
Occ. C	24	781	3.0	112	6199	1.8	0.027	1.7	YES	F
Occ. D	57	923	5.8	223	4815	4.4	0.068	1.3	NO	-
...	...	...	...	...	...	...	...	...	...	...
	b	d		a	c					

• The ratio column shows the proportion of women leaving the occupation divided by the proportion of men also leaving the occupation. In terms of the contingency table values, r*atio = (b/(b+d))/(a/(a+c))*. A ratio value greater than 1 indicates that women are more likely to leave the occupation than men.

• The second last column is simply an interpretation of the p-value. The null hypothesis is rejected with a 95% confidence limit for occupations where the p-value is smaller than 0.05. For example, the null hypothesis is rejected for Occ. C since p-value = 0.027 < 0.05.

• The last column interprets the ratio value in the cases where the proportions are statistically significantly different. It indicates which gender (M or F) has the greatest tendency to leave.

2.5. Limitations

An intrinsic limitation of our results is strictly related to the underlying structure of our raw data and has nothing to do with the methodology itself. As mentioned, our dataset lists the occupations of the members at the end of each FY. Consequently, we are limited to a one-year resolution of the professional activities. Occupational transfers that are occurring at a higher frequency are therefore neglected. However, since it is highly unlikely for RegF members to change occupations more than once a year, this data limitation should have a negligible impact on our results.

3. RESULTS

Tables 3 to 5 present the Fisher's test results for the Case Studies 1 to 3 respectively (see Table 1). For clarity, we only show the p-value and the gender that is more likely to leave [6]. Also, results are sorted by increasing *p*-values, and only occupations where the null hypothesis was rejected (*p*-value < 0.05) are shown.

Case Study 1 - Table 3 presents the list of occupations for which a statistically significant difference was observed between occupation transfer rates for trained RegF male and female members. Our method allowed us to study large (see Occ. A) and very small populations (see transfers for Occ. H). Out of the 83 occupation groups studied, our analysis shows that 12 of them present significantly different transfer rates for men vs. women. Out of these 12 cases, 10 of them indicated that women were more likely to leave the occupation. It is worth noting that for almost all occupations (11 out of 12), the gender with the smallest population is the one that is more likely to transfer. Occ. I. is the only exception where the members that are more likely to transfer from their occupation are also the one with the largest population.

Table 3. Fisher's test results for the analysis of Case Study 1 (Table 1): Are members that are still in the RegF the following year leaving their occupation?

	Female			Male			Fisher Test	
Occupation	Transferred	Not Transferred	Transfer Rate (%)	Transferred	Not Transferred	Transfer Rate (%)	p-value	More Likely to Leave
Occ. A	102	16658	0.0	129	10677	0.0	0.000	M
Occ. B	14	393	0.0	192	18829	0.0	0.000	F
Occ. C	13	592	0.0	46	6804	0.0	0.001	F
Occ. D	6	265	0.0	20	3906	0.0	0.005	F
Occ. E	10	281	0.0	186	13290	0.0	0.009	F
Occ. F	8	225	0.0	46	3729	0.0	0.012	F
Occ. G	12	188	0.1	1434	51475	0.0	0.013	F
Occ. H	3	247	0.0	1	1338	0.0	0.014	F
Occ. I	16	3134	0.0	383	41313	0.0	0.018	M
Occ. J	10	663	0.0	22	3755	0.0	0.021	F
Occ. K	14	1788	0.0	55	13236	0.0	0.040	F
Occ. L	11	382	0.0	142	9847	0.0	0.049	F

Case Study 2 - Table 4 shows a different situation. This table presents the occupations from which the release rate from the RegF differs between men and women. In this case, members leaving one occupation for another are counted as "Not Released" and are summed with the members that are staying in their occupation. We found 17 occupations with a significant difference. For 10 of these cases, men were more likely to be released than women. In order to interpret these results, one might want to look at factors such as the age distribution of males and females in each of these occupations, as this could help to explain the reason for the difference in release rates. This goes beyond the scope of this article; but it is mentioned here to emphasize the fact that our method only identifies cases where the null hypothesis is rejected. A deeper understanding of the situation, and its underlying causes, will require further investigation.

Table 4. Fisher's test results for the analysis of Case Study 2 (Table 1): Are members leaving the RegF?

	Female			Male			Fisher Test	
Occupation	**Transferred**	**Not Transferred**	**Transfer Rate (%)**	**Transferred**	**Not Transferred**	**Transfer Rate (%)**	**p-value**	**More Likely to Leave**
Occ. A	1092	17855	5.8	983	11282	8.0	0.000	**M**
Occ. B	321	6278	4.9	789	11093	6.6	0.000	**M**
Occ. C	243	2402	9.2	2028	28610	6.6	0.000	**F**
Occ. D	49	256	16.1	5828	59280	9.0	0.000	**F**
Occ. E	26	815	3.1	421	6313	6.3	0.000	**M**
Occ. F	306	3778	7.5	3115	48442	6.0	0.000	**F**
Occ. G	32	954	3.2	1053	17916	5.6	0.001	**M**
Occ. H	59	168	26.0	132	649	16.9	0.003	**F**
Occ. I	51	1313	3.7	376	6461	5.5	0.007	**M**
Occ. J	3	8	27.3	62	1241	4.8	0.014	**F**
Occ. K	39	647	5.7	153	1649	8.5	0.019	**M**
Occ. L	20	381	5.0	175	1897	8.4	0.019	**M**
Occ. M	11	316	3.4	104	1430	6.8	0.022	**M**
Occ. N	34	697	4.7	50	629	7.4	0.033	**M**
Occ. O	20	159	11.2	262	3644	6.7	0.033	**F**
Occ. P	82	727	10.1	201	2378	7.8	0.041	**F**
Occ. Q	109	1766	5.8	28	282	9.0	0.042	**M**

Case Study 3 - Finally, **Table 5** shows results for the populations of RegF members that are leaving their occupation while still undergoing training [7]. In this case, 13 occupational groups presented a statistical difference between men and women. Out of these, 11 occupations showed a greater transfer rate for women. Also, the gender that is more likely to transfer from their occupation is systematically the one with the smallest population (as we observed for most occupations in Case Study 1).

Table 5. Fisher's test results for the analysis of Case Study 3 (Table 1): Are untrained members that are still in the RegF the following year leaving their occupation?

	Female			Male			Fisher Test	
Occupation	Transferred	Not Transferred	Transfer Rate (%)	Transferred	Not Transferred	Transfer Rate (%)	p-value	More Likely to Leave
Occ. A	16	36	0.3	380	5106	0.1	0.000	F
Occ. B	22	233	0.1	71	3026	0.0	0.000	F
Occ. C	10	63	0.1	67	2001	0.0	0.000	F
Occ. D	9	68	0.1	44	1707	0.0	0.000	F
Occ. E	5	26	0.2	18	682	0.0	0.002	F
Occ. F	11	1177	0.0	18	638	0.0	0.005	M
Occ. G	16	109	0.1	186	2912	0.1	0.007	F
Occ. H	31	393	0.1	233	4964	0.0	0.012	F
Occ. I	15	86	0.1	196	2378	0.1	0.013	F
Occ. J	6	44	0.1	40	958	0.0	0.019	F
Occ. K	13	75	0.1	75	970	0.1	0.020	F
Occ. L	9	69	0.1	16	356	0.0	0.024	F
Occ. M	11	795	0.0	11	287	0.0	0.026	M

4. CONCLUSIONS AND REMARKS

We proposed an analysis method that can be used to detect workforce mobility pattern differences. One main advantage of our approach is that its validity is not dependent on the population sizes. We demonstrated this feature in our results using small populations, but we could break down our dataset even further (based on rank, for example), and we would still be able to obtain meaningful results for some of the subpopulations.

One key result obtained from the current analysis is that, in all cases studying the propensity-to-leave an occupation, we observed that the underrepresented gender of the occupation is consistently the one that is more likely to leave (when there is a statistically significant difference between the transfer rates).

Our examples in this paper were limited to male/female comparisons leaving an occupation or releasing from the RegF. However, any question that can be addressed using a 2x2 contingency table can be analyzed using our approach. In fact, Fisher's test is not limited to 2x2 tables and our technique could therefore be extended to more complex problems (or mobility questions) that can be summarized using larger contingency tables.

We would like to emphasize the fact that finding a statistically significant difference between two groups does not necessarily mean that there is an actual bias explaining the results. Such results simply indicate that there is a difference between the observed proportions that cannot solely be explained by randomness. These results can potentially be explained by a number of factors, including demographics (e.g. age and years of service profiles of the members). If one is interested in identifying gender-biased situations for example, we suggest using our method as a first mining tool to investigate a large datasets in order to identify areas that warrant further investigation. At this point, more thorough analyses must be carried to establish the underlying reason(s) explaining the observed differences.

ENDNOTES AND REFERENCES

[1] Fisher, R.A. (1954). Statistical Methods for Research Workers. Oliver and Boyd publishing.

[2] Agresti, A., A Survey of Exact Inference for Contingency Tables. In: Statistical Science, Vol. 7, No. 1, 1992.

[3] It is beyond the scope of this article to explain Fisher's test in detail. We simply gave a brief description of its advantages and limitations in the main text. We encourage the interested readers to refer to [1,2] or other textbooks for more details.

[4] Software environment for statistical computing and graphics (www.r-project.org).

[5] The data reported is real, only descriptive name of the occupations were removed to preserve confidentiality. For the same reason, the generic occupation labels (A, B, …) are not consistent from Tables 2 to 5.

[6] The ratios and the yes/no answers presented in Table 2 are straightforward to obtain for the results presented in Tables 3 to 5.

[7] In this context, training refers to early career training. For this study, this includes the training required to become qualified in one's occupation, including subsidized university training where applicable.