## *Methods*

### *Study Participants and Procedures*

Reservists were excluded from the analyses in this paper. Reservists in the Mil-2002 were drawn from all primary reserve forces who had paraded in the six months leading up to the survey. In contrast, the Mil-2013 reservists were drawn from those who had deployed to the Afghanistan mission, because estimating the mental health effects of that mission was an important objective for this most recent military survey. Thus, reservists from the 2002 and 2013 military surveys were not directly comparable.

### *Measures*

**Past year MHSU**. The nurse category was worded as "nurse" in the two civilian surveys, as "nurse including a CF case manager" in Mil-2013 and as "nurse practitioner, physician's assistant, or medic" in the Mil-2002. Nurses are an important provider category in specialty mental health care in CAF (along with psychiatrists, psychologists, and social worker, counselor, or psychotherapists); we therefore decided it was important to capture the nurse category in our tables, despite differences in wording across the surveys. However, we note that any differences across the four surveys in the nurse category must be interpreted with caution.

**Clinical covariates**. The CIDI algorithm for GAD diagnosis was slightly different in the two recent surveys (Mil-2013 and Civ-2012). To have comparable GAD diagnoses, we recalculated the GAD diagnosis for the two recent surveys using the algorithm from the Mil-2002.

### *Statistical Analyses*

In this paper, we used analytic methods that are similar to those used in a previous paper by our group on perceived need for and perceived sufficiency of mental healthcare[1]. Specifically, in both papers, we assessed absolute and relative differences in prevalence rates across different surveys, while controlling for a similar set of socio-demographic and clinical correlates (but looked at a different outcomes). The relevant sections on sociodemographic and clinical covariates and absolute and relative differences in the manuscript and in supplemental online material analytic method sections are partly drawn from our previous paper[1].

***Absolute and Relative Differences***. Absolute and relative differences provide unique information and are complementary in assessing the MHSU prevalence differences across different populations and time periods. Therefore, we assessed both relative and absolute differences to provide a complete picture of differences in MHSU prevalence across different surveys. Absolute differences (ADs) capture the actual impact of interventions by estimating the number of individuals who are (and could potentially be) influenced by interventions. However, ADs are not affected by baseline prevalence and comparing the strength of an intervention across different populations can be difficult using ADs alone. In contrast, relative differences (RDs) take into account baseline prevalence and make it easier to compare differences in the impact of interventions across different populations [2] but can be misleading by themselves.

Considering the magnitude of the AD (or RD) alone can lead to different even opposite conclusions. This can be seen by considering the following two scenarios. In the first scenario where the prevalence of interest is very low in the reference survey, a small AD between two surveys might result in a large RD (e.g., Following an intervention, the prevalence of MHSU increases by 1% from 1% to 2%, yielding an AD of 1% but an RD of 2, i.e., doubling of MHSU prevalence, seemingly a "dramatic" intervention effect if RD is considered alone). In the second scenario, where the prevalence of interest is very high in the reference survey, a large AD might result in a small RD (e.g., Following an intervention, the prevalence of MHSU goes up from 70% to 90%, yielding an AD of 20% but an RD of 1.29). Perhaps even more importantly for our purposes, ADs and RDs can yield seemingly inconsistent results when comparing intervention effects in two or more groups across different time periods (i.e., in interaction assessments). Interaction assessments are sensitive to the type and scale of the interaction being assessed. Absence of interaction on the additive scale (absolute difference) implies the presence of multiplicative interaction for relative difference and likewise, absence of multiplicative interaction for relative difference implies the presence of additive interaction[3-5]. It has been recommended that researchers test and report both additive and multiplicative interactions[6].

**Absolute differences in MHSU**. We explored absolute differences in MHSU using iterative proportional fitting, a post-stratification procedure that creates samples that agree on marginal totals of selected characteristics by adjusting sample weights. Compared to other sample matching approaches for assessing absolute differences for categorical outcomes, such as direct and indirect standardization, iterative proportional fitting has the advantage of including more matching factors. We employed iterative proportional fitting to Mil-2002, Civ-2012, and Civ-2002 separately to adjust their sample weights to calibrate them to have the same marginal totals on selected variables as Mil-2013. The following variables were including in calculating the iterative proportional fitting weights: age, sex, education, marital status, family income, ethnicity/culture original, self-reported mental health status, self-reported physical health, past 12-month major depressive episode, past 12-month suicide ideation and

attempt. We used IHB raking macro for calculating the adjusted weights [7]. To avoid extreme weights that will result in unstable analysis results, we employed the Margin Cap Value method to trim extreme weights[8].

**Relative differences in MHSU**. For each provider type, three binary logistic regression models were constructed. The first model included Mil-2013 and Mil-2002 to assess the temporal change in military. The second model included Civ-2012 and Civ-2002 to assess the temporal change in civilians. The third model included Mil-2013 and Civ-2012 surveys to assess the difference in MHSU between contemporary military and civilian populations. We employed logistic regression standardization and propensity score matching approaches to adjust the survey differences in socio-demographic and clinical variables. These two approaches have an important advantage compared to iterative proportional fitting: they take into account information on the joint distribution of the matching factors, which is ignored in iterative proportional fitting and could be a source of bias in estimating sample differences.

*Logistic regression standardization*. To control for the confounding effects of the selected socio-demographic and clinical variables in PRRs estimation, we constructed binary logistic regression models for MHSU with adjustment for both selected socio-demographic and clinical variables.

*Propensity score matching*. Propensity score matching is an alternative approach to regression standardization. It has the advantage of reducing problems with model convergence and instability. These undesirable analytic issues emerge when many matching factors are included and the number of cases is relatively small. Given a total of 12 matching factors were included in the analyses, it was necessary to have propensity score matching as an additional approach for sample matching to assess if the results obtained from the logistic regression standardization were robust. We calculated each respondent's propensity score for being included in the Mil-2013 from multivariable logistic regression models including all the selected matching variables. We adopted two approaches for propensity score matching: regression/covariate adjustment and stratification. The first approach included the calculated propensity score as a covariate in the model. The second approach first divided the survey sample into five equal strata based on the value of the calculated propensity score, and then included the variable indicating the five strata into the model. Based on existing literature, stratifying respondents into five strata could eliminate more than 90% of the bias.

*Temporal changes in MHSU in military versus the civilians*. In the analyses for assessing absolute differences in MHSU, the interaction between survey type (military versus civilian) and survey time (2012/2013 versus 2002) represents the difference between the two prevalence rate differences (DPRD) - calculated separately in military and civilians - in temporal changes in the

prevalence of MHSU. The DPRD is formulated as: (the MHSU PRD in the military (2013 - 2002)) - (the MHSU PRD in civilians (2012 - 2002)). An interaction of zero (DPRD=0) indicates that the temporal change is the same in military and civilians. A positive interaction (DPRD>0) favors greater temporal change in military than civilians.

In the analyses for assessing the relative differences in the prevalence of MHSU, the interaction represents the ratio of the two PRRs (RPRR) - the MHSU prevalence rate ratios of the contemporary survey to the survey conducted 10 years before, calculated separately in military and in civilians. The RPRR can be formulated as: $PRR_{military}$ (Mil-2013÷Mil-2002) ÷ $PRR_{civilian}$ (Civ-2012÷Civ-2002). A RPRR of 1 indicates that the same temporal change has taken place in both military and civilians. A RPRR of greater than 1 favors greater change in military than civilians.

To account for random error and obtain confidence intervals, we relied on regressions to calculate the absolute and relative differences as well as the temporal changes rather than simple subtractions and divisions.

References

1.      Fikretoglu D, Liu A, Zamorski M, Jetly R. Perceived need for and perceived sufficiency of mental health care in the Canadian Armed Forces: Changes in the past decade and comparisons to the general population. *Can J Psychiatry* 2016; **61 (Supplement 1)**: 36S-45S.

2.      Hirai A. Calculating absolute rate differences and relative rate ratios in SAS/SUDAAN and STATA. http://www.citymatch.org/sites/default/files/documents/MCHEPITraining/Hirai_Risk%20Differences_Risk%20Ratios_5_20.pdf (accessed June 23, 2016.

3.      Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. Philadelphia, PA: Lippincott, Williams, & Wilkins 2008.

4.      VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiologic Methods* 2014; **3**(1): 33-72.

5.      Siemiatycki J, Thomas DC. Biological models and statistical interactions: an example from multistage carcinogenesis. *Int J Epidemiol* 1981; **10**(4): 383-7.

6.      Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. *Int J Epidemiol* 2012; **41**(2): 514-20.

7.        Battaglia MP, Hoaglin DC, Frankel MR. Practical considerations in raking survey data. *Survey Practice*, 2009.

http://www.surveypractice.org/index.php/SurveyPractice/article/view/176/html (accessed August 16, 2016).

8.        Izrael D, Battaglia MP, Frankel MR. Extreme survey weight adjustment as a component of sample balancing (a.k.a.

raking). SAS Global Forum 2009 Conference; Cary, NC: SAS Institute Inc; SAS Global Forum 2009 Conference.

**Supplemental Online Table 1. Raw prevalence of past-year MHSU by provider category among comparable groups from the four surveys[a]**

| Provider Category | Mil-2013 (Regular force) | Mil-2002 (Regular force) | Civ-2012 (Subsample) | Civ-2002 (Subsample) |
|---|---|---|---|---|
| | % (95%CI) | % (95%CI) | % (95%CI) | % (95%CI) |
| **Any professional[b]** | 20.77 (19.73-21.81) | 12.56 (11.64-13.48) | 8.09 (7.24-8.87) | 6.53 (6.02-7.05) |
| **Psychiatrist** | 6.53 (5.89-7.17) | 3.59 (3.02-4.16) | 1.25 (0.91-1.60) | 1.13 (0.92-1.34) |
| **Family doctor or general practitioner** | 9.89 (9.14-10.64) | 6.63 (5.92-7.34) | 4.69 (4.08-5.30) | 4.12 (3.71-4.53) |
| **Psychologist** | 7.90 (7.21-8.58) | 4.04 (3.51-4.57) | 2.09 (1.70-2.47) | 1.83 (1.51-2.15) |
| **Nurse[c]** | 5.69 (5.09-6.29) | 1.68 (1.31-2.05) | 0.55 (0.33-0.77) | 0.19 (0.12-0.26) |
| **Social worker, counsellor or psychotherapist** | 12.72 (11.83-13.60) | 6.85 (6.12-7.58) | 2.60 (2.10-3.10) | 1.68 (1.42-1.95) |

[a] 95% confidence intervals were calculated using 500 bootstrapped weights

[b] Since the nurse provider category was not comparable across the four surveys, we created a second aggregate category for any professional excluding nurse. The results from using the two aggregate variables - including or excluding nurse provider - yielded almost identical results in all of the analyses (results available upon request).

[c] The Nurse category is not fully comparable among the four surveys.

**Supplemental Online Table 2. Adjusted prevalence of past-year MHSU by provider category, calculated using iterative proportional fitting weights[ab]**

| Provider Category | Mil-2013<br>% (95%CI) | Mil-2002<br>% (95%CI) | Civ-2012<br>% (95%CI) | Civ-2002<br>% (95%CI) |
|---|---|---|---|---|
| **Any professional[c]** | 20.77 (19.73-21.81) | 13.61 (12.34-14.89) | 12.05 (10.54-13.57) | 8.57 (7.31-9.83) |
| **Psychiatrist** | 6.53 (5.89-7.17) | 3.94 (3.07-4.81) | 2.73 (1.76-3.71) | 2.13 (1.53-2.73) |
| **Family doctor or general practitioner** | 9.89 (9.14-10.64) | 7.26 (6.29-8.23) | 7.54 (6.20-8.88) | 5.25 (4.29-6.21) |
| **Psychologist** | 7.90 (7.21-8.58) | 4.66 (3.84-5.47) | 3.61 (2.71-4.51) | 2.78 (2.02-3.54) |
| **Nurse[d]** | 5.69 (5.09-6.29) | 1.54 (1.09-2.00) | 1.44 (0.65-2.22) | 0.36 (0.15-0.58) |
| **Social worker, counsellor or psychotherapist** | 12.72 (11.83-13.60) | 6.56 (5.56-7.56) | 4.16 (2.98-5.34) | 2.27 (1.70-2.85) |

[a] The adjusted prevalence for Mil-2002, Civ-2012 and Civ-2002 surveys were calculated using weights to approximate the population of Mil-2013 survey. The following variables were included in calculating iterative proportional fitting weights: age, sex, education, marital status, family income, ethnicity/culture original, self-reported mental health status, self-reported physical health, past 12-month major depressive episode, past 12-month suicide ideation and attempt
[b] 95% confidence intervals were calculated using 500 bootstrapped weights
[c] Since the nurse provider category was not comparable across the four surveys, we created a second aggregate category for any professional excluding nurse. The results from using the two aggregate variables - including or excluding nurse provider - yielded almost identical results in all of the analyses (results available upon request).
[d] The Nurse category is not fully comparable among the four surveys.

**Supplemental Online Table 3. Relative differences in past-year MHSU prevalence – using propensity score matching to adjust for survey differences in sociodemographic and clinical characteristics[ab]**

| Provider Category | With adjustment approach 1 | | | With adjustment approach 2 | | |
|---|---|---|---|---|---|---|
| | Mil-2013 vs. Mil-2002 | Civ-2012 vs. Civ-2002 | Mil-2013 vs. Civ-2012 | Mil-2013 vs. Mil-2002 | Civ-2012 vs. Civ-2002 | Mil-2013 vs. Civ-2012 |
| **Any professional[c]** | 1.62 (1.44-1.82) | 1.14 (0.98-1.36) | 2.08 (1.74-2.49) | 1.62 (1.44-1.82) | 1.15 (0.98-1.34) | 2.33 (1.95-2.77) |
| **Psychiatrist** | 1.60 (1.28-2.00) | 0.95 (0.66-1.36) | 3.06 (1.91-4.91) | 1.60 (1.28-2.01) | 0.94 (0.65-1.36) | 3.31 (2.15-5.09) |
| **Family doctor or general practitioner** | 1.32 (1.13-1.55) | 1.11 (0.91-1.35) | 1.45 (1.14-1.86) | 1.33 (1.14-1.56) | 1.11 (0.92-1.35) | 1.65 (1.32-2.05) |
| **Psychologist** | 1.66 (1.38-1.99) | 0.99 (0.75-1.31) | 2.59 (1.94-3.46) | 1.63 (1.36-1.96) | 0.98 (0.74-1.29) | 2.95 (2.26-3.84) |
| **Nurse[d]** | 3.33 (2.53-4.39) | 2.99 (1.62-5.51) | 5.94 (3.00-11.78) | 3.20 (2.43-4.21) | 2.97 (1.58-5.60) | 5.35 (2.74-10.44) |
| **Social worker, counsellor or psychotherapist** | 1.94 (1.67-2.25) | 1.53 (1.17-2.00) | 3.61 (2.67-4.90) | 1.91 (1.65-2.22) | 1.52 (1.17-1.99) | 3.90 (2.87-5.31) |

[a.] The relative difference was denoted as MHSU prevalence rate ratio (PRR) of one survey to another. Approach 1 included the calculated propensity score as a covariate in the model. Approach 2 first divided the survey sample into five equal strata based on the value of the calculated propensity score, and then included the variable indicating the five strata into the model.

[b.] 95% confidence intervals were calculated using 500 bootstrapped weights

[c.] Since the nurse provider category was not comparable across the four surveys, we created a second aggregate category for any professional excluding nurse. The results from using the two aggregate variables - including or excluding nurse provider - yielded almost identical results in all of the analyses (results available upon request).

[d.] The Nurse category is not fully comparable among the four surveys

**Supplemental Online Table 4. Relative differences between Mil-2013 and Mil-2002 surveys in past-year MHSU prevalence – using logistic regression standardization to adjust for survey differences in the sociodemographic and clinical characteristics[a][b]**

| Provider Category | Original results[c] | Results with additional adjustment variables[d] |
|---|---|---|
| Any professional[e] | 1.84 (1.56-2.16) | 1.82 (1.58-2.10) |
| Psychiatrist | 1.71 (1.17-2.52) | 1.61 (1.23-2.11) |
| Family doctor or general practitioner | 1.56 (1.27-1.93) | 1.28 (1.06-1.54) |
| Psychologist | 1.96 (1.55-2.49) | 1.69 (1.36-2.10) |
| Nurse[f] | 5.39 (3.19-9.09) | 3.81 (2.77-5.25) |
| Social worker, counsellor or psychotherapist | 2.04 (1.65-2.52) | 2.02 (1.71-2.39) |

[a].  The relative difference was denoted as MHSU prevalence rate ratio (PRR) of one survey to another
[b].  95% confidence intervals were calculated using 500 bootstrapped weights
[c].  Calculated from the multiple logistic regression with the adjustment for socio-demographic variables including age, sex, education, marital status, ethnicity/cultural origin, and household income and clinical variables including self-reported physical and mental health status, past 12 months major depressive episode diagnosis, past 12 months suicide ideation and attempt
[d].  In addition to the variables listed above, the multiple logistic regression models adjusted also for past 12-month panic disorder diagnosis, past 12-month generalized anxiety disorder diagnosis, and past 12-month post-traumatic distress disorder diagnosis.
[e].  Since the nurse provider category was not comparable across the four surveys, we created a second aggregate category for any professional excluding nurse. The results from using the two aggregate variables - including or excluding nurse provider - were almost identical (results available upon request).
[f].  The Nurse category is not fully comparable among the four surveys.

**Supplemental Online Table 5. Relative differences between Civ-2012 and Civ-2002 surveys in past-year MHSU prevalence – using logistic regression standardization to adjust for survey differences in sociodemographic and clinical characteristics[a][b]**

| Provider Category | Original results[c] | Results with additional adjustment variables[d] |
|---|---|---|
| **Any professional[e]** | 1.15 (0.96-1.38) | 1.13 (0.94-1.36) |
| **Psychiatrist** | 0.92 (0.62-1.36) | 0.87 (0.58-1.30) |
| **Family doctor or general practitioner** | 1.10 (0.88-1.37) | 1.06 (0.85-1.33) |
| **Psychologist** | 0.98 (0.72-1.32) | 0.96 (0.71-1.30) |
| **Nurse[f]** | **3.01 (1.59-5.69)** | **2.82 (1.43-5.56)** |
| **Social worker, counsellor or psychotherapist** | **1.55 (1.17-2.06)** | **1.52 (1.14-2.02)** |

[a]  The relative difference was denoted as MHSU prevalence rate ratio (PRR) of one survey to another
[b]  95% confidence intervals were calculated using 500 bootstrapped weights
[c]  Calculated from the multiple logistic regression with the adjustment for socio-demographic variables including age, sex, education, marital status, ethnicity/cultural origin, and household income and clinical variables including self-reported physical and mental health status, past 12 months major depressive episode diagnosis, past 12 months suicide ideation and attempt
[d]  In addition to the variables listed above, the multiple logistic regression models adjusted also for past 12-month mania diagnosis.
[e]  Since the nurse provider category was not comparable across the four surveys, we created a second aggregate category for any professional excluding nurse. The results from using the two aggregate variables - including or excluding nurse provider - were almost identical (results available upon request).
[f]  The Nurse category is not fully comparable among the four surveys.

**Supplemental Online Table 6. Relative differences between Mil-2013 and Civ-2012 surveys in past-year MHSU prevalence – using logistic regression standardization to adjust for survey differences in sociodemographic and clinical characteristics[a][b]**

| Provider Category | Original results[c] | Results with additional adjustment variables[d] |
|---|---|---|
| **Any professional[e]** | **2.45 (1.97-3.05)** | **2.49 (1.99-3.13)** |
| **Psychiatrist** | **3.68 (2.24-6.05)** | **4.09 (2.35-7.09)** |
| **Family doctor or general practitioner** | **1.51 (1.17-1.95)** | **1.50 (1.14-1.96)** |
| **Psychologist** | **2.92 (2.14-4.00)** | **2.94 (2.13-4.06)** |
| **Nurse[f]** | **7.39 (3.60-15.19)** | **7.98 (3.45-18.47)** |
| **Social worker, counsellor or psychotherapist** | **3.95 (2.77-5.63)** | **4.11 (2.84-5.96)** |

[a.] The relative difference was denoted as MHSU prevalence rate ratio (PRR) of one survey to another

[b.] 95% confidence intervals were calculated using 500 bootstrapped weights

[c.] Calculated from the multiple logistic regression with the adjustment for socio-demographic variables including age, sex, education, marital status, ethnicity/cultural origin, and household income and clinical variables including self-reported physical and mental health status, past 12 months major depressive episode diagnosis, past 12 months suicide ideation and attempt

[d.] In addition to the variables listed above, the multiple logistic regression models adjusted also for past 12-month generalized anxiety disorder diagnosis and past 12-month alcohol abuse and dependence diagnosis.

[e.] Since the nurse provider category was not comparable across the four surveys, we created a second aggregate category for any professional excluding nurse. The results from using the two aggregate variables - including or excluding nurse provider - were almost identical (results available upon request).

[f.] The Nurse category is not fully comparable among the four surveys.