



---

**The YMCA's Contribution to Well-Being in the GTA  
Technical Report**

**Submitted to Michael Hall, YMCA**

**September 2018**

**Datassist, Data Science for the Non-profit World  
Heather Krause, PStat., Principal Scientist  
heather@idatassist.com**

---

## **Acknowledgments**

We are deeply indebted to YMCA team, particularly to Michael Hall, Heather Alinsangan, Gobi Sriranganathan, and Cheyne Singh. We also have benefitted tremendously from the methodological support of the York University Statistical Consulting Services, particularly Dr. Georges Monette.

## **Introduction**

We aimed to evaluate the health effects of the presence of YMCA locations in neighbourhoods as well as the health effect of YMCA membership. We used data from a randomly collected public survey, a convenience sample YMCA members survey, and administrative Statistics Canada data of neighborhood characteristics. In order to operationalize the concept of health effects, we used four self-reported measures of health. These were Well-being, Mental Health, General Health, and Sense of Belonging. These were all self-rated and the outcomes were recoded to create binary variables. The cut-off points for the sorting of responses into good health and poor health for each outcome were provided by the YMCA research staff.

Two separate sets of models were constructed. One set of models explored the neighborhood level effects of the presence of YMCA locations and the second set explored the individual level effects of membership.

The individual level models were a series of generalized linear models including propensity scores. For the neighborhood level models, a mixed-effects spatial analysis was conducted based on YMCA location geographic data.

## **Data**

For this analysis, we combined four datasets that provide data about the Greater Toronto area. These were: a public telephone survey, a YMCA members survey, administrative data from Statistics Canada 2016 Census, and the geographic location data on YMCA locations.

The public telephone survey was conducted using stratified sampling and was conducted by an independent research firm. The respondents included 8,270 randomly selected residents of the GTA over the age of 15. The members survey was conducted online by the YMCA and included a duplicate subset of relevant survey questions. The members survey included 12,332 responses from YMCA members over the age of 15. The administrative data comes from Statistics Canada 2016 Census. The YMCA provided the dataset containing the geographic locations of the 440 YMCA locations across the Greater Toronto Area.

The measures used in the models include the following.

### *Individual characteristics*

- Gender measured as a self-report with three options.
- Age as a continuous integer.
- Employment Status as a seven-level variable including: Self-employed, Employed full-time, Employed part-time, Currently unemployed, Retired, Homemaker, Student.
- Marital Status as a five-level variable including: Single never married, Married or common law, Separated, Divorced, Widowed.
- Immigrant Status as binary variable of Born in Canada or Not born in Canada.

- Educational Attainment as a four-level variable of Less than high school, High school graduate, Some post-secondary education, Completed post-secondary degree.
- Sexual Orientation as a seven-level variable including Bisexual, Heterosexual, Gay, Lesbian, Queer, Two-Spirit, Other.
- Ethnic Self-Identity in the survey included over 35 responses. These responses were recoded into ten aggregate categories based on the current Statistics Canada recoding recommendations. These ten categories are Black, East Asian, Eastern European, Indigenous, Latin American, North American (not Indigenous), South Asian, West Asian, Western European, and Other.

### *Neighborhood characteristics*

In these analyses, we employ neighborhood classifications that were developed by the YMCA research team for the Greater Toronto-Wellesley Institute GTA Well-Being Monitor.<sup>1</sup> To the extent possible, these neighborhood classifications correspond to the neighbourhood definitions previously defined by the regional and municipal level governments in each appropriate locale.

The average post-tax household income from the 2016 Canadian Census was used to measure income for neighborhoods.

The count of YMCA locations for each neighborhood was developed using YMCA geographic data on YMCA locations.

### *Outcome variables*

Well-being, or Life Satisfaction, was measured by asking respondents to rate their life satisfaction using an eleven-point scale (0 to 10), where zero represented “very dissatisfied” and ten represented “very satisfied.” Good Life Satisfaction was defined as a score of 8 or more.

General Health was measured with the question “In general, would you say your health is excellent, very good, good, fair, or poor?” Good General Health category included participants who reported that their health is “excellent” or “very good” rather than “good”, “fair,” or “poor.”

Mental Health was measured with the question “In general, would you say that your mental health is: excellent, very good, good, fair or poor.” Good Mental Health category included participants who reported that their health is “excellent” or “very good” rather than “good”, “fair,” or “poor.”

Sense of Belonging was measured with the question “How would you describe your sense of belonging to your local community?” Good health in Sense of Belonging category included participants who reported that their sense of belonging to their local community was “Somewhat Strong” or “Very Strong” as distinct from “Somewhat Weak” or “Very Weak.”

---

<sup>1</sup> See Hall et al. (2018). Life in the GTA: A window on well-being. Toronto: YMCA for a description of the GTA Well-Being Monitor and the neighbourhood classifications.

## Methodology

### *Neighborhood level models*

Mixed-effects models were used to calculate neighborhood effects. These models are designed to control for the non-independence of spatial data. These types of models are designed to accurately estimate effects using data that is clustered into neighborhoods. This location data of YMCA locations was used to calculate a raster-based mean center and event counting algorithm. This was then used to mean center and calculate standard distances between each neighborhood and the nearest YMCA location.

These results were then used in mixed-effects models to estimate the effect of having a YMCA within a neighborhood. The models built for this report used spatial autocorrelation as the random effects and the YMCA location counts and neighborhood income as the independent fixed effects. Income was included as an independent covariate to control for the possibility that YMCA locations are placed in specific neighbourhoods at least partially based on income. Including income in the model controls for some of this influence and allows us to more clearly estimate the contribution effect of YMCA locations in neighbourhoods.

Including income as an independent fixed-effect variable also us to compare apples to apples among neighbourhoods as well as to control for the likelihood that characteristics of individuals within neighbourhoods significantly influence the health outcomes. This fixed effect provides a mechanism for disaggregating the effects of the presence from a YMCA location from the effects of neighborhood socioeconomic characteristics. It would theoretically be possible to control for these differences by including a wide selection of sociodemographic indicators such as neighborhood unemployment, neighborhood immigration proportion, etc. However, previous research has shown that income captures much of the same variation as a linear combination of these variables and using income as a single independent predictor rather than multiple alternative predictors provides the advantage of retaining much more statistical power. Two-way interaction effects were tested but found to be insignificant.

Researchers often want to know if there is a way to get an analogue of R-squared goodness-of-fit metric for mixed models This is a challenging question because it turns out that while R-squared gets mixed reviews for assessing traditional regression, it is even more problematic for mixed-effect models. The issues include how to handle sampling variation and how to incorporate the random variation. For the purposes of this report, we implement a Bayesian version of R-squared as developed by Gelman and Pardoe (2006)<sup>2</sup>. However, all R-squared results should be used with care.

The neighborhood level models were built using the R Software for Statistical Programming version 3.5.1 and the packages nlme version 3.1-137, spida version 2.0.1, emmeans version 1.2.4. Estimates from these models were built using model matrices and linear combinations of estimated marginal means. Missing values were handled through listwise deletion.

---

<sup>2</sup> Gelman, A., & Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*, 48(2), 241-251.

### *Individual level models*

The individual level models were a series of generalized linear models including propensity scores. Confounding factors within the individual level models include age, gender, employment status, educational attainment, marital status, immigration status, sexual orientation, and ethnic self-identity. Please note that we did not include individual or household income as a confounding factor in the individual level models. The income variable is considered to be a mediator not a confounder and as such including it in the development of propensity scores would distort the results rather than correct them. Propensity scores for each individual were calculated from logistic models including those covariates. Based on these scores, YMCA members were matched to non-YMCA members. Then, we estimated the effects of YMCA membership using generalized linear models.

Propensity score matching models are quasi-experimental designs that allow researchers to address questions of causal relationship in the absence of an experimental control group. These methods have been used in previous peer-reviewed published research exploring health<sup>3</sup> and life satisfaction<sup>4</sup>.

Establishing a counterfactual in this setting is not practical for ethical and practical considerations. There is no mechanism for requiring people in the GTA to either be on not be YMCA members due to random assignment. To assess the impact of YMCA membership on health, then, requires the use of a quasi-experiment. Traditional regression analysis is not suitable in this situation since significant bias will exist that cannot be accounted for. The approach this research takes to deal with this bias is to match individuals from the treatment group with individuals from a constructed control group to obtain more accurate estimates of the true effect of the treatment (YMCA membership in this case).

The matching method used in this report uses the nearest-neighbor one-to-one matching estimator which finds the nearest neighbour from the control group for each of the vectors. The distance method used was a logit link. The matching was conducted using R Program for Statistical Computing version 3.5.1 and the package MatchIt version 3.0.2. Listwise deletion was used to handle missing data.

### *Limitations*

Design features of the research that should be considered when interpreting the findings primarily include potential bias in the sample, generalizability of the results based on the nature of survey research, and potential instability of results due to single point in time data collection.

The use of propensity scores to reduce bias in observational data is a good solution to the lack of experimental data. However, this method cannot fully remove bias that is outside the ignorable treatment assignment assumption. This means that if there are unobserved factors which are giving rise to differences in the two samples, these are not corrected for by the matching method.

---

<sup>3</sup> Suldo, S. M., Savage, J. A., & Mercer, S. H. (2014). Increasing middle school students' life satisfaction: Efficacy of a positive psychology group intervention. *Journal of happiness studies*, 15(1), 19-42.

<sup>4</sup> Binder, M., & Coad, A. (2013). Life satisfaction and self-employment: A matching approach. *Small Business Economics*, 40(4), 1009-1033.

The study is limited in terms of ecological generalizability to an entire population. Specifically, the intervention may only be generalizable to GTA residents and YMCA members who are available and willing to complete this type of survey. The public survey was a probability-based stratified random sample. The YMCA members survey was a non-probability online survey that was distributed to YMCA members through personalized invitations and included up to three reminders in an effort to reduce non-response bias.

The next steps for future research would be to collect longitudinal data that would allow for more robust estimates of causal impact. As well, a randomized sample for the YMCA members survey would increase the research generalizability.

## Results

### *Individual Level Models*

The individual level models rely on the survey data that includes both YMCA members and non-members. This data is not balanced across key confounders at baseline. The correction of this unbalance is the purpose in employing the propensity score matching technique. Table 1 shows the distributions of the baseline characteristics by cohort. Table 2 shows the distribution of these same characteristics by cohort once the propensity scores have been applied. The matching process has achieved balance between the two cohorts.

Table 1: Baseline characteristics by cohort

		Y Members	Non-Y members
Gender	Male	30	41
	Female	67	59
	Other	3	1
Age	16-34	24	23
	35-44	35	15
	55-64	12	15
	65+	8	27
Education	Less than High School Graduation	4	7
	High School Graduation	10	14
	Some Post-Secondary	15	6
	Completed Post-Secondary or higher	69	71
Immigrant Status	Born in Canada	40	62
	Not born in Canada	58	37

Table 1 (cont'd)

Ethnic Self-Identity	Western European	39	48
	North American (not Indigenous)	16	17
	South Asian	9	6
	East Asian	9	6
	Eastern European	7	5
	Black	7	5
	West Asian	8	3
	Latin American	1	1
	Indigenous	1	1
	Other	2	8
Employment Status	Self-employed	9	10
	Employed full-time	44	28
	Employed part-time	11	7
	Currently unemployed	13	6
	Retired	8	42
	Homemaker	3	3
	Student	6	2
Marital Status	Single and have never been married	21	18
	Married or common-law	63	58
	Separated	5	3
	Divorced	6	8
	Widowed	1	11
Sexual Orientation	Bisexual	6	2
	Heterosexual	63	85
	Gay	3	2
	Lesbian	1	1
	Queer	1	0
	Two-Spirit	0	0
	Other	4	1

Table 2: Characteristics of matched cohorts

		Y Members	Non-Y members
Gender	Male	38	41
	Female	62	59
	Other	0	0
Age	16-34	24	0
	35-44	30	0

Table 2 (cont'd)

	45-54	20	0
	55-64	16	0
	65+	10	0
Education	Less than High School Graduation	1	1
	High School Graduation	8	10
	Some Post-Secondary	15	12
	Completed Post-Secondary+	76	77
Immigrant Status	Born in Canada	23	26
	Not born in Canada	77	74
Ethnic Self-Identity	Western European	50	49
	North American (not Indigenous)	19	19
	South Asian	6	6
	East Asian	5	6
	Eastern European	5	5
	Black	5	5
	West Asian	2	2
	Latin American	2	2
	Indigenous	4	3
	Other	2	3
Employment Status	Self-employed	16	14
	Employed full-time	38	43
	Employed part-time	5	4
	Currently unemployed	10	10
	Retired	29	24
	Homemaker	1	3
	Student	1	2
Marital Status	Single never been married	20	18
	Married or common-law	64	64
	Separated	5	3
	Divorced	7	8
	Widowed	4	6
Sexual Orientation	Bisexual	1	2
	Heterosexual	98	97
	Gay	0.5	0.5
	Lesbian	0.1	0.1
	Queer	0.0	0.1
	Two-Spirit	0.0	0.0
	Other	0.1	0.1

The unmatched cohorts show baseline imbalances in the distribution of gender, employment status, educational status, marital status, and sexual orientation which have been corrected by the propensity score method. The propensity score matching process successfully matched 76% of the sample.

As shown in Table 3, the marginal distribution of health outcomes across the two cohorts is also not balanced at baseline. This is a result of sampling difference and further illustrates the need for statistical adjustment in order to estimate the potential differences a YMCA membership could make in one’s quality of health and life.

Table 3: Percent of each cohort in the “high” category of each outcome

	Y Members	Y Non-Members
High Well-being	72%	62%
High General Health	54%	52%
High Mental Health	71%	62%
High Sense of Belonging	73%	75%

### *Generalized Linear Models*

Using the propensity scores in generalized linear models produced statistically significant differences between the two cohorts in all our models. The YMCA members are more likely to have high levels of health in the four outcomes of interest. Table 4 through 7 show the summary of the model results.

Table 4: Generalized Linear Model for Well-being (Model 1)

	B	S.E.	Wald	Sig.	Exp(B)
YMembership	0.63	0.03	567.25	0.00	1.19
Gender	-0.21	0.03	59.66	0.00	0.81
Age	0.21	0.02	123.98	0.00	1.23
Employment	-0.21	0.01	607.65	0.00	0.81
Immigrant	0.08	0.03	5.98	0.01	1.08
Education	0.03	0.01	16.18	0.00	1.03
Sexual Orientation	-0.01	0.01	1.08	0.30	0.99
Ethnicity	-0.08	0.01	156.90	0.00	0.92
Constant	1.11	0.09	144.22	0.00	3.04

Table 5: Generalized Linear Model for General Health (Model 2)

	B	S.E.	Wald	Sig.	Exp(B)
YMembership	0.03	0.03	1.07	0.30	1.03
Gender	-0.64	0.03	587.95	0.00	0.53
Age	-0.07	0.02	14.34	0.00	0.93
Employment	-0.18	0.01	525.17	0.00	0.83
Immigrant	0.18	0.03	34.02	0.00	1.20
Education	0.23	0.01	1042.94	0.00	1.26
Sexual Orientation	-0.01	0.01	1.14	0.28	0.99
Ethnicity	-0.19	0.01	909.48	0.00	0.83
Constant	1.31	0.09	216.40	0.00	3.72

Table 6: Generalized Linear Model for Mental Health (Model 3)

	B	S.E.	Wald	Sig.	Exp(B)
YMembership	0.23	0.03	75.13	0.00	1.07
Gender	0.11	0.03	15.96	0.00	1.12
Age	0.52	0.02	743.27	0.00	1.68
Employment	-0.04	0.01	19.79	0.00	0.96
Immigrant	0.16	0.03	22.36	0.00	1.17
Education	0.13	0.01	298.44	0.00	1.13
Sexual Orientation	0.00	0.01	0.08	0.77	1.00
Ethnicity	-0.09	0.01	223.51	0.00	0.91
Constant	-0.95	0.09	109.25	0.00	0.39

Table 7: Generalized Linear Model for Sense of Belonging (Model 4)

	B	S.E.	Wald	Sig.	Exp(B)
YMembership	0.07	0.03	6.40	0.01	1.03
Gender	0.33	0.03	130.17	0.00	1.39
Age	0.45	0.02	501.18	0.00	1.57
Employment	-0.10	0.01	115.86	0.00	0.91
Immigrant	-0.23	0.03	45.86	0.00	0.80
Education	0.34	0.01	1826.31	0.00	1.41
Sexual Orientation	0.01	0.01	1.55	0.21	1.01
Ethnicity	-0.09	0.01	200.34	0.00	0.91
Constant	-1.38	0.10	209.86	0.00	0.25

Model 1 shows that matched YMCA members are 19% (95% CI: 15% - 22%) more likely to have a high level of Well-being. This model correctly predicts 78% of the data and Hosmer & Lemeshow test shows a poor fit 4% of the time, indicating that this model is a good fit for our data.

Model 2 shows that matched YMCA members are 2% (95% CI: 1% - 3%) more likely to have a high level of General Health. This model correctly predicts 89% of the data and Hosmer & Lemeshow test shows a poor fit 2% of the time, indicating that this model is a good fit for our data.

Model 3 shows that matched YMCA members are 7% (95% CI: 4% - 9%) more likely to have a high level of Mental Health. This model correctly predicts 80% of the data and Hosmer & Lemeshow test shows a poor fit 3% of the time, indicating that this model is a good fit for our data.

Model 4 shows that matched YMCA members are 2% (95% CI: 1% - 3%) more likely to have a high level of high Sense of Belonging. This model correctly predicts 75% of the data and Hosmer & Lemeshow test shows a poor fit 4% of the time, indicating that this model is a good fit for our data.

### ***Neighborhood Level Models***

These models incorporate both the public survey data and the matched member survey data, as well as the income data from the 2016 Canadian Census, and the geographic location data supplied by the YMCA. We include income as a fixed effect predictor. The model results shown in Tables 8 through 11 estimate the effect of the presence of a YMCA location. The presence of a YMCA location increases the percent of the neighborhood population that reports good health.

38% of the neighborhoods in the GTA do not have a YMCA located within them. 27% of the neighborhoods in the GTA has 1 YMCA located within them. 15% of neighborhoods have 2 YMCA locations within them. Finally, 20% of GTA neighborhoods have 3 or more YMCAs located within them.

There is a small and non-linear difference in average neighborhood income by the number of YMCAs located within the neighborhood. The average neighborhood income reduces as the number of YMCAs increases. The average household income of a neighborhood with no YMCA locations is \$89,000. The average household income of a neighborhood with 1 YMCA location is \$92,000 and the average household income of a neighborhood with 2 YMCA locations is \$89,000. This is further support for including income as a fixed effect.

Collinearity diagnostics showed that while collinearity was observed between the primary variables ( $r < .2$ ). Using "Leave one out" model comparisons confirmed that collinearity did not affect any of the significant effects reported below.

The models show that neighborhoods with a Y location compared to a neighborhood without a Y location is likely to have 3% (+/ 1.5%) more people with High Well-being; 3% (+/ 1.5%) more people with High General Health; 4% (+/ 2%) more people with High Mental Health; and 2% (+/ 1.5%) more people with High Belonging.

Table 8: Mixed Effects Model for Well-being (Model 10)

Constant	0.26	0.17	1.49	0.14
YMCAs	0.04	0.03	2.37	0.02
IncomeN	0.00	0.00	2.26	0.02

Bayesian estimated R-squared: 62%

Table 9: Mixed Effects Model for General Health (Model 11)

	Estimate	Std. Error	z value	Pr(> z )
Constant	-0.19	0.16	-1.23	0.22
YMCAs	0.04	0.01	2.18	0.03
IncomeN	0.00	0.00	1.85	0.06

Bayesian estimated R-squared: 59%

Table 10: Mixed Effects Model for Mental Health (Model 12)

	Estimate	Std. Error	z value	Pr(> z )
Constant	0.15	0.17	0.87	0.39
YMCAs	0.04	0.03	2.32	0.02
IncomeN	0.00	0.00	2.72	0.01

Bayesian estimated R-squared: 61%

Table 11: Mixed Effects Model for Sense of Belonging (Model 13)

	Estimate	Std. Error	z value	Pr(> z )
Constant	0.91	0.19	4.94	0.00
YMCAs	0.02	0.01	1.82	0.04
IncomeN	0.00	0.00	1.12	0.26

Bayesian estimated R-squared: 53%

## Conclusion

The results of our analysis demonstrate that the presence of a YMCA within a neighborhood has an overall positive effect on the health outcomes of the entire neighborhood. The results of our individual models show that controlling for demographics, the membership in the YMCA is related to more positive health outcomes. YMCA membership is very likely having a positive contribution to the well-being of its members. The health outcomes of interest are significantly higher for YMCA members.