Minimal Positive Influence of Helpline Efforts on Reduction of Homeless Death in Toronto*

Sandy Yu

March 12, 2024

This paper analyzes the influence of Central Intake Line staff efforts on the reduction of homeless death count in Toronto. Data on coded number of calls from Central Intake Call Wrap-Up Codes dataset and data on monthly homeless death count from Deaths of People Experiencing Homelessness dataset are used in generalized linear regression models to estimate the influence. We find that the staffs' net efforts create a statistically insignificant influence on the reduction of homeless death, with specific types of efforts appearing counterproductive. This implies that while a positive influence does exist, from a statistical point of view, the impact of the Central Intake Line staff is not significantly different from non-existent.

Table of contents

*Code and data are available at: https://github.com/Jingying-yu/help-line-homeless-death-analysis.git

1 Introduction

The City of Toronto has been consistently providing funding and staff in an effort to reduce the active homeless population within its governing district. Amongst numerous city efforts, the 24/7 helpline — Toronto Central Intake Line — was created to offer referrals to emergency shelters, sleeping accommodations, and provide general information about homelessness and prevention. (Shelter 2020). Starting in 2020, detailed data about the Central Intake Line became available on the OpenDataToronto portal (Sharla Gelfand 2022).

The estimand of the paper focuses on examining the influence of helpline efforts on the reduction of homeless deaths in Toronto. Using Central Intake Call Wrap-up Codes data (Shelter 2020) from OpenDataToronto (Sharla Gelfand 2022), three different aspects of the helpline efforts were isolated: total calls coded, referral to shelter, and homelessness prevention information provided. These three key aspects were measured against the monthly cumulative death count from Deaths of People Experiencing Homelessness Data (Health 2017) documented by the Toronto Public Health (Toronto 2024). The analyses yielded an insignificant negative correlation between the net efforts of the helpline staff. The variable referral to shelter yields a positive correlation with homeless death count, whereas the effort to provide homelessness information yields a negative correlation. This implies that the net influence of helpline efforts is statistically negligible, while individual efforts can be productive or counterproductive. This examination in the net efforts and individual efforts can provide the helpline with a suggestion of which types of services or calls would be more economically efficient to allocate their labor resources on.

Analyses and findings in this paper are structured into several sections: Section 2 – Data, Section 3 – Model, Section 4 – Results, and Section 5 – Discussion. The Data section examines all datasets and variables kept for analysis, followed by an explanation of their data cleaning processes. The Model section defines linear models used for further analysis, explain its components, and presents model justifications. The Result section focuses on visualizing and presenting the model results through data presented in Data section. The Discussion section further evaluate the interpretations behind the model results presented in the previous section, and touches on any weaknesses and next steps.

2 Data

All data used in this paper are obtained through OpenDataToronto Portal (Sharla Gelfand 2022). Two different datasets: Central Intake Call Wrap-Up Codes Data (Shelter 2020) and Deaths of People Experiencing Homelessness(Health 2017), are retrieved to analyze the effect of Toronto's Central Call Line (denoted as *Helpline* for the remainder of this paper) efforts to the death counts of homeless individuals in Toronto. Data is cleaned and analyzed using the open source statistical programming language R (R Core Team 2023) and supporting packages tidyverse (Wickham et al. 2019), janitor (Firke 2023), rstanarm (Goodrich et al. 2022), arrow (Richardson et al. 2024), ggplot2 (Wickham 2016), and knitr (Xie 2023). Detailed description of each dataset can be found in the subsections below.

2.1 Central Intake Call Wrap-Up Codes

On the OpenDataToronto portal, there are several datasets that reflect the City's effort to shelter the local homeless population. The Central Intake Call Wrap-Up Codes Dataset (Shelter 2020), stored in the Central Intake Calls Catalogue (Shelter 2020), is one of the freshest and most detailed. Data is stored and published by the Shelter, Support & Housing Administration since November of 2020 and refreshes on a monthly basis. The latest refresh occurred on January 15th, 2024.

The data set provides a daily summary of the number of calls received, the number of calls classified into distinct wrap-up codes by the nature of its issue, and a count of calls under each wrap-up code. One of the example wrap-up codes in the original data set was: *Code 1A - Referral to a Sleeping/Resting Space*. The original data set includes 13 distinct wrap-up codes; only two codes, *Code 1A — Referral to a Sleeping/Resting Space* and *Code 2C — Information - Homelessness & Prevention Services*, were chosen for our analysis. Code 1A and Code 2C are best suited as measurements for positive impact the Central Intake Line could provide because they provide a count of the number of callers provided with directions or advice. Other non-suitable wrap-up codes include: *Code 1D - Declined Shelter/Resting Space*, *Code 4B - Disconnected - No Outcome*, etc.

The final dataset only includes monthly cumulative data before July of 2023. The rationalization behind the action will be explained in Section 2.3.

Table 1: Sample of monthly summation for Helpline Coded Calls data

Month	Calls Coded	Referral to Shelter	Homelessness Prevention Info
2020-11-30	8367	1433	2029
2020-12-31	10232	2027	2427
2021-01-31	12091	1726	2453
2021-02-28	10525	1662	2094
2021-03-31	12287	1771	2202
2021-04-30	12668	1232	1870

2.2 Deaths of People Experiencing Homelessness

The Deaths of People Experiencing Homelessness Dataset (Health 2017) contains monthly cumulative records of homeless deaths. The dataset is published by Toronto Public Health (Toronto 2024). The earliest data record started in January of 2017, and the latest record ends in June of 2023.

The original dataset contains three columns: *Year of death*, *Month of death*, and *Count*. After careful inspection of raw data, any rows that containing an "unknown" value are excluded. Although this action created a source of error in future analyses, this is still a necessary step because we are plotting death counts against a timeline.

Final dataset only includes data from November 2020 to June 2023 (Table 2). The rationalization behind the action will be explained in **?@sec-data-diff-datasets**.

Table 2: Sample of Cleaned Deaths of People Experiencing Homelessness Dataset

Month	Death Count
2020-11-30	15
2020-12-31	20
2021-01-31	20
2021-02-28	15
2021-03-31	15
2021-04-30	11

2.3 Final Combined Data

During the data cleaning process, code was written to ensure that both the Central Intake Call Wrap-Up Codes Dataset (Shelter 2020) and the Deaths of People Experiencing Homelessness Dataset (Health 2017) are filtered to include only data between 1st November, 2020 to 30st June, 2023.

The decision is reached by taking the common time period between the two datasets. In the latest refresh, the Central Intake Call Wrap-Up Codes Dataset (Shelter 2020) begins on 3rd November, 2020 and end on 31st December, 2023; whereas the Deaths of People Experiencing Homelessness Dataset (Health 2017) begins on January of 2017 and ends on June 2023.

The ultimate purpose of taking the common time period is to ensure that we can have the same amount of observations for analysis. Thus a combined dataset that kept the variables of interest in both datasets is created (Table 3).

Table 3: Sample of monthly homeless death count and Helpline Coding efforts

Month	Total Coded	Referred to Shelter	Homeless Info	Death Count
2020-11-30	8367	1433	2029	15
2020-12-31	10232	2027	2427	20
2021-01-31	12091	1726	2453	20
2021-02-28	10525	1662	2094	15
2021-03-31	12287	1771	2202	15
2021-04-30	12668	1232	1870	11

3 Model

While we can still make claims on the correlations between helpline efforts and homeless death count without modelling, regression analysis is still necessary to provide rigorous backbones for a claim's statistical significance.

Here we briefly describe the Bayesian analysis model used to investigate the correlation between homeless death counts in Toronto and the efforts of the Helpline staff. Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Magnitude of effect created by the helpline efforts can be estimated through linear models. Gaussian regression model is best suited for our analytical purposes. Other generalized linear models such as Logistic and Multilevel are not considered due to the nature of the variables in interest. Logistic regression model requires data that yields either an **yes** or **no** result, whereas multilevel model require more complexity between variables. Due to the nature of the Poisson function, the Poisson model was considered and discarded because model results in extremely small values with no evidently different interpretations.

3.1.1 Model 1: Total Coded Model

$y_i|\mu_i,\, \sigma \sim \text{Normal}(\mu_i, \sigma)$

$\mu_i = \alpha + \phi_i$

$\alpha \sim \text{Normal}(0, 2.5)$

$\phi \sim \text{Normal}(0, 2.5)$

$\sigma \sim \text{Exponential}(1)$

Where:

- y_i is the number of homeless death count in Toronto recorded by Toronto (2024) per month
- ϕ_i is the total number of calls received and coded by the helpline per month

3.1.2 Model 2: Referred Informed Model

$y_i|\mu_i,\, \sigma \sim \text{Normal}(\mu_i, \sigma)$

$\mu_i = \alpha + \beta_i + \gamma_i$

$\alpha \sim \text{Normal}(0, 2.5)$

$\beta \sim \text{Normal}(0, 2.5)$

$\gamma \sim \text{Normal}(0, 2.5)$

$\sigma \sim \text{Exponential}(1)$

Where:

- y_i is the number of homeless death count in Toronto recorded by Toronto (2024) per month
- β_i is the number of homeless individuals referred to shelters through the helpline, per month
- γ_i is the number of individuals provided with information on homelessness prevention through the helpline, per month

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.3 Model justifications

We expect a net negative relationship between the net efforts made by the helpline staff (Recorded by the Total_Coded) and homeless death count, which implies that net effect of helpline calls **reduces** homeless death count in Toronto.

Additionally, we expect a negative relationship between homeless death and the two individual efforts: *number of homeless individuals referred to shelter*, and *number of individuals provided with homelessness prevention information.* Since both factors are coded efforts made by the helpline, both factors should also help to reduce homeless death count in Toronto.

4 Results

Our results are summarized in Table 4.

- there exist a net negative correlation between helpline efforts and homeless death count in Toronto

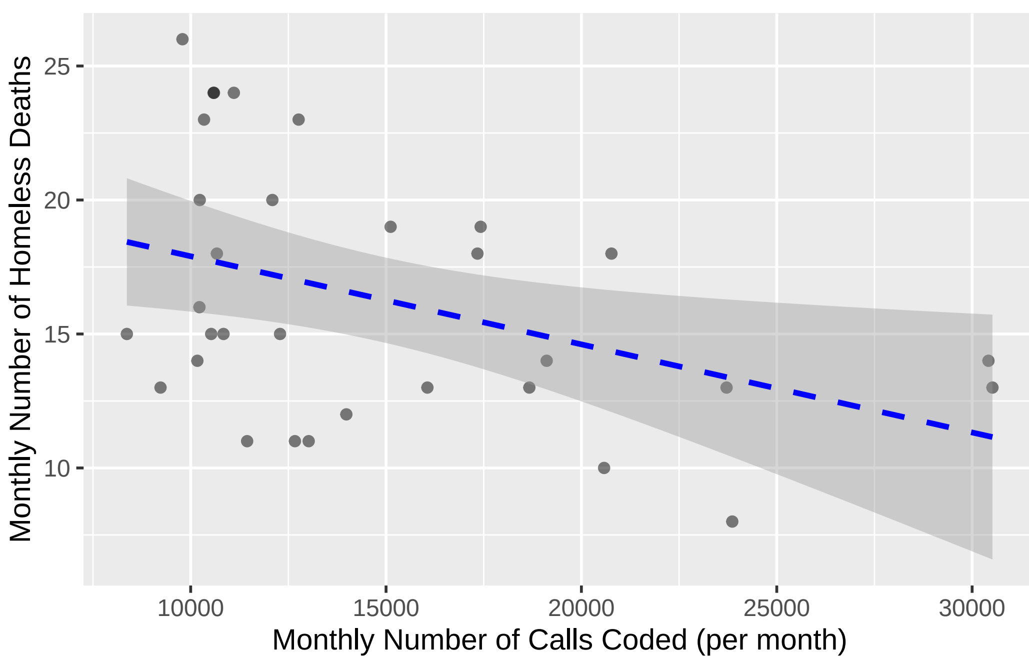


Figure 1: Visualization to show correlation between number of Calls Coded and Homeless Death Counts

Table 4: Explanatory model of homeless death count in relation to different helpline efforts

	Total Coded Model	Referred and Informed Model
(Intercept)	21.0727	16.7375
	(2.2010)	(3.4460)
Total_Coded	−0.0003	
	(0.0001)	
Referred		0.0059
		(0.0022)
Informed		−0.0048
		(0.0027)
Num.Obs.	32	32
R2	0.159	0.228
Log.Lik.	−92.606	−91.554
RMSE	4.27	4.10

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hypothesis 1: - null hypothesis: true effect of total_calls_coded = 0 - use t-stat to check at 95% significance level

Hypothesis 2: - null hypothesis: true effect of referred = 0 - use t-stat to check at 95% significance level

Hypothesis 3: - null hypothesis: true effect of referred = 0 - use t-stat to check at 95% significance level

5 Discussion

5.1 Realistic Interpretation of Net Efforts

- On average, for every increase in calls coded by the helpline, there is a 0.0003 unit decrease in homeless death in Toronto. There is an overall reduction, albeit very small, in homeless death as the number of total coded calls increase. This implies that the helpline need to take at minimum 3333 calls per month to save 1 homeless life in Toronto. On average, the helpline takes 14828.59 calls per month, meaning that 4.45 lives are saved on average.
- 4.45 is not a significant number in both a practical and statistical sense. To improve the return on total number of calls coded, we attempted to identify promising call codes that contribute to the net positive reduction in homeless death. Out of the 13 different call codes, *Referral to Shelter* and *Information on Homeless Prevention* are best suited

as individual efforts of measurement for positive impacts because they provide a count of the number of callers provided with directions or advice.

5.2 Individual Effort: Referral to Shelter

- On average, for every homeless caller referred to shelter, there is a 0.0059 unit **increase** in homeless death in Toronto. Though not significant on a statistical level, this trend is still very counter-intuitive discouraging. In the case where the correlation is more significant, we would want to further analyze factor related to referrals to shelter. Accuracy on the current status of shelters, level of detail on instructed, and misjudgment of shelter eligibility are possible factors of examination. However, since the effect of this call code is statistically insignificant, it is more likely just a fluke caused by the dataset.

5.3 Individual Effort: Information on Homelessness Prevention

5.4 Weaknesses and next steps

Weaknesses: very small sample size, only 32 observations Next steps: investigate into the possible factors that caused a net zero impact on reduction of homeless death count and propose possible solution using data.

Appendix

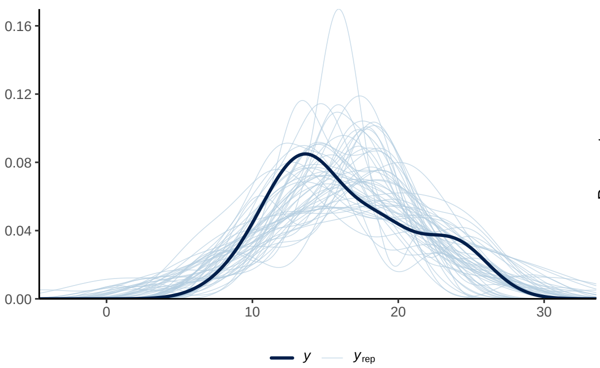
A Additional data details

B Model details

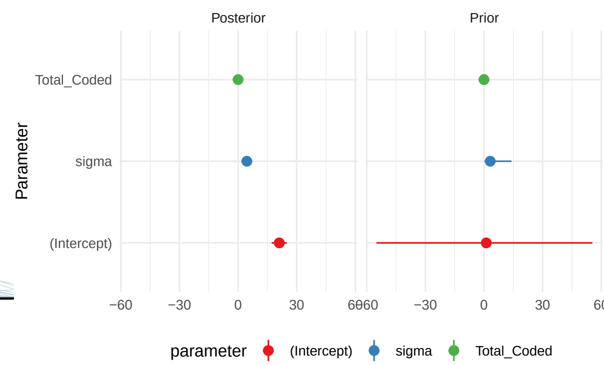
B.1 Posterior predictive check

In Figure 2a we implement a posterior predictive check. This shows...

In Figure 2b we compare the posterior with the prior. This shows...



(a) Posterior prediction check



(b) Comparing the posterior with the prior

Figure 2: Examining how the model fits, and is affected by, the data

In Figure 3a we implement a posterior predictive check. This shows...

In Figure 3b we compare the posterior with the prior. This shows...

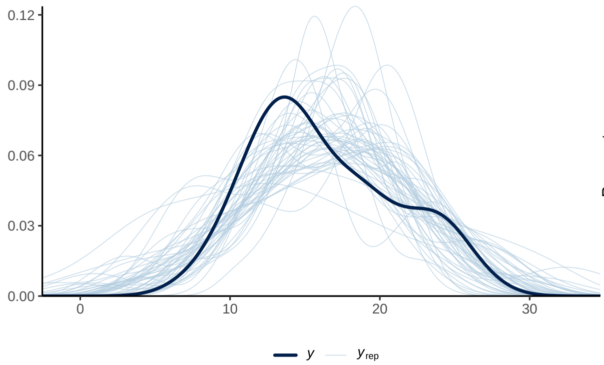
B.2 Diagnostics

Figure 4a is a trace plot. It shows... This suggests...

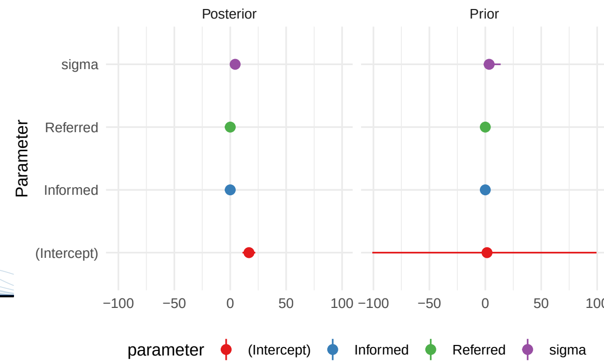
Figure 4b is a Rhat plot. It shows... This suggests...

Figure 5a is a trace plot. It shows... This suggests...

Figure 5b is a Rhat plot. It shows... This suggests...

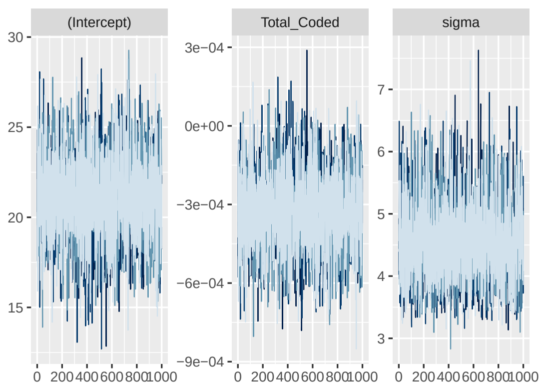


(a) Posterior prediction check

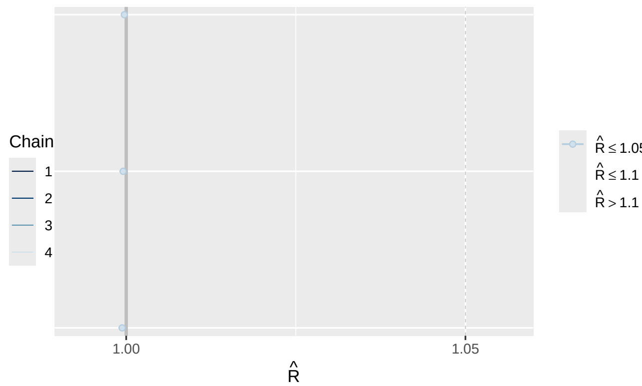


(b) Comparing the posterior with the prior

Figure 3: Examining how the model fits, and is affected by, the data

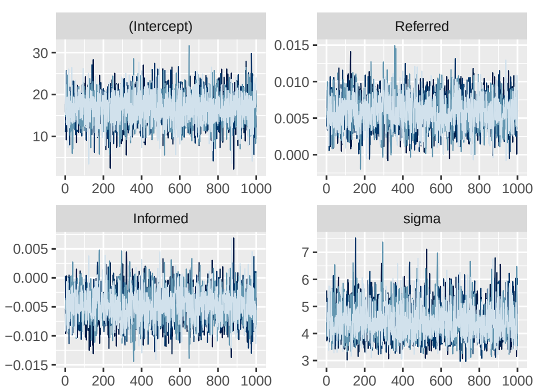


(a) Trace plot

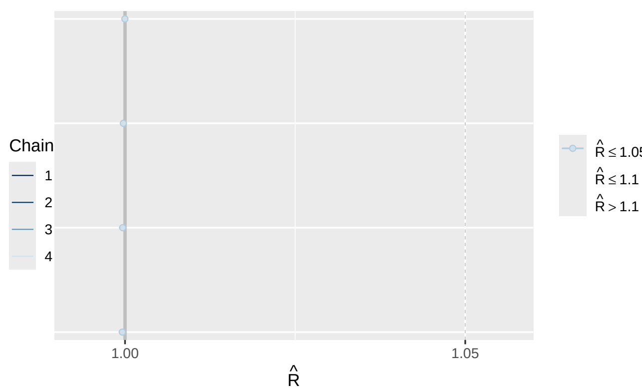


(b) Rhat plot

Figure 4: Checking the convergence of the MCMC algorithm for total_coded_model



(a) Trace plot



(b) Rhat plot

Figure 5: Checking the convergence of the MCMC algorithm for referred_informed_model

References

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Health, Toronto Public. 2017. "Deaths of People Experiencing Homelessness." https://open.toronto.ca/dataset/deaths-of-people-experiencing-homelessness/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://github.com/apache/arrow/.

Sharla Gelfand, City of Toronto. 2022. "Opendatatoronto: Access the City of Toronto Open Data Portal." https://open.toronto.ca.

Shelter, City of Toronto, Support & Housing Administration. 2020. "Central Intake Call Wrap-up Codes Data." https://open.toronto.ca/dataset/central-intake-calls/.

Toronto, City of. 2024. "Toronto Public Health." https://www.toronto.ca/city-government/accountability-operations-customer-service/city-administration/staff-directory-divisions-and-customer-service/toronto-public-health/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.