



LGBTI Exclusion in Serbia: Measuring its Socioeconomic Dimensions



WORLD BANK GROUP

UCLA School of Law
Williams Institute



AMERICAN UNIVERSITY
WASHINGTON, DC



Analytical Framework



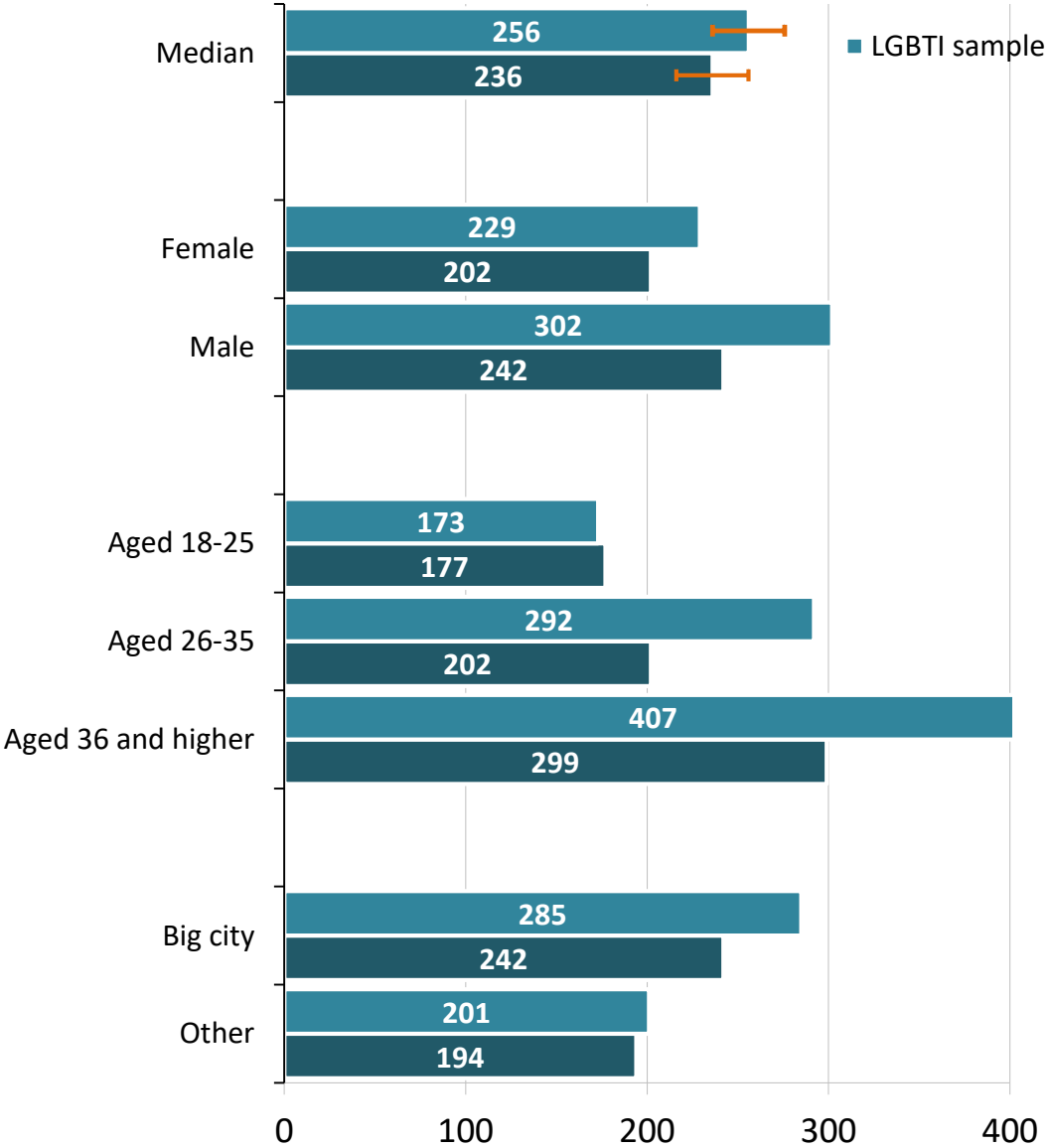
Western Balkans Research Overview

- **Regional:** Replication of the European LGBT Survey implemented by the European Fundamental Rights Agency (FRA) in 2012.
- **Serbia:** Two experimental studies examining the access of LGBTI people to housing and education.
 - ❑ Implemented a revised version of the SILC with specific SOGI related questions.



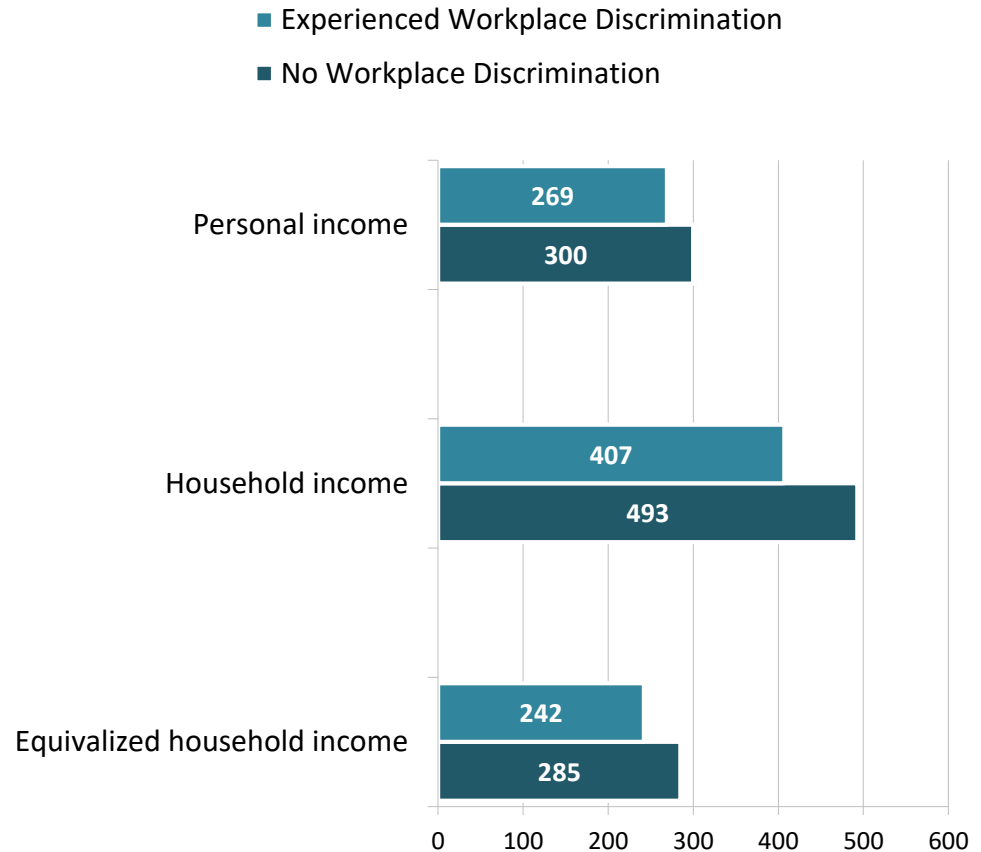
Key Findings

Monthly Personal Income (median)



Understanding the effects of workplace discrimination

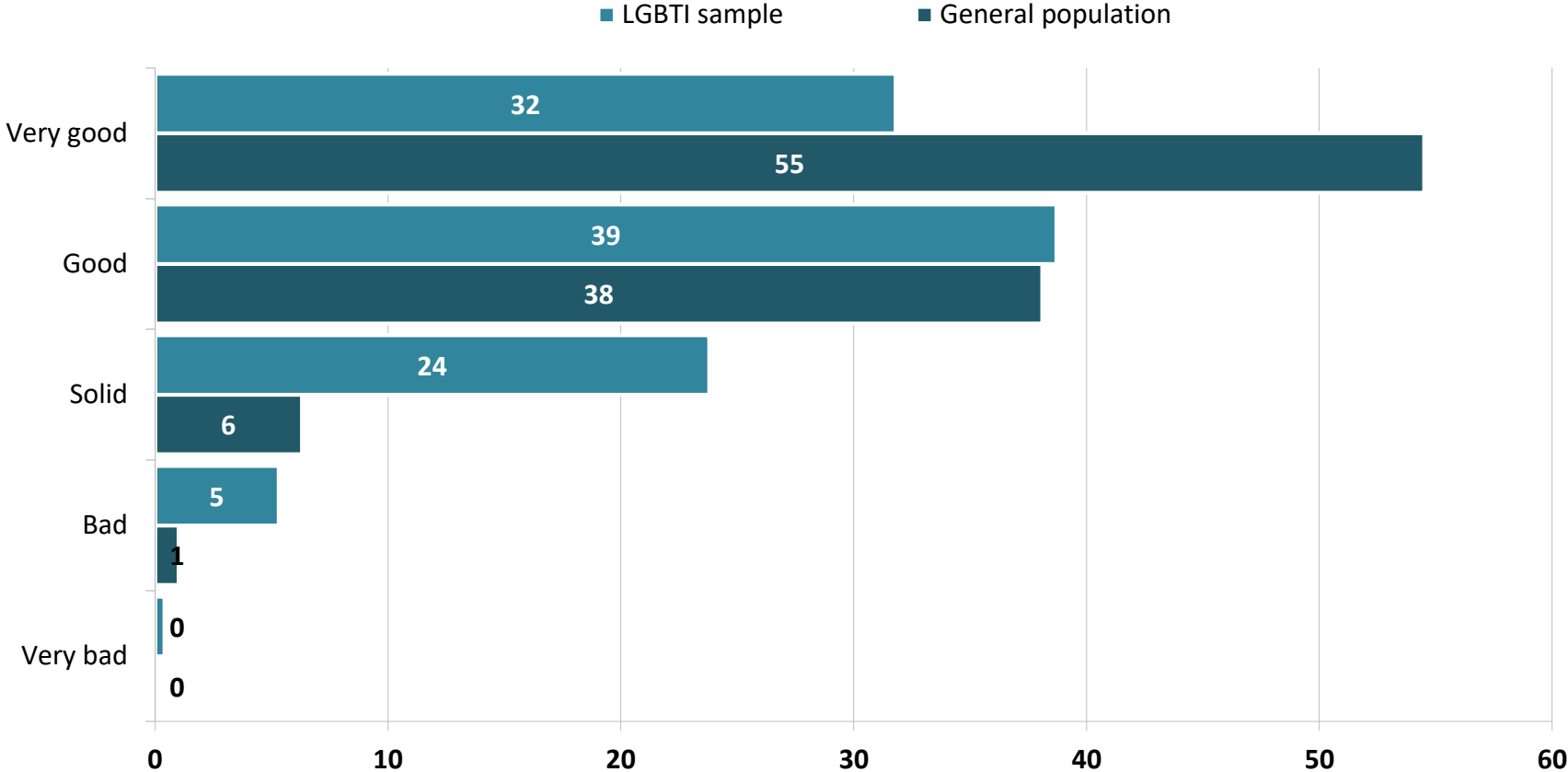
- 15% of respondents have experienced discrimination at work because of their SOGI
- Due to discrimination 10% have quit a paid job in the past
- 7% took unplanned leave because of discrimination they experienced



Increased Risk of Poverty



State of Health





Methodological

Overview

OVERVIEW OF THE SILC

- A survey that collects data on income, poverty, social exclusion, and living conditions
- Survey of approximately 8,008 households (in 2013)
- The survey applies the European Statistical Office (Eurostat) definitions and survey items to operationalize key concepts of income, poverty, social exclusion, and living conditions.
- The survey also contains key sample demographics that may predict economic outcomes

MACHINE LEARNING AND THE SILC

- When designing the LGBTI SILC questionnaire, we wanted to make sure that we included survey items that are predictive of income disparities.
- We were also limited by survey length, knowing that we could not replicate the entire SILC questionnaire.
- We used employment status as our key dependent variable, and selected a whole host of candidate variables that could be predictive of employment status.
- We relied on chi-squared automatic interaction detection (CHAID) – a decision tree method of machine learning – to prune the candidate variable to the key variables that predict this outcome.
- The items predictive in this model were included in the LGBTI SILC questionnaire

CHAID RESULTS AND THE SILC

- The variables that the machine learning analysis selected were:
 - Age (HL7)
 - Educational attainment (OP6)
 - Martial status (OP8)
 - Self-evaluation of health (L10_1)
 - Having an AC unit (DL_6_7)

CHALLENGES OF SAMPLING LGBTI POPULATIONS

- Traditional probability-based ways of surveying LGBTI people is extremely limited because:
 - There is no sampling frame for LGBTI people
 - There are no demographic targets for weighting adjustment
 - Existing estimates in other regions suggest 3%-5% of the population identifies as LGBT. Thus, traditional surveys would have to successfully interview approximately 10,000 people to (luckily) have 300-500 LGBT identified respondents.
 - This problem compounds with survey non-response. In the US, response rates of 8% means that 125,000 people would have to be contacted to obtain an analytic sample of 10,000 to have a sample of 300-500 LGBT identified respondents.
- Operationalizing sexual orientation, gender identity, and people with intersex traits is challenging due to cross-national differences in understandings of the concepts, the concepts themselves are multidimensional and in fluctuation.

WEIGHTING ADJUSTMENTS TO THE LGBTI SAMPLE

- We conducted a systematic review of studies measuring LGBTI populations across a variety of countries across the world. We identified 154 administrative studies and 150 peer-reviewed articles or organizational reports that resulted in 520 estimates of sexual orientation and/or gender identity.
- From this, we conducted a Bayesian multilevel meta-analysis, to evaluate the propensity individuals identify as LGBT, taking into account country and regional characteristics.

$$s^* \sim \text{binomial}(n^*, \pi)$$
$$\text{logit}(\pi) = b_0 + b_1 * \text{Sex} + \alpha^{\text{Sexual orientation}} + b_2 * \text{Transgender} + \alpha^{\text{Age}} + \alpha^{\text{Country}}$$
$$\alpha^{\text{Country}} \sim \text{Normal}(u_0 + u_1 * \text{GDP_PC} + u_2 * \text{LGBT Acceptance} + u_3 * \text{LGBTI rights}, \sigma^2)$$

- This permitted the estimation of a country-specific weight $W = \frac{1}{\pi}$
- This primarily ensures that our sample is appropriately weighted for estimates of LGBTI people (i.e., not overrepresenting gay men by virtue of the sampling & recruitment strategy).

METHODOLOGICAL CHOICES AVAILABLE TO COMPARE THE LGBTI SAMPLE TO THE SILC IN SERBIA

1) Compare LGBTI sample estimates to unadjusted population estimates from the SILC



2) Adjust the LGBTI sample characteristics to match the general population characteristics from the SILC, then draw comparison.

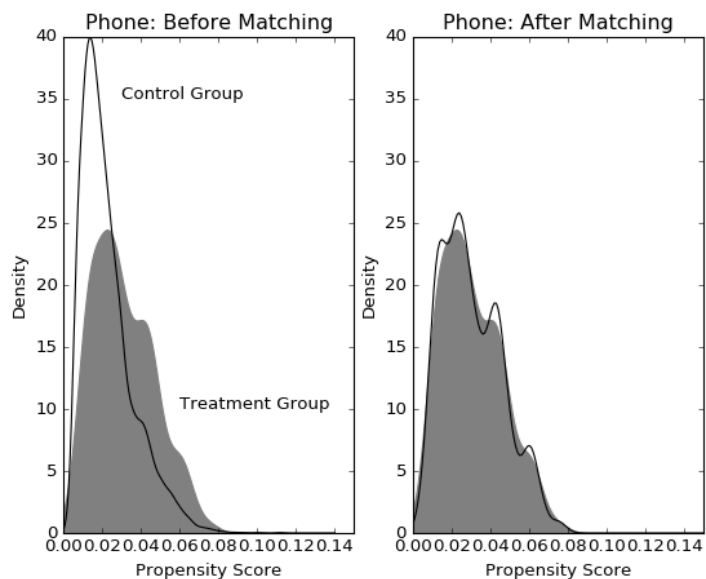


3) Compare the LGBTI sample estimates to adjusted population estimates from the SILC, after matching on key demographic characteristics.



STATISTICAL MATCHING

- Matching is a process that brings two samples that may be different on key characteristics more in line with one another.



- This balance minimizes differences between the two groups. If done appropriately, it can yield valid causal estimates.
- ☐ We used the inverse of a Covariate Balanced Propensity Score to weight the Official SILC to match the characteristics of LGBTI sample.

DATA PRE-PROCESSING & VARIABLES SELECTED FOR MATCHING

- We first subset the Official SILC in Serbia to individuals aged sixty or younger, who had a computer in the household, and had Internet access.
- The matching procedure took the following details into account: age, sex assigned at birth, educational attainment, marital status, region, and urbanity.
- While the CHAID analysis suggested other variables, the team decided that a minimal model might be preferred, if we consider health status to be an outcome as opposed as something for which to be adjusted.

COMPARING SAMPLE CHARACTERISTICS BEFORE-AND-AFTER MATCHING

Age	LGBTI Sample	Unweighted SILC	Weighted SILC
17-29	55.14%	32.16%	53.78%
30-44	39.90%	36.52%	41.26%
45-67	4.95%	31.32%	4.96%

Region	LGBTI Sample	Unweighted SILC	Weighted SILC
RS11	52.78%	23.48%	52.90%
RS12	27.79%	33.08%	27.68%
RS21	11.13%	22.74%	10.89%
RS22	8.30%	20.70%	8.52%

Marital Status	LGBTI Sample	Unweighted SILC	Weighted SILC
Not Married or Parried	84.67%	44.87%	86.28%
Married or Partnered	15.33%	55.13%	13.72%

Urban	LGBTI Sample	Unweighted SILC	Weighted SILC
Densely populated	72.25%	36.05%	72.16%
Medium populated	23.53%	27.81%	23.63%
Sparsely populated	4.22%	36.14%	4.22%

Education	LGBTI Sample	Unweighted SILC	Weighted SILC
Low	3.06%	15.13%	2.82%
Medium	39.54%	62.44%	39.37%
High	56.65%	22.43%	57.81%
Missing	0.75%	0.00%	0.00%

BENEFITS AND LIMITATIONS

- Adjustments overcome inherent limitations in just making comparisons between the LGBTI sample and the general population
- It still requires estimation assumptions that may mean that model misspecification can harm the internal validity of the study
- By weighting the SILC to the LGBTI sample distribution, we increase internal validity but decrease external validity
- Likely the best we can do absent LGBTI measures on the official SILC

CAREFUL CONSIDERATIONS FOR PURPOSEIVE SAMPLES LIKE LGBTI POPULATIONS

- Question wording and operationalizing sexual orientation, gender identity, and intersex traits
- Careful consideration of inclusivity and what one will be able to meaningfully say about subgroups
- Careful consideration of questionnaire design in preparation for data analysis
- Advances in statistical modeling can help combine probability and nonprobability samples for data analysis.

THANK YOU!

Dominik Koehler
Social Inclusion Consultant
dkohler@worldbank.org

Andrew R. Flores
Assistant Professor
aflores@american.edu

Gladys Senderayi
Public Sector Specialist
rsenderayi@worldbank.org