

Surveys, Big Data, and Experiments

How Can We Best Learn about LGBTI Development Outcomes?

Dominik Koehler

Nicholas Menzies



WORLD BANK GROUP

Governance Global Practice Group

July 2017

Abstract

There is little rigorous quantitative data about the lives of lesbian, gay, bisexual, transgender, and intersex (LGBTI) people in developing countries. This makes the development of policy to improve the welfare of lesbian, gay, bisexual, transgender, and intersex people difficult, and it also makes it difficult to know whether lesbian, gay, bisexual, transgender, and intersex-focused policies and programs are working. Filling this data gap is necessary to understand the development outcomes for lesbian, gay, bisexual, transgender, and intersex people. Quantitative

data practices exist that can be drawn on to fill the gap, including household surveys, experiments, and big data analysis. Summarizing existing experience, this paper provides guidance on how to study development outcomes for lesbian, gay, bisexual, transgender, and intersex people, by: paying attention to the different ways to define sexual orientation, gender identity and expression, and sex characteristics; and collecting samples that allow conclusions to be drawn with the broader lesbian, gay, bisexual, transgender, and intersex community, as well as the general population.

This paper is a product of the Governance Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at dkohler@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Surveys, Big Data, and Experiments: How Can We Best Learn about LGBTI Development Outcomes?

Dominik Koehler and Nicholas Menzies

JEL classification: C83 Survey Methods & Sampling Methods; O35 Social Innovation

Keywords:

Lesbian, Gay, Bisexual, Transgender, LGBTI, Sexual Orientation, Gender Identity, SOGI, Development Outcomes, Population Size, Surveys, Big Data, and Experiments.

Acknowledgements

The authors would like to thank the following for thoughtful guidance and comments: Georgia Harley, Kristen Himelein, Chloe Schwenke, M. V. Lee Badgett, Clifton Cortez and the members of the World Bank's SOGI Task Force. All errors are the authors alone, and the views expressed herein do not represent the position of the World Bank or its Executive Directors.

The authors can be contacted at: dkoehler@worldbank.org; nmenzies@worldbank.org

Introduction

Each year in the U.S. we see an increase in ads and TV spots during Pride month or at mega-events, such as the Super Bowl or the Grammys, that feature happy, wealthy same-sex couples. Seeing these ads, [reading newspaper articles about the spending power of same-sex couples](#) and consuming popular images of lesbian, gay and trans people on television or in movies, it would be easy to assume that LGBTI people are better off than most. However, appearances can be deceiving and what the media portrays is often far from the lived reality of most LGBTI people, in the United States and beyond. There is a tendency to conflate the experience of one subgroup, wealthier gay men, with the experience of the others tied up in the “LGBTI” acronym, when their opportunities and outcomes are quite different, and what we know about each subgroup varies significantly.

The little rigorous data that are available suggest that [LGBTI people are often among the most vulnerable in society](#) and on average fare worse than non-LGBTI people.¹ Unfortunately, these data are patchy and so it is hard to say for certain - though theory and qualitative research support this overall finding. Of course, large differences exist across contexts. The scarce quantitative data that do exist are mostly limited to developed countries, and focus on urban and cisgender² male populations. [The invisibility in the data](#) makes it more difficult to persuade policy makers that LGBTI people should be specifically included in development programs, know how to target such programs to best address their needs, and measure whether such programs are working.

To encourage development professionals to address the data gap, this paper provides an overview of, and some reflections on, existing quantitative data collection efforts.³ The goal is to stimulate and inform future knowledge creation and avoid some obstacles previous studies faced. The paper addresses issues of **defining LGBTI communities, estimating the LGBTI population, the representativeness of samples**, (including comparing outcomes to non-LGBTI populations), using **big data** and **experimental techniques**. Primarily, we hope to reach empirical experts and researchers who are new to the topic to encourage more SOGIE inclusive, rigorous data collection. Secondly, we hope that those with a background in SOGIE issues might read this and thus be more open to placing an emphasis on quantitative data. By highlighting the specific development outcomes of LGBTI people, policy makers who otherwise would not have recognized LGBTI people as an area of concern might be convinced to design more inclusive development programs.

¹ Valfort, Marie-Anne. 2017. [LGBTI in OECD Countries: A Review](#). OECD Social, Employment and Migration Working Papers, No. 198.

² Cis gender refers to a person whose self-identity conforms to their biological sex, that is, not transgender.

³ Both qualitative and quantitative data are necessary to make informed policy decisions, as well as create the political space to act. Existing data collection efforts have focused more heavily on qualitative data, and we therefore address quantitative data here as one way of encouraging researchers to address this imbalance.

1. Definitions: What are “lesbian”, “gay”, “bisexual” “transgender” and “intersex”?

In order to conduct effective research into a particular group of the population, it is necessary to define it. This can be very challenging for LGBTI people for many reasons. LGBTI – as the acronym suggests – is an agglomeration of a number of groups, each of which has their own characteristics and definitional issues (as well as development challenges and outcomes).⁴ The individual experiences of each subgroup tend to be overshadowed by the more visible and more studied experience of gay men. In particular, the experiences of trans, intersex, and bisexual people tend to be the most understudied.⁵ The ways in which the terms are understood, expressed and labeled also significantly varies within and across countries. None of these are unique to LGBTI people as subjects of research, though together they present a special challenge when it comes to conducting quantitative research that by its very nature emphasizes standardized questions and aggregate answers.

Somewhat hesitantly – for fear of seeming to diminish the incredible diversity of sexual orientation, gender identity and expression, and sex characteristics – we present here some broad, higher-level definitions. While we have tried to formulate inclusive definitions, we recognize that some may work in some contexts and not in others. We do this in an attempt to guide common thinking in the drafting of surveys.⁶ One of the benefits of quantitative data is the ability to collect information on the experience of a large number of people and to compare across contexts. This benefit is enhanced with the use of comparable questions, methods, and analysis. But standardization does come with costs, including glossing over the richness and uniqueness of individual lives. Another risk is that the utilization of common definitions can undermine the collection of data in any given context if people do not understand the terms employed (see below). When it comes to formulating questions, each would need to be thoroughly context-tested prior to any survey and a case-by-case balance struck between local needs and cross-country comparability.

Sexual Orientation, Gender Identity and Expression, and Sex Characteristics

Sexual orientation consists of three conceptual elements: (1) attraction, which refers to a person’s enduring emotional, romantic, or sexual attraction to a person of same and/or

⁴ Each of these intersects with other markers of identity that impact development outcomes, such as age, race, ethnicity, gender etc.

⁵ We also note that there are other (interchangeable and additional) words that people use which are not directly reflected in the “L”, “G”, “B”, “T” and “I” – which has led to the more recent adoption of “LGBTI+” (along with many other formulations). We celebrate this diversity and fluidity but given the empirical focus of this paper limit ourselves “LGBTI” as this is where (relatively) more data exist.

⁶ For further reference: [Yogyakarta Principles on the Application of International Human Rights Law in relation to Sexual Orientation and Gender Identity](#).

opposite gender; (2) behavior, which relates to the sex of sex partners (individuals of the same sex, different sex, or both sexes); and self-identification of a person.⁷ *Lesbian (women) and gay (men)* refer to people who have an enduring emotional, romantic, or sexual attraction primarily or exclusively to people of the same gender. *Homosexuality* is another term frequently used to describe this attraction (compared to *heterosexual* or *straight*). *Bisexual* individuals can have the same emotional, romantic, or sexual attraction to another person regardless of gender.⁸

For the purposes of data collection by way of a quantitative survey, it's important to take into account that these elements do not always align. For example, it is common to find men who have sex with men (MSM), who might also be married to women and who would not self-identify as gay or bisexual. Here one behavior (sex) is seemingly in contrast to another (marriage to an opposite sex spouse) as well as identity. It is therefore important to include questions related to all three dimensions of sexual orientation in a survey.⁹

Gender identity refers to a person's deeply felt, internal identification as a man, woman or some other gender, which can be different to the sex assigned at birth. *Gender expression* refers to a person's outer appearance, speech, social interaction and how others may perceive this. *Transgender* refers to an individual whose gender identity is different from the biological sex that was assigned at birth.¹⁰ By contrast, *cisgender* refers to people whose gender identity and sex assigned at birth are the same.

Intersex is a scientific term which describes a variety of chromosomal, hormonal, and anatomical conditions or *sex characteristics* in a person that do not fit the typical definitions of male and female.

In countries around the world, there are many different understandings of sexual orientation, gender identity and expression.¹¹ Facebook's 51 options for gender identity is just one indicator of this variety.

Further complexity is added to the above definitions for a number of reasons:

⁷ Park, Andrew. 2016. [Reachable: Data collection methods for sexual orientation and gender identity](#).

⁸ Other terms include pansexual: attraction to all sexes, genders and gender identities; and asexual: attraction to none.

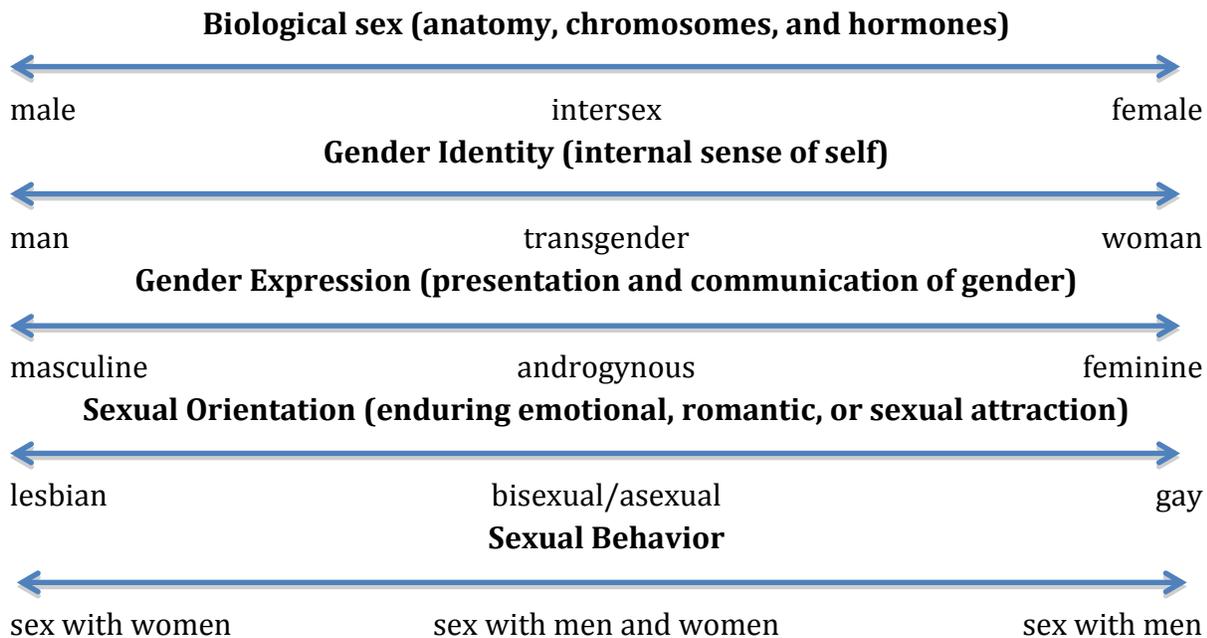
⁹ "Identity, behavior, attraction, and relationships all capture related dimensions of sexual orientation but none of these measures completely addresses the concept." (Gates, 2011)

¹⁰ There are other terms used to describe gender identity, most commonly transsexual. Terms such as transvestite should not be confused with transgender, since it only refers to a person who dresses in clothes traditionally associated with the other gender. The term also has a negative connotation, with cross-dressers often preferred.

¹¹ These include (among many, many others): third gender, a term common in Asia; two-spirited, a term used by Native Americans to describe a person that can fulfill both gender roles; Hijra a term used in India to describe transgender people mostly transgender-women. Because of its scientific character this does not necessarily apply to intersex which can be considered as a more static concept compared to sexual orientation, gender identity and expression.

Spectrums

Sexual orientation, gender identity, and expression are commonly understood not to be binary, but rather exist in a series of spectrums.¹²



Source: World Bank modeled after Center for Gender Santiy 2009

Where a person finds themselves on the various spectra can change over time and only taken together do they compose a person's sexual orientation, gender identity, and expression.¹³ Some surveys have asked respondents to place themselves on one or more of these spectrums, while others simply offer a binary set of options for each e.g. male/female or transgender yes/no. Sexual orientation, gender identity and expression, and sex characteristics also interact with each other. For instance, a transman, a transgender individual who identifies as a man¹⁴, can be attracted to men, women or both and thus might be straight, homosexual or bi (or might adopt some other term).

Being out

In addition, people often hide their sexual orientation, gender identity and expression, and/or sex characteristics wholly or in part (i.e., from certain people and/or in certain

¹² The following diagram is adopted from: Center for Gender Santiy. 2009. [„Sex and Gender diagram”](#).

¹³ There might be other dimensions which are relevant to a person's sexual orientation, gender identity and/or expression based on cultural difference or other circumstances that have not been taken into consideration for this paper.

¹⁴ National Center for Transgender Equality. 2014. [Transgender Terminology](#).

circumstances). This is known in some places as being ‘closeted’ (not revealing your status) - in contrast to being ‘out’. ‘Coming out’ is the process in which an individual discloses their sexual orientation, gender identity and expression, and/or sex characteristics within their social surroundings.¹⁵ Someone who is ‘out’ does not necessarily disclose their identity to everyone, everywhere, or all the time. Being out or not, and in which contexts, can affect the way in which a person is treated by their family, friends, employer, and society and thus has an impact on their lived experience and development outcomes. There are several ways of trying to measure this, including asking respondents to whom they have revealed their status and asking them how they think others perceive them.

2. Operationalizing the definitions

Defining the population that is the target of any data collection or knowledge generation effort is a necessary step in designing research projects but the operationalization of definitions can pose a significant challenge especially for hard to survey/hidden populations, such as LGBTI people. Researchers will need to decide which approach to adopt, there is no generalized right answer other than to be driven by the particular research questions, development problems, and context at hand.¹⁶ Following, we will use the example of two surveys to show potential was of addressing this challenge.

2.1. Asking about SOGI in surveys

[The 2012 EU LGBT Survey](#) by the European Union Agency for Fundamental Rights (FRA) recognized that “(...) the probability that a person identifies as LGBT (considering their sexual behavior or general preferences and/or identifying themselves as being gay/lesbian/bisexual) may vary across countries and social contexts, as well as change over time (during the life-course, for example).” (FRA, 2012) They developed the following questions to ascertain sexual orientation, covering identification [A4], behavior [A7] and then attraction [A8].¹⁷

A4. Would you say you are ...

PLEASE SELECT THE ONE ANSWER THAT FITS YOU THE BEST

1	Lesbian
2	Gay
3	Bisexual
4	Heterosexual/Straight
6	Other, please write here:
9	Don't know

Source: FRA 2012

¹⁵ Hiding sexual orientation and gender identity are not directly comparable forms of the ‘closet’ and need to be examined and researched on their own terms.

¹⁶ Tourangeau, Roger; Edwards, Brad; Johnson, Timothy P.; Wolter, Kirk M.; Bates, Nancy. 2014. Hard-to-Survey Populations.

¹⁷ FRA. 2012. [EU LGBT Survey Technical Report](#).

A7. In the last five years, you have had sex with:

We refer to sex as any behaviour which involves genital contact. We do not refer only to sexual intercourse involving penetration.

PLEASE SELECT THE ONE ANSWER THAT FITS YOU THE BEST

1. Only women (or with one woman)
2. Mainly women
3. Both men and women
4. Mainly men
5. Only men (or with one man)
6. No one
9. Don't know

A8. Whom are you sexually attracted to?

PLEASE SELECT THE ONE ANSWER THAT FITS YOU THE BEST

1. Males
2. Females
3. Both males and females
4. I am not sexually attracted to anyone
9. Don't know

Source: FRA 2012

A5. Do you agree or disagree with the following statements?

PLEASE RESPOND TO EACH LINE

	Strongly disagree	Disagree	Agree	Strongly agree	Don't know
A. I feel feminine	1	2	3	4	9
B. I feel masculine	1	2	3	4	9
C. I look feminine	1	2	3	4	9
D. I look masculine	1	2	3	4	9
E. I wish I was more feminine	1	2	3	4	9
F. I wish I was more masculine	1	2	3	4	9

Source: FRA 2012

In addition, respondents were asked whether they identify as transgender. If they answered yes, they were asked to select between Transgender; Transsexual; Woman with a transsexual past; Man with a transsexual past; Gender variant; Crossdresser; Queer; or Other (with the option to briefly explain the response). This was followed by a set of questions which addressed gender identity on a more subconscious level, allowing researchers to gain a more complete picture of a person's experience.

The second example comes from a survey conducted by the Williams Institute in cooperation with UNDP in Nepal.¹⁸ It uses a similar set of questions to the FRA survey to cover the different dimensions of SOGIE, however, it shows one significant variation as a result of taking into account the local context. In the question, 16.1 researchers decided to merge the two concepts, sexual orientation and gender identity into one question reflecting the local context. The other questions (16, 17 and 18) cover self-identity, attraction ("feelings") and sexual behavior similar to the FRA.

¹⁸ UNDP; Williams Institute. 2014. [Surveying Nepal's Sexual and Gender Minorities: An Inclusive Approach.](#)

16	There are many terms people use to identify their sexual orientation and gender identity. Now I am going to read some terms to you and I'd like to know to which degree do the following terms apply to you? For each term you must answer "not at all""somewhat" or "strongly" (circle answers)	Meti not at all somewhat strongly मेटी हुँ होइन केही हदसम्म हो धेरै नै हो	16.1	What is the primary term you use to refer to your sexual orientation/gender identity? १६.१ आफ्नो यौनिक अभिमुखीकरण/लैङ्गिक पहिचानलाई बुझाउन तपाईंले प्रयोग गर्ने मुख्य शब्द कतल हो ?	
		Ta not at all somewhat strongly टा हुँ होइन केही हदसम्म हो धेरै नै हो			
		Kothi not at all somewhat strongly कोथी हुँ होइन केही हदसम्म हो धेरै नै हो			
		Pantheni not at all somewhat strongly पन्थी हुँ होइन केही हदसम्म हो धेरै नै हो			
		Hijara not at all somewhat strongly हजिडा हुँ होइन केही हदसम्म हो धेरै नै हो			
		Fulumulu not at all somewhat strongly फुलुमुलु हुँ होइन केही हदसम्म हो धेरै नै हो			
		Transgender not at all somewhat strongly तसुरोलङ्गी हुँ होइन केही हदसम्म हो धेरै नै हो			
		Gay or lesbian not at all somewhat strongly पुरुष समलङ्गी वा महिला समलङ्गी हुँ होइन केही हदसम्म हो धेरै नै हो			
		Heterosexual or straight not at all somewhat strongly वपिरति लङ्गी वा स्ट्रेट हुँ होइन केही हदसम्म हो धेरै नै हो			
		Bisexual not at all somewhat strongly द्वलङ्गी हुँ होइन केही हदसम्म हो धेरै नै हो			
१६	आफ्नो यौनिक अभिमुखीकरण तथा लैङ्गिक पहिचानलाई चिनाउनका लागि मानसिहरले घरे शब्दहरू प्रयोग गर्दछन्। अब केही शब्दहरू पढ्छु र जान्नु चाहन्छु कि तपाईंको लागि कतल हदसम्म यी शब्दहरू लागू हुन्छन्। हरेक शब्दको लागि तपाईंले «हुँ होइन», «केही हदसम्म हो» वा «धेरै नै हो» भन्ने जवाफ रोज्नु पर्नेछ। (उत्तरमा गोली घेरा लगाउनुहोस्)	१७	17	People are different in their sexual attraction to other people. Which best describes your feelings? Are you attracted to (multiple answers possible) मानसिहरू यौनिकरूपमा अन्य मानसिहरूपरता फरक-फरक किसिमले आकर्षित हुन्छन्। तपाईंको भावनालाई सबभन्दा उत्कृष्ट तरिकाले कुन कुल बझाउँछ ? तपाईं यनीहरूपरता आकर्षित हुनुहुन्छ ? (एकभन्दा धेरै उत्तरहरू सम्भव छन्)	Metis Kothis Panthenis Hijaras Males Females Transgender men Transgender women Other
		१८	18	In the past 12 months, who have you had sex with? (multiple answers possible) वगत १२ महिनामा तपाईंले कोसँग यौन सम्पर्क राख्नुभयो ? (एकभन्दा धेरै उत्तरहरू सम्भव छन्)	Metis Kothis Panthenis Hijaras Males Females Transgender men Transgender women Other I have not had sex

Source: UNDP, Williams Institute 2014

Source: UNDP, Williams Institute 2014

Two very useful guides on how to survey for both sexual orientation ([SMART guide](#)) and gender identity ([GenIUSS group](#)) give an in-depth overview of how definitions might be operationalized and outline a number of challenges when surveying LGBTI communities. Both guides were developed for the U.S. and as such mainly use language that reflects sexual orientation, gender identity and expression in that context. Below, we give an overview of how surveys outside of the U.S. have dealt with these difficulties and what lessons can be learned.

3. Estimating the population size

The truth is that we do not know how many LGBTI people there are. Estimating the total number of LGBTI people who live in a given city, region or country is difficult, not only

because of the definitional issues raised above but also because LGBTI people are subject to stigma and marginalization in many countries and therefore often live in hiding.¹⁹

As with many hard to survey populations, it is challenging to tell whether LGBTI respondents are in a position to reveal the truth about these sensitive issues. Their trust to any outside entity, especially if it is a government sponsored survey, can be low and hence they are unlikely to disclose their sexual orientation, gender identity, and/or sex characteristics for fear of repercussions. Even in countries where LGBTI people face lower levels of discrimination they might not be willing to disclose such deeply personal information. Collecting data in a way that does not inadvertently 'out' people and being explicit to respondents about privacy measures to protect the confidentiality of participants and their data are critical in encouraging higher response rates and accurate answers.

It is important to measure how many lesbian, gay, bisexual, transgender and intersex persons live in any given location to fully understand the size of the development challenge and the magnitude of responses needed. In simple political terms, it can also be important for making the case that LGBTI people deserve a proportional slice of public and private resources.²⁰ Understanding the size (and other characteristics) of the total LGBTI population is further useful for establishing the ground upon which detailed surveys of a subset of the population can be done.

Data from the United States

Reproduction of the Question on Relationship to Householder From the 2010 Census

2. How is this person related to Person 1? Mark ONE box.

<input type="checkbox"/> Husband or wife	<input type="checkbox"/> Parent-in-law
<input type="checkbox"/> Biological son or daughter	<input type="checkbox"/> Son-in-law or daughter-in-law
<input type="checkbox"/> Adopted son or daughter	<input type="checkbox"/> Other relative
<input type="checkbox"/> Stepson or stepdaughter	<input type="checkbox"/> Roomer or boarder
<input type="checkbox"/> Brother or sister	<input type="checkbox"/> Housemate or roommate
<input type="checkbox"/> Father or mother	<input type="checkbox"/> Unmarried partner
<input type="checkbox"/> Grandchild	<input type="checkbox"/> Other nonrelative

Source: U.S. Census Bureau 2010

According to the [Gallup](#) for example, 4.3% of Americans self-identify as L, G, B, or T. This is an increase of 0.8% based on the last estimates from 2012.²¹ Washington DC has the highest percentage of LGBT-identifying residents at 8.6%, while in South Dakota only 2% self-identify as LGBT.²²

¹⁹ Tourangeau, Roger; Edwards, Brad; Johnson, Timothy P.; Wolter, Kirk M.; Bates, Nancy. 2014. Hard-to-Survey Populations.

²⁰ Of course, this can cut both ways in situations where the population is numerically small. This is not to undermine the human rights arguments that no matter the size of a marginalized population, their human rights need to be protected, respected and promoted.

²¹ Romero, Adam P. 2017. [1.1 Million LGBT Adults Are Married to Someone of the Same Sex at the Two-Year Anniversary of Obergefell v. Hodges.](#)

²² Gates, Gary J. 2017. [Vermont Leads States in LGBT Identification.](#)

In a study from 2011, the Williams Institute estimated there are roughly 8 million Americans who *self-identify* as lesbian, gay, or bisexual and 700,000 as transgender.²³ Almost twice as many, an estimate of 8.2% (19 million) of the US population, report that they have engaged in same-sex *behavior* and 25.6 million or 11% have acknowledged some sort of same-sex *attraction*.²⁴ This shows the sharp contrasts between LGBT *identity* (4.3%), *sexual behavior* (8.2%) and *sexual attraction* (11%) components within the above definition.

A [Google U.S. consumer survey](#) found that 5.7% of respondents from a nationwide sample of 10,000 identified as LGBTI.²⁵ It is particularly interesting that more than 10% of people aged 18-24 identified as LGBT compared to only 2-3% of people older than 45.

The 2010 U.S. Census asked questions on a person's household, which allowed the identification of some same-sex households. After accounting for reporting issues, around 650,000 same-sex households were counted or around 0.6% of all US households.²⁶ More recent data (2017) on same-sex marriages shows that there are nearly 1.1 million married LGBTI people in the US or over 547,000 same-sex couples.²⁷ This provides an alternate way to identify LGBTI people, yet it has significant limitations, including not accounting for single people and couples not sharing the same household.

Data from other countries

According to data from a [2016 online survey in 9 EU countries](#), 5.9% of Europeans self-identify as LGBT.²⁸ Germany has the highest percentage (7.4%), followed by Spain with 6.9% and the UK with 6.5%, while in Hungary only 1.5% self-identify as LGBTI.²⁹ An older [survey from the UK](#) (2010) suggests that roughly only 1.6% of the British population identifies as LGB with an additional 3% not able or willing to answer the question. A Norwegian survey from 2010, showed only 1.2% of the population identifies as lesbian, gay or bisexual.³⁰ A 2014 Australian survey came to the conclusion that 3% of men and 4% of women identified as non-heterosexual.³¹ A [survey from ten Brazilian cities](#) indicates that on average 7.8% of the surveyed men identified as gay and 2.6% as bisexual, totaling to 10.4%

²³ Gates, Gary J. 2011. [How many people are lesbian, gay, bisexual, and transgender?](#)

²⁴ Ibid.

²⁵ Only half of these reported being out at work.

²⁶ Lofquist, Daphne; Lugaila, Terry; O'Connell, Martin; Feliz, Sarah; 2010. [Households and Families: 2010](#).

²⁷ Romero, Adam P. 2017. [1.1 Million LGBT Adults Are Married to Someone of the Same Sex at the Two-Year Anniversary of Obergefell v. Hodges](#).

²⁸ Based on a census-representative survey of 11,754 people conducted in August 2016 by Dalia Research.

²⁹ Dalia Research. 2016. [Counting the LGBT population: 6% of Europeans identify as LGBT](#).

³⁰ Gulløy, Elisabeth; Normann, Tor Morten. 2010. [Sexual identity and living conditions](#).

³¹ Richters, Juliet; Altman, Dennis; Badcock, Paul B.; Smith, Anthony M. A.; de Visser, Richard O.; Grulich, Andrew E.; Rissel, Chris; Simpson, Judy. 2014. [Sexual identity, sexual attraction and sexual experience: the Second Australian Study of Health and Relationships](#).

of the urban population. Of the surveyed women 4.9% identified as lesbian and 1.4% as bisexual, resulting in a total of 6.3% lesbian and bisexual women. The National Socio-Economic Characterization Survey (CASEN) from Chile provides one of the few population estimates for transgender people (in this case specifically the gender category 'other'). When comparing the responses for 'sex assigned at birth' and 'gender identity', 2.7% of the Chilean population can be considered transgender.³²

Because of its medical nature, population estimates for intersex people are somewhat more reliable. Widely accepted estimation by Anne Fausto-Sterling in 2000, estimates that 1.7% of all live births are intersex.³³

These examples highlight the broad range of population estimates for LGBTI people in different countries and even within a given country estimations can vary significantly over time as the example from the UK shows. To fully understand the development outcomes of LGBTI people it is important for future research to address this knowledge gap.

Nepal's 2011 census

In 2011 the [Nepalese government included "third gender", along with male and female in its census](#). "Third gender" was thought to be a locally resonant umbrella term for sexual orientation, gender identity, and expression. Unfortunately, a number of challenges were encountered which prevented a reasonable estimation of the population size.³⁴

The biggest problem turned out to be definitional ambiguity. Neither respondents nor enumerators had sufficient information on what the term "third gender" meant or who was supposed to be covered under it. Further, there was insufficient training provided to enumerators on how to properly ask the question and respond to questions from citizens. While some of the challenges could have been avoided by providing a guidebook and/or better training there was also a bigger conceptual problem, revealed through a follow-up survey conducted by UNDP and the Williams Institute in 2014. As mentioned above, the respondents to the later survey – which allowed open-ended identification - used 21 terms to describe their sexual orientation and gender identity. Only 51.4 % self-identified as 'third gender', posing the question whether other respondents would have chosen to identify as 'third gender' if just that option was provided like it was the case in the census.³⁵

After the Nepalese census attempt, censuses in [India](#) and [Pakistan](#) have moved forward with similar approaches to include transgender populations. [The next census in the UK](#) will

³² Valfort, Marie-Anne. 2017. [LGBTI in OECD Countries: A Review](#). OECD Social, Employment and Migration Working Papers, No. 198.

³³ Fausto-Sterling, Anne. 2000. Sexing the Body.

³⁴ Park, Andrew. 2015. [An inclusive Approach to Surveys of Sexual and Gender Minorities](#), Report of Meeting.

³⁵ UNDP; Williams Institute. 2014. [Surveying Nepal's Sexual and Gender Minorities: An Inclusive Approach](#).

likely be the first one in a developed country asking questions about sexual orientation and gender identity. In [Uruguay](#), an effort is underway to conduct a census among transgender people.

4. Methodological overview: How developmental outcomes for LGBTI people have been measured through surveys

In the following section, we will discuss four surveys of LGBTI people that have attempted to assess development outcomes: the previously mentioned survey in Nepal, a survey in India; the FRA survey from Europe; and the largest survey of intersex people from Australia. For each survey, we will focus on the methodologies, constraints, and lessons.

Nepal

The Williams Institute and UNDP conducted one of the [most recent surveys on sexual orientation, gender identity and/or expression](#) in cooperation with a local Nepalese NGO, Blue Diamond Society.³⁶ Respondents were asked a set of questions, including open-ended questions, on their sexual orientation, gender identity and/or expression covering the self-identification, behavioral, and attraction dimensions of all three. A number of survey enumerators from the Blue Diamond Society were recruited, trained and ultimately able to reach 1,178 respondents.³⁷

Most of the respondents were recruited through local LGBTI networks based on drop-in centers the Blue Diamond Society manages throughout the country. This venue-based snowball convenience sample³⁸ survey was developed to fit local circumstances. The sampling strategy is a combination of three distinct recruiting methods: Venue-based, snowball, and convenience. Venue-based refers to the drop-in centers that were managed by Blue Diamond Society that served as a primary venue to recruit respondents. The survey furthermore employed a snowballing approach where an initial number of subgroup members serve as ‘seeds’ and recruit further respondents of the subgroup until the required sample size is reached. These ‘seeds’ were selected on a non-probability basis simply because they were conveniently accessible. All three methods are commonly used to survey hard to reach populations but they have important limitations.

³⁶ This included a feedback on the methodology, survey tool and data analysis.

³⁷ UNDP; Williams Institute. 2014. [Surveying Nepal’s Sexual and Gender Minorities: An Inclusive Approach](#).

³⁸ This sampling strategy is a combination of three distinct recruiting methods: Venue-based refers to the drop-in centers that were managed by Blue Diamond Society; the snowballing is a nonprobability approach to sample design where respondents help recruit new respondents; and lastly a convenient sample which in this case refers to the initial group of respondents that were interviewed because they could be easily accessed through Blue Diamond Societies existing networks.

In this case for example, the sampling strategy led to very few women being interviewed. Respondents report an average income that is almost twice as high as the national average in Nepal and they were also much better educated than the national population (18% had completed higher education compared to 10% nationwide). Further, only 14% undertook agricultural work³⁹ compared to 75% of the national population.⁴⁰ Leading to a sample that can hardly be considered representative.

There are several explanations for this misrepresentation. Most critically, the survey only covered 32 of 75 Nepali districts and was not able to reach the most remote areas. The highest response rates were achieved in Kathmandu, the capital, explaining why the sample is skewed toward the more educated and wealthy. Further, the sampling strategy was built around a community network which is heavily focused on men, contributing to the fact that the survey only reached very few women (87.6% of respondents were assigned male at birth, 12% female and 1% intersex).⁴¹

Methodological takeaways:

- Extensive consultation and training are useful when surveying a population that is being marginalized and excluded. In this case, it helped avoid the problems that the Nepalese government encountered when conducting their census. One of the main findings of the survey was the diversity of terms used in Nepal to describe sexual orientation, gender identity, and expression. Providing respondents with locally relevant options to questions about sexuality and gender identity was essential. Designing the survey in a way that allows such diversity needs to be balanced with a way to aggregate data and allow some sort of generalization during data analysis.
- Snowball surveys like this provide an opportunity to get an in-depth understanding of discrimination and exclusion in a community (even if extrapolation to the population at large is difficult) that would be impossible to cover in a census, given space limitations.
- Sampling remains a critical challenge. Many LGBTI people in more rural parts of the country were unable to participate in the survey like women who were only partially reached. It's also unlikely that the survey reached many sexual and gender minorities that hide their sexual orientation, gender identity and expression, and intersex characteristics since they would avoid the networks that were used to distribute the survey. Similar issues have been faced with other hard to survey populations; in Section 6 we provide an overview of some of the lessons learned from those populations.

³⁹ *ibid.*

⁴⁰ CIA. [The World Factbook](#).

⁴¹ UNDP; Williams Institute. 2014. [Surveying Nepal's Sexual and Gender Minorities: An Inclusive Approach](#).

While this survey is a milestone in large-scale LGBTI data collection, it illustrates a number of challenges that should be taken into consideration when developing future surveys.

India

In India, the World Bank developed a survey with the goal to explore the interconnection between sexual orientation, gender identity and development outcomes, specifically the relationship between discrimination and socioeconomic status.⁴²

The implementing partner, Amaltas, used a similar community-based approach as employed in Nepal to recruit respondents. Due to a lack of organizations with a rural membership and the overall difficulties of data collection in remote areas, the survey was carried out in urban areas only. Within the L, G, B and T communities⁴³ the survey administrators found it particularly difficult to reach bisexual and lesbian respondents. In India, most bisexuals tend to hide their sexual orientation, especially since many of them are married with families. They live out their bisexuality mainly through same-sex behavior, rather than identification. Lesbians, on the other hand, were hard to reach because of a paucity of lesbian-focused community-based organizations that could help carry out the survey.

The survey reached 943 respondents in eight Indian cities. The majority of the respondents were ‘men who have sex with men’ (MSM). In India, the term is often used as an umbrella term to describe gay identities, but also includes people who engage in same-sex behavior yet would not identify as gay men.⁴⁴ The survey also reached a high number of hijras⁴⁵ and a much smaller number of ‘female-born’, one local term for lesbians.

The results show high levels of discrimination in the daily lives of the respondents. They are excluded from key services such as banking, housing, and health care. Around one-third of the respondents did not have a bank account and thus could not access any financial services. Similarly, many reported discrimination in the housing market, 13% said they were unfairly denied accommodation and 15% claimed they were evicted from their homes because of their sexual orientation, gender identity and/or expression. When attempting to access health care facilities, 15% of the respondents were denied access and another 15%

⁴² Singh, Suneeta; Krishan, Anjali; Mishra, Chittaranjan; Ghai, Kavya. Experienced discrimination and its relationship with life chances and socio-economic status of sexual minorities in India.

⁴³ Intersex people were not part of the data collection.

⁴⁴ Singh, Suneeta; Krishan, Anjali; Mishra, Chittaranjan; Ghai, Kavya. Experienced discrimination and its relationship with life chances and socio-economic status of sexual minorities in India.

⁴⁵ Hijra is one South Asian term for male to female transgender identities; it does not cover all trans identities.

were removed from a facility. One-fourth had been mistreated in government medical facilities.⁴⁶

Many of the sampling issues that were faced in Nepal, were also faced in India but one additional issue becomes evident in this case:

- Given the constraints of the sampling method, only limited comparison to the general population is possible. Meaning that it is nearly impossible to conclude whether LGBTI people are worse off than non-LGBTI people. For example, the study found that one-third of the respondents did not have access to a bank account but in order to understand whether this a problem specific to the LGBTI community a comparable data set for the general population would be needed. One way to respond to this is to include questions on SOGIE status in larger population surveys and thus reach both LGBTI and non-LGBTI people at the same time. But given the likely small number of LGBT people within an overall population, a large sample (of LGBTI and non-LGBTI) people would be needed to pick up enough LGBTI people to be able to undertake a meaningful analysis of the results – especially of the component L,G,B,T,I parts. Further, problems around having respondents reveal themselves and truthfully respond will remain and perhaps be heightened if the survey is not run by trusted local LGBTI partners. Respondents need to be sure that their privacy and security is ensured.⁴⁷ Hence, it could be more effective to include non-LGBTI people in research projects specific to the LGBTI community, using the same sampling strategy and similar questionnaires to generate comparable data sets. This approach was recently tested by the World Bank in [a survey in Thailand](#) (results forthcoming in Fall 2017).

Europe

In 2012, the European Union Agency for Fundamental Rights (FRA) [collected data on sexual orientation, gender identity and expression](#) in 28 European countries.⁴⁸ The online survey reached 93,079 respondents in total. The overarching goal was to explore how LGBT⁴⁹ people experience fundamental rights.⁵⁰

FRA designed the survey instrument in 27 languages in close cooperation with a scientific advisory board, made up of representatives from civil society as well as local and

⁴⁶ Singh, Suneeta; Krishan, Anjali; Mishra, Chittaranjan; Ghai, Kavya. Experienced discrimination and its relationship with life chances and socio-economic status of sexual minorities in India.

⁴⁷ Ibid.

⁴⁸ At that time 27 EU member countries as well as Croatia, which officially joined in 2013, participated.

⁴⁹ Intersex people were not specifically targeted by this survey.

⁵⁰ FRA. 2012. [EU LGBT Survey Technical Report](#).

international experts. A consortium of Gallup Europe⁵¹ and International Lesbian Gay Bisexual Trans and Intersex Association (ILGA) Europe⁵² implemented the survey. Over a period of four months any person over 18 years⁵³ that identified as LGBT could participate in the anonymous online survey. The online questionnaire and sampling through self-identification provided a number of advantages such as easy distribution across 28 countries, a collection of a large data set and the development of a tool which could be readily applied in other contexts.⁵⁴

The questionnaire covered the following areas:

- Public perceptions and responses to homophobia and/or transphobia;
- Discrimination;
- Rights awareness;
- Safe environment;
- Violence and harassment;
- Social context of being an LGBT person; and
- Personal characteristics, including age and income group.

The survey discovered that despite anti-discrimination laws across the EU, one-third of the respondents felt discriminated against in at least one of the following areas: housing, health care, education, social services, and access to goods and services. Transgender respondents were even more vulnerable, with 35% reporting being attacked or threatened with violence in the previous 5 years. Together with bisexual women, they are also most likely to report income levels in the lowest quartile.⁵⁵

As with the other surveys, extrapolating from the results to the total LGBT population or non-LGBT people is difficult given the sampling method. Respondents were a segment of the population that had access to the Internet, and frequent LGBT friendly websites, events or other services where the survey was advertised. Overall, the vast majority of respondents were younger than 40, mostly gay and well educated. An unusually high number of students answered the survey.⁵⁶

FRA opted to use an online survey for two reasons: First, the necessary sample size for a representative sample in each country would have needed to be extremely large in order to achieve a robust analysis. They estimated that around 800,000 screening interviews would

⁵¹ Gallup is an international survey and consultancy company.

⁵² ILGA-Europe is the European umbrella organization of civil society organizations working on LGBTIQ issues.

⁵³ FRAs focus on respondents over 18 years is very common among these types of surveys. Unfortunately, this makes invisible the experiences of younger LGBTI+ people.

⁵⁴ FRA. 2012. [EU LGBT Survey Technical Report](#).

⁵⁵ Ibid.

⁵⁶ Ibid.

be needed in order to find 1,000 LGBT respondents in each member state. Second, FRA noted that LGBT people are likely to disguise their sexual orientation or gender identity in a screening interview with an unknown enumerator. Hence the self-identifying online survey was chosen, even though the sample would not allow a comparison with the larger population.⁵⁷

Anonymity was an important aspect of the study design, respondents were not asked for their names or any other personal information. No information about the computer itself was collected (for example through cookies). While the anonymity provided the respondents with security it had two major implications. Respondents had to take the whole survey in one go because there was no way of saving progress and continuing at a later stage. Furthermore, it is possible that respondents took the survey more than once since no data about them or their computer was stored to prevent multiple entries. These constraints are inherent of online surveys and should not diminish the overall potential these surveys have especially in areas where the internet is widely accessible. The involvement of local experts helped in reaching out to large numbers of LGBT people and facilitated the distribution of the survey. The implementing consortium of Gallup Europe and ILGA Europe prepared background notes on the local LGBT communities to target advertisement accordingly.⁵⁸

Australia

In 2015, Australia conducted a survey of 272 intersex people, one of the largest surveys of its kind.⁵⁹ The survey was structured to be on, with and for the intersex community, as one means of redressing the historical role of intersex people as subjects of (often medical) inquiry. An online data collection tool was used with both forced-choice (quantitative) and open-ended (qualitative) questions. The survey was circulated through online networks, social media, and general advertising. The survey asked participants what term participants preferred to use to describe themselves to themselves, to family and friends, and when seeking medical assistance. The survey revealed high levels of mental health issues among respondents (60% had thought about suicide and 19% tried it; compared to less than 3% - combined - for the Australian population at large). The intersex respondents also reported having a much higher level of none-completion of secondary school (18%) compared to the general Australian population (2%). Further, a 41% of respondents earned less than \$20,000 per year (minimum wage is \$34,000).⁶⁰

⁵⁷ Ibid.

⁵⁸ Ibid.

⁵⁹ Jones, Tiffany; Hart, Bonnie; Carpenter, Morgan; Ansara, Gavi; Leonard, William; Lucke, Jayne. 2016. [Intersex: Stories and Statistics from Australia](#).

⁶⁰ Ebd.

5. Lessons learned from other ‘hard to survey’ populations

Sexual orientation, gender identity and/or expression, are complex, deeply internal processes.⁶¹ Surveys assume that the respondents ‘know’ the information that is being asked about and is willing to reveal it. For sexual orientation, gender identity and expression this will often not be the case. That said, there are other identities that are highly personal (e.g. faith) and in many contexts, these are also subject to stigma and discrimination and are therefore sensitive questions that could be potentially dangerous to disclose. Many of the above-described methodological problems in reaching LGBTI communities and comparing their experiences to those of the non-LGBTI population are familiar to researchers looking at other ‘hard to survey’ populations – such as victims of domestic violence, drug users or sex workers.

Past studies and surveys of sexual and gender minorities have built on experiences with other hard to reach populations. What is often needed is a set of culturally sensitive questions that allow respondents to articulate their own self-identification. This, of course, makes meaningful quantitative analysis more complicated. The most challenging part of data collection for any of these groups is defining the target population and sampling them. The Nepal and India surveys used a venue-based snowball convenience sampling method, in which key actors of the community distributed the surveys and each respondent was encouraged to do the same. This method has been used for communities where other standard probability sampling methods are problematic. Respondent-driven sampling,⁶² another form of chain-referral sampling, takes into account the nature of the referrals and has been used in some cases to correct for bias in the overall network. Online surveys, such as the FRA, are another approach with its own challenges.

Surveys of vulnerable groups require a high amount of privacy and trust in order to overcome barriers to non-response, inaccuracy, and access. Many LGBTI people face violence as result of their sexual orientation, non-conforming gender identity and expression, and/or sex characteristics similar to many victims of gender-based-violence. Analogous to surveys on sexual and gender-based violence, protocols, and training programs need to be developed to ensure that respondents feel safe in answering surveys, are not put at the greater physical, emotional or psychological risk by doing so and can be appropriately referred to counseling and welfare services as needed. This will also lead to better data as people are more likely to disclose their true experiences when they feel safe.

⁶¹ In this sense, Intersex might be viewed differently since it is a relatively clear defined medical term and while Intersex people might face similar discrimination and exclusion they also have unique subset of challenges that LGBT people do not face.

⁶² Similar methodology to snowballing. A mathematical model of who recruitment process is than used to weight the sample. [More information](#)

Research also frequently faces issues around motivation – not just for LGBTI respondents but for a useful comparison group. If respondents’ do not see the benefits of taking the time to respond, researchers need to apply different techniques to engage. Community involvement in all stages of the research can, therefore, be an important tool to reach the desired number of respondents.

6. The potential of big data in building LGBTI knowledge

Surveys are only one tool to collect data on LGBTI people, in recent years so-called ‘big data’ has become an important source for data. The information that every user is providing by surfing the internet and using digital services can be a powerful tool to better understand any given population.

Social media platforms like Facebook, for example, give users the opportunity self-identify as LGBTI. According to the companies own [research from 2015](#) more than 6 million American’s have come out on Facebook. While this estimation is based on self-identification Facebook, like many other companies, collects much more data about their users. Peoples likes, the places they frequently visit and the post they write could all be used to develop a much more accurate estimation of how many LGBTI users the social network has.

According to an [article in the New York Times](#) from December 2013, 5% of all [Google porn searches](#) in the US is for gay porn. Looking at these (gay) porn searches around the world, Google data shows they range from 6% of all porn searches in Central America, South America, Europe, Australia, Pacific Islands and the Caribbean to 2.5% in Africa, Eastern Europe, and the Middle East. This data is useful, for instance in showing the simple but often contested fact that LGBT people exist around the world. Yet it does not allow much more substantial conclusions to be drawn, as for example, women search for gay porn as well.

Definitional issues remain a problem for big data like for any other source of data. Is Facebook’s self-identification enough or should other user data be used to identify those who do not publicly state they are LGBTI? Is someone LGBTI just because they search for gay or lesbian porn? While the underlying issues are similar, ‘big data’ does face its own unique challenges when used to better understand the LGBTI population. Nevertheless, it provides important additional sources of data, especially as Internet access becomes more widely available.⁶³

⁶³ One innovative new technique is random domain intercept technology whereby Internet users who incorrectly type in a domain name, are directed to a survey. In 2015, this technique was used by [ILGA Global for the first global online survey](#) of attitudes toward LGBTI people, which attracted over 96,000 responses and will be repeated each year.

7. Experimental approaches to measure discrimination and exclusion

So-called ‘audit studies’, provide an alternative approach to gathering evidence on the development outcomes, discrimination, and exclusion of LGBTI people.⁶⁴ Experimental methods can overcome some of the challenges noted above yet face their own challenges in terms of generalizability and the general downsides of experiments. Such methods have been used to highlight disparities on the basis of race, ethnicity, and gender, as well as for sexual orientation, gender identity, and expression.

Experimental methods have been used to detect discrimination in employment opportunities and access to housing for LGBTI people.⁶⁵ These experiments have for example focused on the job application process and revealed significant bias against openly lesbian and gay applicants, resulting in between a 5% to 40% lower callback rate for interviews, as well as lower wage offers. One [2011 study](#) focused on gay male job seekers and their chances of being invited to an interview. Two fictitious counterfeit resumes were constructed to show applicants with the same qualifications and sent to over 1,700 job postings in the United States. One of those resumes was randomly assigned leadership experience in a gay university campus organization, with the ‘control’ resume listed leadership experience in a non-gay organization.⁶⁶ The study found that “While heterosexual applicants had an 11.5% chance of being invited for an interview, equally qualified gay applicants only had a 7.2% chance of receiving a positive response” ([Tilcsik, 2011](#)).

A similar study was conducted in the United Kingdom. Over 140 ‘applicants’ were in correspondence with 5,549 companies. This study focused on the job prospects of gay men and lesbians. Gay men and lesbians had a 5% lower probability of being invited to an interview. Further, the companies that they received call backs from were paying on average less than those that invited heterosexual applicants.⁶⁷ In Greece, a study unveiled not only a 26% lower chance for a call back for gay and lesbian applicants but also 1.5% lower initial wage offer.⁶⁸

⁶⁴ Pager, Devah; Shepherd, Hana. 2008. [The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets](#).

⁶⁵ Studies for other minority groups have applied similar models for housing, credit and school applications.

⁶⁶ Tilcsik, András. 2011. [Pride and Prejudice: Employment Discrimination against Openly Gay Men in the United States](#).

⁶⁷ Drydakis, Nick. 2014. [Sexual Orientation Discrimination in the United Kingdom’s Labour Market](#).

⁶⁸ Drydakis, Nick. 2009. [Sexual Orientation Discrimination in the Labour Market](#), [Labour Economics](#) and Drydakis, Nick. 2011. [Women’s Sexual Orientation and Labor Market Outcomes in Greece](#).

8. Conclusion

Without greater certainty about the development status of LGBTI people and a rigorous comparison of their welfare to the population at large, it is more difficult to make the case for effective policies and programs – and to know whether any interventions are effective.

Rigorous, quantitative data on LGBTI people are scarce, even in developed countries, and even on the most basic issues – such as the size of the population. Surveying LGBTI people has challenges, not least defining the population and the unwillingness of many LGBTI people to reveal themselves to interviewers due to stigma. Lessons can be learned from experience with other ‘hard-to-survey’ populations, where issues of definition, stigma, and access also apply. Through consultation, cultural sensitivity and training it is possible to collect good quality data.

Probability sampling presents a particularly challenging issue. As the LGBTI population is relatively small, a general population survey would have to be very large (and costly) to pick up large enough samples of the L, G, B, T and I subpopulations (even if the issues with getting people to reveal their identities can be overcome). Experimental studies and big data can help triangulate information on some issues and address some of the constraints of population surveys.

We encourage survey experts, data analysts and LGBTI development professionals and the LGBTI community to work together to collect better data and keep focus particularly on issues of definition, sensitive interview protocols and collecting data that can be readily compared to non-LGBTI populations.