DISCUSSION PAPER

data2x

Intersectionality and SOGIESC Data: Opportunities and Challenges

April 2024

Prepared for Data2x by
M. V. Lee Badgett
Koppa LGBTI+ Economic Power Lab

Introduction: LGBTI people and Intersectionality

The recognition that individuals might have multiple marginalized identities that interact to create vulnerabilities is at the heart of a growing interest in intersectionality within the economic development field (Badiee, Borrowman, and Buvinic, 2024). Early discussions about intersectionality often focused on the varying experiences of women whose racial or ethnic identities placed them in multiply marginalized positions. A broader view of intersectionality now pulls in more identities, including age, disability, migration status, and others, which have increased the demands on data collection and analysis in the development context (Badiee and Buvinic, 2024). This discussion paper discusses some opportunities and challenges of generating data on an often-overlooked set of identities: sexual orientation, gender identity and expression, and variations in sex characteristics, commonly abbreviated as SOGIESC.

A global human rights movement led by lesbian, gay, bisexual, transgender, and intersex (LGBTI) people has drawn on many sources of data to document the violence, stigma, and discrimination that drive the development challenges faced by those with minority SOGIESC identities.[1] According to existing research, LGBTI people experience discrimination, poverty, poorer mental and physical health, food insecurity, violence, school bullying, family rejection, and other challenges, particularly in developing countries (Badgett 2020; Park and Mendos 2019; O'Malley and Holzinger 2018). The COVID pandemic both revealed and exacerbated these gaps in countries where research exists (Herman and O'Neill 2020; Outright International 2020; Pereira et al. 2021). This body of research makes a strong case for seeing SOGIESC as an important determinant of development outcomes that should be included as an analytical dimension of intersectionality.

[1] Much of the data used in the LGBTI human rights context might be considered citizen-generated data, also an important source for studying intersectionality (Badiee and Buvinic 2024).

Although the existing studies have been important, many are based on limited and unrepresentative samples, so they are not sufficient to reveal inequalities and development needs in every country or to understand development intervention impacts for LGBTI people. Alongside citizen-generated data, it is equally important to collect data through more official and systematic channels, expanding both the quantity and the quality of such data, as well as its usefulness for making comparisons within and between countries. Accordingly, this paper focuses on SOGIESC data collected by national statistics offices (NSOs), multi-national development data programs, and civil registration and vital statistics data. After showing why SOGIESC identities matter in the context of intersectionality, the bulk of this paper then presents an overview of the status of data collection by national statistics offices in low- and middle-income countries and by other development data programs, along with current knowledge about best practices for such data collection.[2] The intention is to provide some starting points for future discussions and collaborative efforts to generate a greater quantity and quality of SOGIESC data.

Why SOGIESC identities matter for intersectional development data

Much of the more academic research on LGBTI people focuses on analyzing whether differences exist for LGBTI people compared to non-LGBTI people, using this single dimension of comparison. However, SOGIESC identities intersect with other identities in at least two ways that are crucial for creating and analyzing data for development purposes. First, most LGBTI people have other identities that place them in lower positions on hierarchical social or economic orderings. For example, the economic and social status of lesbian, bisexual, and transgender women will be shaped by the fact that they are women. Their gender identity as women exposes them to barriers created by gendered norms, laws, and social treatment in addition to any disadvantages related to being lesbian, bisexual, or transgender. Similarly, LGBTI people from marginalized castes, ethnicities, or racial groups will likely experience disadvantages when compared to heterosexual cisgender people with the same caste/ethnic/racial identity.

Second, analyses of the influence of categories such as gender or caste typically contrast the position of a privileged group, such as men or people of high-status castes, with that of lower status categories while leaving out the possibility of counterbalancing influences from other social identities. A gay white man or gay man of a favored caste position might not experience the same degree of expected privilege as others of his race or caste if he is rejected by his family or is shut out of economic opportunities because of his sexual orientation. Some research shows that white gay men might have economic outcomes more like those of people in more marginalized race or ethnic groups, for example (Badgett et al., n.d.).

Generally, many analysts assume that intersecting layers of marginalization reduce the opportunities and economic outcomes of individuals by creating a double or triple disadvantage. Some examples from the United States, which has intersectionality-friendly data, show development-related outcomes that take into account gender, race, sexual orientation, and gender identity (intersex data is not available). As expected, across different racial or ethnic groups, being LGBT increases the risk of poverty and reduces earnings compared with heterosexual cisgender people of the same racial-ethnic-gender group (Badgett, Choi, and Wilson 2019; Douglas and Steinberger 2015; Del Río and Alonso-Villar 2019).

[2] Other parts of the "Data Value Chain" that focus on usage of the data for policy analysis and impact monitoring are not addressed in depth even though they have begun to be discussed in the research community (Guyan 2022; Daly, Crehan, and Grzywnowicz 2022).

In some cases, though, SOGIESC identities might offset the effects of other marginalized identities, such as race or gender. For example, white female lesbians appear to have unobserved characteristics related to stronger labor market attachment that elevate their earnings over those of white heterosexual women in many countries, but that pattern is not present for comparisons of lesbians in other racial/ethnic groups to their heterosexual counterparts in the U.S. (Douglas and Steinberger 2015). An experimental study found that black gay men are more likely to get job offers than either black heterosexual men or white gay men, perhaps because of offsetting stereotypes (the gay stereotype of weakness and effeminacy offsetting the stereotype of the threatening black man) (Pedulla 2014). The potential complexity of intersectional analyses shows the importance of intersectional theorizing along with other parts of the data value chain, especially careful analysis of data (Badiee and Buvinic 2024).

Overall, data on the SOGIESC identities of individuals serve two important needs when incorporating intersectionality. First, data on LGBTI people make visible the challenges that they face, including unemployment, poorer health, employment discrimination, and high poverty rates. Second, having information on the SOGIESC identities of survey respondents, alongside data on gender, disability, caste, ethnicity, race, and other important characteristics, makes possible a better understanding of the interaction of those characteristics. The first reason—and possibly both reasons—are behind efforts at the World Bank, Interamerican Development Bank, the United Nations, the European Union, and the Asian Development Bank to collect new data on LGBTI people through national surveys and other kinds of data collection. In addition, the UN Development Programme's LGBTI Inclusion Index intends to include data-based measures of the lived experience of LGBTI people in every country alongside legal measures of inclusion, creating a greater demand for SOGIESC data globally.[3] The remainder of this discussion paper explores how best to achieve the goals of statistical inclusion and intersectionality with examples of existing survey measures of SOGIESC, and good practices for addressing the challenges in collecting SOGIESC data.

SOGIESC and LGBTI concepts

To discuss the collection of SOGIESC data, we first need some clarity about what these concepts mean. Actually, at least three distinct concepts are embedded in the SOGIESC acronym, each of which would be associated with a measure on development-oriented surveys: sexual orientation, gender identity, and sex characteristics. A recent report by the National Academies of Sciences, Engineering, and Medicine (NASEM) in the United States evaluated measures of SOGIESC used on surveys in the U.S. and other countries and made recommendations about specific measures to use (National Academies of Sciences Engineering and Medicine 2022). This section draws on some of the concepts and findings presented in that report.

As a concept, sexual orientation is generally considered to have at least three dimensions. The first is **attraction**, or the sex or gender of the people that an individual is romantically or sexually attracted to. Having an attraction to or romantic interest in people of the same sex might also be related to the second dimension of **sexual behavior**, that is having sexual partners with the same or different gender identities. The third dimension is **identity**—thinking of oneself as heterosexual, gay, lesbian, bisexual, or some other identity related to sexuality. (Note that being transgender or intersex is not a sexual orientation option.)

[3] Furthermore, statistical inclusion of SOGIESC measures is one of the indicators of national inclusion in the LGBTI Inclusion Index (Badgett and Sell 2018).

Sexual attraction to people of the same sex and/or behavior with same-sex partners sometimes leads people to embrace a personal identity as lesbian, gay, or bisexual.[4] However, research has shown that these different dimensions are not always aligned. Some people with same-sex attraction might not have same-sex sex partners. Those with same-sex partners might not use those gay, lesbian, or bisexual identity terms, or they might have a different self-identity related to sexuality. A common example is that HIV key populations include a behavioral categorization—MSM—for men who have sex with men regardless of their sexual orientation identity.

Before defining gender identity and sex characteristics, it is helpful to clarify the distinction between **sex and gender**, two concepts that are often used on surveys as if they were identical. The NASEM report (among many other sources) defines sex as a biological concept "based on a cluster of anatomical and physiological traits that include external genitalia, secondary sex characteristics, gonads, chromosomes, and hormones" that is usually assumed to be binary, that is male or female. In contrast, gender is a socially constructed concept that links three general elements:

> ...**Gender identity**, which is a core element of a person's individual identity; **gender expression**, which is how a person signals their gender to others through their behavior and appearance (such as hair style and clothing) and cultural expectations about social status, characteristics, and behavior that are associated with sex traits (NASEM, p 1-3).

Seeing the distinctions between sex and gender allows for a clearer understanding of transgender people and people with intersex traits.

As noted in the NASEM report, **gender identity** is defined as an individual's core sense of self as a man or woman, or neither or both. Within this definition, transgender people may be broadly defined as those whose gender identity does not match the sex they were assigned at birth. Transgender people defined in that way might identify with a binary gender identity category or with a non-binary gender identity.

Sex characteristics are the anatomical and physiological traits that are used to assign sex at birth. People with variations in characteristics are those whose sex characteristics or traits are not aligned with those of only one sex. Such individuals might identify as being intersex or as having differences in sex development (DSD).

As might be inferred from the concept definitions, each of them could require more than one measure if used on a survey. In practice, choices about which dimensions to measure are often related to the purpose of the survey (Sexual Minority Assessment Research Team 2009). For example, a survey about health and HIV might include questions on sexual behavior. A survey about education and young people might ask about attraction, since young people might not have had sexual partners or have developed sexual orientation identities. A survey about employment experiences and income might use an identity measure to capture a characteristic that is more likely to be disclosed in the workplace.

[4] Other kinds of sexual orientation identities exist but are not explicitly included in the measures recommended by the NASEM report, such as queer or pansexual. Those identity terms are often used by people whose sexual attractions encompass people with a range of gender identities, not just the two binary gender identities that are generally embedded in the lesbian, gay, and bisexual identity context.

Measures of SOGIESC concepts recommended by NASEM

Those concepts have led to the development of many different measures of SOGIESC. Enough different measures now exist in some contexts (mostly North America and Europe) to require identifying the **best** measures of sexual orientation, gender identity, and sex characteristics for surveys, administrative forms, and other uses. This task has been the subject of several studies, government reports, and expert panels.[5] The NASEM report focused on measures of identity for both sexual orientation and gender identity, and this section summarizes their recommendations and rationale. Although the NASEM study was commissioned to focus on US data, they studied the experiences of other countries in using similar questions as well.

The NASEM study's recommendations about survey measures for these concepts was guided by several important principles related to data collection:[6]

- People deserve to count and be counted (inclusiveness).
- Use precise terminology that reflects the constructs of interest (precision).
- Respect identity and autonomy (autonomy).
- Collect only necessary data (parsimony).
- Use data in a manner that benefits respondents and respects their privacy and confidentiality (privacy).

The NASEM panel also carefully considered other methodological criteria and research, such as comprehensibility and having been tested in the general population and among LGBTI people. Based on these on different criteria, the NASEM report made recommendations about survey questions related to sexual orientation identity (one question) and gender identity (a two-step set of questions) for use in survey research, nonsurvey research, and administrative systems.

NASEM-Recommended sexual orientation question

Which of the following best represents how you think of yourself? [Select ONE]:

☐ *Lesbian or gay*
☐ *Straight, that is, not gay or lesbian*
☐ *Bisexual*
☐ *[If respondent is American Indian or Alaska Native:] Two-Spirit*
☐ *I use a different term [free-text]*
☐ *(Don't know)*
☐ *(Prefer not to answer)*

[5] Some examples include (Sexual Minority Assessment Research Team 2009; Gender Identity in U.S. Surveillance Group (GenIUSS) 2014; National Academies of Science, Engineering 2020; European Commission High Level Group on Non-Discrimination Equality and DIversity 2023).

[6] These bullet points are headers that were pulled verbatim from the report (National Academies of Sciences Engineering and Medicine 2022).

NASEM-Recommended gender identity two-step question

Q1: What sex were you assigned at birth, on your original birth certificate?

- ☐ *Female*
- ☐ *Male*
- ☐ *(Don't know)*
- ☐ *(Prefer not to answer)*

Q2: What is your current gender? [Mark only one]

- ☐ *Female*
- ☐ *Male*
- ☐ *Transgender*
- ☐ *[If respondent is AIAN:] Two-Spirit*
- ☐ *I use a different term: [free text]*
- ☐ *(Don't know)*

The NASEM report did not make specific recommendations of questions on intersex status or nonbinary gender identities given the limited amount of research on those measures. However, for intersex status, the report recommended that any such question be an additional stand-alone question and that "intersex" should not simply be added as a response option to sex questions. Their preferred stand-alone question that has been tested to some extent is the following:

Have you ever been diagnosed by a medical doctor or other health professional with an intersex condition or a difference of sex development (DSD) or were you born with (or developed naturally in puberty) genitals, reproductive organs, or chromosomal patterns that do not fit standard definitions of male or female?

- ☐ *Yes*
- ☐ *No*
- ☐ *(Don't know)*
- ☐ *(Prefer not to answer)*

Experience with SOGIESC survey data in developing countries

Given the experience with using similar questions in the US and other high-income countries (UK, Canada, and New Zealand), these recommendations are likely to be useful starting points for consideration by NSOs in other countries. This section discusses the developing countries that are known to have used very similar measures in NSO surveys and other administrative datasets, which will be a rich source for future analysis and methodological research.

Table 1 lists surveys from NSOs or international development agencies that are known to include questions that measure some component of sexual orientation or gender identity, and intersex

status.[7] This list is based on reports from NSOs, news media reporting, and empirical studies in the academic literature.

- ***Demographic and Health Surveys (DHS):*** Six countries have included questions on same-sex sexual behaviors of men at least once (Dominican Republic, Philippines, Cambodia, Bangladesh, Myanmar, Colombia); South Africa has allowed reporting of same-sex couples; India, Nepal, and Pakistan have included an "other" gender category. In a March, 2023, webinar, officials from the DHS program reported that they are working on recommendations related to SOGI questions on the DHS.
- ***National surveys or censuses:*** NSOs have included one or more SOGIESC questions in several different ways. Several countries include questions that allow identification of same-sex partners. Four countries in South Asia (Bangladesh, India, Nepal, and Pakistan) include a third gender option on their census gender question or on labor force surveys. Several Latin American countries have added a wider range of SOGIESC questions to existing population surveys.
- ***Surveys of nonprobability samples:*** NSOs in Peru, Ecuador, Mexico, and Colombia have undertaken separate surveys of nonprobability samples of the LGBTI population. NSOs in Mexico and Colombia conducted the nonprobability surveys to supplement surveys of probability samples.
- ***Surveys identifying same-sex couples:*** In addition to countries including direct questions about sexual orientation and gender identity on surveys, other surveys allow the identification of people in same-sex relationships, who are presumed to be lesbian, gay, or bisexual (and might also be transgender) and who can be compared to people in different-sex relationships in research using the same datasets. Of course, research using same-sex couples leaves out single people and does not allow measurement of a transgender or intersex gender identity, but studies often use the couple data where better data are not available (Badgett et al., n.d.). That research has provided insights into the experiences and economic outcomes of at least part of the LGBTI community.

We know little about why these countries added SOGIESC questions. One recent study analyzed the decision by NSOs in Argentina, Canada, and the UK to add gender identity measures to their censuses. That study highlighted the importance of positive social norms related to gender identity, political pressure, the legal recognition of gender identity, champions, and NSO resources (Garland et al. 2023), and Kevin Guyan made similar points about censuses within the UK (Guyan 2022). Future research might conduct similar case studies about the factors influencing the adoption of SOGIESC measures in the countries found in Table 1.

It is also useful to look at development-related surveys that do not included SOGIESC questions. Another common source of development data is UNICEF's Multiple Indicator Cluster Surveys (MICS). An inspection of the MICS6 and MICS7 questionnaires at the household level and the base questionnaires for men and women reveal a binary sex question and no questions on gender identity or sexual orientation. In the Reproductive Care module, there is a Sexual Activity section but no questions in the module ask about the sex of sexual partners. However, the Safety and Discrimination module includes a set of questions about whether the respondent has faced discrimination or harassment int the last 12 months (MICS6) or three years (MICS7), including one question about whether the discrimination was based on sexual orientation (MICS6) or, for

[7] Mexico's National Survey on Sexual and Gender Diversity is the only one with a question about intersex status: "Were you born with any variation in your body regarding your sex, such as genitalia, hormonal levels or other?"

MICS7,"Your sexual orientation or gender identity, such as being attracted to a person of the same sex, self-identifying differently from sex assigned at birth or as being sexually, bodily, or gender diverse?" Given the reference to an instance of discrimination and the response options (yes, no, don't know), the question does not allow for inferences about a respondent's sexual orientation or gender identity.

Civil registration and vital statistics systems (CRVS) and SOGIESC measures

Civil registration and vital statistics systems (CRVS) in some developing countries have begun to adapt to the needs of LGBTI individuals, although detailed research on this topic remains scarce. The clearest systemic adaptations to date involve gender markers, which we can track with the ILGA World database, and the implied changes to marriage registries.[8] These examples, described further in this section, show that CRVS systems can be made more inclusive of gender minorities, in particular. An alternative approach to changing gender markers that has been suggested would be to remove them from identification documents altogether. However, as discussed later in this section, that approach is in tension with development strategies that focus on making sure women have IDs that allow them to fully participate in many important settings.

According to ILGA World's database, 67 countries allow changes in gender markers on official documents, such as birth certificates, civil registries, or identification documents, although such processes are not without controversy among LGBTI civil society. Individuals must meet certain country-specific requirements (such as surgery or other medical requirements) in order to change their gender marker in most of those countries. Only 18 countries clearly allow self-identification as the only requirement for changing a gender marker.

Another concern is that the ability to change a gender marker is often limited to choosing from binary options. There are only 18 countries, including some developing nations like Argentina, Bangladesh, Brazil, India, Kenya, Nepal, and Pakistan, that offer nonbinary gender marker options in at least some situations, and sometimes that is the only option for all transgender people seeking a gender marker change. It is important to note that some transgender or intersex people may prefer to opt for one of the binary gender categories over a nonbinary marker. As a result of the existing rules, neither the counts of people changing their gender marker nor the counts of those using a nonbinary gender marker will provide an accurate count of transgender and intersex people in most countries.

Some of these complexities extend to legal recognition of same-sex marriages and civil partnerships. As of 2023, 34 UN member countries permit same-sex marriage and 36 allow civil partnerships (and some countries allow both). Some of the marriage equality countries are developing countries: Argentina, Brazil, Colombia, Costa Rica, Cuba, Ecuador, and South Africa. Although it appears that civil registration systems must have adapted to having two people of the same (binary) genders marrying, the extent to which civil registration systems accommodate nonbinary markers for marriage purposes remains unclear, warranting further investigation.

These developments underscore the evolving nature of CRVS systems in addressing the diverse needs of LGBTI individuals, yet also highlight ongoing gaps and opportunities for enhanced inclusivity and data representation. The incremental strategy of expanding gender markers and

[8] ILGA World Database, https://database.ilga.org/en, last accessed April 3, 2024.

making it easier to change gender markers is one path to greater inclusion. Another path that has been suggested is eliminating gender markers in identity documents, making changes unnecessary and preventing situations where transgender people may face discrimination because of a mismatch between gender identity and gender marker. However, some questions have arisen about whether there are competing development objectives that require gender markers.[9]

This question contrasts two potentially competing needs and suggests a tension between two gender equality goals. First, development officials have identified the need to include a gender marker in civil registration and identification in order to track progress toward greater levels of registration for women. Having an official identification document facilitates access to social entitlements, the ability to travel, and to access to development-related resources, for example, so women's civil registration is essential. Having a gender marker on the document is essential for accessing gender-segregated health or other services. Second, for transgender and intersex people, gender markers on formal documents can create barriers to getting access to resources if someone's gender identity or gender expression are not consistent with their official gender marker (Lebbos, Esquivel-Korsiak, and Clark 2021). A potential trade-off could exist, since the removal of gender markers might help equalize the playing field for transgender and intersex people, but it might also make inclusion of cisgender (that is, non-transgender) women more difficult.

In this context, it is helpful to note that foundational principles in the ID-for-Development efforts explicitly acknowledge that inclusion and nondiscrimination are priorities for all identification systems:

> All identification systems should be free from discrimination in policy, in practice, and by design. This includes ensuring that legal frameworks; requirements and procedures to register, obtain, or use identification; and the data that are collected or displayed on credentials do not enable or reinforce discrimination against particular groups, such as those who may face increased risks of exclusion for cultural, political, economic or other reasons. (World Bank and signatories 2022, p 12).

The groups singled out for protection from discrimination in these principles clearly include sexual and gender minorities, as well as women.

In practice the potential trade-off between seemingly competing goals (inclusion of women and gender minorities) seems less stark. LGBTI civil society advocates remain willing to pursue many avenues to inclusion, and a primary strategy continues to be the incremental inclusion approach of allowing gender marker changes and providing a nonbinary option. Those changes would be consistent with both goals, assuming that transgender and intersex people have the knowledge, resources, and autonomy to choose whichever category that describes their gender identity. Concerns about privacy and disclosure would remain for those changing their marker, however.

The discussion about the need for gender markers has also revealed some gaps in the rationale for those markers. Some stakeholders, including the NASEM report authors, have questioned the need for gender markers as being necessary for all administrative documents, and often the purpose of including sex or gender markers is not clearly articulated (National Academies of Sciences Engineering and Medicine 2022). The NASEM report also points out that for administrative data, a distinction can be made between data for statistical purposes and data for personal identification. An partial way out of the development dilemma posed in this section was implied by the NASEM

[9] The discussion of this tension draws on a presentation by Alan Gelb at the Center for Global Development on January 25, 2024, and his paper in progress on this topic. He also points out many of the subtleties of identification systems that are important to consider in this discussion.

authors for situations where the primary reason offered for having a gender marker is to collect data. For an important document such as a birth certificate (or perhaps for an ID card), the registration form might collect data on items that will be included on the document, such as date of birth and parents' names, while other items collected "below the line" on the registration form would only be used for statistical purposes and would not appear on a document. For identity documents, putting gender "below the line" would provide needed data to assess gender inclusion while not putting transgender and intersex people at risk.

Clearly, additional research that includes and goes beyond gender markers is needed to get a better understanding of how CRVS systems can become more LGBTI inclusive.

Good practices to move SOGIESC data forward in NSOs and elsewhere

Overall, given the small number of countries collecting SOGIESC data in a limited number of surveys or censuses, there are clearly some gaps in knowledge and some barriers to address. NSOs, researchers, LGBTI civil society, development agencies, and other data stakeholders should collaborate to identify opportunities to design, test, and implement new SOGIESC measures on population surveys. This section describes some common issues and strategies that have emerged so far.

Consultation with LGBTI civil society: One of the ways to address each of the concerns outlined in this section is consultation and active collaboration with LGBTI civil society. They are the experts in LGBTI people's lives and can offer crucial perspectives on the other concerns described in this section. They can help with assessing the level of risk in a given situation and how to manage it, including identifying whether conditions are serious enough that no data should be collected. They are experts in local terms and conceptions of SOGIESC that should be considered in designing survey questions. They may be important partners in recruiting survey respondents and in convincing local LGBTI communities to participate in data collection efforts. Their voices are vital for encouraging NSOs to include SOGIESC questions (Guyan 2022; Garland et al. 2023). There have been many meetings about data collection attended by researchers and activists who were affiliated with LGBTI communities in different parts of the world.[10] For these reasons (and others), human rights advocates and others connected with the LGBTI communities have strongly pushed for consultations at each step of data collection and research with those data (United Nations Independent Expert 2019).

Methodological research by NSOs: As with any survey questions, we should be attentive to the quality of SOGIESC measures and of the data produced, so continued methodological research is essential. NSOs have conducted such research in the US, UK, Canada, Argentina, and Mexico, and possibly other countries seen in Table 1, as well. Efforts to build support for data collection amongst policymakers in different branches of government may be enhanced by conducting methodological research.

Some persistent concerns that can be addressed by methodological research include the following:

- "Construct validity," which asks how well a survey item measures the underlying concept, is one of the key methodological criteria used by the 2022 NASEM report.

[9] See (Brown and Herman 2020; Park 2015).

- There are well-known methods for testing the comprehensibility of measures for the general population and for LGBTI people by using cognitive testing and field testing.
- Many household surveys and some administrative records require "proxy reporting" by respondents for other household members. If household members are LGBTI but that status is not known by the respondent, there will be undercounts and perhaps biases in the LGBTI data that will make it challenging to analyze. Some research on proxy reporting has been conducted in the US (and perhaps other countries) and shows that it is feasible to ask SOGI questions to proxies, but there is not yet enough research to assess the accuracy and quality of the data that is produced (National Academies of Sciences Engineering and Medicine 2022).
- Intersex status is included in the SOGIESC umbrella term, but we have very little research on questions about intersex status. Some measures of intersex status have been used in research settings and could be studied, including Mexico's National Survey on Sexual and Gender Diversity (Table 1).
- Training of survey enumerators is essential to get accurate data, and training procedures could be further studied.

Legal concerns: Somewhere around 65-70 countries still criminalize homosexuality, and others criminalize some gender expression. In those countries, there may be political or legal barriers to adding SOGIESC questions onto government surveys. There may also be reluctance to put LGBTI people at greater risk if they disclose their identities to either governmental or nongovernmental survey teams. As noted earlier, consultation with LGBTI civil society is essential in these situations, since vibrant LGBTI communities may exist in places where being LGBTI is criminalized.

Ethical concerns: In particular, many stakeholders in the data collection discussions express concerns about risk of harm to LGBTI people that result from collecting data on SOGIESC. What is required to enhance benefits and minimize risks and what minimum standards should developing countries consider for SOGI data collection? A set of SOGIESC data principles has begun to emerge from discussions in many different contexts (United Nations Independent Expert 2019; Park 2022; European Commission High Level Group on Non-Discrimination Equality and Diversity 2023). These are some of the core principles:

- Do no harm: Where LGBTI people live in contexts that put them at great risk, such as countries with highly punitive laws or negative public attitudes, research and data collection must be approached with caution. This principle also extends to thinking about how data are used and whether some research might be done that is harmful to the community.
- Privacy and confidentiality are essential in all countries, since the disclosure that someone identifies as a sexual or gender minority could lead to harms such as discrimination, violence, or, in some places, arrest. These concerns are relevant in all countries.
- Strong data security standards must be used to ensure privacy and confidentiality.
- In addition to the need to protect the privacy of data on individuals, international human rights principles also require states to make clear how data will be used for legal and legitimate purposes.
- Data should be analyzed in ways that will benefit the well-being of members of SOGIESC communities, including in the areas of health, education, economic well-being, financial inclusion, and social protection. This part of the data value chain should not be underestimated as an ethical principle given the risks taken on by LGBTI people in offering their data on surveys (Guyan 2022).

Cultural concerns: Cultural concerns question the use of survey questions and concepts from other countries, particularly from the Global North. Local and indigenous mappings of behavior or attraction onto identities or terms might vary from that seen in the countries that have produced the survey questions and concepts offered above. As commentators have noted, the sharp distinction between sexuality and gender that is embedded in the SOGIESC concepts discussed in this paper may vary considerably in non-Western countries (e.g. Altman 2001). Sexuality and gender might be combined in local identities rather than pulled apart into very different categories.

At the same time, though, Altman and others have pointed out several globalizing tendencies that have led to greater adoption of LGBTI terms in many countries and regions: popular culture, global gay travel, and especially the expanding global LGBTI human rights movement. The rise of the internet has likely accelerated the globalization of identities. So while it is essential to involve local LGBTI civil society in the design of survey questions that are appropriate for different cultural contexts, it is possible that the resulting questions might not vary as much from the NASEM items as one might expect.

Moving forward

Although relatively little SOGIESC data has been collected by NSOs, a growing number of stakeholders have begun to see the need for it, especially as a dimension of intersectionality. We have learned a lot in some countries about how NSOs can include SOGIESC measures on surveys, with some of that knowledge noted here. The topic of data has become a much more important one for LGBTI communities as data is increasingly seen as a practical advocacy tool to highlight inequalities and the need for policy action. The number of ethical discussions in prominent forums attests to the investment of attention and effort by those communities to work through some important concerns about how data can be collected in an ethical and safe way. And although it was not the main subject of this paper, we have also learned how to analyze data in ways that speak to the life needs of LGBTI people (Badgett et al., n.d.; Guyan 2022; Daly, Crehan, and Grzywnowicz 2022).

Although this short discussion paper could not summarize all that has been written, it opens the way to more informed discussions about and workplans for expanding SOGIESC data in new places. This is the moment for collaborative efforts drawing together stakeholders from statistical agencies, academia, development institutions, and LGBTI civil society. Knowledge sharing both regionally and globally can accelerate the pace of change and result in the SOGIESC data that we will need to further the study of intersectionality.

References

Altman, Dennis. 2001. *Global Sex*. Chicago: University of Chicago Press.

Badgett, M. V. Lee. 2020. *The Economic Case for LGBT Equality: Why Fair and Equal Treatment Benefits Us All*. Boston: Beacon Press.

Badgett, M. V. Lee, Soon Kyu Choi, and Bianca D. M. Wilson. 2019. "LGBT Poverty in the United States: A Study of Differences between Sexual Orientation and Gender Identity Groups." Los Angeles, CA. https://williamsinstitute.law.ucla.edu/wp-content/uploads/National-LGBT-Poverty-Oct-2019.pdf.

Badgett, M. V. Lee, and Randall Sell. 2018. "A Set of Proposed Indicators for the LGBTI Inclusion Index." New York.

Badgett, M.V. Lee, Christopher S. Carpenter, Maxine J. Lee, and Dario Sansone. n.d. "A Review of the Economics of Sexual Orientation and Gender Identity." *Journal of Economic Literature*, 1–52.

Brown, T. N. T., and J. L. Herman. 2020. "Exploring International Priorities and Best Practices for the Collection of Data About Gender Minorities: A Focus on South America." Los Angeles, CA. https://williamsinstitute.law.ucla.edu/wp-content/uploads/WPATH-English-Mar-2020.pdf.

Daly, Felicity, Phil R. Crehan, and Micah Grzywnowicz. 2022. "The LGBTI Inclusion Index: An Innovative Tool to Incentivize Human Rights and Development Data." *Journal of Human Rights Practice* 14 (2): 600–621. https://doi.org/10.1093/jhuman/huab057.

Douglas, Jamie H., and Michael D. Steinberger. 2015. "The Sexual Orientation Wage Gap for Racial Minorities." *Industrial Relations* 54 (1): 59–108. https://doi.org/10.1111/irel.12077.

European Commission High Level Group on Non-Discrimination Equality and DIversity. 2023. "Guidance Note on the Collection and Use of Data for LGBTIQ Equality." Vol. 7. Luxembourg: Publications Office of the European Union. https://commission.europa.eu/document/download/66adbc7e-99cb-4d88-a653-d7fbfba9d7e8_en?filename=Guidance note on the collection and use of data for LGBTIQ equality_FINAL.pdf.

Garland, Katie, Emma Pottinger, Johanna Choumert Nkolo, Galina Lapadatova, Gary Collins, Sara Litke-Farzaneh, and So O'Neil. 2023. "Everyone Counts: Lessons about Collecting Gender Data to Improve Health Outcomes of the Transgender and Non-Binary Population for the U . S . Census Bureau and Other Population-Based Survey Instruments." https://www.mathematica.org/publications/everyone-counts.

Gender Identity in U.S. Surveillance Group (GenIUSS). 2014. "Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys." Los Angeles, CA.

Guyan, Kevin. 2022. *Queer Data: Using Gender, Sex and Sexuality Data for Action*. Bloomsbury. London: Bloomsbury Academic.

Herman, Jody L., and Kathryn O'Neill. 2020. "Vulnerabilities to COVID-19 Among Transgender Adults in the U.S." https://williamsinstitute.law.ucla.edu/publications/transgender-covid-19-risk/.

Lebbos, Toni Joe, Victoria Esquivel-Korsiak, and Julia Clark. 2021. "ID Systems and and SOGI Inclusive Design." Washington, DC. https://documents1.worldbank.org/curated/en/803981634587620467/pdf/ID-Systems-and-SOGI-Inclusive-Design.pdf.

National Academies of Science, Engineering, and Medicine. 2020. *Understanding the Well-Being of LGBTQI+ Populations.* Washington DC: The National Academies Press. https://doi.org/10.17226/25877.

National Academies of Sciences Engineering and Medicine. 2022. *Measuring Sex, Gender Identity, and Sexual Orientation.* Washington, D.C.: National Academies Press. https://doi.org/10.17226/26424.

O'Malley, Jeffrey, and Andreas Holzinger. 2018. "The Sustainable Development Goals: Sexual and Gender Minorities." https://doi.org/10.4324/9781315527093.

Outright International. 2020. "Vulnerability Amplified: The Impact of the COVID-19 Pandemic on LGBTIQ People." New York.

Park, Andrew. 2015. "An Inclusive Approach to Surveys of Sexual and Gender Minorities: Report of Meeting, Kathmandu, Nepal." Los Angeles, CA. https://williamsinstitute.law.ucla.edu/wp-content/uploads/Inclusive-Survey-SGM-Nepal-Mar-2015.pdf.

———. 2022. "Principles for the Safe and Ethical Collection of Data about SOGIESC in State Efforts to Monitor Progress toward the Sustainable Development Goals." https://www.apark.net/s/G7Outright.pdf.

Park, Andrew, and Lucas Ramon Mendos. 2019. "For All: The Sustainable Development Goals and LGBTI People." https://doi.org/10.13140/RG.2.2.23989.73447.

Pedulla, David S. 2014. "The Positive Consequences of Negative Stereotypes: Race, Sexual Orientation, and the Job Application Process." *Social Psychology Quarterly* 77 (1): 75–94. https://doi.org/10.1177/0190272513506229.

Pereira, Henrique, Jéssica Pedro, Cindy Mendes, Mariana Duarte, and Patrícia G. Silva. 2021. "Psychosocial Impacts of COVID-19 Pandemic on Lesbian, Gay, and Bisexual People Living in Portugal and Brazil—A Qualitative Study." *Journal of Psychosexual Health* 3 (2): 146–59. https://doi.org/10.1177/26318318211017466.

Río, Coral Del, and Olga Alonso-Villar. 2019. "Occupational Segregation by Sexual Orientation in the U.S.: Exploring Its Economic Effects on Same-Sex Couples." *Review of Economics of the Household* 17: 439–67. https://doi.org/10.1007/s11150-018-9421-5.

Sexual Minority Assessment Research Team. 2009. *Best Practices for Asking Questions about Sexual Orientation on Surveys.* Los Angeles. http://williamsinstitute.law.ucla.edu/wp-content/uploads/SMART-FINAL-Nov-2009.pdf.

United Nations Independent Expert. 2019. "Data Collection and Management as a Means to Create Heightened Awareness of Violence and Discrimination Based on Sexual Orientation and Gender Identity." https://www.ohchr.org/en/documents/thematic-reports/ahrc4145-data-collection-and-management-means-create-heightened.

World Bank and signatories. 2022. *Principles on Identification for Sustainable Development: Toward the Digital Age.* Washington, DC. https://www.idprinciples.org/%0Ahttps://documents1.worldbank.org/curated/en/213581486378184357/pdf/Principles-on-Identification-for-Sustainable-Development-Toward-the-Digital-Age.pdf.

Table 1: Developing Country Surveys with SOGIESC Data

Country	Dataset	Years	Sexual Orientation				Gender Identity		
			Identity	Behavior	Attraction	Partnership	Trans	3rd Gender	Intersex
Bangladesh	DHS	2004		x (men)					
Bangladesh	Census	2022						x	
Brazil	Brazilian National Health Survey	2019							
Brazil	Census	2010				x			
Cambodia	DHS	2005, 2010, 2014, 2021		x (men)					
Colombia	DHS	2015	x			x	x		
Colombia	Great Integrated Household Survey "Gran Encuesta Integrada de Hogares" o GEIH				x		x	x	
Colombia	Multipurpose Household Survey of Bogotá ("Encuesta Multiproposito de Bogotá")	2014, 2017, 2021	x		x		x	x	
Dominican Republic	DHS	2002		x (men)					
Ecuador	Encuesta Sobre las Condiciones de Vida de la Polación "GLBTI"	2012-2013							
India	DHS	2019-2021						x	
India	Census, Periodic Labor Force Survey	2017-2022 (PLFS)						x	
Mexico	National Survey on Sexual and Gender Diversity (ENDISEG)	2021	x	x	x		x		x
Myanmar	DHS	2015-16		x (men)					
Nepal	Census	2011, 2021						x	
Pakistan	Census	2017						x	
Peru	Primera Encuesta Virtual para Personals LGBTI en el Perú, 2017	2017							
Philippines	DHS	2003		x (men)					
South Africa	DHS	2003, 2016				x			
South Africa	Census	2011				x			