A big ask
Sex and data collection

What is happening with sex-based data collection? **Alice Sullivan** addresses contemporary confusions and controversies

Sex is a fundamentally important variable in social statistics. And yet, in recent years, data collection on this vital statistic has become controversial.[1,2]

The push against sex-based data collection is a growing trend internationally. Administrative data collection exercises have removed data on sex, sometimes by conflating sex with gender identity (see, for example, the US Census Bureau: bit.ly/3TWbeVp). This has affected data on a wide range of questions, including crime, education, employment and health. In the UK, sex appears to have become a devolved matter, with practice in Scotland diverging from England and Wales and Northern Ireland. Decisions on the sex questions in both the England and Wales 2021 Census and Scotland's Census a year later in 2022 have been resolved in the judicial review courts, with contrasting outcomes in each case. In Scotland the guidance accompanying the sex question stated that the answer given may diverge from the respondent's legal sex, whereas in the rest of the UK the guidance stated that those "considering how to answer" should provide their legal sex as denoted by their birth certificate or gender recognition certificate.

Attempts to remove sex-based language and sex-based data collection share a common root in queer theory and gender identity activism originating in the USA. This is part of a larger political project aimed at the erasure of

Alice Sullivan is professor of sociology and head of research at the UCL Social Research Institute.

Definitions

- ***Sex*. In humans, sex is a binary biological category. Individuals are classified by reproductive function as male or female. Sex is determined at conception, and is immutable.[6]**
- ***Gender*. The term "gender" refers to the stereotypes and social roles that are associated with each sex. Gender is a social category, rather than an individual one, and refers to how society sees girls and boys and women and men, based on their sex. Gender refers to the hierarchical power structure between men and women.[7]**
- ***Gender identity*. The term "gender identity" refers to some people's sense that they identify psychologically as a member of the male or female sex, particularly when this identity clashes with their biological sex. It refers to how individuals see themselves, rather than how society sees them. Gender identity is not clearly defined in conceptual terms, and has not been operationalised as a single agreed variable.**

sex categories (bit.ly/3i5lFIG). A feature of this activism has been calls for "no debate" leading to radical changes in policy and practice, including changes to data collection, being brought in without input from expert and critical voices.[3]

Conceptual confusion underpins much current discussion regarding data collection on sex and gender. The view that "sex is a spectrum", a commonplace of postmodern gender theory, has attained considerable influence. For example, in a recent article for *Significance*, Thornton *et al.* cast doubt on the idea that sex is binary, and describe sex as the sum of a set of sexed characteristics.[4] They suggest that it is better to ask specifically about these sexed characteristics (for example, "do you have ovaries"?) rather than asking a study participant's sex. There are a number of problems with this perspective: first, it neglects the fact that sex has systematic effects on health and social life; second, it leads to language which many people, especially women, may find dehumanising; and third, it leads to questions which are less readily understood and more likely to mislead study participants than a simple question on sex.

Alongside muddled thinking about sex, much commentary and official guidance displays a fundamental misunderstanding about the concept of gender, conflating gender, understood as a social structure which affects people according to their sex, with gender identity, which relates to an individual subjective perception of self which may clash with one's sex. This elision of gender and gender identity obscures the social importance of sex. This confusion has led to the claim, from sources including Scotland's chief statistician (bit.ly/3gtjGNW) and the US National Academy of Sciences, Engineering and Medicine,[5] that "gender" should be collected as a default, and sex should only be asked in limited circumstances.

Sex is more than the sum of its parts

Thornton *et al.* claim that sex is not binary, and is defined by a constellation of primary and secondary sex attributes.[4] Yet how would we know which attributes related to each sex if this was the case? The fact is that sexed attributes do not randomly cluster in individuals, and neither can they be seen as a spectrum.[8,9] There is a reason why the people with ovaries do not also have testicles, and vice versa. Sex is an integrated system. Every cell in the body is sexed, and male and female cells have different characteristics and responses (bit.ly/3gv9dlh).

Humans reproduce sexually, and the two sexes are defined by their roles in sexual reproduction. The claim that more than two sexes exist would imply that there exist additional classes of people, beyond male and female, who play a necessary role in the production of a new human.

Sex has systematic effects on both health[10] and social experience. Sex is not a mere proxy for its correlates. It is because we want to know people's sex that we ask for this information, rather than some alternative such as "do you have a uterus?" (though of course we might need to know that too in specific medical circumstances). The significance of not having a uterus is rather different in the case of a woman than in the case of a man. By the age of 60, 30% of women in the USA have had a hysterectomy, and thus are women without a uterus (bit.ly/3ACyS2g).

Asking about body parts rather than sex is both unclear and dehumanising.[11] Tellingly, this deconstruction of humans into sexed body parts is applied far more often to women than to men.[12] Men seem to have evaded redefinition as prostate-bearers or ejaculators.

An illustration is provided by *The Lancet*'s notorious statement that: "Historically, the anatomy and physiology of bodies with vaginas have been neglected".[13] As one correspondent pointed out: "Although the intent of the wording might have been noble, language that risks deeply offending the majority cannot be considered inclusive".[14] In a similar vein, Democratic congresswoman Alexandria Ocasio-Cortez responded to the threat to the legal right to abortion in the USA by stating: "The gutting of Roe v Wade imperils every menstruating person in

the US." This desexed language has become prevalent in official information about women's health (bit.ly/3i5sgCW).

Using plain language which is widely understood, including by people with limited education or English language skills is a basic principle of questionnaire design as well as medical communication. To ask, "Do you have ovaries?" instead of "Are you male/female?" as Thornton *et al.* suggest,[4] is likely to lead to error and non-response which will disproportionately affect marginalised groups. The population which is likely to be confused by desexed language may be expected to be considerably larger than the population which such language is aimed at, so the damage to data quality may be substantial.

Queering the binary

Proponents of the view that sex is not binary typically invoke people with differences/variations of sexual development (DSDs/VSDs). Yet it is clearly a fallacy to suggest that the existence of a small minority of anomalous cases invalidates the existence or usefulness of a categorical variable. For example, the boundary between life and death can, in exceptional cases, be difficult to ascertain. Yet we still collect mortality statistics.

The proportion of births where there is any ambiguity in ascribing sex at birth is tiny, less than 0.02%.[15] Queer theorists often grossly exaggerate this figure in pursuit of their goal of "queering the binary". Queer theorists typically prefer the term "intersex", which is seen as offensive by some with DSDs, as it implies that people with DSDs do not have a sex. Activists in the cause of gender self-identification treat "intersex" as an identity as opposed to an umbrella term encompassing a set of developmental conditions, and some people may "identify as" intersex without having such a condition or diagnosis. This is reflected in the inclusion of "I" in acronyms such as LGBTQI+. The appropriation of DSD conditions in the service of a distinct ideological cause is deemed insensitive by many individuals and families affected by DSD conditions (bit.ly/3EsQPl6).

Thornton *et al.* state that "although intersex and transgender identities may overlap, they are not necessarily co-occurring identities".[4] In reality, there is no evidence of overlap as far as diagnosis is concerned – DSDs are just as rare in children referred to gender identity development services as in the general population.[16]

Survey questions which include "intersex" as a possible response category to a sex question (e.g. male/female/intersex/other) are both inaccurate and insensitive. The recent US National Academies report acknowledges that DSD status should not be conflated with sex.[5] However, it recommends asking stand-alone questions about DSDs. It is difficult to see how this could be justified except in exceptional cases, and therefore this recommendation may be seen as tokenistic or performative. As an ethical principle, general data collection exercises do not request information on rare conditions, because this would be intrusive and may potentially identify individuals. Nor can it be justified, since there will not be enough cases for any useful analysis.

Sex, gender and gender identity

Thornton *et al.* rightly argue that it is important to be clear in one's use of terminology, in order to avoid confusion.[4] Gender identity is not the same thing as sex. The acknowledgement that these distinct variables cannot be captured accurately in a single item is welcome.

However, Thornton *et al.* use the term "gender" as a synonym for "gender identity", and appear to be unaware of the sociological usage of gender as a social structure.[4] Their decision flow-chart for researchers suggests a naïve view of a rigid boundary between biology and social life, taking no account of the interactions between the two.

It is a basic principle of questionnaire design that a question should not be open to widely different interpretations by different respondents. "Gender" is commonly used to refer to all of the three concepts in the "Definitions" box above: sex, gender and gender identity. Therefore, using "gender" on its own is inevitably unclear, and as such should be avoided in questionnaires.

In common usage, "gender" is simply a synonym for sex. In fact, as the linguist Deborah Cameron (bit.ly/3ERf51E) points out, people have been using the term "gender" as a fancy term for sex for over 500 years. Some researchers may claim that this use of gender is "wrong", but good questionnaire design takes its lead from common parlance rather than assuming knowledge of technical or disciplinary usage.

For many sociologists, "gender" is a social structure which affects people according to their sex. On this definition, individuals do not have a gender as such, but we need to know their sex in order to understand how gendered roles and power structures affect their lives.

Finally, gender identity refers to individual psychology rather than a social structure. In this sense "gender" is treated as an individual characteristic.

As the transgender population has increased rapidly, particularly among youth, those with responsibility for data collection are rightly concerned to gather accurate data on this group. In order to achieve this, we need accurate data on both sex and gender identity. People with trans identities, including non-binary identities, are also affected by their sex. The lack of high-quality sex-disaggregated longitudinal evidence on medical transition has raised particular concerns, for example in the Netherlands.[17] and the UK (bit.ly/3VhoVPv).

Collecting useful data on gender identity, at a time when this relatively unfamiliar phenomenon is in rapid flux, is challenging, and requires careful consideration of the intelligibility and meaningfulness of the questions used within the general and gender-diverse populations. Yet relatively little attention has been paid to considering how best to collect data on gender identity. Instead, gender identity advocates have focused on attempting to deconstruct and erase sex as a category.

The confusion between gender (as a social structure) and gender identity (as an individual self-definition) pervades much

As the transgender population has increased rapidly, particularly among youth, those with responsibility for data collection are rightly concerned to gather accurate data on this group

A good question requires categories which are clearly defined and communicated, rather than porous and obscure

recent commentary from official bodies. The argument that "gender" rather than sex should be the default variable appears to be based on the premise that sex only matters where there is a direct biological cause of difference – often combined with a refusal to acknowledge the salience of biology even where it is obvious, for example in sports.[18] This attempt to minimise the importance of sex as a sociological variable ignores the fact that gendered social structures affect people according to their sex. Gender identity is a distinct concept, and there is no evidence to suggest that gender identity is more important than sex across all the domains that social and health scientists may be interested in (bit.ly/3ABtHQ7).

Conclusions

The attempt to "queer" categories is antithetical to good questionnaire design. A good question requires categories which are clearly defined and communicated, rather than porous and obscure. Sex is not a difficult concept. Failure to collect data on sex means that we fail to monitor sexism and to capture sex differences.

To speak of sex differences does not imply that these differences are straightforwardly biologically determined.[19] The idea that collecting data on sex somehow implies biological determinism is simply a fallacy. If men and women are treated differently according to their sex, we cannot capture this difference without data on sex.

Sex should be collected as a variable by default, because being male or female affects people both physically and socially in a systematic way. Data on gender identity should also be collected in contexts where it may be relevant and useful.

It is time to call a halt on the erasure of sex in data collection. ■

References

1. Sullivan, A. (2020) Sex and the census: Why surveys should not conflate sex and gender-identity. *International Journal of Social Research Methodology*, **23**(5), 517–524.

2. Sullivan, A. (2021) Sex and the Office for National Statistics: A case study in policy capture. *Political Quarterly*, **92**(4), 638–651.

3. Suissa, J. and Sullivan, A. (2021) The gender wars, academic freedom and education. *Journal of Philosophy of Education*, **55**(1), 55–82.

4. Thornton, S., Roy, D., Parry, S., LaLonde, D., Martinez, W., Ellis, R. and Corliss, D. (2022) Towards statistical best practices for gender and sex data. *Significance*, **19**(1), 40–45.

5. National Academies of Sciences, Engineering and Medicine (2022) *Measuring Sex, Gender Identity, and Sexual Orientation*. Washington, DC: National Academies Press.

6. Kashimada, K. and Koopman, P. (2010) Sry: The master switch in mammalian sex determination. *Development*, **137**(23), 3921–3930.

7. Oakley, A. (1998) Science, gender, and women's liberation: An argument against postmodernism. *Women's Studies International Forum*, **21**(2), 133–146.

8. Sullivan, A. (2020) Response to Fugard and Hines. *International Journal of Social Research Methodology*, **23**(5), 539–540.

9. Wright, C. M. and Hilton, E. N. (2020) The dangerous denial of sex. *Wall Street Journal*. 13 February.

10. Viveiros, A., Rasmuson, J., Vu, J., Mulvagh, S. L., Yip, C. Y. Y., Norris, C. M. and Oudit, G. Y. (2020) Sex differences in COVID-19: Candidate pathways, genetics of ACE2, and sex hormones. *American Journal of Physiology – Heart and Circulatory Physiology*, **320**(1), H296–H304.

11. Gribble, K. D., Bewley, S., Bartick, M. C., Mathisen, R., Walker, S., Gamble, J. *et al.* (2022) Effective communication about pregnancy, birth, lactation, breastfeeding and newborn care: The importance of sexed language. *Frontiers in Global Women's Health*, 3, 818856.

12. Dahlen, S. (2021) Do we need the word "woman" in healthcare? *Postgraduate Medical Journal*, **97**(1150), 483–484.

13. Davis, S. (2021) Periods on display. *The Lancet*, **398**(10306), 1124–1125.

14. Dahlen, S. (2022) Inclusivity starts with language. *The Lancet*, **399**(10323), 434.

15. Sax, L. (2002) How common is intersex? A response to Anne Fausto-Sterling. *Journal of Sex Research*, **39**(3), 174–178.

16. Butler, G., De Graaf, N., Wren, B. and Carmichael, P. (2018) Assessment and support of children and adolescents with gender dysphoria. *Archives of Disease in Childhood*, **103**(7), 631–636.

17. Biggs, M. (2022) The Dutch Protocol for Juvenile Transsexuals: Origins and evidence. *Journal of Sex & Marital Therapy*. doi.org/10.1080/0092623X.2022.2121238

18. Devine, C. (2022) Female Olympians' voices: Female sports categories and International Olympic Committee transgender guidelines. *International Review for the Sociology of Sport*, **57**(3), 335–361.

19. Richardson, A. (2022) Biologisms on the left and the right. *Ethnic and Racial Studies*, **45**(13), 2496–2518.