

OSR draft guidance: collecting and reporting data about sex in official statistics
Fair Play For Women response, April 2021

1. Who we are

Fair Play For Women Ltd is a campaigning and consultancy organisation which raises awareness, provides evidence and analysis and works to protect the sex-based rights of women and girls in the UK. Founded in 2017, our work is focused on understanding when and how gender-and sex-based rights conflict in law and policy-making. Our aim is to ensure that everyone's needs are fairly balanced and that women and girls are not forgotten in good policy-making.

We believe in compassion and fairness for all. We support the rights of trans people to live in safety and to be treated fairly. We also support the rights of women and girls, and this is our focus.

2. Summary

It is because the material reality of sexed bodies matters that sex is collected as a variable. This draft guidance on collecting and reporting data about sex is a welcome development. We endorse this move as progress towards ensuring that official data can be trusted to be of high quality and value.

We share the concerns that the conflation of sex and gender, which is already occurring, compromises official statistics in every way: trustworthiness, quality and value. This is apparent from the example cited, relating to crime statistics, and the explanation in the MoJ's Technical Guide on such data which states that the data are a mix of sex and self-identified gender which cannot be disaggregated. This means the sex-based crime data being reported are already contaminated.

The pursuit of data harmonisation is also fatally undermined by such conflation. This is a problem for administrative data sets such as those in the criminal justice system and in public health. It will become even more problematic if these data sets are to be combined as the basis for a national data overview as an alternative to the Census.

Official statistics are not personal records. They must be accurate if they are to be useful. If a personal preference is to be recorded and addressed as if they are the sex opposite to their sex registered at birth, then both attributes should be recorded separately. To record claimed gender identity as if it is someone's sex is to perpetuate the conflation which this guidance seeks to avoid.

Therefore we recommend the guidance goes further, and specifies that:

- a. "Sex" means sex observed and registered at birth, and this variable is paramount in collection of sex data. To avoid confusion, it may be qualified as birth sex, or as sex on a birth certificate or gender recognition certificate. No other version of "sex" as a variable can exist in official statistics.
- b. It may sometimes be necessary to request birth sex. This is lawful and acceptable.
- c. If any related variable such as gender identity is to be collected, then sex should also be collected.
- d. Gender should not be used as a synonym for sex. This leads to confusion and conflation.
- e. The distinction between data records for individual use and for official statistics must be properly understood and maintained, so that personal preference does not drive recorded data.

3. Importance of sex as a variable for official statistics

Biological sex is a material reality. While most of us might struggle to define what makes a person male or female, the ability to distinguish between the sexes is acquired early in childhood and appears to be instinctive and highly accurate. It does not rely on sight of birth records or intimate body parts. It exists even if it is not recorded. It is because of the material reality of sexed bodies that biological sex, observed and recorded at birth, is collected as a variable. There is ample evidence that birth sex is an important factor affecting people's lives. There are significant differences in experience, opportunities and outcomes between males and females across most aspects of our society. This is why there is no debate about the need to collect sex as a variable in most, perhaps all, official statistics. As the ONS says,

“Sex, as biologically determined, is one of the most frequently used and important characteristics the census collects as it is used in most multivariate analysis of data and feeds into the UK population projections. It is critical that the collection of information on gender identity for a small population (estimated to be less than 1%) does not jeopardise the quality of data collected on sex for the population who don't have trans identities or the protective characteristics of gender reassignment.”¹

There is no evidence yet that any other sex-adjacent concept such as gender identity can substitute for sex in such data and analyses. If gender identity or some other variable representing how someone lives is thought to be important, it should also be collected. In time, the inclusion of both variables would enable comparison to be done so that it could be determined whether some alternative factor should replace birth sex. This would be possible only if both variables were collected in parallel. Until such time as this has been demonstrated, the replacement of sex with an alternative is not justified and cannot provide official statistics of equal value.

4. Language and definitions

4.1 “Sex” usually means sex observed and registered at birth

It is only through the shared meaning of words that data collection and interpretation can have any value whatsoever. Sex has a material reality and a meaning defined in law (e.g. in the Equality Act, and elsewhere). The question, “What is your sex?” is universally understood to refer to this definition of sex, i.e. birth or biological sex. This was a clear finding in the ONS research into the sex and gender identity questions in the Census. Like age, another key variable, sex does not need to be defined for people to understand the question. Even people who don't like to answer know what is being asked.

4.2 UK law allows for some people to legitimately change their recorded sex

In limited circumstances, the law permits some people, those who have satisfied the requirements of the Gender Recognition Act, to change the sex recorded on their birth certificate. This does not affect their bodies and is not dependent on changing their bodies. The concept of legal sex, as written on a GRC, would have no meaning if sex was not itself an independent variable, determined in the body and not by documentation.

4.3 The GRA generates an error in the data, which may change over time

It must be recognised that recorded or registered sex is not a fixed variable, but one which can vary over time if the law changes, while sex recorded at birth is fixed. Those whose legally-recorded sex is no longer a reflection of their biological sex observed at birth are at present a very small proportion

¹ ONS Census Transformation Project – The 2021 Census: Assessment of Initial User Requirements on Content for England and Wales: Gender Identity Topic Report, referred to as ‘Gender Identity Topic Report,’ May 2016

of the population (c. 6000). This generates an error in any data set which asks, “What is your sex?” This error may be currently tolerable in official statistics (although it is not clear that it is being measured or monitored). However, if this were to change, such that many more people had a legal sex different from their birth sex, then the error could become significant. This is an impending risk in Scotland, given current Scotland government policy to reform the GRA legislation to permit anyone to change their legal sex based on self-identification. If enacted, it would be difficult to identify the extent and impact of such errors, unless birth sex is recorded alongside “legal sex”.

4.4 Any other “concept of sex” is gender identity, not sex

Sex is defined in law; gender or gender identity are not. Sex is observed and recorded at birth; gender is not. Sex is not an umbrella term. If used without a qualifier, it is widely understood to mean biological sex, normally observed and recorded at birth, just as age is understood to mean time elapsed since birth. The GRA permits sex as recorded on a birth certificate to be amended in certain cases, which indicates the primacy of recorded birth sex as the target variable.

4.5 Where “sex” is collected, the target variable is invariably birth or biological sex

As section 3 explains, it is the material reality of sexed bodies that the sex variable seeks to record. Therefore we believe that this guidance must go further. It should remind data collectors that the sex variable required is normally sex registered at birth, as a reflection of biological sex. Under current laws, sex on a birth certificate or gender recognition certificate is a good enough proxy for this, despite the c.6000 GRA holders whose recorded sex has changed, but this may not always be so.

There may be instances where birth sex is the required variable; the OSR should make clear in this guidance that it is not unlawful or inappropriate to specify and collect such data if necessary.

No other variable can be called “sex”. It would be appropriate for the OSR to specify this for the avoidance of doubt and data confusion.

5. Personal vs public good

The draft guidance says that “Collection and reporting of data about sex is a sensitive and potentially divisive topic and there may be times when producers are unable to meet the requests of everyone who has an interest in their statistics.” There is no evidence for this assertion, or to believe that sex is a more sensitive or divisive topic than many other variables collected, such as age, disability, or marital status. There are many more people who are disabled or widowed or divorced than are sensitive about their sex. Not wishing to answer is not a basis for redefining the target variable. There is no reason to compromise official statistics quality and value on this critical variable.

The recent error by the ONS in Census2021, resulting in unlawful guidance on the sex question, appears to have resulted from putting personal preference ahead of data integrity. The conflation of sex and gender identity not only fails the public by compromising official statistics. It also risks failing the individual, by misrecording important information.

This is a particular risk when administrative data systems are used both to store operational records for reference to named individuals and as data sets to provide official statistics (e.g. employment records used for gender pay gap data). The draft guidance does not acknowledge this issue. It may well be entirely appropriate that some administrative systems record something other than (birth) sex even for people who do not have a GRC, so that they will be addressed using their preferred title, for example. The draft guidance needs to recognise that this is a different purpose which records a different variable, and therefore requires a different field.

The approach ultimately taken by the Census2021 which captures both sex and gender identity is a good example of this. The NHS inclusion of fields for both sex and “gender” in electronic patient records is another, which reflects their dual use for medical reference by clinicians who need to know a person’s sex and by customer-facing medical staff who wish to address individuals according to customer preference. However, it is reportedly failing in practice because staff understand “gender” to be synonymous with “sex”, and are therefore recording data in one field only.²

The OSR should clarify that these are two distinct variables, and how they are determined; that gender identity may be self-identified, and may have more than two possible answers, while sex is a simple binary which humans have in common with other mammals, where there are only two possible responses, female or male. (Intersex conditions are developmental disorders of one sex or the other, not a third sex.)

6. Risks of continuing conflation of sex and gender identity

6.1 Data harmonisation

The ability to compare and integrate data sets is at risk if this conflation of sex and gender identity continues. The 2021 Census in England, Wales and Northern Ireland was saved from this conflation through legal intervention. As it stands, the Scottish Census in 2022 will not be comparable with the Census taken in the rest of the UK in 2021. The opportunity to replace the Census with a data set created by integrating large administrative data sets will also be at risk and may prove unworkable, if up to 1 in 135 of the adult population have different sexes recorded in different data sources. This is entirely possible, given the claim by Stonewall that the UK trans population is up to 500,000, while only 6000 GRCs have been issued in total since the GRA in 2004. Those without a GRC may be registered as the other sex, or using a first name indicative of such, for some or all of their official presence in government records such as passport, driving licence, NHS number, HMRC reference. Integrating these data sets will be difficult. As soon as these data are cut to represent a subset, the inaccuracies and mismatches may render them unworkable.³

6.2 Public understanding

A person’s biological sex does not rely on documentation. The differences in life experiences between the sexes are well-documented and remain an area of public concern, academic study and policy attention. Some differences, like propensity to commit violent and sexual crimes, are particularly striking. Yet the conflation of sex and gender identity is already eroding public understanding of these issues, and may lead to a shift in public support for important safeguarding measures such as female-only spaces that exist precisely because of sex differences. See, for example, media reports of female rapists, women sex abusers, women using pornography, etc, many of which are in fact reporting on male activity by perpetrators claiming a female gender identity.⁴

6.3 Code breaches

The Code says data should not be “materially misleading”. Given the standard meaning of sex is being born male or female, it is misleading to present gender identity as if it were biological sex.

It says data should “remain relevant, and support understanding of important issues.” Biological sex remains one of the most important variables (as per ONS previously cited). This has not changed. Human biology has not changed.

² <https://www.bmj.com/content/372/bmj.n735/rr-0>

³ <https://fairplayforwomen.com/small-errors-in-sex-data-have-large-distortive-effects/>

⁴ https://fairplayforwomen.com/sex_data_wrong/

The OSR guidance says that “Producers need to ensure data and statistics stay relevant to a changing society.” One change in society affecting data collection is the rise of the concept of gender identity, and the growth in numbers of people who present as if they are the other sex, contrary to their sexed bodies. The only way in which this change can properly be recorded and understood is by maintaining the definition of sex as biological sex, and capturing one or more separate variables when appropriate. These might include “gender identity”, as per the voluntary question in Census2021. All of these would need to be defined so that the variation to (birth) sex is understood by those responding and by those using the data. To do otherwise is to replace sex with a different variable.

7. Usage and norms for data collection

The draft guidance does not go far enough. Given the critical importance of sex as a variable, and the need to be able to disaggregate data sets by sex, it is not sufficient to recommend that data collectors should specify their version of “sex”. If sex matters as a variable then it must be collected as such, as birth sex. If it does not matter then there can be no basis for collecting some sex-adjacent variable as a proxy for birth sex. Only the collection of both sex and another sex-adjacent variable could, in time, lead to the understanding that the other variable is more significant. We are very far from that conclusion, and indeed no one is claiming it.

The guidance says that “it is hard for the statisticians to change the administrative systems”. But if these administrative data sets are to become official statistics then it is necessary that they meet the standard required of official statistics. If they are to be an input to national statistics then this is the standard that must be applied. Being hard to change should not be used as the reason standards are lowered for an important variable such as sex. Clear and unambiguous meanings for data collection should always apply.

The code says that “Producers should be clear about definitions or terminology they use, and these should be harmonised to be consistent and coherent with related statistics and data where possible. The terms ‘sex’ and ‘gender’ should not be used interchangeably in official statistics.” We agree. The OSR therefore has to choose whether it will be possible to intervene to stop the terms being used interchangeably in administrative data sets such as NHS records, or acknowledge that such a data set cannot meet the standard required for official statistics.

It is also questionable whether data collectors who nominate some sex-adjacent variable are acting lawfully in terms of data privacy, for example, employers collecting “gender” or gender identity instead of sex. Further, it may not be lawful to ask for one piece of personal information and record it as another.

These datasets are used as the basis of official statistics such as the “gender pay gap” reporting. Guidance is needed as to what is appropriate and lawful to record, and how to ensure the correct data are collected. The gender pay gap is intended to, and understood to, refer to male vs female, as per the protected characteristic of sex in the Equality Act; the data collected and reported should reflect this. Individual employees’ identities and preferences should be recorded as a separate field.

The OSR Code requires that there is good reason to collect and retain personal data. Where something similar to, but not identical to, sex registered at birth is being recorded and stored, the guidance should require that a sufficient case has been made to justify such personal information, and that it has been established as being an important variable. In such cases, we would expect that birth sex would remain at least as important, and should be collected alongside, and not be replaced by, a novel or different “concept of sex” which is in reality a gender identity.