



2024

Study on self-identification of origin in Belgium

Improving Equality Data Collection in Belgium II



Equal
— Opportunities **.be**

Table of contents

1	Project: Improving Equality Data Collection in Belgium (IEDCB)	4
2	Introduction	5
3	Equality data on origin	7
3.1	Why collect equality data on origin?	7
3.2	How are data on origin collected?	7
3.3	What is self-identification?	8
3.4	Advantages and disadvantages of self-identification	9
3.5	International practices	13
4	The situation in Belgium	15
4.1	What data do we have on origin in Belgium?	15
4.2	Self-identification of origin in Belgium	15
4.3	The legislative framework: data protection	16
5	Methodology	18
5.1	Advisory group	18
5.2	Expert consultation	18
5.3	Focus groups	19
5.4	Online survey	19
6	Results	22
6.1	Respondents' views on the collection of data on origin in Belgium	22
6.2	Specific conditions for collecting data on origin via self-identification	23
6.3	Questions on self-identification of origin	24
6.4	Comparison between proxy variables and self-identified origin	34
6.5	Evaluation and discomfort with survey questions	40
7	Self-identification: a practical guide	44
7.1	Good ethical practices	44
7.2	Which questions to use	44
7.3	Which response categories to use	45
8	Recommendations	48
8.1	Use self-identification as a useful measurement tool to combat discrimination and structural racism, possibly in combination with proxy variables	48
8.2	Actually use equality data in developing and monitoring public policy	48
8.3	Act from a human rights-based approach to data: data collection through self-identification should not harm	49
8.4	Use self-identification in accordance with ethical principles and legal conditions, with a specific goal and transparent communication	49

8.5	Consult the involved groups at different stages of the data collection process through a participatory process	49
8.6	Establish a permanent platform for exchange and consultation on equality data.....	50
8.7	Further research on self-identification of origin is needed.....	50

List of tables and graphs

Chart 1: Distribution of geographic response categories in the sample (N=370)	26
Chart 2: Distribution of ethnocultural response categories in the sample (N=370)	27
Table 1: Comparison between geographic and ethnocultural response categories	29
Table 2: Comparison between ethnocultural self-identification and auto-hetero perception.....	31
Chart 3: Percentage that reported having already experienced discrimination, by geographic response category (N=190).....	33
Chart 4: Percentage that reported having already experienced discrimination, by ethnic-cultural response category (N=190).....	33
Chart 5: Percentage that reported having already experienced discrimination, by auto-hetero perception (N=190).....	34
Chart 6: Distribution of the sample according to the proxy variable for origin (N=370)	35
Table 3: Self-identification with geographic response categories for those considered 'Belgian with Belgian origin' according to the proxy variable (N=139)	37
Table 4: Self-identification with ethnocultural categories for those considered 'Belgian with Belgian origin' according to the proxy variable (N=139)	37
Table 5: Auto-hetero perception for those considered 'Belgian with Belgian origin' according to the proxy variable (N=139).....	38
Table 6: Classification into proxy groups for those who answered only 'Belgium' (or only 'Belgium' + 'Europe in the EU') to the geographic self-identification question (N=109)	39
Table 7: Classification into proxy groups for those who answered only 'White' to the ethnocultural self-identification question (N=139)	40
Table 8: Evaluation of the questions: percentage of people who indicated discomfort (N=370).....	41
Chart 7: Percentage that reported experiencing discomfort with the questions, by ethnocultural response category.....	43

1 Project: Improving Equality Data Collection in Belgium (IEDCB)

This study is part of the second edition of the *Improving Equality Data Collection in Belgium* (IEDCB) project¹, funded by the Equal Opportunities Department (FPS Justice) and implemented by Unia. The project aims to improve the collection and use of data on (in)equality and discrimination (or equality data) in Belgium and ran from January 2023 to June 2024.

The IEDCB II project consists of two major parts:

1. A section focusing on equality data on the protected discrimination criteria of disability, health status and physical characteristics. For more info on this part of the project, please visit the website page and the report '*Improving Equality Data Collection in Belgium II: Final report*'².
2. A second part focusing on the principle of self-identification of origin and its use in equality data in Belgium. This report is the result of this second part.

This report deals with origin-based discrimination and the mapping of origins. These topics can be sensitive for some people and may raise difficult questions or traumatic experiences.

We would like to express our heartfelt gratitude to the individuals and organizations that participated in the research process leading to this report. It is thanks to their time, feedback and knowledge that we have been able to realise this work. This collaboration and the discussions that came with it were extremely valuable.

¹ Unia, [Data on discrimination and \(in\)equality: project equality data](#).

² Unia, [Improving Equality Data Collection in Belgium II: Final report](#)

2 Introduction

Thomas, Marie, and Ahmed are all Belgians, as are their parents. Yet each of them has experienced discrimination: Thomas based on his Jewish origin, Marie based on the black colour of her skin, and Ahmed based on his name which does not sound 'Belgian'. If we want to collect data on discrimination and inequality based on origin, in their case nationality is not relevant information because it does not point to the discrimination they have experienced. **And if we asked them to define their origin themselves, what would they answer?** Could we better identify who in the population is at risk of being affected by discrimination and inequality and thus make anti-discrimination policies more targeted and effective?

This study examines the method of self-identification, building on several international recommendations that advocate for the use of the self-identification method in the collection of data on origin.³ The study also follows the recommendation from IEDCB I which states that "wherever possible and appropriate, the use of self-identification should be applied, subject to finality and in accordance with the legal framework"⁴. The use of self-identification in the collection of equality data is also a fundamental principle of the human rights-based approach to data⁵.

However, the principle of self-identification for measuring origin is currently not often used in Belgium. Instead, proxies based on administrative data, such as nationality or nationality at birth, are often used to determine people's origin. Although the nationality of the individuals and their parents can be an indicator of origin, it does not always make it possible to identify the groups discriminated based on origin. Indeed, certain groups are treated unequally independently of their nationality, but linked to their origin (such as Roma and Jews) or to visible or phenotypic characteristics linked to origin (such as skin colour and appearance). **Self-identification is a method that could fill these gaps.**

However, there are still many challenges to the use of self-identification of origin in Belgium. By studying whether and how self-identification could be applied, this study aims to respond to the recommendations for the use of self-identification. This study therefore constitutes a first step in a process towards a broader application of the principle of self-identification of origin in Belgium.

The question we answer with this study is: **what is the acceptability and added value of using self-identification of origin as a method to collect equality data in Belgium?**

This study has several objectives:

- Probing the **views of the groups involved** on the data collection of origin and the self-identification method.
- Defining the **conditions** for its use in the fight against discrimination and racism in Belgium.
- Studying the **added value** of self-identification to identify groups at risk of discrimination, using a comparison with the widely used method with proxy variables.
- Formulating **good practices and recommendations**.

³ The *Subgroup on Equality Data* of the European Commission's High Level Group on Non-Discrimination, Equality and Diversity invites member states to use self-identification in the '[Guidelines on improving the collection and use of equality data](#)' in 2018 and in the '[Guidance note on the collection and use of equality data based on racial or ethnic origin](#)' in 2021.

⁴ Unia (2021), [Improving equality data collection in Belgium](#)

⁵ OHCHR (2018), [A human rights-based approach to data. Leaving no one behind in the 2030 agenda for sustainable development](#), p.13.

Origin(s)?

We focus in this study on '**origin**' in the **broad sense** to consider different forms of discrimination related to so-called 'racial' discrimination grounds. By origin, we therefore mean nationality, ethnic or national origin, so-called 'race', skin colour, and descent (Jewish or Roma).

Origin is an ambiguous concept that is the subject of much debate in the social sciences. Without going into the conceptual debate - which is beyond the scope of this project - we want to stress that origin is a **social construct** of which the meaning is not fixed. Therefore, we do not use a strict definition in this report. The origins of individuals and groups are complex and multiple.

Origin can be an element of a person's identity. Identity is changeable and can vary over time and depend on the situation the person is in. Although these concepts (identity and origin) are strongly linked, in this report we focus on origin **as an element of discrimination and inequality** rather than on its meaning for people's identity.

In this report, we use the singular form when talking about the concept of 'origin' in the abstract sense and as a discrimination ground. When we talk about a person's origins, for example in survey questions, we use the plural form to respect the plural nature of a group's or individual's origins.

3 Equality data on origin

3.1 Why collect equality data on origin?

To effectively combat the various forms of racism and discrimination related to origin, **reliable data are needed on which public policies can be based and against which they can be evaluated.** *"The main purpose of collecting and analysing equality data as part of the fight against systemic discrimination is to improve the situation and experiences of groups exposed to such discrimination, and to ensure that the specific needs of these groups are properly addressed."*⁶

Systemic or structural discrimination is often invisible at the individual level but becomes visible in figures that capture a particular domain or the entire society. These inequalities transcend individual discrimination and take place at the level of society, institutions and the state.⁷ Unia's Socio-Economic Monitoring, for example, allows structural inequalities on the labour market in Belgium to be mapped.⁸ The report is therefore used in the fight against discrimination and the promotion of inclusion in Belgium. The figures are used, for example, to develop positive actions. The government and other sectors use these figures to implement targeted actions and ensure that their sector reflects the population. These data can also be used to carry out targeted checks using field tests⁹.

Data disaggregated by origin for different domains make it possible to understand the **mechanisms of inequality and discrimination by making the invisible visible.** They give an idea of which groups are affected by structural inequalities and discrimination and the extent to which they are, and they make it possible to map evolutions.¹⁰ It is important to point out that Unia, as an equality body, is only interested in data related to origin that have the aim of combating discrimination and inequality.

While many international bodies recommend governments to pursue the collection of disaggregated data for origin, there are different trends in governments' practices. Some governments are very reluctant. Belgium is one of them.¹¹ We discuss the Belgian situation regarding the collection of data on origin in more detail in part 4 of this report.

3.2 How are data on origin collected?

Data on origin can be collected from different types of sources: censuses, administrative data, surveys, etc. Different methods can be used in these sources to collect data on origin. We will focus here on two methods, although others also exist:

1. **Using proxy variables:** these are indirect variables to measure personal characteristics using available information about that person¹². Nationality, nationality at birth, or country of birth of a person and sometimes of their (grand)parents are most often used, but also other

⁶ Committee of Experts on Intercultural Integration of Migrants (ADI-INT) (2024), [Training manual on equality data collection and analysis to prevent and address systemic discrimination](#), p.15, 67

⁷ Unia, [Understanding racism](#)

⁸ Unia (2022), [Socio-economic Monitoring 2022: labour market and origin](#)

⁹ Unia (2020), [Tests de situation 2.0 : quelques nouvelles propositions d'Unia](#)

¹⁰ Chopin, I., et al. (2014), [Policy Report: Ethnic Origin and Disability Data Collection in Europe: Measuring Inequality-Combating Discrimination](#)

Jacobs, D., et al. (2009), The challenge of measuring immigrant origin and immigration-related ethnicity in Europe.

Simon, P. (2005), [La mesure des discriminations raciales: L'usage des statistiques dans les politiques publiques. Revue internationale des sciences sociales.](#)

¹¹ Galonnier, J., et al. (2020). [Faire avec ou contre la race ? Les dilemmes des organisations internationales](#), p.23.

¹² High Level Group on Non-discrimination, Equality and Diversity, Subgroup on Equality Data (2021), [Guidance note on the collection and use of equality data based on racial or ethnic origin](#), p.39.

information such as language spoken at home, postal code, surname, etc. In this way, groups are often captured by means of an a posteriori reclassification (for example: people of 'Belgian origin', 'EU origin', 'non-EU origin'). Individuals cannot therefore indicate themselves to which category they belong.

2. **Self-identification:** The person concerned defines their origins themselves if they so wish. This method is explained in more detail in the next section.

These methods can also occur together in the same data collection process¹³. More information on the different methods of measuring origin can also be found in Chapter 6 of the IEDCB I final report.¹⁴

3.3 What is self-identification?

Self-identification is a method of data collection in which **data on personal characteristics are provided by the people to whom they relate**. Respect for self-identification is a basic principle of the United Nations Office of the High Commissioner for Human Rights' **human rights-based approach to data**: *"populations of interest should be self-defining, which means that the parameters of the population cannot be imposed by an external party"*¹⁵.

This human rights-based approach to self-identification gives individuals the free choice to self-declare information about their personal characteristics or not to do so. Moreover, the proposed response categories should be **developed in a participatory way**. All questions on personal characteristics should **also offer an open response** and allow **multiple response categories to be indicated**.

A central element of this approach is the principle that **affected groups should not be disadvantaged or harmed**, for example by using inappropriate terminology, reproducing stereotypical categories, or reinforcing the stigmatisation of certain groups.

The method of self-identification can also be applied for data collection on other characteristics, such as gender identity, sexual orientation, or disability. In this study, we focus only on the use of self-identification when collecting data on origin.

The method of self-identification of origin differs from other methods that assign an origin to an individual without them being able to confirm or deny it. In the principle of self-identification, an individual is considered the best judge of their own origin.

Specifically, the method of self-identification of origin can take different forms:

- The most common form is a **direct question** such as "What are your origins?" or "How would you define your origins?".

¹³ European Commission (2016), [European Handbook on equality data](#)

Each of the methods identified to measure diversity in public services has advantages and disadvantages, which is why it is ideal to combine the methods. This is also one of the conclusions of the FedDiverse study currently being conducted by researchers at VUB, UGent and UCL.

In the '[Guidance Note for Implementation of Survey Module on SDG Indicator 16.b.1 & 10.3.1 \(Discrimination\)](#)', OHCHR proposes a list of standard questions that combine several characteristics, including nationality and ethnic origin.

Unia (2021), [Improving equality data collection in Belgium](#), p.41.

OHCHR (2018), [A human rights-based approach to data. Leaving no one behind in the 2030 agenda for sustainable development](#).

¹⁴ Unia (2021), [Improving equality data collection in Belgium](#), p.41.

¹⁵ OHCHR (2018), [A human rights-based approach to data. Leaving no one behind in the 2030 agenda for sustainable development](#).

- It can also take the form of a **question probing experiences of discrimination**. For example, "Have you ever been treated unequally or discriminated based on your origins? If yes, which ones?" or "Do you think you belong to a minority at risk of racial or ethnic discrimination? If yes, which one?"
- **Finally, auto-hetero perception** is a variant on self-identification in which a person can indicate how others or society see that person. For example, "How do you think others see you?" This method allows one to measure how one thinks one is seen by others, which can be independent of how the person sees themselves. How one is seen by others plays an important role in discrimination.

3.4 Advantages and disadvantages of self-identification

The literature review conducted for this study revealed **certain advantages and disadvantages** of each of the two origin data collection methods described above: **self-identification and proxies based on (birth) nationality**. Below, we elaborate on the advantages and disadvantages. In doing so, we rely mainly on the '*Guidance Note on the collection and use of equality data based on racial or ethnic origin*' of the *European Union Agency for Fundamental Rights* (hereinafter 'FRA').¹⁶ More information can also be found in the report IEDCB I¹⁷.

Respect for the individual and their ability to act

One of the advantages of the method of self-identification is that it respects individual freedoms. Giving people the choice of whether to share information about their origin prevents them from being categorised by others against their will, gives them the power of choice and makes them the **spokesperson for their own origin**.¹⁸

Another advantage is that self-identification **respects the multiple, complex, and situational nature of origin**.¹⁹ It offers the possibility of indicating different origins and reflecting the mix of origins in the family. Or, on the other hand, to also denote an origin that is different from the family origin and better reflects one's identity.²⁰

Moreover, when self-identification is used in line with the human rights-based approach to data, it is **intrinsically linked to a participatory process**. The groups involved can participate in all stages of the research process: question design, data collection and analysis. If this participation is actually applied, it provides a great advantage for self-identification. It increases the groups' respect and recognition of the question and terminology used. This will **increase the acceptance of the self-identification question** and, consequently, the level of participation of the groups involved in data collection.

This also shows the disadvantages of using the proxy method based on information on (birth) nationality: in this method, an individual's origin is defined and inferred by external sources. The proxy

¹⁶ High Level Group on Non-discrimination, Equality and Diversity, Subgroup on Equality Data (2021), [Guidance note on the collection and use of equality data based on racial or ethnic origin](#), p.39-41.

¹⁷ Unia (2021), [Improving equality data collection in Belgium](#), p.47-48

¹⁸ Fourot, A.-C., et al. (2005), [L'enquête sur la diversité ethnique: L'autodéfinition ethnique et la comparaison intergénérationnelle, vers une meilleure compréhension de la diversité ?](#), p.65.

Ringelheim, J. (2010), [L'identité culturelle à l'épreuve du droit international des droits de l'homme](#), p.6 et 31

¹⁹ Verhaeghe, F., et al. (2020), [Identificational assimilation patterns in young first, second, 2.5 and third generation migrants](#), p. 518

Simon, P. (2004), [Etude comparative de la collecte de données visant à mesurer l'étendue et l'impact de la discrimination aux états-Unis, Canada, Australie, Grande-Bretagne et Pays-Bas](#), p.57

²⁰ Simon, P. (2005), [La mesure des discriminations raciales: L'usage des statistiques dans les politiques publiques. Revue internationale des sciences sociales](#), p.24

method for determining origin therefore much less captures the multiple, changing, and complex nature of origin. We elaborate on this in the following sections. Moreover, most proxy methods for origin in Belgium do not involve participation in the creation of the method.

Objectivity/subjectivity

A disadvantage of self-identification is that it is considered more subjective because the individual influences the allocation of origin.²¹ In contrast, proxy variables are often described as more objective information because they do not vary depending on individual perception. Unlike verifiable data such as nationality, data obtained through self-identification are not verifiable.²² **The fundamentally subjective nature of responses** must be considered when using this method.

However, it should be noted that **the creation of proxies also involves a series of subjective decisions**. For example, choices must be made about what information is used (such as nationality at birth of (grand)parents) and how people are grouped into categories based on this information.²³

This is also reflected in the different proxy methods for measuring origin currently used by different institutions in Belgium. These methods differ in the information used (e.g., birth nationality or first registered nationality) and the categories of origin used. Although we can argue that the proxy methods use verifiable information, the choices made to arrive at a final origin variable are not objective and certainly not homogeneous across institutions.

Regrouping into categories

The grouping of origins into different origin categories that happens in both methods is non-trivial and often even **a subject of controversy**. This applies both to the grouping of origins into origin groups for proxy variables and to the response categories proposed in self-identification questions.

When using self-identification, providing only an open-ended response field would render the data unusable for analysis²⁴. Therefore, it is often necessary to propose certain answer categories and limit their number. These response categories should also be carefully designed to minimise the number of responses in the offered open response field.

The challenge of creating response categories in self-identification is therefore a drawback: it includes a technical and statistical dimension, but above all a social and political one. Creating response categories has to do with the way we look at society and its diversity: how are different groups in society seen²⁵? This approach is often associated with the fear of essentialising these categories: that this classification of society is performative and will further divide society.²⁶

The proposed categories obviously do not capture the nuances of the diversity of origins in the population and reduce its complexity. Nevertheless, it is crucial that these categories reflect **as closely**

²¹ Simon, P., Clément, M., (2006), [Rapport de l'enquête " Mesure de la diversité ". Une enquête expérimentale pour caractériser l'origine](#), p.32.

²² Ringelheim, J. (2010), [L'identité culturelle à l'épreuve du droit international des droits de l'homme](#), p.8.

²³ Ringelheim, J., et al. (2010). [Ethnic Monitoring - The Processing of Racial and Ethnic Data in Anti-Discrimination Policies: Reconciling the Promotion of Equality with Privacy Rights](#), p.149.

²⁴ If everyone gave a different answer, a considerable amount of reclassification work would be needed to statistically analyse the data. Moreover, this reclassification would be contrary to the purpose of self-identification.

²⁵ Martiniello, M., Simon, P., (2005), [Les enjeux de la catégorisation](#), p.2.

Jacobs, D., Rea, A., (2005), [Construction et importation des classements ethniques](#), p.2.

²⁶ Jacobs, D., et al. (2009), [The challenge of measuring immigrant origin and immigration-related ethnicity in Europe](#), p.71.

Jacobs, D., Rea, A., (2005), [Construction et importation des classements ethniques](#), p.18.

as possible the population's diversity, collective identities, historical and social context, and the groups potentially discriminated based on their origins.²⁷

The challenge of creating categories is also a drawback of proxy methods. The proxy method also regroups different origins into origin groups. This also carries the risk of homogenisation, as a wide variety of origins are grouped into single categories such as "foreign origin" or "non-EU origin". This grouping inherently forces to reduce the complexity of the social reality of these groups.

Stability and comparability

A disadvantage of self-identification is that this method is **more difficult to standardise**. The idea is that the response categories proposed for self-identification can change over time and according to the specific context (country, institution, organisation, etc.). Thus, they can continue to reflect the changing diversity of groups in society and the specific context. The history of migration, the groups represented, and the terminology used vary widely between countries and evolve over time. This complicates the standardisation and comparability of data across institutions and countries and over time.²⁸ Moreover, individuals may define themselves differently over time and depending on context, which also makes it difficult to compare data. An advantage of self-identification is that this flexibility allows the method **to be better adapted to the social realities** of individuals and groups in changing societies.

An advantage of the proxy method is that the information used and the categories created are more stable, making comparisons over time and between regions or countries easier. **However, proxy methods and categories can also differ between different data producers**. The use of proxies therefore requires the different data collectors to coordinate on the methods and categories used to produce comparable analyses. As mentioned above, this is currently not the case in Belgium. At the very least, it requires full documentation to be made available so that the methods can be reproduced.

Ability to identify discriminated groups: risk of under- or over-reporting

A drawback of both methods of measuring origin is that they can never fully identify exactly the groups being discriminated against.

There is a risk of underreporting when collecting data using proxies. First, there is a risk that proxies using **information linked to (grand)parents across generations lose their relevance** for identifying discriminated groups and are no longer an adequate tool to measure the diversity of a population.²⁹

The second drawback of proxies is that they do not always make it possible to identify groups disadvantaged **on the basis of visible characteristics** such as skin colour, name, accent or other elements associated with 'origin'.³⁰ For example, after several generations, the criteria of nationality or country of birth of parents no longer make it possible to identify a person of African origin who is

²⁷ Ringelheim, J. (2018), [Nommer les groupes discriminés pour mieux combattre la discrimination: La fin d'un tabou ?](#), p.91. Martiniello, M., Simon, P., (2005), [Les enjeux de la catégorisation](#), p.2.

²⁸ European Commission (2016), [European Handbook on equality data](#), p.78.

²⁹ Jacobs, D., et al. (2009), [The challenge of measuring immigrant origin and immigration-related ethnicity in Europe](#), p.80. Ringelheim, J., De Schutter, O., (2010). [Ethnic Monitoring - The Processing of Racial and Ethnic Data in Anti-Discrimination Policies: Reconciling the Promotion of Equality with Privacy Rights](#), p.14.

Chopin, I., et al. (2014), [Policy Report: Ethnic Origin and Disability Data Collection in Europe: Measuring Inequality-Combating Discrimination](#), p.49.

³⁰ Simon, P., Clément, M., (2006), [Rapport de l'enquête " Mesure de la diversité ". Une enquête expérimentale pour caractériser l'origine](#), p.19-20.

discriminated based on skin colour, since the person and their parents have Belgium as their nationality and place of birth.

This is also the case for **people whose origin are not linked to nationality**, such as Roma, Travellers, or Jews.³¹ In a context of growing diversity, the group of people of 'mixed' origins also poses a challenge which may lead to over- or under-reporting. Also, for people who are adopted, information on nationality of parents may be an unrepresentative indicator of experiencing discrimination.

An advantage of self-identification is that it can address this issue by being **better suited to capture the multiple and complex nature of origin**. Self-identification can also complement the proxy method that is now often used unilaterally. In this way, a more complete picture of discriminated groups can be obtained.

But there is also a risk of underreporting when using the self-identification method. The groups involved can be underestimated in various ways. It can take the form of a **boycott**: individuals refuse to define themselves using these categories, either because they are against the concept of self-identification itself, because the categories do not correspond to their social reality, or because they fear the consequences that would be associated with this data collection. The latter is particularly relevant for historically stigmatised groups, such as Jews and Roma, who may shy away from defining themselves as belonging to these groups and may choose to define themselves differently. Therefore, it is very important **to clearly communicate** objectives and methods, work in a participatory way from the beginning, and build trust.³²

There may also be under- or over-reporting if individuals report a different origin than others generally attribute to them: the data are then insufficient to measure potential discrimination, which is generally based on how someone is perceived rather than on how someone perceives themselves.³³ We emphasise here the relevance of using either auto-hetero perception or questions that focus on experiences of discrimination.

Sensitive character

Historically, data on origin have been used to oppress, exclude, or eradicate certain groups. As a result, the issue of data on origin or on discrimination based on 'racial' criteria is **polarising** and the collection of such data evokes **fear**. A disadvantage of the self-identification method is that it is more touchy than the proxy method. Indeed, self-identification directly asks about a person's origin, while the proxy method based on nationality uses pre-existing administrative data.

Proxy variables are less touchy in the public debate. Moreover, this information is not subject to the same legal restrictions. Proxy variables are often based on administrative data that already exist in the National Register.³⁴

³¹ Pavee Point (2017), [Revised ethnic equality question needed in Census 2021](#).

³² Simon, P., Clément, M., (2006), [Rapport de l'enquête " Mesure de la diversité " . Une enquête expérimentale pour caractériser l'origine](#), p.11 et 55.

³³ European Commission (2016), [European Handbook on equality data](#), p.6.

Simon, P. (2005), [La mesure des discriminations raciales: L'usage des statistiques dans les politiques publiques. Revue internationale des sciences sociales](#), p.26.

³⁴ Chopin, I., et al. (2014), [Ethnic Origin and Disability Data Collection in Europe: Measuring Inequality - Combating Discrimination](#).

3.5 International practices

Collecting data on the origin of the population through self-identification is an Anglo-Saxon-inspired practice. Many countries have long used ethnocultural categories in their **censuses** to produce **ethnic statistics** that can be used to develop specific anti-discrimination measures.³⁵ Paradoxically, some of these data collections make use of a category used in the past to dominate and exclude, and are now used to redress the harm done to discriminated groups³⁶.

Among the countries that have historically collected data on the origin of their inhabitants are the United States³⁷, Australia³⁸, Brazil³⁹, Canada (which uses the term '*visible minorities*')⁴⁰, Ireland⁴¹, and the United Kingdom⁴². These countries collect data disaggregated by origin in various public domains, and more specifically through an '*ethnic question*' in national censuses.

Ireland is a particularly interesting example: it is the first European country to **have developed a coordinated national strategy for equality data**, with different levels of government at the table. This is the result of a political determination reflected in a number of inspiring practices⁴³. In 2019, for example, the country held a public consultation on the content of the future census. This enabled civil society, and the '*Pavee Point*' association representing Roma and Travellers, to provide input for developing the response categories in the self-identification question on origin. The answer categories thus better reflect the reality of the population and the terminology used by stakeholders, especially by the Roma population⁴⁴. Raising awareness among Roma and involving them in this process also greatly increased their participation in the self-identification question in the census.

Several European countries prohibit or severely **restrict** the collection of data on origin, no doubt due to the trauma of the persecution of Jews during World War II. Some countries collect data on origin through their census using a self-identification question, for example Malta (in the national census since 2021)⁴⁵ but also some Eastern European countries. There are many practices on *equality data* on origin in Europe, some of which are included in the FRA's compendium of good practices⁴⁶.

The use of self-identification is more common in **specific surveys on discrimination** or integration, sometimes with a focus on minorities. Examples include a survey on people of African origin in Spain⁴⁷,

³⁵ Balestra, C., L. Fleischer (2018), [Diversity statistics in the OECD : How do OECD countries collect data on ethnic, racial and indigenous identity?](#), p.22.

³⁶ Simon, P. (2005), [La mesure des discriminations raciales: L'usage des statistiques dans les politiques publiques. Revue internationale des sciences sociales](#), p.16.

³⁷ United States Census Bureau (2021), [Measuring Racial and Ethnic Diversity for the 2020 Census](#)

³⁸ Australian Bureau of Statistics (2014), [Ancestry Standard](#)

³⁹ Instituto Brasileiro de Geografia e Estatística, [Social Inequalities due to Colour or Race in Brazil](#)

⁴⁰ Visible minorities are defined as "persons, other than Aboriginal peoples, who are non-Caucasian in race or non-white in colour". Statistics Canada (2021), [Visible minority of person](#)

⁴¹ Central Statistics Office, [Question 14](#)

⁴² Gov. UK, [List of ethnic groups](#)

⁴³ Changes in the Irish labour market were also measured in 2004, 2011 and 2014, using self-identification and experiences of racism. It is also important to note that an attempt was made to harmonise the response categories with those of the UK, to make the data comparable.

⁴⁴ Pavee Point Traveller and Roma Centre (2017), [Towards an Ethnic Equality Question in Census 2021](#).

⁴⁵ High Level Group on Non-discrimination, Equality and Diversity, Subgroup on Equality Data (2021), [Guidance note on the collection and use of equality data based on racial or ethnic origin](#), p.36.

Sansone, K., (2021), [Census to collect data on race, sexual orientation and religion for first time](#), Malta Today (13 May 2021).

⁴⁶ European Union Agency for Fundamental Rights, [Compendium of practices for equality data collection](#)

⁴⁷ European Union Agency for Fundamental Rights, [Examination of the African and Afro-descendant population in Spain: Identity and access to rights](#)

a survey on discrimination in Italy⁴⁸, a survey on discrimination and unequal access to rights in France⁴⁹, a survey on integration in Estonia⁵⁰, a survey on living conditions, origins and population trajectories in Portugal⁵¹, a monitoring of discrimination and racism in Germany⁵², and a survey on discrimination experiences in the Netherlands⁵³.

In several European countries, self-identification of origin is used in surveys collecting data **on the Roma population or Travellers**. Examples can be found in Croatia, Italy, and the Czech Republic, with policies targeting Roma.⁵⁴ As noted in the IEDCB I report, this is a specific target group for which little data is available in Belgium.

By citing these various examples of the use of self-identification, we do not claim that the methods used are ideal, nor have we verified the specific methodologies used for all these studies. With these examples, we mainly want to draw attention to the intention of measuring origin-based discrimination on a large scale in different countries to serve **as inspiration**. It is instructive to learn about the process of designing and conducting these surveys and to draw methodological lessons from them.

⁴⁸ High Level Group on Non-discrimination, Equality and Diversity, Subgroup on Equality Data (2021), [Guidance note on the collection and use of equality data based on racial or ethnic origin](#), p.36.

⁴⁹ Défenseur des Droits (2019), [Inégalités d'accès aux droits et discriminations en France. Contributions de chercheurs à l'enquête du défenseur des droits : tome I](#), p.177.

⁵⁰ European Union Agency for Fundamental Rights, [Integration Monitoring of the Estonian Society](#)

⁵¹ European Union Agency for Fundamental Rights, [Pilot survey on living conditions, origins and trajectories of the resident population in Portugal](#)

⁵² Deutsches Zentrum für Integrations- und Migrationforschung (2022), [National Monitoring of Discrimination and Racism \(NaDiRa\)](#)

⁵³ European Union Agency for Fundamental Rights, [Experiences of discrimination in the Netherlands](#)

⁵⁴ High Level Group on Non-discrimination, Equality and Diversity, Subgroup on Equality Data (2021), [Guidance note on the collection and use of equality data based on racial or ethnic origin](#), p.37.

4 The situation in Belgium

4.1 What data do we have on origin in Belgium?

In Belgium, we **mainly have data on origin based on proxy variables**. These proxies are often compiled from administrative data such as those from the National Register or the Crossroads Bank for Social Security. However, censuses have not been conducted in Belgium since 2011.⁵⁵

Using proxies, **several institutions have developed variables to measure origin**, including Statbel's origin variable⁵⁶ or the origin variable used in the Socio-Economic Monitoring by Unia and the FPS Employment⁵⁷. Both use administrative data on a person's and their parents' (birth) nationality. Public institutions also use proxies based on self-reported nationality to map origin in certain surveys, such as the 'Living Together in Diversity' survey (SID-survey)⁵⁸.

In Belgium, nationality, together with nationality at birth, remains the main indicator to identify origin, although **there is no uniform definition or operationalisation**.

4.2 Self-identification of origin in Belgium

We note that there are **few public data sources in Belgium that use the principle of self-identification of origin**.

Until now, the few examples of data collection on origin using self-identification in Belgium were either **more academic in nature or at the European level**. Examples include the *European Social Survey* (hereafter ESS)⁵⁹, or the FRA's survey on minorities and discrimination in the European Union⁶⁰ and on experiences and perceptions of anti-Semitism.⁶¹ In 2018, this question on self-identification appeared in the ESS survey: "How would you describe your origin?", accompanied by a list of response categories from the '*European Standard Classification of Cultural and Ethnic Groups*' adapted to the Belgian context.

Both the experts we interviewed and the literature highlight **the difficulties** specific to the Belgian context for using self-identification: it is controversial for several reasons that are explained in detail in section 3.4 of this report. First and foremost, this controversy has to do with the sensitivity of origin data and the fear of misuse. The fear that this kind of data could be used for other purposes puts the various positions on edge. In addition, the **terminology used** for response categories can also be the subject of fierce discussions. Finally, this controversy relates to the legal restrictions on the collection of this type of data, which we discuss in more detail in the next section.⁶²

⁵⁵ Odasso, L., (2020), [Controversial Approaches to Measuring Mixed Race in Belgium: the \(in\)visibility of the Mixed Race Population](#), p.287.

⁵⁶ Statbel (2023), [Diversity according to origin in Belgium](#)

⁵⁷ Unia (2022), [Socio-economic Monitoring 2022: labour market and origin](#)

⁵⁸ Agentschap Binnenlands Bestuur, Vlaamse Overheid, [Samenleven in Diversiteit](#)

⁵⁹ The 'European Social Survey' (ESS) is an international scientific survey conducted every two years across Europe since 2002. The ESS measures people's attitudes, opinions and behaviour in more than 20 countries. [ESS website](#).

⁶⁰ European Union Agency for Fundamental Rights, (2017), [Second European Union Minorities and Discrimination Survey](#)

⁶¹ European Union Agency for Fundamental Rights, (2018), [Experiences and perceptions of anti-Semitism - Second survey on discrimination and hate crime against Jews in the EU](#), p.7.

⁶² Odasso, L., (2020), [Controversial Approaches to Measuring Mixed Race in Belgium: the \(in\)visibility of the Mixed Race Population](#), p.289.

Jacobs, D., Rea, A., (2005), [Construction et importation des classements ethniques](#), p.13.

On the other hand, a number of actors, especially academics, stress the **importance** of having data on origin in Belgium. While there is always a risk of misinterpretation, the absence of such data is potentially even more damaging and may condemn us to powerlessness in the face of discrimination.⁶³

But even though collecting data on origin is a subject of debate in Belgium, it is **becoming less and less taboo**⁶⁴. In 2005, Unia (then the 'Centre for Equal Opportunities and Opposition to Racism') consulted several civil society organisations about collecting ethnic statistics to better combat discrimination in recruitment and at work. Despite their fears about the instrumentalisation of these data, the participants recognised even then that these statistics were a useful measurement tool to "become aware of the extent or persistence of such discrimination and facilitate the monitoring of targeted policies to combat it".⁶⁵

The Belgian legislative framework also restricts the collection of data on origin and consequently the use of self-identification of origin. This is the subject of the next section.

4.3 The legislative framework: data protection

In this report, we briefly discuss the legislation applicable to the collection of data on origin.

Data on a person's origin are so-called 'sensitive' personal data and their processing is therefore subject to European and Belgian legislative frameworks. These clearly state what is permitted and what is not. The collection and processing of such data are subject, among others, to the GDPR⁶⁶, the Belgian Law of 30 July 2018⁶⁷ regarding the processing of personal data, and the Belgian Law of 4 July 1962⁶⁸ regarding public statistics.

This legislative framework for the protection of **'sensitive' personal data**, such as data relating to 'racial' or 'ethnic' origin, is often interpreted very restrictively, as if it is absolutely forbidden to collect data on origin in Belgium. However, the legislative framework for collecting and protecting personal data contains a number of exceptions and thus allows different types of collection, **under certain conditions and in strict circumstances**, as specified by the GDPR and explained in detail in the IEDCB I report.

We do not promote the collection and use of equality data at any cost. Compliance with this legislative framework and with the specific conditions for collecting and using equality data is essential, especially in a context where people fear that their data will be misused⁶⁹. For more information and

⁶³ J.B., (2020), [Données ethniques dans la recherche scientifique : encourager ou combattre la discrimination ?](#), RTBF, 11 May 2020.

Ringelheim, J., et al. (2010). [Ethnic Monitoring - The Processing of Racial and Ethnic Data in Anti-Discrimination Policies: Reconciling the Promotion of Equality with Privacy Rights](#), p.149.

Ringelheim, J. (2018), [Nommer les groupes discriminés pour mieux combattre la discrimination: La fin d'un tabou ?](#), p.83-84.

⁶⁴ Torbeyns, A., (2021), [Taboe rond het verzamelen van etnische gegevens vervaagt](#), De Standaard (17 April 2021)

⁶⁵ Centre for Equal Opportunities and Opposition to Racism (2005), [Consultation 2005 : Statistiques ethniques](#)

⁶⁶ [Regulation \(EU\) 2016/679 of 27 April 2016](#) on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Official Journal of the European Union 04 May 2016, L119/1.

⁶⁷ [Loi du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel](#).

⁶⁸ [Loi du 4 juillet 1962 relative à la statistique publique](#).

⁶⁹ Ringelheim & Wautelet (2022), [Commission d'évaluation des lois fédérales tendant à lutter contre la discrimination - rapport final](#), p.55.

Unia (2021), [Improving equality data collection in Belgium](#), p.25, 56.

advice, we recommend reading the IEDCB I report⁷⁰, as well as the page on this subject on the online platform eDiv⁷¹, or the Data Protection Authority (GBA) website⁷².

⁷⁰ Unia (2021), [Improving equality data collection in Belgium](#)

⁷¹ eDiv par Unia, [Vie privée \(protection des données\)](#)

⁷² [Data Protection Authority website](#).

5 Methodology

The methodology of this study is based on the guidelines of the *Subgroup on Equality Data*, set up by the European Commission's *High-Level Group on Non-Discrimination, Equality and Diversity* and facilitated by the FRA⁷³. Following the principle '*Nothing about us without us*', we applied a **participatory approach**, implemented through several methods that we describe below. These participatory methods allowed us to involve a large number of stakeholders in the project, in particular people who face discrimination based on their origin.

The research started from a need distilled from dialogue with both researchers and advocacy groups during the IEDCB I project. We then organised interviews and focus groups with both researchers and people with lived experience of discrimination and inequality. We discussed their perspectives on the possibility of applying self-identification. We used this input to design an online survey with self-identification questions. Throughout the research we also worked with an advisory group consisting of civil society organisations.

Some research participants pointed out that the research team consisted mainly of 'white' people and that this may affect the design of the study and the analysis of the results. The team was aware of the specific position and perspective of the researchers and used participatory methods to highlight the expertise of different groups involved and to valorise the views of ethnocultural minorities.

5.1 Advisory group

A French-speaking and a Dutch-speaking advisory group were set up, consisting of **civil society organisations** working on origin-related topics or representing different groups affected by origin-based discrimination. These advisory groups met three times: in March 2023, in September 2023, and in February 2024. During these meetings, we presented the project's progress and future steps, before asking participants for feedback. We also asked them for feedback on the final report before we published it. In each case, the feedback was considered, and the research project was adjusted.

5.2 Expert consultation

Several experts were consulted during this study:

- Between February and June 2023, the research team conducted seven semi-structured interviews with French- and Dutch-speaking **social science academics** working on the topic of origin. We used an interview guide to question these experts on several themes identified in the literature, including collecting data on origin, self-identification and its relevance in Belgium, the sensitivities of the Belgian context, the formulation of response categories and the participatory methodology. These interviews were recorded and transcribed. We incorporated many of their comments into the reflections and approach of the study.
- During the same period, we conducted four online interviews **with different agencies in European countries** that had experience using self-identification of origin when collecting equality data: Ireland, Germany, and Portugal. We questioned these actors about the methodology that they used, participation of the groups involved, and the process of creating self-identification questions and response categories. These experiences served as a source of inspiration for our different research phases.

⁷³ High Level Group on Non-discrimination, Equality and Diversity, Subgroup on Equality Data (2021), [Guidelines on improving the collection and use of equality data](#)

- Finally, in February 2024, we invited a number of experts from different **statistical institutions in Belgium** to a meeting to present our initial research findings. We received feedback on the study and a better understanding of the challenges involved in the operationalisation of self-identification.

5.3 Focus groups

Between May and September 2023, we realised **ten focus groups**: six were conducted in French, one in English, and three in Dutch. We organised these in collaboration with organisations that are in close contact with the field and that represent different groups at risk of discrimination.

The purpose of the focus groups was to consult a **diversity of people with foreign origins** or people who are part of minority groups at risk of discrimination based on their origins.

A total of 77 participants of different origins and ages took part in the ten focus groups. Most of them were women: 18 men and 59 women.

The focus groups did not allow us to consult certain minority groups in all their diversity. This is important, for example, for Roma and people of Jewish origin. These are very diverse groups for whom the issue of self-identification is particularly relevant and sensitive. We paid specific attention to consulting these groups, but some subgroups were not represented during the focus groups. For example, we failed to reach Travellers.

We conducted these focus groups using a questionnaire structured around four main questions, which were addressed during a two-hour discussion:

- "Have you ever been discriminated against or treated differently based on your origin? If so, on what do you think it was based?"
- "How do you define your origin?"
- "Here is the question on self-identification of origin and the answer categories presented to Belgian respondents in the 2018 *European Social Survey (ESS)*. What do you think about this question and the answer categories?"
This ESS question, used to give participants a concrete example, was our starting point during the focus groups for discussing the response categories. This allowed us to adjust the survey questions and answers of the online survey.
- "What do you think about the fact that data on the origin of individuals are collected to better identify inequality and discrimination?"

Finally, we asked everyone which elements they still wanted to put forward.

With the consent of each of the participants, an audio recording of each conversation was made, based on which we fully transcribed and anonymised the discussion. We analysed the transcribed focus group discussions using the software 'ATLAS.ti' for **qualitative data analysis**. These analyses allowed us to design and modify our online survey, both in terms of the questions asked and the terminology and response categories used.

5.4 Online survey

Finally, we published **an online survey** from 14 November to 22 December 2023 (a pdf of the full survey is available [on this webpage](#)). We analysed the data derived from the survey using the open-source programme R.

This survey had the following objectives:

- **Testing different ways of measuring origin:** through different forms of self-identification, through auto-hetero perception, and through the proxy method based on the nationalities of the respondent and their (grand)parents.
- **Testing different lists of response categories:** one with geographic response categories and one with ethnocultural response categories.
- **Poll people's opinions** on these questions and categories to study their relevance and acceptability.
- **Compare the proxy method and the self-identification method** to assess whether self-identification allows identification of groups at risk of discrimination that proxies based on nationality do not identify, and vice versa.

Survey distribution

We distributed the survey through various channels, such as Unia's newsletter and LinkedIn page. We also distributed the survey very widely in civil society. For some organisations with a low-literacy audience or with many people learning the language, we offered to come and explain and distribute the survey in person. Through these different avenues, we eventually **reached 370 individuals who completed the survey in full**.

The advisory group argued that the response rate should also be seen in light of the fact that the survey was preceded by focus groups. This created a climate of trust and encouraged the participation of certain groups that might otherwise have been more reluctant to respond. Moreover, as a national human rights institution, Unia enjoys a certain legitimacy among people of foreign origin.

Nevertheless, a sample of 370 people obviously remains limited and can **by no means be seen as representative of the entire Belgian population**. Especially since we specifically targeted individuals from specific groups and, in addition, some groups were under- or overrepresented, as we discuss below.

Sample composition and bias

We describe here the composition and diversity of our sample for different personality characteristics. 67.8% of respondents define themselves as women, 27.8% as men, and 3.25% as not belonging to either of these groups. In both the focus groups and the survey, we observe **unequal participation in terms of gender**: more women than men participated in this study. This is a common bias in the social sciences. Women tend to be better represented than men in the public, voluntary or activist sectors.⁷⁴

In terms of age, we see a **good distribution across the different age groups**. 21% of respondents are between 18 and 30 years old, 54% are between 30 and 50 years old and 25% are over 50 years old. The average age is 42.

In terms of the educational level of respondents, there is considerably less diversity in the sample. A vast majority have obtained a higher education degree (53.5% a master's degree and 24.3% a bachelor's degree), 7.6% have a secondary education degree, 3% have a primary education degree,

⁷⁴ Dodson, K., (2015), [Gendered Activism: A Cross-national View on Gender Differences in Protest Activity](#)
Lanfranchi, J., Narcy, M. (2015), [Female Overrepresentation in Public and Nonprofit Sector Jobs: Evidence From a French National Survey](#)

and 2.7% indicate having no degree. **Thus, there is a clear bias in terms of gender and level of education**⁷⁵ in our sample. It is important to take this into account when interpreting the results.

Regarding the education level bias, we received a lot of feedback on the complex, conceptual, and inaccessible nature of the study, and especially the online survey. In order to still try and reach an audience as diverse as possible in terms of educational level, we visited two organisations (*'Lire et Écrire'* and *'Ligo, Centres for Basic Education'*) in person to collect additional responses. This allowed us to explain the survey verbally. Participants could complete it with our help and had the option to fill it out on paper. We reached about 30 people who experienced difficulties with language and/or digital tools when completing such a survey.

39% of respondents live in Flanders, 33% in Brussels, 21% in Wallonia, and 5% in the German-speaking community. The **relatively high proportion of Brussels residents** in the sample can be explained in two ways. First, for this study we worked closely with civil society organisations and many of them have a strong connection to Brussels. Second, we deliberately sought respondents who were as diverse as possible, and Brussels residents have a greater diversity of origins than the rest of Belgium⁷⁶.

⁷⁵ Statbel (2023), [Level of education](#)

⁷⁶ Statbel (2023), [Diversity according to origin in Belgium](#)

6 Results

This chapter summarises the results obtained using the different research methods. The results of the quantitative survey should be interpreted with caution as it is a small, non-representative sample.

The results we use in this report are a selection from the large number of analyses we have done. Sometimes we have adapted the tables and graphs for this report (e.g. by selecting specific categories) to present the data as comprehensibly as possible.

For more details on sample characteristics and a more comprehensive analysis (in Dutch/French), please refer to the appendices available on Unia's [website](#). The full English version of the survey is also available as an [appendix](#).

6.1 Respondents' views on the collection of data on origin in Belgium

Research participants highlighted the benefits and risks of collecting data on origin. The main benefit was that **figures on diversity and discrimination make it possible to combat stereotypes and disinformation**. It forces people to see things as they really are and makes it possible to take concrete political action to combat racism and discrimination.

Some participants **were against** collecting data on origin or felt uncomfortable with it. They found it inappropriate, essentialising, and inherently discriminatory. They do not see it as an effective tool to fight discrimination, but rather as a polarising approach that focuses on differences between people rather than similarities. Moreover, it could harm communities who are the victim of racism. For many, collecting data on nationality is sufficient.

This opposition was also largely related to the **risks** participants associate with collecting data on origins. First and foremost is the **fear that these data could be used** to fuel discriminatory or stigmatising discourse or to reinforce stereotypes. Several focus group participants mentioned their fear of a form of ethnic census, a fear often linked to **past traumas** such as the Holocaust. A number of potential pitfalls were pointed out: misinterpretation of the data in the press or in politics, stigmatisation of a community, failure to respect the right to privacy. There were also fears that the data would fall into the wrong hands, especially if the far right came to power.

In the survey, we addressed **different contexts** in which data collection on origin could take place. Specifically, we asked for ten different contexts (census, scientific surveys, education, employment, health care, etc.) whether people **would be willing to answer** questions on origin in that context **and under what conditions** (confidentiality of data, supervision by an independent body, etc.).

However, we received feedback from various quarters that this question was very difficult to understand. This is partly due to the question format: the listed contexts are quite abstract, even though an explanation was provided for each context, but the conditions are also abstract and rigid. It can be difficult to put oneself in these abstract fictional situations. However, the method of a survey does not allow for more nuanced answers and sufficient depth.

On the other hand, the question can also be perceived as difficult because respondents must get used to the idea of data collection in certain contexts and are not familiar with how this can help fight discrimination. The interpretation of responses is also challenging because people could give multiple answers. Therefore, we consider the results of this question with caution and will only cover them in a limited way here.

We find that the proportion of people who would be willing to answer questions on origin **without objection** is highest for the '**scientific surveys**' and '**census**' contexts. The focus groups and open-ended responses to the survey question show that respondents generally found it acceptable for data on origin to be collected when the purpose is to combat discrimination, for example by equality bodies.

For other contexts such as **housing, employment, health care, and education**, participants had mixed opinions. The survey results show that the proportion of people who would not answer a self-identification question was highest for the contexts '**housing**', '**justice and police**', '**employment**', and '**media**'.

This is consistent with the results of the focus groups: although these are contexts where many inequalities persist, it seemed instinctively inappropriate for participants to collect data in these contexts. Consequently, participants would be more inclined to respond **provided that such data are collected for the purpose of measuring and combating discrimination and that this purpose is clearly explained**. Participants also mentioned other conditions that were important to them when collecting data, and these are discussed in the next section.

6.2 Specific conditions for collecting data on origin via self-identification

In this section, we discuss the specific conditions that participants considered essential when collecting data on origin through self-identification. They are structured around three main themes. For example, for many of those consulted, **collecting data on origin through self-identification is acceptable and opportune only under the following conditions**:

Respect for privacy and the data protection legislative framework

- Absolute compliance with the data protection legislative framework
- **Anonymity and confidentiality**: it must be impossible to identify individuals based on the data
- **Participation** in this type of data collection should be **voluntary**
- Data must be destroyed after a certain period
- Provision of **complaint procedures** in case of non-compliance with this legislative framework

Communication and transparency

- **Proactive communication** about the data collection, **adapted to the target audience**
- Absolute **transparency** on every stage of data collection and use
- Clarification of the concepts used, including what is meant by terms such as 'origin', 'race' or 'ethnic origin'
- Making the interpretation and analysis of data clear in order to reduce the risk of wrong conclusions being drawn from them

The legitimacy of the data collector and the goal of the data collection

- Transparency about **which institution or organisation** collects and/or processes the data
- Precise and clearly explained goal: it is necessary to ensure that the data are used responsibly, for the purposes of equality and the fight against racism and discrimination. The goal must be

considered **legitimate** by respondents. If not, participants may be suspicious and perceive this type of question as a 'trap'

- **Trust** in the institution asking the self-identification question: this is necessary before information about origins can be shared. A participatory approach can facilitate this

6.3 Questions on self-identification of origin

Opinion of focus group participants on self-identification of origin

The participants consulted during the focus groups generally regarded self-identification as a respectful method of collecting data because it allows them to choose how to express their origin. However, it was often pointed out that the concept of origin is very broad and abstract. For many people, **origin is difficult to grasp** because it consists of different facets and varies over time. **Each person defines themselves based on a combination of elements** such as the country of birth, nationality (original or acquired), the city where they live, the time they have been living in Belgium, the continent or region of origin, the culture or values they hold dear, etc. Yet participants most often defined their origin in terms of their nationality/nationalities or country of origin, often in the form of 'Belgian of ... origin' or 'Belgian ...'.

One observation that came up repeatedly in the focus groups was that **identification differs greatly depending on generation**: parents who arrived in Belgium may feel a strong bond with their country of origin, while their Belgian-born children feel fully Belgian. The reverse also occurs: after years of living in Belgium, a parent feels fully Belgian, but their child chooses to reaffirm their link with the country of origin where they have not lived.

Another important aspect that emerged from our focus groups is that **self-identification also depends on context**: depending on where you are, who you are talking to, what country you are in, what question you are asked and what group you belong to, the answer may vary. For example, someone of Moroccan origin may define themselves as 'Belgian of Moroccan origin' when in Belgium, as 'Moroccan from Rabat' when in Morocco, or as 'Belgian' when abroad.

In the following sections, we present the different self-identification questions we tested in the online survey, as well as an analysis of the responses to these questions. In drafting the questions and answer categories, we considered the concrete input from the focus groups.

Self-identification question with geographic response categories

The first self-identification question for origin in the survey (see box 1) presented respondents with a **list of geographic response categories**. In the remainder of the text, we will refer to this as the 'geographic self-identification question'. The categories consisted of the six largest nationalities of origin in Belgium, as well as the various continents. In doing so, we tried to strike a balance between sufficient detail so that respondents could identify with the categories on the one hand, and a simple and accessible list on the other.

BOX 1: SELF-IDENTIFICATION QUESTION WITH GEOGRAPHIC RESPONSE CATEGORIES

How do you define your origin(s)?

**The (alphabetical) list is made up of the six national origins most present in Belgium as well as the different continents.*

**You can select up to 4 answers, and/or add your own if none of the options suit you.*

- | | |
|--|--|
| <input type="checkbox"/> Asia | <input type="checkbox"/> North Africa |
| <input type="checkbox"/> Belgium | <input type="checkbox"/> North America |
| <input type="checkbox"/> Europe in the European Union | <input type="checkbox"/> Oceania |
| <input type="checkbox"/> Europe outside the European Union | <input type="checkbox"/> South / Central America |
| <input type="checkbox"/> France | <input type="checkbox"/> Sub-Saharan Africa |
| <input type="checkbox"/> Italy | <input type="checkbox"/> Turkey |
| <input type="checkbox"/> Morocco | <input type="checkbox"/> I don't know |
| <input type="checkbox"/> Near/Middle East | <input type="checkbox"/> I would rather/also identify myself as: _____ |
| <input type="checkbox"/> Netherlands | <input type="checkbox"/> I prefer not to answer |

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

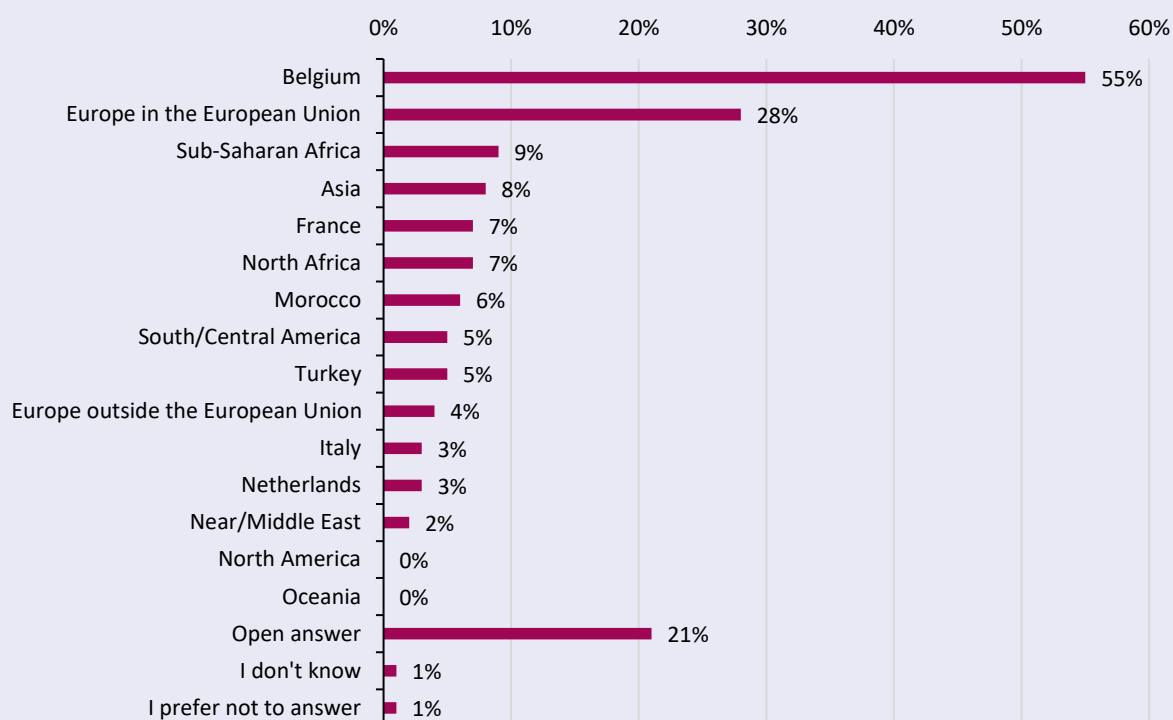
Respondents could indicate up to four response categories. 52% of respondents indicated one answer. 32% indicated two answers, 14% three answers. **Very few respondents went up to four origins.** In chart 1 below, we present the distribution across the different geographic answer categories. The total adds up to more than 100% because people could indicate multiple answers at the same time. **We note that 55% of respondents indicated 'Belgium'.** 'Europe in the EU' is the second most common answer category.

'Belgium' and 'Europe in the EU' were often indicated together: 27% of people who indicated Belgium also indicated 'Europe in the EU'. This could include people of Belgian origin who also identify with Europe (as a broader category), but also people with origins in both Belgium and another EU country. 'Belgium' and the open category also occur together often: 23% of the people who indicated 'Belgium' also indicated the open category. The analysis of these open answers shows that these are often people who indicate having origins in Belgium and a specific other country that is not in the list.

In total, the open category was indicated 76 times, of which 66 times by people who had already indicated another category. The analysis of these open answers shows that some of them can be reduced to categories that were also in the multiple-choice list. Sometimes the open answers indicated a combination of one of the categories in the list and another origin that was not included in the list.

The open-ended responses made many references to specific countries that were not included in the list. This indicates that many people define their origins by reference to a **specific country**. We see a similarity with the self-identification of **focus group participants**, most of whom defined their origins in terms of their **nationality/nationalities or country of origin**, sometimes in combination with other elements.

Chart 1: Distribution of geographic response categories in the sample (N=370)



People were allowed to indicate multiple answers, so the sum of the above percentages is more than 100%.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

A number of open responses in the survey referred to terms that were not listed but were listed in the (next) self-identification question based on **ethnocultural** response categories, such as Berber, Jewish, Roma, Black.

A few answers referred to an alternative, such as being 'human' or 'global citizen'. This group therefore did not seem to want to place themselves in a particular category. Finally, some answers also referred to other regions of the world or to specific regions of Belgium.

Self-identification question with ethnocultural response categories

The second self-identification question in the survey was worded slightly differently and offered a **different list of response categories based on ethnocultural elements** (see box 2). In the remainder of this report, we will refer to this as the 'ethnocultural self-identification question'. The list is based **on input from the focus groups, the expert interviews, and the literature review with good practices from other countries**. This showed, among other things, that it is important to provide separate categories for 'Roma' and 'Travellers' because these are two different groups. Moreover, it showed that people of Jewish origin are an important group to include in this study. Indeed, there is little data on the discrimination this group experiences, but there may be a great fear of defining themselves as 'Jewish'.⁷⁷ The focus groups showed that it is important to include the response category separately

⁷⁷ European Union Agency for Fundamental Rights, (2018), [Experiences and perceptions of anti-Semitism - Second survey on discrimination and hate crime against Jews in the EU](#), p.7.

in the self-identification question because it is an origin that helps define people's origin in addition to (for example) their nationality.

BOX 2: SELF-IDENTIFICATION QUESTION WITH ETHNOCULTURAL RESPONSE CATEGORIES

Which of these groups do you consider yourself to belong to?

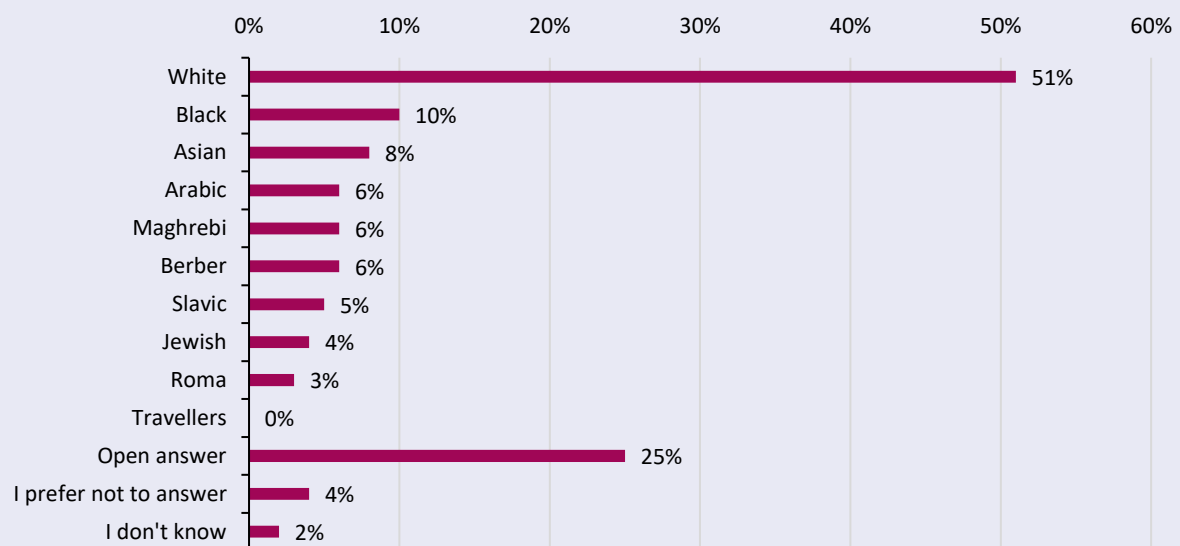
**You can select up to 4 answers, and/or add your own if none of the options suit you.*

- | | |
|-----------------------------------|--|
| <input type="checkbox"/> Arabic | <input type="checkbox"/> Slavic |
| <input type="checkbox"/> Asian | <input type="checkbox"/> Travellers |
| <input type="checkbox"/> Berber | <input type="checkbox"/> White |
| <input type="checkbox"/> Black | <input type="checkbox"/> I don't know |
| <input type="checkbox"/> Jewish | <input type="checkbox"/> I would rather/also identify myself as: _____ |
| <input type="checkbox"/> Maghrebi | <input type="checkbox"/> I prefer not to answer |
| <input type="checkbox"/> Roma | |

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

In chart 2 below, we present the distribution across the different ethnocultural response categories. We see that 51% of respondents indicated 'White'. The open answer is the second most common answer category.

Chart 2: Distribution of ethnocultural response categories in the sample (N=370)



People were allowed to indicate multiple answers, so the sum of the above percentages is more than 100%.

Source: 'Measuring discrimination and inequality using self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

76.5% of respondents indicated only one answer to this question. This is significantly more than for the geographic response categories. The most common combinations are:

- 'White' + 'open category'
- 'White' + 'Slavic'
- 'Black' + 'open category'
- 'Asian' + 'open category'
- 'Arabic' + 'Berber' + 'Maghrebi'

Other analyses of this question also show a **large overlap between these last three response categories**. We see, for example, that of those who indicated 'Maghrebi', 52% also indicated 'Arabic' and 48% also indicated 'Berber'. Of those who indicated 'Berber', 50% also indicated 'Maghrebi'. Similarly, of those who indicated 'Arabic', 50% indicated 'Maghrebi'. So here the question arises whether it is necessary to include all three categories in the list or whether we could perhaps simplify it by omitting one of the three. Those participating in the focus groups who were of North African origin were more likely to define their origin by nationality or country of origin rather than by these different terms.

In the **open answers** to this question, there were many references to **terms that were not listed but were listed in the geographic self-identification question**, such as 'Belgian' or 'European', sometimes in combination with another term.

Some of the open answers referred to **'mixed'** (18 people in total). 'Métisse' was almost always used in French, while 'mixed' or 'gemengd' was used in Dutch. This seems to indicate that people do not always find it sufficient to indicate two response categories (e.g. 'Black' + 'White') when they want to reflect their mixed origins.

Some answers referred to other regions of the world not included in the list, such as 'Latin America' and 'Mediterranean'. Eleven answers referred to another discrimination criterion, including three that mentioned 'Muslim'.

Comparison between geographic and ethnocultural response categories

To better understand the appropriateness of these self-identification questions and response categories, **we put both questions side by side and analysed the response patterns**. Table 1 shows what percentage of those who indicated a particular geographic response category subsequently indicated a particular ethnocultural response category. For readability, this table presents only a selection of the different response categories (the full table is available on [this webpage](#) in Dutch/French).

We find that there is a large overlap between people who answered 'North Africa' to the question with geographic response categories on the one hand and people who answered 'Maghrebi', 'Berber' or 'Arabic' to the question with ethnocultural response categories on the other. We can note that **'Maghrebi' has the largest overlap with 'North African'**. In addition, we find that 89% of those who answered **'Sub-Saharan Africa'** to the question with geographic categories answered **'Black'** to the question with ethnocultural categories. We also see a large overlap between 'Belgium' and 'White' and between 'Asia' and 'Asian'.

Of those who indicated **'South/Central America'** in the geographic self-identification question, we see that 58% indicated the open category in the ethnocultural self-identification question, which is remarkably high (see table 1). This could indicate that they **could not identify with one of the proposed ethnocultural categories**. This is confirmed by the analysis of the open-ended responses,

where 'Latin American' appeared several times. Identification with Latin America had also emerged in one of the focus groups.

Among those who answered 'Near/Middle East' or '**Turkey**' in the geographic self-identification question, we also see a striking pattern in the ethnocultural self-identification question. Of the 19 respondents who answered 'Turkey' as geographic origin, **26% refused to answer in the ethnocultural self-identification question**. We must be careful with the interpretation here because this is a small group, but this could possibly indicate that people of Turkish origin found it difficult to agree with any of the proposed ethnocultural response categories. When designing the questionnaire, we chose to use as few geographic categories as possible in the self-identification question with ethnocultural response categories so as not to mix the two. Hence, we decided not to include 'Turkish' in the ethnocultural self-identification question. **The analysis shows that it may be opportune to mix geographic and ethnocultural categories in the answer options.**

Table 1: Comparison between geographic and ethnocultural response categories

Ethnocultural response categories	Geographic response categories				
	North Africa (n=25)	Sub-Saharan Africa (n=35)	South / Central America (n=19)	Near/Middle East (n=6)	Turkey (n=19)
Arabic	32%	0%	0%	50%	0%
Asian	0%	0%	5%	17%	47%
Berber	44%	0%	0%	0%	0%
Black	4%	89%	5%	0%	0%
Maghrebi	64%	0%	0%	0%	0%
Open answer	32%	31%	58%	0%	32%
I prefer not to answer	0%	0%	0%	0%	26%
I don't know	0%	0%	5%	33%	5%

For readability, this table contains only a selection of all origin groups. For the full table (in Dutch/French), see [this webpage](#).

The above table should be read **column by column**. For example: of those who indicated 'North Africa' when asked about geographic origins, 32% indicated 'Arabic' when asked about ethnocultural origins, 0% indicated 'Asian', etc.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

Open question on origins

The third self-identification question in the survey was a **non-mandatory open question on origins** (see box 3). It was filled in by 302 people and hence not by 68 people. Of the 15 people who refused to answer one or both of the previous self-identification questions, seven did fill in the open question.

BOX 3: SELF-IDENTIFICATION QUESTION WITH OPEN ANSWER

How do you yourself usually define your origin(s)? _____

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

Most responses referred to **geographic categories**, with 'Belgian' being used often (86 times). In addition, 'European' (13 times), 'Moroccan' (12 times) and 'Turkish' (9 times) were also common. 'Belgian' was often combined with another origin. In total, 59 times a specific country was mentioned that did not appear in the multiple-choice lists of the previous questions. As the results of the focus groups also showed, **geographic origins are therefore a common way of defining oneself**.

26 responses could be brought down to the **ethnocultural categories suggested earlier**. The most common was White (13 times). However, the following ethnocultural categories were also cited: Jewish, Roma, Black, Berber, Maghrebi and Slavic.

Some answers again referred to **regions of the world** not included as such in the lists of previous questions, such as 'Latin America' and 'Africa'. Some answers also referred to a **specific region of Belgium**, such as 'Flemish', 'Brussels' or 'Walloon', or to provinces or cities. A few answers referred to another discrimination criterion such as 'social origin', e.g. 'white Belgian middle class'. Finally, there were six people who indicated that they do not want to or cannot define themselves in this way.

It is important to note that **the location of the question in the survey** is important. Here, we chose to place this question after two self-identification questions in which different answer categories were proposed, so as to test the acceptance of these different categorisations without prior reflection. This choice creates a **bias**: participants' answers to the open-ended self-identification question may be influenced by the pre-proposed answer categories. So, the opposite choice could also be interesting: letting participants freely define their origin first before answering questions with categories.

The auto-hetero perception question compared with the self-identification questions

Next, we asked respondents how they think they are perceived by others (see box 4). This is called 'auto-hetero perception'. The suggested answer categories for this question were the same as for the self-identification question with ethnocultural response categories, in order to compare the two questions. We chose the **ethnocultural response categories** because they **are more linked to visible characteristics that are often the source of discrimination**. Testing both types of response categories in an auto-hetero perception question would unfortunately have made the survey too long.

On the auto-hetero perception question, **73% of respondents indicated only one answer**. This is similar to the ethnocultural self-identification question, and a higher proportion than for the geographic self-identification question. What is striking from the analysis of the open-ended responses to the auto-hetero perception question is that eight people indicated that they were perceived as 'Muslim' by others.

For certain response categories, we see a **large overlap between responses to the two questions** (see table 2). Of those who defined themselves as '**Black**', 92% said they were also seen as 'Black' by others. We also see large overlap between the two questions for the categories '**White**' (83%), '**Arabic**' (75%) and '**Asian**' (66%).

BOX 4: AUTO-HETERO PERCEPTION QUESTION

In this question, we are considering how you think you are perceived by others, even if this is not how you feel. Discrimination is often based on how a person is perceived by others.

How do you think others see you?

**You can select up to 4 answers, and/or add your own if none of the options suit you.*

- | | |
|-----------------------------------|--|
| <input type="checkbox"/> Arabic | <input type="checkbox"/> Slavic |
| <input type="checkbox"/> Asian | <input type="checkbox"/> Travellers |
| <input type="checkbox"/> Berber | <input type="checkbox"/> White |
| <input type="checkbox"/> Black | <input type="checkbox"/> I don't know |
| <input type="checkbox"/> Jewish | <input type="checkbox"/> Other people see me rather/also as: _____ |
| <input type="checkbox"/> Maghrebi | <input type="checkbox"/> I prefer not to answer |
| <input type="checkbox"/> Roma | |

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

For other response categories, however, we see a **very low overlap**. Of those who defined themselves as '**Berber**', only 18% said they were seen as 'Berber' by others. 68% of them said they were seen as 'Maghrebi', and 68% said they were seen as 'Arabic'. 'Berber' thus appears to be a category by which people define their own origin, but to which they are not often linked by others. Among people who said they were '**Jewish**', only 31% said they were seen as 'Jewish'. 69% of them reported being seen as 'White' and 23% as 'Maghrebi'.

Table 2: Comparison between ethnocultural self-identification and auto-hetero perception

Perception by others	Self-identification with ethnocultural response categories					
	Arabic (n=24)	Asian (n=29)	Berber (n=22)	Black (n=38)	Jewish (n=13)	White (n=188)
Arabic	75%	7%	68%	3%	8%	3%
Asian	0%	66%	0%	3%	0%	2%
Berber	8%	0%	18%	3%	8%	2%
Black	0%	7%	0%	92%	0%	3%
Jewish	0%	0%	0%	0%	31%	3%
Maghrebi	54%	10%	68%	3%	23%	4%
White	12%	10%	9%	3%	69%	83%
Open answer	17%	34%	27%	21%	8%	13%

For readability, this table contains only a selection of all origin groups. For the full table (in Dutch/French), see [this webpage](#).

The above table should be read **column by column**. For example: of those who indicated 'Arabic' on the self-identification question on ethnocultural origins, 75% indicated 'Arabic' on the question on ethnocultural origins, 0% indicated 'Asian', etc.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

Experiences of discrimination and self-identification questions

We then probed for experiences of discrimination based on origin (see box 5). We note that in the survey evaluation questions, several respondents were critical of the way this question was asked. Specifically, this question would suggest that white people cannot be discriminated based on their origin. However, we mention in this question that **discrimination is often experienced by people perceived as non-white, non-Belgian or of foreign origin**, as also confirmed by the literature. **Thus, the formulation of the question does not exclude the possibility that white people can also be discriminated based on their origin.** This formulation is inspired by the German National Monitoring of Discrimination and Racism (NaDiRa)⁷⁸. Moreover, the focus groups indicated that people are often discriminated against because they have a non-white skin colour or for their physical characteristics that are seen as 'non-Belgian'. However, a different wording of the question might have made this even clearer.

BOX 5: QUESTION ABOUT EXPERIENCES OF DISCRIMINATION ON THE BASIS OF ORIGIN

Equality is a fundamental right. Unfortunately, some people are subject to racial discrimination. Discrimination is the unfair or unequal treatment of a person on the basis of personal characteristics. Among other things, the law prohibits behaviours that target certain people or groups of people on the basis of their nationality, so-called "race", skin colour, ancestry, national origin or ethnic origin.

Discrimination is often experienced by people who are perceived as non-white, foreign or of foreign origin. Have you ever been treated in a discriminatory way on the basis of these characteristics?

- ☐ Yes
- ☐ No
- ☐ I don't know
- ☐ I prefer not to answer

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

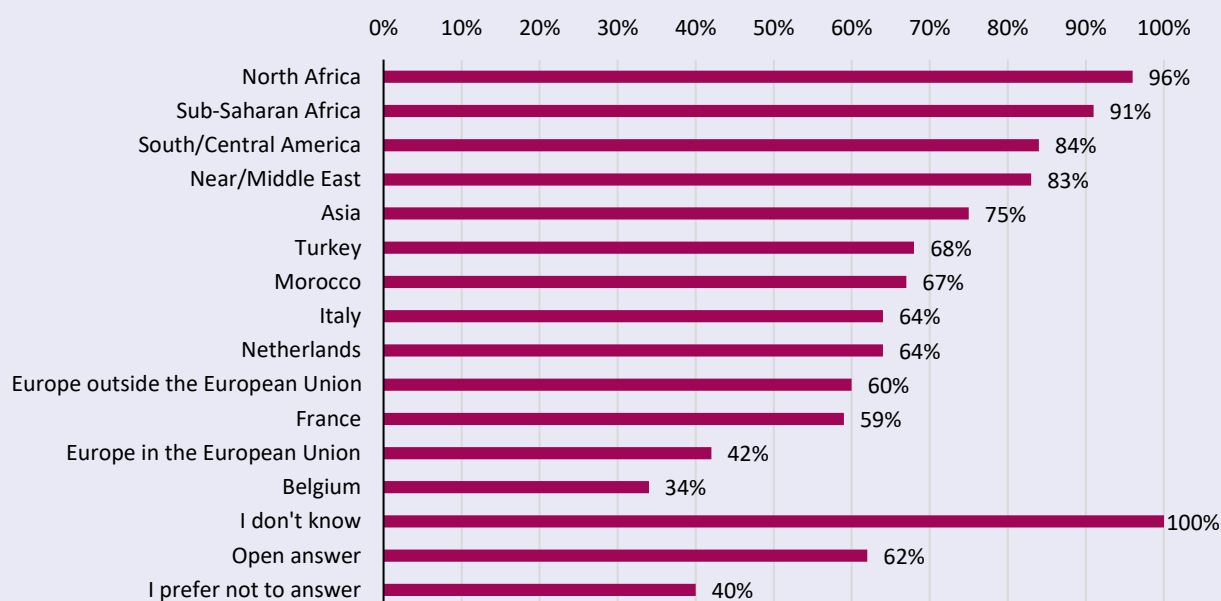
To the question "Have you ever been treated in a discriminatory way on the basis of these characteristics?", **51.4% of respondents** answered "**Yes**", 43% "No" and 4.6% "I don't know". We must be careful when crossing this data with reported origin, as it sometimes involves small groups in our sample. But it does give us an indication of which groups in our sample frequently face discrimination.

If we look at the intersection between responses to geographic categories and who reported having 'already experienced discrimination' (see chart 3), we see that **a high percentage of those defining themselves as 'North African', 'Sub-Saharan African', 'South/Central American' or 'Near/Middle Eastern' reported having experienced discrimination.**

When comparing with the ethnocultural self-identification question (see chart 4), we see especially that **of those defining themselves as 'Asian', 'Maghrebi' or 'Black', a high proportion reported having experienced discrimination.**

⁷⁸ DeZIM, (2021), [Methodenbericht zur Studie "Rassistische Realitäten"](#).

Chart 3: Percentage that reported having already experienced discrimination, by geographic response category (N=190)

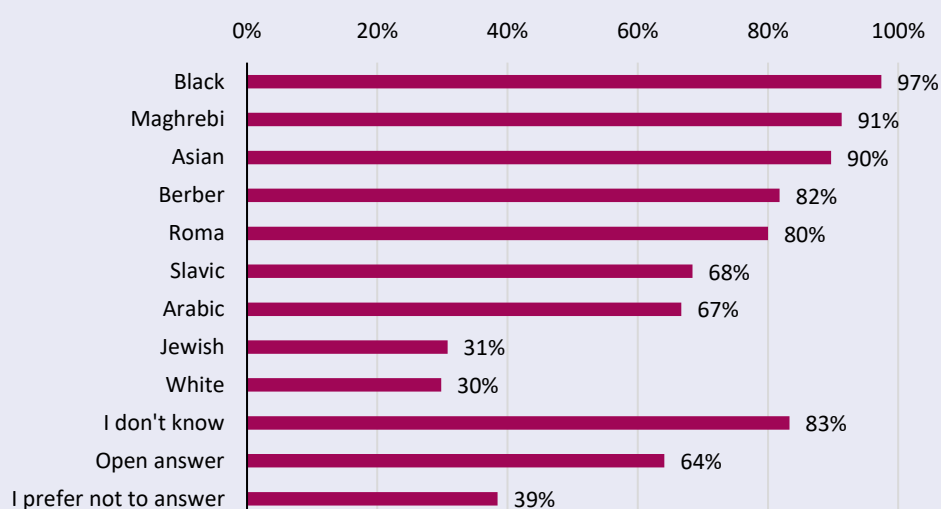


This chart contains only the percentages of those who answered 'yes' to the question about discrimination experiences (N=190). For the full table with all response categories (in Dutch/French), [see this webpage](#).

Note that some subgroups are small. These percentages should therefore be interpreted with caution.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

Chart 4: Percentage that reported having already experienced discrimination, by ethnic-cultural response category (N=190)



This chart contains only the percentages of those who answered 'yes' to the question about discrimination experiences (N=190). For the full table with exact all response categories (in Dutch/French), [see this webpage](#).

Note that some subgroups are small. These percentages should therefore be interpreted with caution.

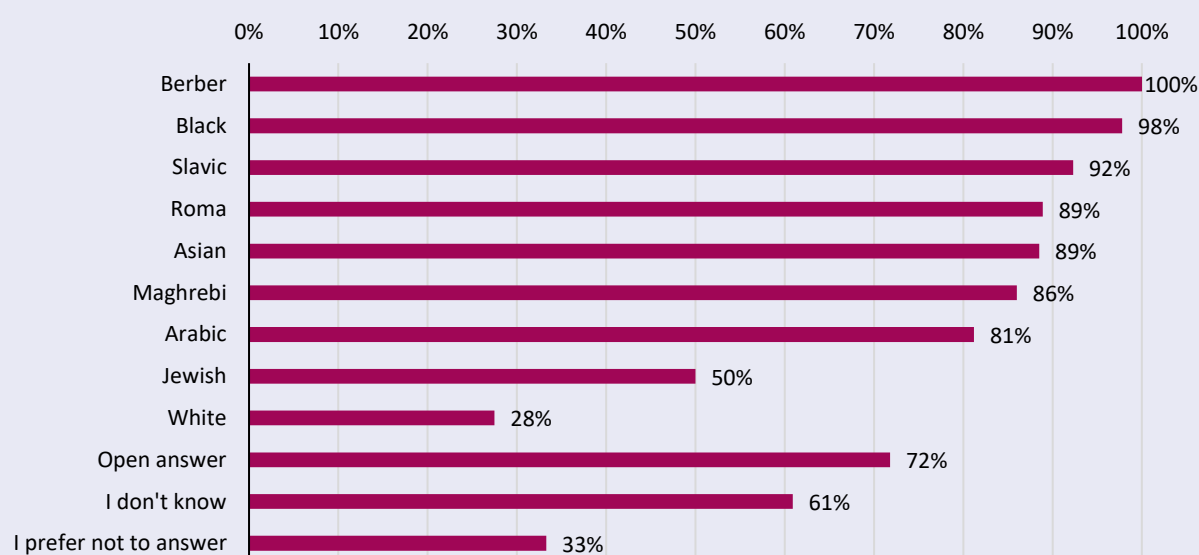
Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

Finally, we put the auto-hetero perception question alongside the question on experiences of discrimination (see chart 5). In the **auto-hetero perception question**, people could indicate how they think they are perceived by others. Thus, the comparison with this question is **interesting** because **discrimination is also based on how others see us**.

For several categories of the auto-hetero perception question, the proportion of people who have experienced discrimination appears to be **slightly higher than for those same categories of self-defined origin**. This is particularly the case for people who think they are perceived as 'Arabic', 'Berber', 'Black', 'Roma' or 'Slavic'.

For people who **define themselves as 'Jewish'**, the proportion who reported experiencing discrimination was low (31%); however, the proportion is much higher when we look at people who **reported being perceived as 'Jewish'** (50%). These are often small groups so we must be careful in our interpretation, but these analyses do suggest that **it may be pertinent to use the auto-hetero perception question if the aim is to identify the groups that experience discrimination**.

Chart 5: Percentage that reported having already experienced discrimination, by auto-hetero perception (N=190)



This chart contains only the percentages of those who answered 'yes' to the question about discrimination experiences (N=190). For the full table with exact percentages and numbers for all response categories (in Dutch/French), [see this webpage](#).

Note that some subgroups are small. These percentages should therefore be interpreted with caution.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

6.4 Comparison between proxy variables and self-identified origin

One of the aims of this study is to compare the self-identification method for origin with the proxy method as commonly used in Belgium. In this way, we can study what the differences are between the two methods, which different groups they identify and what the added value of the self-identification method is. In this section, we thus seek answers to the following questions:

1. Does the **self-identification method** make it possible to **identify groups at risk of discrimination** based on visible or ethnocultural characteristics, that are **not identified** by the **proxy method**?
2. Conversely, does the **proxy method** make it possible to identify groups at risk of discrimination that are **not identified by the self-identification method** (e.g. due to under-reporting)?

The proxy method for measuring origin in our survey

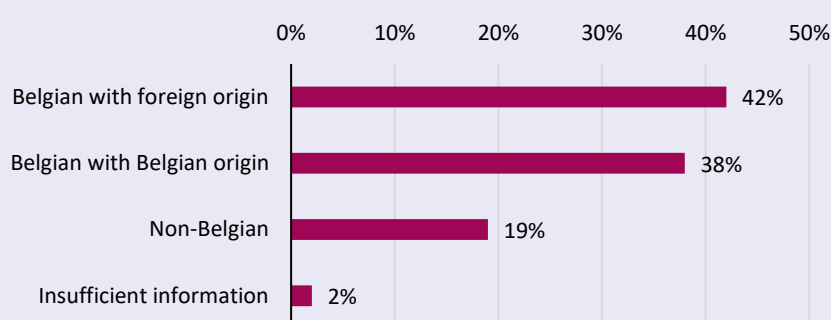
To make a comparison between the proxy method and the self-identification method, **we created a proxy variable based on the survey questions about (birth) nationality**. For this, we followed the method used by Statbel to create their origin variable, to the extent possible⁷⁹. For more details on the method used to create this proxy variable, see the extensive documentation available in Dutch/French [on this webpage](#).

For this variable we use the respondent's current nationality, their birth nationality, and the birth nationalities of their parents, as completed by respondents in the survey. In this way, we divide the individuals in our sample into the four categories below (in line with the Statbel categories):

1. Non-Belgian
Current nationality is not Belgian
2. Belgian with foreign origin
Current nationality is Belgian, but respondent was either born with non-Belgian nationality or one or both parents were born with non-Belgian nationality
3. Belgian with Belgian origin
Current nationality is Belgian, and both parents were born with Belgian nationality
4. Insufficient information

The distribution of these four groups in our sample is shown in chart 6.

Chart 6: Distribution of the sample according to the proxy variable for origin (N=370)



Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

⁷⁹ We derived the Statbel method from the decision trees in the Excel files on [this page](#). We deviate from this method in the following ways:

- The Statbel method uses administrative data from the National Register. Since we do not have this, we use the data that respondents filled in the survey.
- Instead of 'first registered nationality', we use 'birth nationality'.
- If there are dual (birth) nationalities, and one of them is Belgian, the Belgian (birth) nationality always takes precedence for determining origin. We have no insight into the method Statbel uses to deal with dual nationalities.
- We explicitly distinguish groups for which insufficient information is available.

Which groups at risk of discrimination are not identified with a proxy variable but are identified with the self-identification method?

For this section, we focus on the group belonging to the '**Belgian with Belgian origin**' category according to the proxy variable discussed above. The analyses below are done only for this group, which **constitutes 38% of the total sample**. Specifically, we look at who within this group indicated a non-Belgian or non-white response category to the self-identification questions.

The purpose of this is to check the following: according to the proxy method, who among the group seen as Belgian of Belgian origin - and thus **presumed to be less at risk of being discriminated against based on their origin** - does belong, according to the self-identification method, to a group at greater risk of being discriminated against, i.e., the group that indicated a non-Belgian or non-white answer category?

We realise that we are simplifying things here and that the reality is more complex. Moreover, we do not want to state here that people of Belgian origin or white people cannot be discriminated against because of their origin. But the cases opened by Unia to further investigate a report of discrimination, as well as various scientific studies and field tests, show that people of non-Belgian origin or people with a non-white skin colour are at greater risk of discrimination in Belgium.⁸⁰

We first compare the proxy variable with the geographic self-identification question. To do this, we regroup respondents into four groups based on their response patterns to this self-identification question. This is a simplified representation that serves only to enable comparison with the proxy variable:

- People who indicated only 'Belgium' and whom we could consider as 'persons with **Belgian origins**'.
- People who indicated only 'Belgium' and 'Europe in the EU'. This can mean two things: (1) 'I define my origin in Belgium and therefore in Europe in the EU'. (2) 'I define my origin in Belgium and in another European country in the EU'. We consider this group here as 'persons with **Belgian/EU origins**'.
- People who indicated 'Belgium' as well as one of the other categories that is not 'Europe in the EU'. We could consider these people as 'persons with **mixed origins**'.
- People who indicated only one or more categories other than 'Belgium'. We could consider these people as 'people with **non-Belgian origins**'.

How are the respondents who were categorised as 'Belgian with Belgian origin' according to the proxy variable distributed among these four groups? As is shown in table 3, 24,5% of this group of respondents have mixed origins according to the above categorisation and 5% have foreign origins.

Taken together, this means that **29.5% of people who are considered 'Belgian with Belgian origin' based on the proxy variable are considered being of mixed or non-Belgian origin based on a self-identification question**. This is an important finding: a fairly large proportion of people **identified by the self-identification method** as being at risk of discrimination **are not identified using the proxy variable**.

⁸⁰ Unia (2022), [Annual Report](#)

Lippens L., Vermieren S., Baert, S., (2023), [The state of hiring discrimination: A meta-analysis of \(almost\) all recent correspondence experiments](#)

Stad Gent, [Praktijktesten op de private huurmarkt](#)

Table 3: Self-identification with geographic response categories for those considered 'Belgian with Belgian origin' according to the proxy variable (N=139)

Geographic origins grouped together	Number	Percentage
Only indicated Belgium = ' Belgian origins '	72	51,8%
Indicated Belgium + EU = ' Belgian/EU origins '	26	18,7%
Indicated Belgium + a category other than EU = ' Mixed origins '	34	24,5%
Only indicated a category other than Belgium = ' Non-Belgian origins '	7	5,0%

The percentages in this table are calculated for the subgroup of people considered as 'Belgian with Belgian origin' according to the proxy variable (n =139), not for the full sample.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

Next, we go through the same exercise for the ethnocultural self-identification question. We regroup responses to this self-identification question into the following three categories:

- People who indicated only 'White'. These people could be seen as **belonging to a white group**.
- People who indicated 'White' as well as another category. These people could be seen as **belonging to a mixed group**.
- People who only indicated a category other than 'White'. We could think of these people as **belonging to a non-white group**.

In table 4 we look at the distribution across these three categories of the group of respondents who were categorised as 'Belgian with Belgian origin' based on the proxy variable. 19% of this group only indicated a category other than 'white' on the self-identification question, while 12% indicated 'White' as well as another category.

Taken together, this means that **31% of those seen as 'Belgian with Belgian origin' based on the proxy variable indicated belonging to the mixed or non-white group when they answered the self-identification question**.

Again, **a fairly large proportion of the individuals who are identified with the self-identification method as being at risk of discrimination are not identified with the proxy variable**. Interestingly, this proportion is similar to the proportion we saw for the other self-identification question.

Table 4: Self-identification with ethnocultural categories for those considered 'Belgian with Belgian origin' according to the proxy variable (N=139)

Ethnocultural categories grouped together	Number	Percentage
Only indicated 'White' = ' White '	94	67,6%
Only indicated a category other than 'White' = ' Non-White '	26	18,7%
Indicate 'White' + another category = ' Mixed '	16	11,5%
I prefer not to answer/I don't know	3	2,2%

The percentages in this table are calculated for the subgroup of people considered as 'Belgian with Belgian origin' according to the proxy variable (n=139), not for the full sample.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

Finally, we repeat the same exercise for the auto-hetero perception question, using the same classification as for the ethnocultural self-identification question (see table 5). This shows that, **taken together, 21,6% of those categorised as 'Belgian with Belgian origin' according to the proxy variable indicated that others perceive them as belonging to a non-white or mixed group.** This is clearly less than in the self-identification questions, but still quite a lot.

We can therefore conclude that **a significant proportion of the groups at increased risk of discrimination in our sample are not captured when we use the proxy variable.** This may include, for example, people whose origins go further back than their parents, namely third- or fourth-generation individuals. It could also concern people who are discriminated against based on their non-white skin colour but whose parents had the Belgian nationality at birth, for example adopted persons.

It may also concern certain groups not captured by proxies based on (birth) nationality, such as **people of Jewish origin or Roma.** Due to the limitations of our sample, we cannot simply generalise the percentages to the entire population, but the **results show that the self-identification method could be a good complement to the proxy method.** We therefore call for further research on this in our recommendations.

Table 5: Auto-hetero perception for those considered 'Belgian with Belgian origin' according to the proxy variable (N=139)

Auto-hetero perception categories grouped	Number	Percentage
Only indicated 'White' = 'White'	105	75,5%
Only indicated a category other than 'White' = 'Non-White'	17	12,2%
Indicated 'White' + another category = 'Mixed'	13	9,4%
I prefer not to answer/I don't know	4	2,9%

The percentages in this table are calculated for the subgroup of people considered as 'Belgian with Belgian origin' according to the proxy variable (n=139), not for the full sample.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

Which groups at risk of discrimination are not identified with the self-identification method but are identified with a proxy variable?

We then analyse it in the opposite direction: who is considered to be of Belgian origin according to the self-identification questions that were used, but is not considered to be 'Belgian with Belgian origin' according to the proxy variable? In other words, the aim is to find out whether there are **groups that would be identified as less at risk of being discriminated against based on their origin according to the self-identification questions, but that are considered to be more at risk of being discriminated against according to the proxy method.**

We first make the comparison between the geographic self-identification question and the proxy variable. Here, we only analyse the group that could be seen as 'persons with Belgian/EU origins' based

on this self-identification question, as they indicated only Belgium or only Belgium and 'Europe in the EU'⁸¹.

For this group only, we look at the distribution across the three groups of the proxy variable in table 6. **Of those who indicated on the geographic self-identification question that they are of Belgian origin or of Belgian/EU origin, 10% are not considered 'Belgian with Belgian origin' according to the proxy variable.** This percentage is a lot lower than the 30% we saw in the analysis in the opposite direction. We should note here, however, that we are talking about a small number of people, namely 11 respondents.

Table 6: Classification into proxy groups for those who answered only 'Belgium' (or only 'Belgium' + 'Europe in the EU') to the geographic self-identification question (N=109)

Proxy variable	Number	Percentage
Belgian with Belgian origin	98	89,9%
Belgian with foreign origin	9	8,3%
Non-Belgian	2	1,8%

The percentages in this table are calculated for the subset of people who answered only Belgium or only Belgium + Europe in the EU (n=109), not for the full sample.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

Next, we execute the same analysis for the ethnocultural self-identification question. Here, we look specifically at those who answered only 'White' for this self-identification question and would therefore be considered less at risk of discrimination based on their origin. For this group, we look at the distribution across groups of the proxy variable (see table 7). Here we find that of those who defined themselves as 'White', 19,4% are in the 'Belgian with foreign origin' group based on the proxy variable and 12,9% are in the 'Non-Belgian' group.

Taken together, 32,3% of those who answered 'White' in the self-identification question were categorised as a person with foreign origin based on the proxy variable.

This is a fairly large proportion of the persons suspected to be at risk of discrimination who are not identified by the self-identification method but who are identified by the proxy method. This is also considerably more than the 10% based on the geographic self-identification question.

However, when we **break down in more detail** the group who has a foreign origin according to the proxy variable but who answered 'White' in the self-identification question, we get a slightly different picture⁸². Indeed, of the 45 people who defined themselves as 'White' but who are considered to have

⁸¹ We made the choice here to include both those who indicated only 'Belgium' and those who indicated only 'Belgium' and 'Europe in the EU' in the comparison category. As explained earlier, indicating both 'Belgium' and 'Europe in the EU' can mean two things: either people define their origins as Belgian and thus by definition also as European, or they define their origins as both Belgian and from another EU country that is not France, the Netherlands or Italy. Due to the way the response categories were constructed, we cannot determine what respondents meant. So, we had to make a choice in this analysis as to which (non-perfect) category we would take as a comparison group. We decided to include those who indicated both 'Belgium' and 'Europe in the EU' because we suspect that this group consists largely of people of Belgian origin who also identify with the EU. But this choice is debatable, and this should therefore be kept in mind when interpreting the results.

⁸² Again, using the Statbel method, we divide the respondents considered as 'Non-Belgian' and 'Belgian with foreign origin' according to the proxy variable in more detail according to their specific nationality of origin. Based on this more detailed

a foreign origin according to the proxy variable, **the vast majority turn out to be of European origin** according to the proxy variable based on their nationality of origin. This somewhat nuances the conclusion that a fairly large proportion of those at higher risk of discrimination are not identified by the self-identification methods.

Table 7: Classification into proxy groups for those who answered only 'White' to the ethnocultural self-identification question (N=139)

Proxy variable	Number	Percentage
Belgian with Belgian origin	94	67,6%
Belgian with foreign origin	27	19,4%
Non-Belgian	18	12,9%

The percentages in this table are calculated on the subgroup of people who only answered 'White' (N=139), not on the full sample.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

We can conclude that **some of the groups at increased risk of discrimination in our sample are not identified via the self-identification method whereas they are identified via the proxy method. However, this is a smaller proportion than in the other direction.** This may include, for example, people who feel Belgian but who don't have the Belgian (birth) nationality and who don't have parents with that (birth) nationality either. As described in section 3.4 'Advantages and disadvantages of self-identification', there may also be other reasons for this under-reporting. Again, we cannot generalise these percentages and conclusions to the entire population.

6.5 Evaluation and discomfort with survey questions

In this section, we look at respondents' evaluation of the various questions on origin and how (un)comfortable they felt with these questions. At the end of the survey, respondents were asked to evaluate the questions that were asked earlier in the survey. It is important to keep in mind that this evaluation question was perceived as difficult by several participants, not only because of its form, but also because the evaluation came at the end of the questionnaire and not after each question.

Table 8 shows the percentage of respondents who felt 'very comfortable', 'somewhat comfortable', 'somewhat uncomfortable', or 'very uncomfortable' when answering that question. It is notable that **the questions on nationality and birth nationality of (grand)parents involved the least discomfort.** But even for the geographic self-identification question, 60% of respondents were 'very comfortable' and 19,7% were 'somewhat comfortable'. The open-ended question was also evaluated quite positively.

Three questions were evaluated less positively: the ethnocultural self-identification question, the auto-hetero perception question, and the questions on experiences of discrimination.

proxy variable by nationality of origin, we can then divide people into groups in a different way, namely 'people with Belgian origin', 'people with European origin' and 'people with non-European origin'. The latter group are people with or without the Belgian nationality who have origins in a non-European country. Of those who defined themselves as white, 67,6% were of Belgian origin according to the nationality of origin proxy, 28,1% of European origin and 4,3% of non-European origin.

Table 8: Evaluation of the questions: percentage of people who indicated discomfort (N=370)

Question	Very comfortable	Somewhat comfortable	Somewhat uncomfortable	Very uncomfortable	I don't know
Self-identification of geographic origin	60,0%	19,7%	8,4%	4,6%	7,3%
Self-identification of ethnocultural origin	44,9%	20,3%	18,9%	7,8%	8,1%
Open question on origin	54,6%	24,6%	9,5%	3,8%	7,6%
Auto-hetero perception	43,5%	24,3%	17,0%	5,9%	9,2%
Experiences of discrimination	41,1%	25,7%	12,2%	8,9%	12,2%
Elements of discrimination	39,7%	24,3%	13,5%	7,0%	15,4%
Current nationality	70,0%	14,3%	4,9%	3,8%	7,0%
Birth nationality of parents	68,6%	15,1%	4,1%	4,3%	7,8%
Birth nationality of grandparents	67,6%	13,2%	5,9%	4,9%	8,4%

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

If we start looking at the response patterns to the ethnocultural self-identification question and the auto-hetero perception question, we see that there were fewer people who refused to complete (i.e. who indicated 'I prefer not to answer') the auto-hetero perception question (1,6%) than there were people who refused to complete the ethnocultural self-identification question (3,5%). On the other hand, the answer 'I don't know' was indicated more often for the auto-hetero perception question (6,2% vs. 1,6%). It is difficult to determine whether the discomfort was caused by the question itself, the proposed (ethnocultural) response categories, or a combination of both.

The largest proportion of people who felt 'very uncomfortable' is found for the question on experiences of discrimination (8,9%). With this question, it appears that people who indicated experiencing discrimination felt less comfortable with the question than people who indicated not experiencing discrimination. The advisory group and focus groups also indicated that questions about origin, and certainly questions about experiences of discrimination, **are very important but can also be difficult and evoke traumatic experiences.**

In the subsequent open-ended response fields, the respondents told us that they were uncomfortable with questions about their origins for various reasons:

- Because these questions **reopen old wounds** related to experiences of racism
- Because they **are not sure of their origin(s)** and/or have never asked themselves that many questions about it
- Because **many origins intertwine in their family history**
- Because **adoption** is not explicitly considered, which greatly complicates the concept of origin

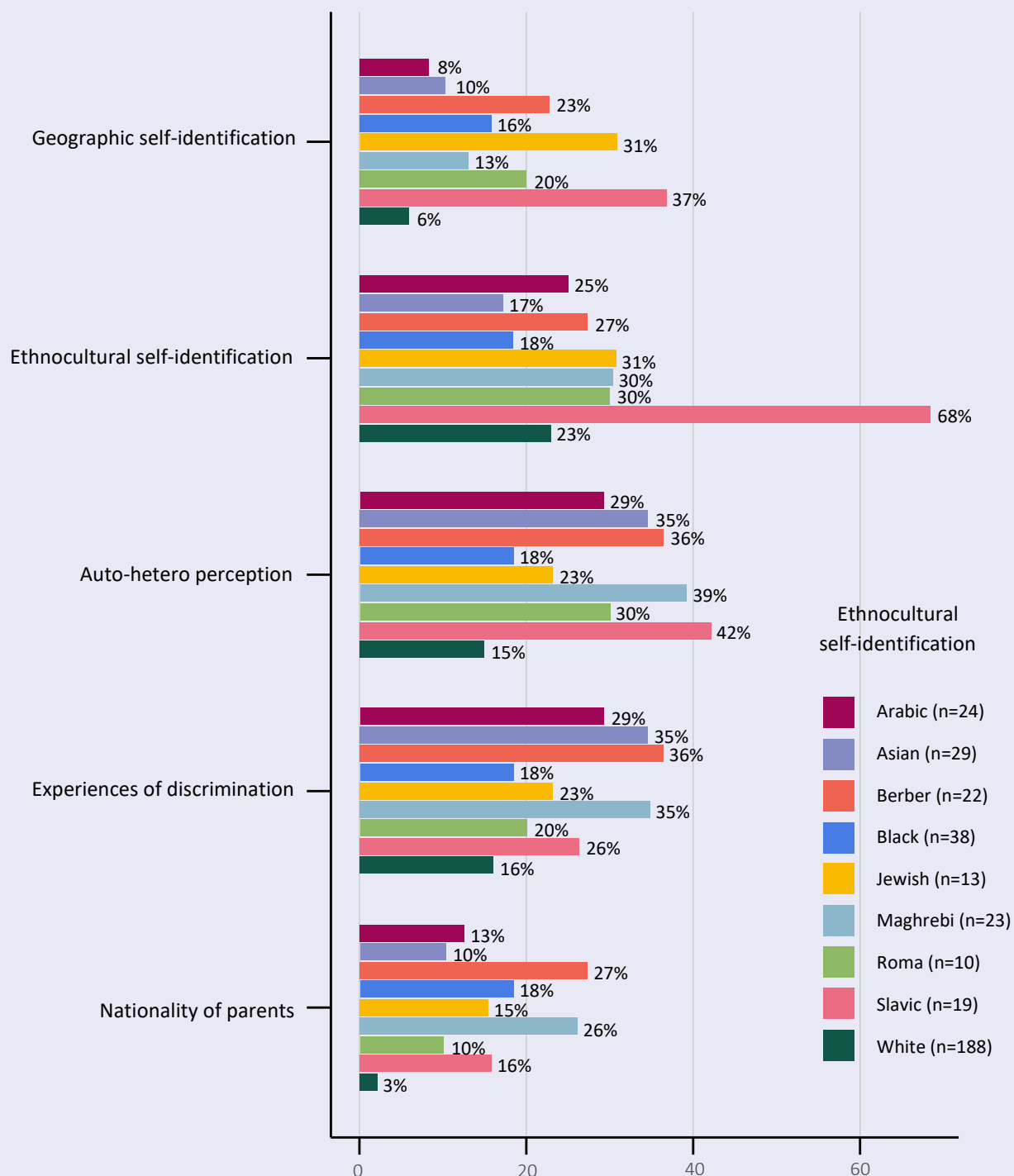
- Because it is difficult to define themselves as belonging to a group and yet be perceived in a different way: **their feelings do not match the image projected onto them** (for example, they may feel Belgian but be perceived as foreigner, Moroccan, African, Muslim, Jew, etc.)
- Because **it is difficult to determine the origin that others project onto you** and that may be the source of the discrimination

Finally, we look at what percentage of respondents **experienced discomfort** with the different questions as a function of the **(self-defined) ethnocultural categories** (see chart 7). In other words, this is the percentage of respondents who indicated feeling 'very uncomfortable' or 'somewhat uncomfortable' per ethnocultural response category.

We find that **people who indicated being 'Black' or 'Asian' report relatively less discomfort when asked about ethnocultural response categories**. People who define themselves as 'White' experience slightly more discomfort than these two groups, but less than most other ethnocultural groups.

When asked about **experiences of discrimination**, we see that a relatively high proportion of those defining themselves as 'Asian', 'Maghrebi' or 'Berber' experienced discomfort when asked this question, while the smallest proportion reported feeling uncomfortable among those defining themselves as 'White'. This is in line with the earlier finding that **respondents who experience less discrimination also felt less uncomfortable with questions that ask about this**.

Chart 7: Percentage that reported experiencing discomfort with the questions, by ethnocultural response category



The percentage who reported 'discomfort' = the percentage of people who indicated 'somewhat uncomfortable' or 'very uncomfortable'. In each case, the percentages were calculated relative to the number of people who defined themselves as being of that origin, so there are people who are included in multiple of these categories.

Source: 'Measuring discrimination and inequality through self-identification of origin in equality data in Belgium?' Survey realised by Unia, 2023.

7 Self-identification: a practical guide

This section contains good practices and practical tips for working with the self-identification of origin method, in three parts: (1) good ethical practices, (2) which questions to use, and (3) which answer categories to use.

These good practices are not exhaustive and certainly not definitive. We focus on the conclusions and outputs of this study conducted by Unia. The good practices described are thus based on the results of the literature review, the interviews, the focus groups, and the online survey. However, this study is only a first step in a longer research process on the implementation of self-identification in Belgium.

Good practices will therefore evolve in the future, but here we already give an outline of how best to work with self-identification of origin in Belgium. For more practical guidance on measuring origin and other characteristics, you can also consult Unia's 'Diversity monitoring: do it yourself!' tool⁸³. Moreover, this tool will be updated based on the results of this study.

A general principle when designing methodology for measuring origin is that the **method should be adapted and designed according to the precise measurement objectives**. Therefore, it is currently not possible to develop a 'one-size-fits-all' method that can be applied everywhere by default. Each data collection effort will require different choices depending on the objectives. In any case, be transparent about the choices made and the methodology used.

7.1 Good ethical practices

When collecting and processing data, **the legislative framework** (see section 4.3 'The legislative framework: data protection') must of course be strictly respected. In addition, there are a number of **good ethical practices to consider**:

- Ensure that the **goal** is clearly defined and legitimate
- Obtain fully voluntary and informed **consent** from the respondent
- Guarantee **anonymity** and **confidentiality**
- Create **trust** and reflect on how to prevent misuse of the data
- Be **transparent** and accessible in all communication
- Ensure that respondents - even if they agreed to participate in the data collection - **can refuse to answer any question** and that they can decide to stop their participation at any time
- Seek **participation** of concerned groups when developing the data collection methodology
- Consider the possible traumas and emotions that the data collection may evoke and provide a **'trauma-sensitive approach'**. Specifically, a *'trigger warning'* and a reference to available psychological help may be added to the survey, for example.

7.2 Which questions to use

In terms of choosing **the type of question** and the use of terminology, these guidelines may help:

- When formulating the self-identification questions, use **'origins' in the plural**, as people can have multiple origins.

⁸³ eDiv par Unia, [Monitoring de la diversité : do it yourself!](#)

- Try not to mix origin and identity in the question. We therefore opted for the question "**How do you define your origins?**" without explicitly talking about identity, because identity is much broader than origin.
- An **auto-hetero perception question** can be a good way to measure origin when the aim is to identify discrimination. This question gives a better picture of how people are seen by others, which is often the source of discrimination. We opted in the survey for the question "**How do you think others see you?**" followed by ethnocultural response categories. The same question could also be asked with more geographic response categories or a mix of both. Do take into account that the survey results showed that respondents did not find this question easy.
- Another appropriate question could be whether someone feels they belong to a discriminated minority, such as "**Do you think you belong to a minority at risk of racial or ethnic discrimination? If so, which one?**"
- **Consider questions that focus on the experience of discrimination linked to origin.** Several experts recommended avoiding questions that are too confrontational in terms of identity, confronting the person with an existential question: "Who am I?". Focus instead on actual or perceived origins that have been a source of discrimination. We concluded that the wording of our survey question was not ideal. We therefore suggest a wording such as "**Have you ever been discriminated based on your origin? If so, which one?**" However, it is crucial to consider the possible trauma these questions can evoke in people who have been victims of racism or discrimination. These types of questions may cause some discomfort, but it remains an option that focuses more on origin in the context of discrimination than on the identity of the person.
- Consider - depending on the objectives - also asking questions about (discrimination experiences based on) other personal characteristics such as gender (identity)⁸⁴, disability⁸⁵, age, social origin, and religious or philosophical beliefs. Consider in advance which **intersectional analyses** are important and necessary and adapt the questionnaire accordingly. For more information on intersectionality in equality data, see part 6.4 in the report '*Improving Equality Data Collection in Belgium*' by Unia⁸⁶.

7.3 Which response categories to use

Response categories should be relevant and appropriate, and may vary depending on the purpose of the measurement: is there a particular target audience or terminology that should be considered in this context? Are there strong collective identities that should not be overlooked? To find out: **consult the groups involved and work with them.** The closer the list of response categories is to the collective identities of the respondents, the higher the response rate and the more reliable the data.

We formulate some guidelines here:

- Offer the possibility of indicating **multiple response categories/origins**. In our survey, we allowed a maximum of four answers. We find that very few people indicated four answers. Since multiple answers complicate analysis and interpretation, **a maximum of three answer categories may also be sufficient.** However, both our focus groups and survey showed that two is insufficient.

⁸⁴ Motmans, J., Burgwal, A. en Dierickx, M. (2020), [Adviesnota: Het meten van genderidentiteit in kwantitatief onderzoek](#), Transgender Infopunt.

⁸⁵ Unia, [Improving Equality Data Collection in Belgium II: Final report](#)

⁸⁶ Unia (2021), [Improving equality data collection in Belgium](#)

- Always leave the option for an **open answer**. Be careful with the wording of the open answer category and do not simply use 'Other'. 'Other' can have a negative connotation and evoke the feeling of being seen as 'the other' or deviating from the norm. A good option for the wording of the open response category is: 'I would rather/also define myself as...'.
- **Offer the option to answer 'I don't know'**. Possibly in combination with another option, although this complicates the analyses. After all, this is a difficult question to which people do not always have a ready-made answer.
- **Order the answer categories alphabetically**. If not, it may suggest a certain 'hierarchy' in which 'Belgium' is often at the top. 'Belgian' may then seem to be considered more valuable than another origin that appears lower down the list.
- **Geographic response categories** such as 'Belgium' seem to be a common and accepted way in which people spontaneously describe their own origins and are therefore important to include. Many open-ended answers in this study referred to a **specific country** that was not included in the list. Depending on the target audience, the inclusion of a specific country in the list may be relevant.
 - Consider including the following national origins given their significant presence in the Belgian population: Belgian, Dutch, French, Italian, Moroccan, and Turkish. Depending on the exact context and objectives, other national origins may also be relevant.
 - For geographic response categories, it may be useful to **make sure that the different categories are 'mutually exclusive'**. In other words, make sure that certain categories do not also fall within other categories. In our survey, this was the case with 'Belgium' and 'Europe in the EU', which complicated the analyses. A solution could be to formulate the category, for example, as 'an EU country other than Belgium, France, the Netherlands, or Italy'.
- Regarding **ethnocultural response categories**:
 - Most people seem to be most comfortable with the geographic origin question, and relatively less so with the ethnocultural question. While we do not have a clear idea of the exact reasons for this, it is important **to tread carefully when using ethnocultural response categories** and to always keep the objectives in mind.
 - **Include ethnocultural categories not linked to nationality, such as 'Jewish', 'Roma', 'Traveller', etc.** These were considered to be missing if we only used geographic origins. Although the Roma who were consulted expressed many subtleties in terms of self-identification, the term 'Roma' seems to be an umbrella term.
 - **Be attentive to large overlaps between certain categories** that may complicate the analysis. For example, in the ethnocultural response categories, there appeared to be a large overlap between these three categories: 'Arabic', 'Berber', and 'Maghrebi'.
 - In the question with ethnocultural response categories that we used, it seemed that mainly the geographic categories 'Latin American' and 'Turkish' were missing. These could potentially be added to the list of ethnocultural response categories.
 - Consider more **regional categories** such as 'Slavic'.
 - Consider **categories that refer to skin colour** such as 'Black' and 'White' because they allow identification of physical markers that are often a source of discrimination.
 - The category 'Gypsy' which is part of the *'European Standard Classification of Cultural and Ethnic Groups'* is considered inappropriate and pejorative.
- It can be relevant to **mix ethnocultural response categories and geographic categories** in the same question. When only one of these frameworks is used, there are always people who do not identify with the categories.

- Create clarity for people who define themselves as 'mixed' or 'métisse'. This could be done by adding a category referring to people with mixed origins such as '**métisse**' or '**mixed**', potentially broken down further. However, this addition may complicate analyses because it does not allow us to understand the different origins on which the discrimination is based. **Another option is to give clear instructions on how people of mixed origins should fill in the question.** This proved especially important in the question on ethnocultural origins, where many respondents indicated mixed origins in the open-ended response options.

8 Recommendations

The following recommendations were drawn up based on the results discussed above. They relate to the development of equality data in Belgium using the method of self-identification of origin. These recommendations are of different natures: principle-based, methodological, and institutional. They are aimed at policymakers and at the various bodies that collect data on the origin of individuals.

8.1 Use self-identification as a useful measurement tool to combat discrimination and structural racism, possibly in combination with proxy variables

There is a need for **more equality data related to origin** and intersectional analyses of these data, **across different domains**, to measure diversity and to identify groups that are victims of racism and discrimination.

Such data make it possible to highlight **the extent of inequality**, discrimination, and structural racism and to **take appropriate action**. In contrast, the absence of such data hinders evidence-based political action.

Depending on the goal of each data collection effort, the groups involved, and the status of the actor collecting the data, **the most appropriate method** for collecting data on origin should be determined. However, **we recommend the use of self-identification, where possible and appropriate**, so that individuals can decide for themselves whether to provide information on their personal characteristics, according to a human rights-based approach to data.

The **method of self-identification of origin** should be used **more often** if the objectives and context warrant it, potentially in combination with the proxy method. Neither method can perfectly identify groups at risk of discrimination. However, the results of this report show that self-identification can identify groups at risk of discrimination based on their origin that cannot always be identified with proxy variables based on nationality. Self-identification thus makes it possible to fill the gaps linked to the exclusive use of proxy variables based on nationality.

We therefore advocate **using a combination of different methods, whenever possible and desirable, to get a complete picture of a given issue**. In practice, this means, for example, using both administrative data to create proxy variables and including a self-identification question in a questionnaire presented to the same sample. This will enable the best possible identification of target groups and problem areas.

In part 7 'Self-identification: a practical guide' of this report, we offer practical advice to guide users through the reflective and methodological processes involved in using self-identification of origin.

8.2 Actually use equality data in developing and monitoring public policy

We call for the available equality data to **actually be used as a basis for developing policies** and actions to combat inequality and discrimination and to monitor this.

These data can also be used to highlight situations specific to certain population groups or domains. By using these equality data as a basis for developing actions to combat discrimination, political efforts can focus on improving the situation and experiences of groups exposed to discrimination.

Over time, these data make it possible to **measure the effectiveness of anti-racism policies** and adjust these policies as the situation evolves.

8.3 Act from a human rights-based approach to data: data collection through self-identification should not harm

In line with the '*do no harm principle*⁸⁷, the collection of equality data, and in particular data using self-identification, should **avoid harming the groups concerned**. The goals for which data are collected should contribute to greater equality and should in no way hurt people or expose them to further discrimination or stereotyping. Instead, they should bring about beneficial structural policy changes.

When using self-identification, **the trauma that such questions can provoke** should be taken into account. It is also important to avoid asking people unnecessarily about their origin and to think carefully about how to conduct this work.

8.4 Use self-identification in accordance with ethical principles and legal conditions, with a specific goal and transparent communication

The collection of equality data on origin must comply with a number of ethical principles, respect for privacy, and the data protection legislative framework. The collection, processing, and analysis of data related to origin must have a **specific, legitimate, and clearly explained objective**.

There should also be **transparency in terms of the objectives** and **communication in an educational way to the target group** and participants, explaining very clearly who the data collector is and what goal it is pursuing with the data collection.

It is also essential to be **transparent about how the data will be processed** and about the existence of a complaints procedure in case of a breach of privacy. In this way, the trust of individuals can be gained. Indeed, we find that affected groups show a certain **fatigue** with regard to participating in research and speaking out on the subject. Moreover, they also lack **trust** in public institutions. Therefore, it is necessary to engage in dialogue about the goal and processing of any data collection.

If these principles are followed, self-identification can be seen as a useful and ethically sound way to measure inequality and ask about personal characteristics.

8.5 Consult the involved groups at different stages of the data collection process through a participatory process

In all future data collection efforts using self-identification, we recommend **consulting the involved groups at each stage through a participatory process**. This consultation is especially essential during the process of creating the response categories. Special attention should be paid to **the connotations of the terms that are used**, how they are perceived by the groups involved, and the possible consequences of their use. **The participatory process should be reprised regularly** so that these categories continue to reflect social realities and changing sensitivities. During this participatory process, the expertise and time of those involved should be considered and valorised.

⁸⁷ OHCHR (2018), [A human rights-based approach to data. Leaving no one behind in the 2030 agenda for sustainable development](#), p.11.

Ideally, this process is carried out by **research teams with diverse origins** to build and maintain trusting relationships with the audiences involved.

Specifically for governmental actors, we recommend initiating a **reflection process** to make greater use of participatory processes on the one hand, and paying attention to the diversity in origins of data collection teams on the other. This **change in mentality** towards more participation of involved groups will enhance the acceptability of data collection and actual participation in it.

8.6 Establish a permanent platform for exchange and consultation on equality data

We advise the development of a **consultative body** on equality data to develop a **coordinated and structural policy on equality data in Belgium**. Through this platform, different actors working on equality data from near and far can meet, exchange information and good practices and develop coordinated actions. This platform should bring together different political actors, government departments, statistical institutions, civil society, equality bodies, and academics, among others.

It is very important that, through this consultative platform on equality data, the various stakeholders initiate a **political debate on the measurement of racism and discrimination**. The aim of this debate should be to remove taboos and dare to engage in a democratic debate on issues on which there is no consensus, such as structural racism, how to name it, how to measure it, how to think about it, and how to act on it in society.

We also recommend organising a **dialogue on the use of self-identification as a method for collecting equality data** via this consultative platform on equality data. The harmonised use by different bodies collecting data requires political will in the first place, but also clarification of the legal room for manoeuvre that different bodies have in this regard. **No form of data collection through self-identification is entirely uncontroversial from the outset**. It is a process that requires willpower and time. This study contributes to this debate by taking a first step. It offers interesting starting points to be completed, refined, and discussed.

This consultative platform on equality data should also ensure **regular dialogue between the various entities that collect data**. By data collectors, we mean statistical institutions, academia, public institutions and administrations, and the private sector that collect data on origin, among others. To produce comparable data, it is necessary to **harmonise practices and agree on both definitions and methods** used in data collection. This is all the more important if self-identification is used more extensively, as this will produce new data that can be compared with each other as well as with existing data.

8.7 Further research on self-identification of origin is needed

This study on self-identification of origin was only a **first step** and also has some limitations, as described in the methodological chapter. **The application of self-identification in equality data in Belgium requires further research**. We suggest some avenues here:

- The data of our online survey could be used for **further analyses**. For example, data on age and generation of migration could be used for intersectional analyses.
- Conduct more quantitative studies to test the use of self-identification with **larger samples representative** of the Belgian population.

- **Continue to engage different communities** present in Belgium through qualitative research methods such as focus groups. Through this survey, we have consulted some of the groups involved, but many other communities could be consulted on self-identification of origin.
- Take a closer look at methods for **collecting and analysing intersectional data** such as origin and gender/age/migration generation/disability/religion/other discrimination criteria, in order to better understand the self-identification of specific populations.
- Further study **the statistical feasibility and concrete operationalisation of self-identification in Belgian equality data**. For instance, how can multiple origins be analysed? To what level of detail can origin be analysed while still maintaining anonymity? How can proxy variables and self-identification be combined while respecting data protection?

Colophon

Study on self-identification of origin in Belgium Improving Equality Data Collection in Belgium II

Brussels, June 2024

Publisher

Unia, Interfederal center for equal opportunities and the fight against racism and discriminations
Place Victor Horta 40 (bte 40), 1060 Brussels
T +32 (0)2 212 30 00

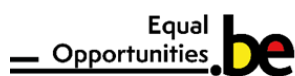
Writing: Unia

Final editing: Unia

Translation: Unia

Graphic concept and layout: Unia

Responsible publisher: Patrick Charlier



This project was funded by the Equal Opportunities Department of the FPS Justice.

Ce rapport est aussi disponible en français.

Deze publicatie is ook verkrijgbaar in het Nederlands.

You can find this publication at www.unia.be in the 'Knowledge & Recommendations' section.

Prefer to read on paper? On the publication page of this report you can request a printed version, via the 'Order printed version' button. We will then send you a free copy; only for large quantities or an express shipment, postage may be charged.

Unless we have stated otherwise, the information in this report is royalty-free. Provided you mention the source and author, you may therefore use this information for personal and non-commercial use without costs. If you want to reproduce, distribute, or otherwise make available any information (texts, photos, images, ...), please first contact Unia via the contact form at info@unia.be.