Distr.: General
19 October 2022

English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Migration Statistics

Geneva, Switzerland, 26−28 October 2022
Item A of the provisional agenda
Improvements in use of administrative data for migration statistics

Finding family relations: About quality issues regarding family immigration statistics and their potential solution using administrative data

Note by Statistics Norway*

Abstract

The implementation of Directive 2004/38/EC into the EEA-agreement and the establishment of a registration scheme in the Norwegian immigration administration meant that citizens of the EEA-area are no longer required to obtain residence permits to stay in Norway. This has led to data loss in administrative records concerning reason for immigration and statistics derived from this. This paper will outline the background of a project, financed by Ministry of Labour and Social Inclusion (AID), initiated to address the issue of missing data regarding the persons of reference for family immigrants. The project is still in its initial, explorative stage, and the paper will therefore mainly limit itself to a description of the data sources used in the production of Statistics Norway's (SN) statistics about family immigration and the consequences of the introduction of the registration scheme for citizens of EEA-countries for the quality of this data. The paper will also describe already implemented measures and provide some tentative considerations of how administrative data can be applied to link family immigrants with their actual person of reference.

*Prepared by Christian Soerlien Molstad

I.   Introduction

1. Data regarding reason for immigration constitute the foundation for several statistics published by Statistics Norway (SN). Among these are numbers on family immigration, as presented in the monitor for family migration which is published annually as alternately article and report (Gulbrandsen & Molstad, 2020; Molstad et al., 2022; Molstad & Steinkellner, 2020).

2. The primary source of data for reason for immigration is information derived from decisions regarding residence permits, taken by the Norwegian Directorate of Immigration (UDI). Until the late 2000s such permits were required for all non-Nordic immigrants. However, the passing of Directive 2004/38/EC and its implementation in the EEA-agreement meant that citizens of the EEA-area have no longer been required to obtain residence permits to stay in Norway. Instead such citizens must register with the police within three months of arrival.

3. A consequence of the simplified procedure associated with this registration scheme is gradual loss of data. Among the data increasingly lost is information regarding person of reference in the case of family immigration. Data regarding the person of reference, the person already living in Norway and to which the family migrant is immigrating, is of great value, as it is used to categorize family immigrants according to the background of the family member.

4. To address this information loss, a project, financed by Ministry of Labour and Social Inclusion (AID), has been initiated. The outcome of the project, a new method for identifying persons of reference, along with an internal memo documenting the process, is to be published in March 2023.

5. The project is still in its early, explorative stage. This paper will outline the background of the project, including a description of the data sources, their use in production of official statistics about family immigration and the consequences of the introduction of the registration scheme for citizens of EEA-countries for data quality. The paper will furthermore describe some of the measures already in place to counter the challenges associated with the loss of data regarding persons of reference and tentatively suggest an approach through which administrative data can be used to further increase the quality of the mentioned variable.

II.   Family immigration statistics and its data sources

6. The registration system for immigration and emigration in Norway is based on three main areas of administration: the adoption authorities, the immigration administration and the national population register (Vassenden, 2015, pp. 129-130), for which only the two latter are relevant for the topic at hand. Statistics regarding reason for immigration is mainly based on data from the Aliens Register (UDB), administered by the Norwegian Directorate of Immigration (UDI), and the Central Population Register (DSF), administered by the Directorate of Taxes (Skatteetaten).

7. While data from both sources are used in producing the data file from which the annual statistics are generated, assigning each immigrant (with non-Nordic citizenship) a reason for immigration is first and foremost reliant on the values in the variable reason for decision in UDB. If information is missing in this variable, supplementary sources are used. Among the supplementary sources are data from administrative records regarding participation in labor market and education.

8. On the finished file reasons for immigration are grouped in five main categories in variable inngrunn1:

- Labour

- Family

- Refuge

- Education

- Other

Alongside these categories there is also a category for "unknown", in which immigrants that have no data for reason for immigration is assigned.

The latest file of which published statistics has been made covers immigration in the period 1990-2020.[1] Figure 1 shows immigration by reason among immigrants with non-Nordic citizenship in this period. Out of total 932 000 persons about 334 000 (36 percent) immigrated due to family reasons. This makes family the single most common reason for immigrating to Norway. As we can see the volume of immigration due to family increased markedly from the mid-1990s until the early 2010s. There are several reasons for this. One of the reasons is the increased influx of refugees from the former Yugoslavia, Syria, Eritrea, Somalia, Afghanistan, Iraq and other countries with armed conflicts and/or systematic persecution. Another is the enlargement of the EU in 2004 and 2007, after which labour migration from new member states (primarily Poland, Lithuania and Romania) increased dramatically. Most family migrants have immigrated to reunite with either refugees or labour migrants.

Figure 1. Immigrations to Norway, by reason, 1990-2020

---
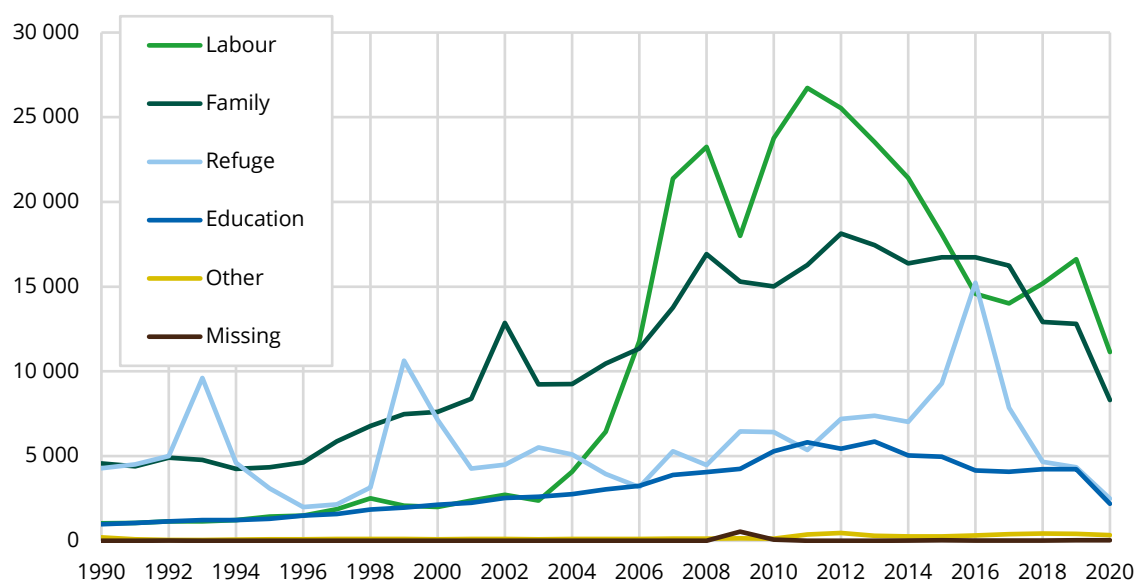
[1] Statistics updated with data for immigrations in 2021 have not been published yet. Publication of updated statistics is set for the 25th of October this year.

9. Family immigration has in other words to be considered in light of other types immigration, as it to a considerable extent constitutes a consequence and extension of these. Data regarding the so-called person of reference, the person to which family migrant is immigrating, is therefore of central importance as it provides much needed context regarding the migration. A key distinction can be made between immigration to immigrants and Norwegian-born children of immigrants and immigration to persons without immigrant background (most often males). Another important distinction, within the first category, can be drawn between those immigrating to labor migrants and those immigrating to refugees. An analysis of outcome with regards labor market, income and housing in a ten year-period from arrival reveals significant differences among family immigrants according to the background of the person of reference (Molstad et al., 2022).

10. With the aid of the social security number registered by UDI in the Aliens Register (UDB), background information regarding the person of reference can be added from data derived from the Central Population Register (DSF). The final data file therefore includes a range of variables regarding the individual, including his/her country of birth (rpfodeland), gender (rpkjoenn), first citizenship (rpforststatsborg), immigrant background (immigrant, Norwegian-born to immigrant parents, Norwegian-born to Norwegian-born parents etc.) (rpinvkat), country background (rplandbak3gen), registry status (resident, emigrated, dead etc.) (rpregstatus) and municipality of residence (rpkommnr). For persons of reference that are themselves immigrants, additional variables concerning their age at time of immigration (rpinnalder) and reason for immigration (rpinngrunn and rpinngrunn1).

11. These variables have been frequently utilized in Statistics Norway's publications regarding family migration. Cross tables displaying family immigration by year according to immigrant background and gender and reason for immigration formed, until recently, a part of the package of tables published annually in the monitor for family immigration as a part of Statistics Norway's framework agreement with the Ministry of Labour and Social Inclusion (AID).

12. Figure 2 shows a simplified graphic representation of the first cross table, excluding the gender dimension. The overall tendency as glanced from the figure confirms the statement made above about family migration to a large extent being an extension of labour migration and the immigration of refugees. A third important category of family immigrants can also be discerned: family immigrants to other family immigrants.

Figure 2. Family immigrations to Norway, by reason for immigration of person of reference, 1990-2020
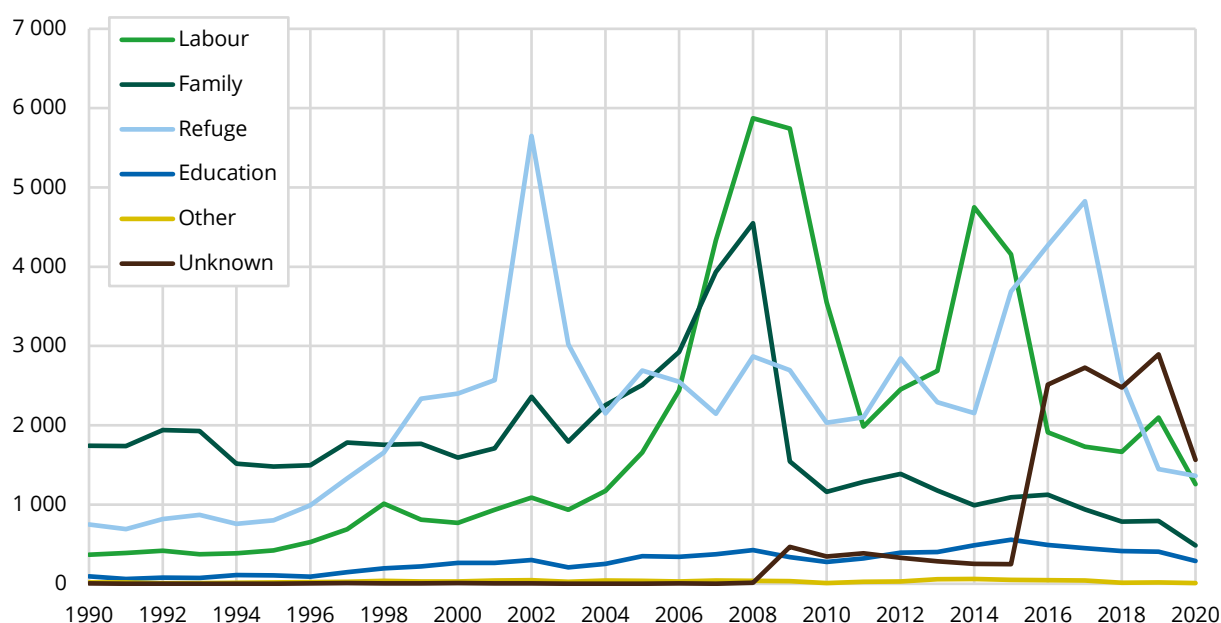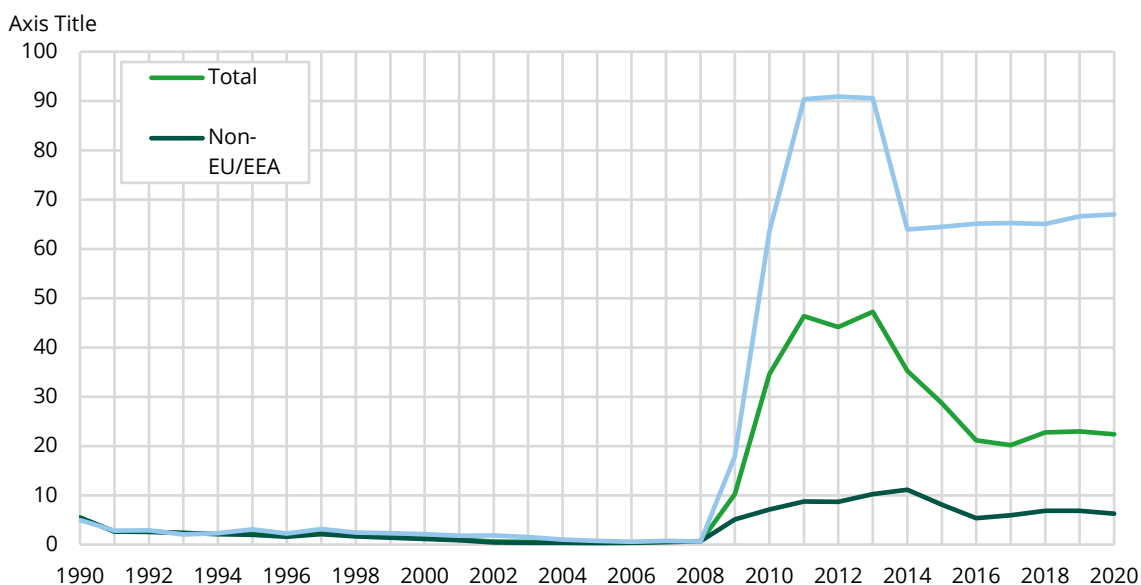

III.   2009: Introduction the registration scheme and loss of data

13. Publications made since 2021 have not included these cross tables or statistics derived from them. This is due to the loss of data regarding the identity of the person of reference and the gradual increase of this loss since the introduction of the so-called registration scheme in 2009.

14. The registration scheme was introduced as a result of the adoption of Directive 2004/38/EC into the EEA agreement in 2007 (Regjeringen.no, 2022). The directive grants citizens of the European Union "the right to move and reside freely within the territory of the Member States". This right is "also granted to their family members, irrespective of nationality" (EUR-Lex, 2022, pp. 78-79). One of the main purposes of the directive was to simplify the process of changing country residence within the EU/EEA-area for citizens and their family members. However, while EU/EEA-citizens do not need permits to reside in member states, the directive states that "for periods of residence of longer than three months, Member States should have the possibility to require Union citizens to register with the competent authorities in the place of residence, attested by a registration certificate issued to that effect" (EUR-Lex, 2022, p. 81). In Norwegian immigration administration a registration scheme was established by UDI, through which EU/EEA-citizens and family members of these have been required to register within three months of arrival. The establishment of a registration scheme was not stipulated in the directive and therefore not a given. Its creation instead stemmed from domestic administrative concerns.

15. A consequence of these procedural changes has been the loss of data with regards to immigrants from EU/EEA-countries. The codes used to classify immigrants in the registration scheme are not as extensive as those applied for residence permits. While the

main categories for reasons of migration (labour, family, education etc.) may still be discerned from the data, the quality of finer subdivisions has become more questionable (Dzamarija, 2013, pp. 9-10). A further deterioration has taken place concerning data on the relation between the family migrant and his/her person of reference (Dzamarija, 2013, p. 21), as well as a more general loss of data about the person of reference. The latter developments have led to an increase in family immigrants with "unknown" person of reference, starting from 2009. This situation is evident in figure 2, where the number of family migrants with person of reference with "unknown" reason for immigration increases from this year.[2]

16. A major division has in other words formed between family immigration from areas inside and outside the EEA-area in terms of quality of data. From a total 334 000 family immigrants arriving Norway in the period 1990-2020, 57 500 (17 percent) did not have a person of reference registered. A majority of these, 47 700 (83 percent), were citizens from EU/EEA-countries. However, the loss of data is not exclusive to family immigration from the EU/EEA-area. As figure 3 shows, there has also been an increase in family immigrants from countries outside the EU/EEA-area for which the social security number (rpfnr) of the person of reference is not available or valid.

Figure 3. Family immigrations to Norway, percentage missing person of reference, 1990-2020

---

[2] However, while the marked increase after 2009 of family immigrants coming to immigrants with "unknown" reason for immigration clearly coincides with the introduction of the registration scheme, the scheme does not account for the sudden burst of missing values from 2016 and on. At first glance the graphs in figure 2 and 3 may seem to present clashing narratives. While in figure 3 the main increase in missing values occur in the aftermath of the introduction of the registration scheme in 2009, in figure 2 the main increase is in 2016, well after the procedural changes. An explanation for the conflicting patterns is to be found in differences of population. Figure 3 includes all family migrations in the period from 1990 until 2020, while figure 2 only includes family migrations in which enough information was available about the person of reference to indicate that he/she was an immigrant. The increase of missing values regarding reason for immigration from 2016 henceforth is therefore a sign of additional loss of data, of which cause we still have not been able to ascertain.

17. A second division in terms of data quality may be made between the two main types of family immigration: a) family reunions and b) family establishments. Family reunions includes family immigration where the family relation existed before the person of reference was resident in Norway. Family establishments refers to those cases where the family migrant immigrated in order to establish a new relationship (in most cases a marriage) with a person resident in Norway (Molstad et al., 2022, p. 9).

18. In the case of family establishments, absence of social security numbers for persons of reference is low on the finished data file. In less than one percent of the family establishments in the period 1990-2020 there was no valid social security number available for the person of reference. This was the case regardless of whether the family immigrant was an EU/EEA-citizen or not (figure 4). The reason behind this difference in the data quality between family reunions and establishments is to be found in already existing measures implemented to cover the gaps in the data. These measures, along with some tentative suggestions for an approach for further steps, will be outlined briefly below.

Figure 4. Percentage family immigrants with missing person of reference, by type of family immigration
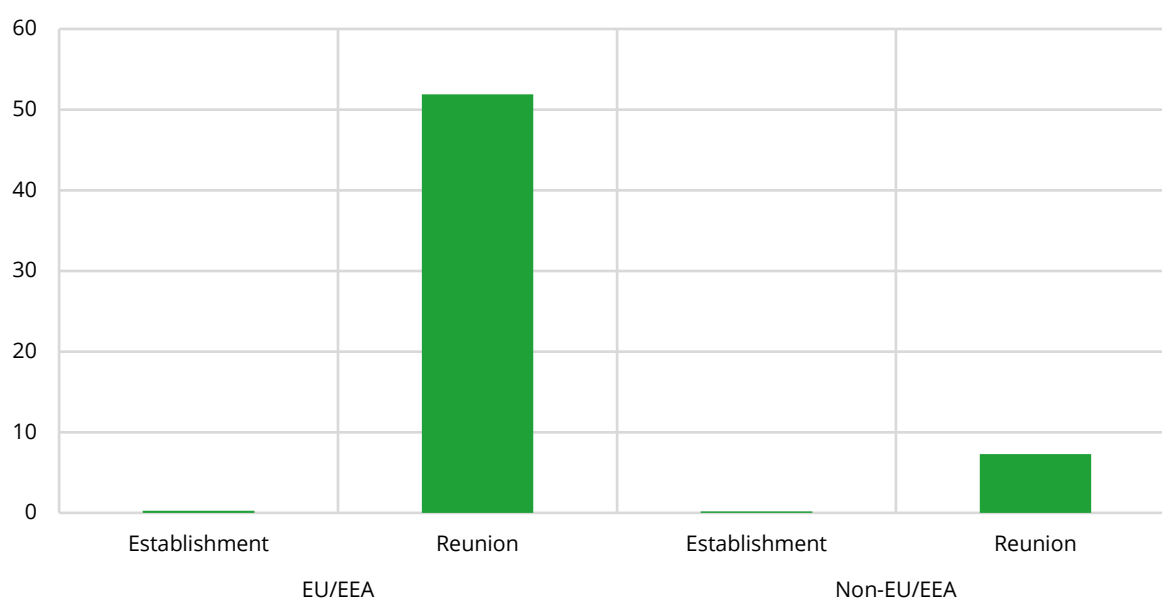
IV. Some preliminary considerations on solutions

19. Missing data is not a new challenge for statistics based on administrative data. So is not the task of establishing linkage between related units in different data sources. An example of the latter can be found in register-based census data, in which data from the Central Population Register (DSF) and the Ground Parcel, Address and Building Register (Matrikkelen) have been integrated. In order to account for incomplete linkage between households in DSF and dwellings in Matrikkelen, a double nearest-neighbour imputation method (DNNI) has been applied to impute dwellings for households that do not

successfully link with a dwelling, using other, statistically similar households as "donors" for picking a non-linked dwelling in the same geographical area (Zhang & Hendriks, 2012). An imputation method also applying a near-neighbour methodology has furthermore been used to impute missing data among immigrants in the Register of the Population's Level of Education (BU) (Jentoft, 2014).

20. As for the missing data regarding persons of reference for family immigrants, we are currently hoping imputation will not be a necessary step. A central goal of the project is to establish linkages that a) do not rest on the assumption of statistical similarity between family immigrants with missing values and those without and b) reflect actual relations between family immigrants and persons of reference, not relations based on statistical prediction of similarity. The preconditions for succeeding with this goal is in principle present. A central premise for family immigration is the residence of a person of reference with who the family migrant may either reunite or start a new family with. We can therefore assume that all persons of reference are present in administrative records mirroring the resident population. The challenge is to find a method (or, as will be suggested below, methods) to establish a link between the family migrant and what, with reasonable probability, can be assumed to be the actual person of reference.

21. As a family relation constitutes the precondition for family immigration, data regarding this relation should form the basis for a method for which to link the migrant to his/her person of reference. The first task in such a method is to discern the family relation between the family immigrant and his/her person of reference. The second is to find a variable or combination of variables which can lead this relation to be associated with an identifiable person.

22. As mentioned above such a method already exists for family establishment. The family relation in this case is the marriage between these individuals. Family migration due to spouse is possible to identify through the variable inngrunn derived from data from UDB, which provides a detailed classification of reason for migration. In this variable family migration to spouse constitutes a foundation for a subgroup of variations (FAM-MAKE-, FAMMAKE-) within a larger group of family migration-related codes (FAM-) (Dzamarija, 2013, p. 51). Marital data concerning the identity of married individuals' spouses is present in data Statistics Norway receives from the Central Population Register (DSF) through the variable ekt_fnr. In the production of the annual file data containing the social security number of spouses (in the variable ekt_fnr), is used to identify the partner in cases of family migrants that have immigrated to such a family member. Other data regarding the successfully identified reference persons are subsequently added.

23. The result of this can be seen in figure 4. As pointed out very few persons of reference are missing in the case of family establishments on the finished data file. As for family reunions, the number of migrations with missing information concerning person of reference remains high.

24. A similar method, using the relation between the family migrant and the person of reference as indicated in the variable inngrunn, may not be utilized for the remaining, overwhelmingly family reunion-related, cases with missing information. In 79 percent of these migrations reason for migration is registered as "EEA-citizen who is family member of an EEA-citizen who is employed etc. or who is spouse/partner or dependent child under 21 year of age of an EEA-citizen accepted at an approved institution of education" (EØSFAMBORG). A further 3 percent includes non-EEA-citizens with the same alternatives of relations to an EEA-citizens (EØSFAMIKKEBORG). It is not possible to deduce from these codes whether the family migrant is a) married or in another form of partnership with the person of reference, b) is the child of the person of reference or c) has another relation to the him/her. The

process of identifying the actual relation between family migrant and person of reference must hence rely on using additional, indirect indices.

25. One such indicator may be age at time of immigration of the family migrant, as already available in the data file in the variable innalder. Breaking down the family immigrants with missing persons of reference according to age reveals that 69 percent were below 18 years of age. About half were furthermore below ten years of age and nearly two thirds less than 15 years old (table 1). It is natural to assume that many (but not all) of these family immigrants of young age have immigrated to a parent and that the person of reference is therefore either their father or mother.

Table 1. Family immigrants with missing person of reference, by five year-interval age at time of immigration

Age at time of immigration	Frequency	Percent	Cumulative Percent
0 - 4	17800	30,9	30,9
5 - 9	10649	18,5	49,5
10 - 14	8032	14,0	63,4
15 - 19	4381	7,6	71,0
20 - 24	2087	3,6	74,7
25 - 29	3542	6,2	80,8
30 - 34	3242	5,6	86,4
35 - 39	2122	3,7	90,1
40 - 44	1286	2,2	92,4
45 - 49	792	1,4	93,7
50 - 54	732	1,3	95,0
55 - 59	793	1,4	96,4
60 - 64	828	1,4	97,8
65 - 69	572	1,0	98,8
70 - 74	347	0,6	99,4
75 - 79	182	0,3	99,8
80+	142	0,2	100,0
Total	57529	100,0	

26. For family immigrants above age of 18 at the time of immigration, a larger array of probable alternatives is available. It is probable that in many cases the immigration will be of spouses reuniting. However, it is also likely that at least some of the immigrations involve parents reuniting with a son or daughter, older children immigrating to reunite with a father or mother or siblings accompanying parents moving to reunite with a son or daughter. The mentioned alternatives are among the ones mentioned in chapter 6 of the act relating to the admission of foreign national into the realm (the immigration act) as legal options for family immigration (Lovdata, 2022) and do not constitute an exhaustive list of possible alternatives.

27. Age at time of immigration does in other words not provide sufficient information, at least not alone, to assign a family relation between family immigrants and the person of reference. Finding a method for establishing linkages between family immigrants and a person of

reference will most probably have to rely on the use of a range of auxiliary variables in combination. In order to identify these auxiliary variables, we will have to perform the following preliminary steps:

i. Identify the possible constellations of family relations that qualify for family immigration

ii. Identify data available, either in UDB or DSF or in other administrative data sources, that may be utilized to distinguish family migrants according to their probable relation to their person of reference as identified in step i

iii. Identify data that can act to utilize this relation to link the family migrant to a person registered in DSF

While the first step is taxonomic, the second and third constitute the start of developing methods that by default will have to be different according to each "class" of family relation.

V. Conclusion

28. The project described in this paper is only in its preliminary, explorative phase. In the coming months a systematic mapping of possible data sources and of potential methods to utilize these to identify linkages between family immigrants and their persons of reference will be conducted. Statistics Norway has available a range administrative data that can provide auxiliary variables for these methods.

29. An important challenge is the plurality of possible family relations that may legally qualify for family immigration. A central task will therefore be to find devising a way of mutually excluding family immigration according to the type of family relation between the immigrants and his/her person of reference.

VI. References

Dzamarija, M. T. (2013). *Innvandringsgrunn 1990-2011, hva vet vi og hvordan kan statistikken utnyttes* (34/2013). Statistisk Sentralbyrå. https://www.ssb.no/befolkning/artikler-og-publikasjoner/_attachment/135482?_ts=140bf5cf4f8

EUR-Lex (2022). Directive 2004/38/EC of the European Parliament and of the Council of 29 April 2004 on the right of citizens of the Union and their family members to move and reside freely within the territory of the Member States, 47 77 (2022). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32004L0038&from=EN

Gulbrandsen, F. B., & Molstad, C. S. (2020). Færre familieinnvandringer i 2019. *Statistisk Sentralbyrå*. https://www.ssb.no/befolkning/artikler-og-publikasjoner/faerre-familieinnvandringer-i-2019

Jentoft, S. (2014). *Imputation of missing data among immigrants in the Register of the Population's Level of Education (BU)* (Notater/Documents 2014/27). Statistisk Sentralbyrå. https://www.ssb.no/utdanning/artikler-og-publikasjoner/_attachment/183364?_ts=146a98197c8

Lovdata (2022). Act relating to the admission of foreign nationals into the realm and their stay here (Immigration Act), chapter 16, (2022). https://lovdata.no/dokument/NL/lov/2008-05-15-35/KAPITTEL_6#%C2%A742

Molstad, C. S., Gulbrandsen, F. B., & Steinkellner, A. (2022). *Familieinnvandring og ekteskapsmønster 1990-2020.* (2022/03). Statistisk Sentralbyrå. https://www.ssb.no/befolkning/innvandrere/artikler/familieinnvandring-og-ekteskapsmonster-1990-2020

Molstad, C. S., & Steinkellner, A. (2020). *Familieinnvandring og ekteskapsmønster 1990-2018* (2020/05). Statistisk Sentralbyrå. https://www.ssb.no/befolkning/artiklerogpublikasjoner/_attachment/411225?_ts=1 6ffbcb2098

Regjeringen.no. (2022). Direktivet om fri personbevegelighet. EØS-notatbasen. https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2006/mai/direktivet-om-fri-personbevegelighet/id2431482/

Vassenden, K. (2015). *Om kvaliteten på den norske inn- og utvandringsstatistikken* (Notater/Documents 2015/17). Statistisk Sentralbyrå. https://www.ssb.no/befolkning/artikler-og-publikasjoner/_attachment/225080?_ts=14ce62dab98

Zhang, L.-C., & Hendriks, C. (2012). Micro integration of register-based census data for dwelling and household. *United Nations Economic Commission for Europe (UNECE): Conference of European Statisticians. Work Session on Statistical Data Editing* https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2012/16_Norway.pdf