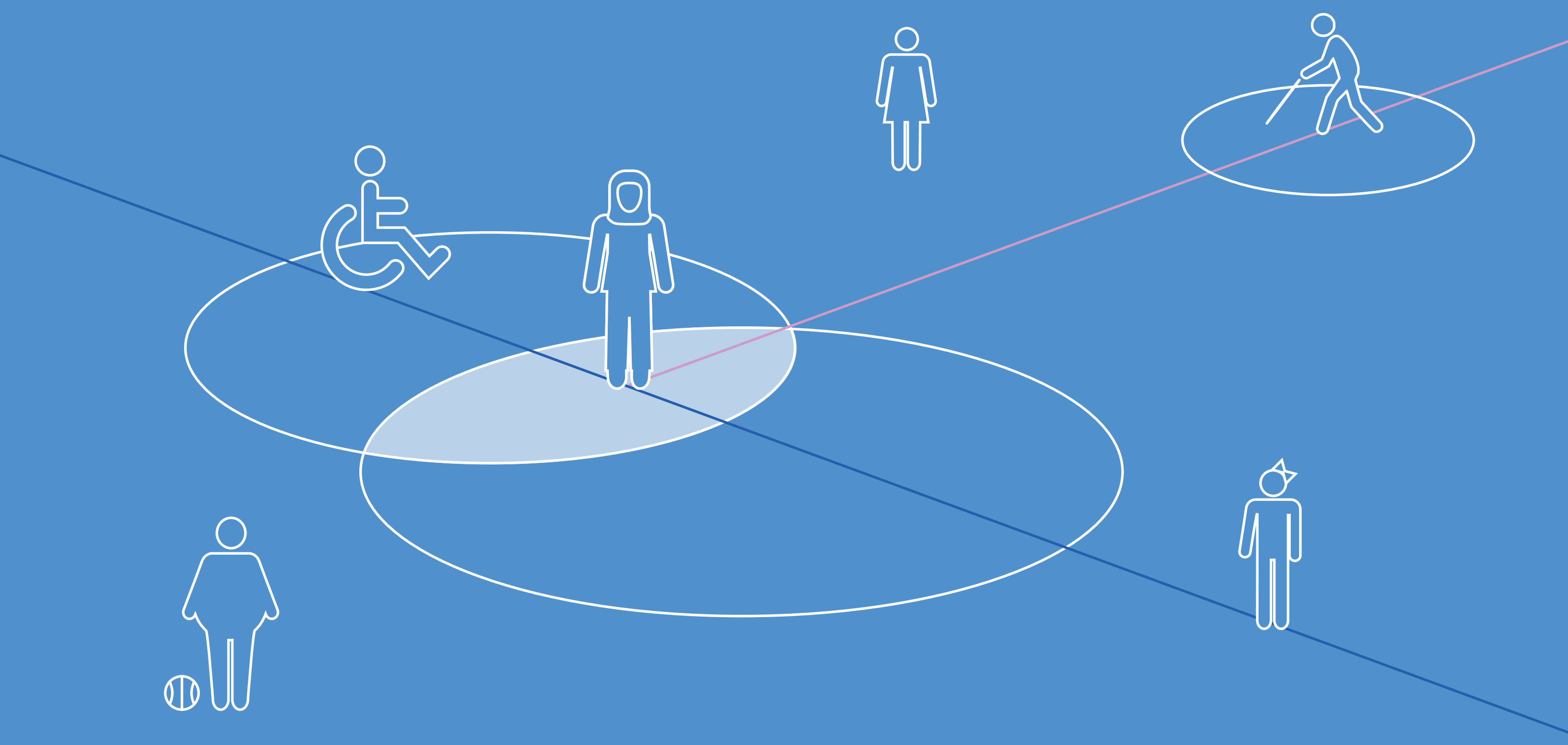
UNODC
United Nations Office on Drugs and Crime

MONITORING HUMAN TRAFFICKING PREVALENCE THROUGH MULTIPLE SYSTEMS ESTIMATION

A United Nations manual for policymakers, practitioners and researchers engaging with sustainable development goal 16.2



ACKNOWLEDGEMENTS

The preparation of this Manual was coordinated by the Trafficking in Persons and Smuggling of Migrants Research Team at the Crime Research Section of UNODC. The contributions of UNODC staff members such as Raggie Johansen, Salome Flores Sierra Franzoni and Fabrizio Sarrica are gratefully acknowledged.

Particular thanks are owed to Peter G.M. van der Heijden, Maarten Cruyff and Jan van Dijk supported by Joris Beijers for the preparation of the final draft of the present Manual.

This Manual benefited from the expertise and advice of Annette Brunovskis, Davina Durgana, Ieke de Vries, Ruth King, Bernard Silverman and Kyle Vincent.

The UNODC Crime Research Section would also like to thank the Demographic and Social Statistics Branch of the Statistics Division at the United Nations Department of Economic and Social Affairs who made the development of this Manual possible.

Suggested citation:

UNODC, *Monitoring human trafficking prevalence through multiple system estimation 2022*

Comments on the Manual are welcome and can be sent to:

Vienna International Centre, PO Box 500, 1400 Vienna, Austria

Tel: (+43-1) 26060-0, Fax: (+43-1) 26060-5866.

Email: askresearch@unodc.org

Disclaimer

Non-sales edited

UNITED NATIONS PUBLICATION

MONITORING HUMAN TRAFFICKING THROUGH MULTIPLE SYSTEMS ESTIMATION

A United Nations manual for policymakers, practitioners and researchers engaging with sustainable development goal 16.2

EXECUTIVE SUMMARY

Effective policies to address human trafficking as one of the Sustainable Development Goals start with solid and up to date knowledge on prevalence and trends. Since statistics on recorded cases largely reflect the efforts of institutions involved, they cannot be used to measure true prevalence and are therefore unfit for monitoring progress in achieving SDG 16.2. The option to carry out regular population surveys on self-reported victimization by human trafficking is not easily achievable in all countries. A cost- saving alternative is the triangulation of available administrative records on trafficking victims from state and non-state actors to estimate the proportions of unrecorded victims, a statistical method known as Capture-Recapture or Multiple Systems Estimation (MSE).

In this manual the basic principles and assumptions of MSE are explained for the general reader. Next, the manual explains which types of administrative data are needed for MSE and how pre-existing, cross-system datasets of police, labour inspectorates or service-providing NGO's can be used for MSE or made fit for that purpose. In several European countries integrated data systems on trafficking victims exist as part of National Referral Mechanisms. These can usually be utilized for MSE after minor adaptions. In countries where no integrated datasets on trafficking victims exist, stand-alone datasets from relevant institutions can be consolidated by dedicated research teams, as has been done successfully in MSE studies in the USA and Australia. This part of the manual also addresses how consolidation of sensitive personal data can be done in compliance with data protection regulation, for example through data encryption or the involvement of a National Statistical Office. It

finally discusses how MSE not only produces prevalence estimates but can also provide insight in the categories of human trafficking victims which are most hidden in the population, such, as sadly seems to be the case in some countries, child victims. In this way results of MSE provides guidance on how national trends in human trafficking compare internationally, which national anti-trafficking efforts need reinforcement and which types of trafficking in particular deserve higher priority in policy plans.

CONTENTS

Readers' guide

This manual provides practical guidance on how to apply the statistical technique of Multiple Systems Estimation (MSE) to generate better estimates of the levels of human trafficking through extrapolation from administrative data of recorded cases. It is meant for a mixed audience of policymakers and practitioners in the field of anti-trafficking committed to achieving Sustainable Development Goal 16.2 and researchers involved in the collection and/or analysis of data on victims of human trafficking to that end.

The first introductory section explains why MSE is a promising, cost-effective approach to monitor trafficking prevalence and progress towards the achievement of Target 16.2 of the Sustainable Development Goals. This section is primarily meant for policymakers and others interested in putting the fight against human trafficking on a stronger evidential footing. It aims to explain the theory behind MSE for the general reader. It then discusses the low costs of MSE and its many uses for policy planning and evaluation. It also addresses concerns about data protection issues concerning sensitive personal data.

The second and third sections are directed at those responsible for collecting, storing, and analyzing data on human trafficking victims, for example, those working for specialized state institutions, NGOs, observatories or research institutes at universities. The second section sets out the data requirements for the execution of MSE, including data protection considerations, and presents the theory behind MSE in more technical language. The third section demonstrates step by step how MSE was applied to an Irish dataset consisting of three lists and four covariates, using a tailor-made MSE package in the language R openly available on the internet. This section also demonstrates how the suitability of an available dataset for MSE can be checked.

In the final section, the manual presents an overview of results of MSE studies conducted in recent years and discusses the lessons learned and challenges remaining.

SECTION 1

Introducing Multiple Systems Estimation as a new way to estimate trafficking prevalence and monitor progress towards SDG 16.2

While human trafficking is not a new phenomenon, the adoption of the United Nations Trafficking in Persons Protocol[1] (henceforth *The UN Trafficking in Persons Protocol*) in the year 2000 in Palermo, Italy marked a quantum leap in the international community's commitment to address it. The UN Trafficking in Persons Protocol provided the first internationally agreed definition of this much-occurring heinous crime and a solid legal basis for effective anti-trafficking action and collaboration across the world. Its entry into force in 2003 spurred a wave of action, as countries introduced new or tailored existing legislation, created victim assistance initiatives, drew up national action plans and upskilled their criminal justice system officials. A broader anti-trafficking community in civil society also started to form, comprised of NGOs, faith-based initiatives, research centres and advocacy groups, to mention just a few.

The UN Trafficking in Persons Protocol definition of trafficking provides common ground for concerted responses to this crime. The full definition reads:

> '...the recruitment, transportation, transfer, harbouring or receipt of persons, by means of the threat or use of force or other forms of coercion, of abduction, of fraud, of deception, of the abuse of power or of a position of vulnerability or of the giving or receiving of payments or benefits to achieve the consent of a person having control over another person, for the purpose of exploitation. A/RES/55/25, 15 November 2000, Art. 3 (a)."

1. The Protocol to Prevent, Suppress and Punish Trafficking in Persons, especially Women and Children, supplementing the United Nations Convention against Transnational Organized Crime.

As is evident from the protocol's definition of the concept of exploitation, it applies to a wide range of different forms of exploitation, including sexual exploitation and forced labour:

> 'Exploitation shall include, at a minimum, the exploitation of the prostitution of others or other forms of sexual exploitation, forced labour or services, slavery or practices similar to slavery, servitude or the removal of organs.'

Article 28 of the parent convention against Transnational Organized Crime calls on State Parties to develop, in consultation with the academic communities, 'common definitions and methodologies' to analyse 'trends in organized crime'. In 2015, the quest for sound and reliable data on human trafficking was invigorated by the adoption of the 2030 Sustainable Development Agenda by the United Nations General Assembly. In this agenda, trafficking in persons is specifically referred to in three SDG targets, namely 5.2, 8.7 and 16.2. Target 16.2 calls on countries to 'End abuse, exploitation, trafficking and all forms of violence against and torture of children'.[2] The UN Statistical Commission's Interagency and Expert Group on SDG Indicators has adopted sets of indicators for each SDG target. For target 16.2, one of the indicators is: 'number of victims of human trafficking per 100,000 population, by sex, age and form of exploitation' (SDG indicator 16.2.2.). UNODC has been designated as the custodian agency for this indicator, responsible for methodological development and facilitation of country-level reporting. In view of lingering doubts about the utility of official statistics on recorded cases as indicator of prevalence, UNODC organized in December, 2013 an expert workshop of researchers specialized in the measurement of elusive populations to discuss alternative measurement options. One of the recommendations coming out of this gathering was to explore the feasibility of Multiple Systems Estimation as a method to estimate human trafficking prevalence through an analysis of data on recorded victims (Van der Heijden et al., 2015).

This manual aims to assist countries in estimating the national prevalence of trafficking in persons by applying Multiple Systems Estimation (MSE) to their administrative data on trafficking victims. MSE is a statistical method for estimating the size of elusive populations. Through the linking of available registers from various institutions of recorded persons belonging to such population, the part of the population observed by any institution is obtained. Subsequently, the overlap between the registers is used to estimate the unobserved part. In the case of human trafficking victims, examples of possible registers are those of police forces, labour inspectorates and NGO's providing services to victims. The aim of this manual is to explain

2. For more information on the Sustainable Development Goals, please see: http://sustainabledevelopment.un.org.

to policymakers responsible for ending human trafficking why MSE is a promising tool to monitor progress in these efforts and to provide technical guidance to practitioners and researchers how MSE can actually be implemented using national data available to them.

The manual is based on experiences and lessons learned from a first wave of ten MSE-based studies regarding human trafficking, several of which were supported by UNODC[3]. Building on this knowledge base, the manual outlines the basic principles underlying MSE and the kind of data required. It also provides a practical guide for the actual estimation procedure. It, finally, sums up the lessons learned from earlier MSE studies on human trafficking data and presents examples of successfully completed studies which can serve as models for those seeking to apply MSE to consolidated datasets on human trafficking from their own countries.

Why better national statistics on human trafficking victims are urgently needed

Statistics on the extent and nature of human trafficking are used to raise awareness of the size and gravity of the phenomenon, help governments, international organizations and non-governmental organizations to develop evidence-based responses and to monitor progress with their implementation (De Vries & Dettmeijer-Vermeulen, 2015). International statistics on human trafficking also allow benchmarking of national achievements against those of relevant peer countries and the identification of best practices across countries as envisaged by the UN Sustainable Development Agenda.

State agencies across the world are collecting data on the numbers of detected victims of human trafficking. These national statistics of recorded victims are regularly collated by the United Nations Office on Drugs and Crime and published in the

3. The first MSE study using trafficking data was carried out in the United Kingdom in 2014 (Silverman et al./Home Office, 2014). This study was followed by a series of studies initiated by UNODC, using data from The Netherlands (2017), Ireland, Romania and Serbia (2018). Reports are available at: https://www.unodc.org/unodc/en/data-and-analysis/tip.html. Subsequent studies include Australia (Lyneham, Dowling & Bricknell, 2019) and Slovakia (Walk Free Foundation, 2019), as well as an unpublished MSE study on data from Belarus. Studies have also been carried out in several locations in the United States, including in New Orleans (Bales, Murphy and Silverman, 2020) and three non-specified study sites (Farrell et al., NIJ, 2019). Using police data on arrests and rearrests for Commercial Sexual Exploitation, estimates of the total numbers of victims in Kansas were made using Capture-Recapture analysis (Phillips, 2017).

biennial *Global Report on Trafficking in Persons*.[4] The collection of these international statistics on detected victims is facilitated by two circumstances. First, trafficking in persons is one of the rare offense types for which, as said, an international legal definition has been adopted, namely the definition given in the United Nations Trafficking in Persons Protocol. This definition is now broadly reflected in the criminal codes of 169 countries across the world (UNODC, 2020). Secondly, an increasing number of countries no longer solely rely on law enforcement to reach out to victims. In line with the victim-centric spirit of the Palermo Protocol, they have set up arrangements for the systemic identification and referral of victims of trafficking in persons by, inter alia, police, labor inspectorates and service-providing NGO's. Some countries have formalized this multi-agency cooperation through special mechanisms such as a National Referral Mechanism (NRM) that stipulate the roles and responsibilities of each entity involved in the identification process (OSCE/ODHIR, 2004). Where such institutional arrangements are in place comprehensive, multi-source statistics on victims of human trafficking can be collected from the various public and non-public institutions involved. This is now the case in a substantial and rapidly expanding number of UN Member States.

In spite of these improvements, the collection of internationally comparable statistics on trends in human trafficking as requested by the SDG agenda, leaves much to be desired. The reports of international organizations show that the collected statistics give rise to questions about their validity, reliability and comparability[5]. For example, the rates of trafficking victims per 100.000 inhabitants in the UN Global Report are up to 100 times higher in some countries than in others. This huge cross-national variation could in theory reflect existing differences in trafficking prevalence. It is more likely, though, to reflect the varying effectiveness of national agencies in arresting traffickers, rescuing victims and recording cases. Tellingly, some countries give little or no priority to detecting victims of this crime at all. Others limit their efforts to certain subcategories, for example trafficking of women for sexual exploitation and ignore other forms. As a consequence, numbers of recorded victims in such countries remain comparatively low while their 'true numbers' may be very high.

The rates of recorded victims do not only vary hugely across countries. The numbers of countries also often show great instability over time, sometimes going up or down by 50 per cent or more from one year to the other. This volatility could, again,

4. Available at: www.unodc.org/glotip.

5. See, for example, UNODC, *Global Reports on Trafficking in Persons*; Council of Europe (2019), *9th General Report on GRETA's Activities* (GRETA, Group of Experts on Action against Trafficking in Human Beings); European Commission, *Data collection on trafficking in human beings in the EU,* 2020.

reflect changes in actual prevalence, for example because in a particular year some unusually large trafficking operations involving many victims, were dismantled. However, such fluctuations are more likely to reflect changes in identification and/or recording practices. The 2018 UNODC *Global Report* provides several examples of significant increases in the number of recorded victims in a country per year after the introduction of new legislation, institutions dedicated to the fight against trafficking or anti-trafficking programs.[6] The inverse also occurs, such as a sudden drop in the numbers of identified victims due to reduced detection efforts by law enforcement agencies undergoing restructuring as happened in The Netherlands after 2014 (CoMensha, 2019). For these reasons, changes in the numbers of identified victims over time may not be indicative of changes in actual trafficking prevalence.

In sum, statistics on identified trafficking victims show only the variable part of the phenomenon which is detected by national authorities and/or NGOs at a certain time. Such statistics cannot be used to reliably assess inter-country variation in the true prevalence of human trafficking or changes over time per country, and thus of progress towards achieving the goal of 'ending exploitation and trafficking' as defined in SDG target 16.2.

The option of Multiple Systems Estimation

The development and testing of methods to better estimate the prevalence of human trafficking has for some time been a priority for several researchers and international organizations. The International Labour Office (2009), in collaboration with Walk Free Foundation, has applied the methodology of sample surveys to measure self-reported rates of victimization by human trafficking. Results from surveys in forty, mainly developing, countries, were used to calculate global and regional rates of 'modern slavery' victimization, as well as rates of hundred or so other individual countries through extrapolation (Walk Free Foundation, 2018).[7] In a subsequent wave in 2020 surveys were conducted in an additional thirty countries, bringing the total to seventy. The survey methodology, however promising in many respects, has inherent limitations[8]. It is difficult to reach international consensus on the pre-

6. UNODC, *Global Report on Trafficking in Persons 2018* (United Nations publication, Sales No. E.19.IV.2), pp. 21-22.

7. The Global Slavery Index presents estimates of the prevalence of 'modern slavery', a concept that does not have an internationally agreed legal definition and is described as follows: "Modern slavery refers to situations of exploitation that a person cannot refuse or leave because of threats, violence, coercion, deception, or abuse of power".

8. For a review of different approaches to measuring human trafficking, see Zhang & Joudo-Larsen, 2021, and Barrick & Pfeffer, 2021.

cise questions used to measure trafficking in persons, as defined in the UN Protocol, and on the appropriate sampling frames and modes of interviewing.[9] To determine statistically meaningful changes over time in the prevalence of human trafficking requires regular repeats of large-scale surveys with concomitantly large budgetary implications.

In 2014 Silverman and others applied MSE to the database on trafficking victims compiled by the National Crime Agency of the United Kingdom (Silverman, 2014; Bales, Hesketh & Silverman). Next, a pilot study, initiated by UNODC, applied MSE to the available dataset on human trafficking victims maintained by a dedicated NGO in the Netherlands (CoMensha, formerly La Strada Netherlands) (Van Dijk & Van der Heijden, 2016). A follow up study differentiated between various categories of victims including victims of trafficking for sexual exploitation or forced labour. Results were co-published by UNODC and the Dutch National Rapporteur on Trafficking in Human Beings in 2017.[10] Since then similar studies have been conducted in other European countries, as well as in Australia and the USA. Results have been published in the UNODC Global Reports on Human Trafficking, and in a special issue of the Journal on Crime and Delinquency (2021). In 2021 the Dutch study was repeated with funding from the Ministry of Justice and Security, using data over the years 2016-2019 (Van Dijk, Cruyff & Van der Heijden, 2021).

The ILO Guidelines concerning the Measurement of Forced Labour, adopted by the 20th Conference of Labour Statistics in 2018, mention MSE as a cost-effective method to estimate prevalence besides population surveys[11].

In sum, MSE has over the last years been developed into a well-tested, promising method to estimate the prevalence of human trafficking through an analysis of available statistics on recorded cases.

9. The challenges of survey research in this field include the skewed distribution of human trafficking among the populations of many countries with a concentration among hard-to- reach segments such as irregular migrants, and the sensitivity of questions on sexual exploitation and forced criminality. For a study on human trafficking among migrant communities, see Zhang, 2012.

10. Available at: https://www.unodc.org/documents/data-and-analysis/tip/TiPMSE.pdf . See also Cruijff, Van Dijk and Van der Heijden, 2017.

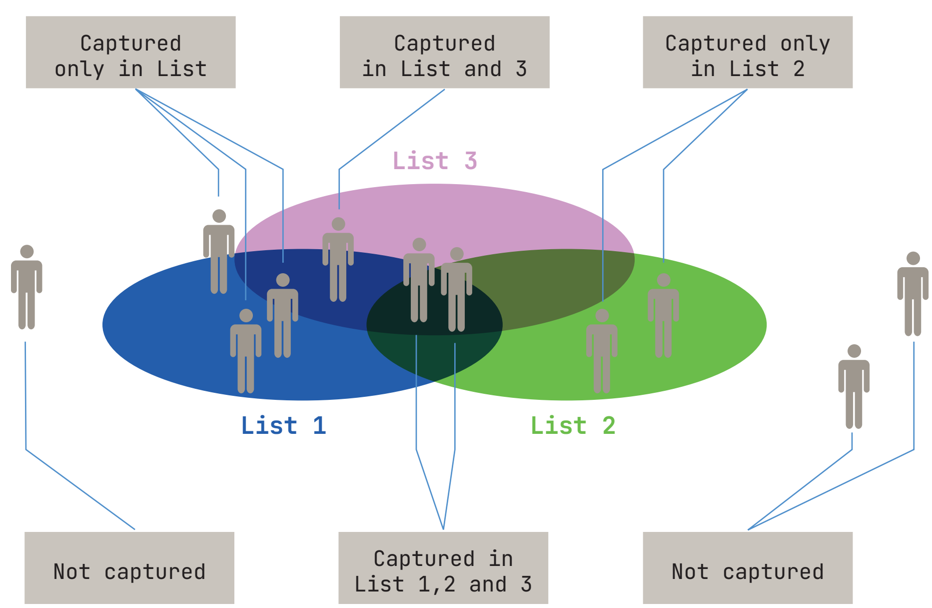
11. "Administrative records, such as lists of persons in forced labour compiled by local authorities or police records, or by non-governmental organizations and other service providers, may be useful for producing estimates of the prevalence of forced labour at relatively low cost. Where there are different administrative sources that refer to a common reference periods and can be confronted against each other so as to measure their overlap with reasonable accuracy, estimates of the prevalence of forced labour may be derived under certain assumptions, known as multiple systems estimations" (ILO, 2018).

What is MSE?

In the third section of this paper a more formal introduction will be given to MSE as a statistical method. Here we will try to explain the basic principles behind MSE for the general reader.

According to some sources, MSE goes back to the Capture-Recapture procedure through which the marine biologist Petersen estimated the size of populations of fish in ponds or fjords. According to others the method has already been applied much longer ago (Goudie & Goudie, 2007). Be that as it may, the procedure to estimate populations of fish by biologists went as follows. The researchers first captured, for example, ten fishes, marked them and set them back in the water. At a later time, they came back fishing in the same pond to see which part of a new batch of fishes captured was marked. When this amounted to, for example, two fishes out of the ten, they knew that of the ten marked fishes left in the pond, one in five had been caught. Under the assumption that the other fishes in the pond had roughly the same chance to be caught, they concluded that there were around fifty fishes in the pond of which ten marked and forty not. This estimation procedure is known as the Petersen estimator.

Figure 1 *Illustration of how MSE works by assessing overlaps between lists of observed victims to estimate the total population*



Fuente: UNODC elaboration

MSE can be understood as an advanced version of Capture-Recapture, whereby the size of a hidden population of humans is estimated by analysing the overlap between three or more administrative lists on which persons belonging to that population appear. Persons belonging to the hidden population of trafficking victims can, for example, be registered by several governmental agencies such as police, immigration, labour inspectors as well as by private providers of legal, medical or psychological assistance or child or youth care. By modelling the distribution of the recorded victims over these lists, an estimate can be made of those victims who do not appear on any of the lists from either police, NGOs or other institution. This exercise provides an estimate of the unrecorded (or uncaptured) numbers, as illustrated in the figure below

Some fundamental assumptions behind MSE

The fundamental assumptions of MSE are best explained by discussing the fundamental assumptions of the Dual Systems Estimation (DSE), i.e. MSE with only two lists (or Capture-Recapture). According to an authoritative review (IWGDMF, 1995), data used for DSE must meet four fundamental assumptions:

1) the data must relate to persons belonging to a closed population;
2) each person must be uniquely identifiable in order to be matched across lists;
3) each person must have the same chance to be included on the lists;
4) when only two lists are used, placement of a person on one of the lists must be statistically independent from placement on the other.

The first assumption can be readily understood in the example of fishermen returning for a second catch. If the location where the catches take place is a fast streaming river, marked fishes would have a near zero chance to be recaptured. In DSE regarding human populations the condition can be approximated by restricting the analysis to data on persons present in a geographical area during a limited period of time. In the case of data on recorded trafficking victims the assumption can be approximated by using a dataset on persons recorded as (presumed) victims of trafficking on the territory of a country during a certain year.

The second assumption can technically easily be met when the data on victims are derived from administrations of, for example, law enforcement agencies, NGO's providing support and so on. Such administrations usually contain personal data on the persons included. However, privacy protection regulations may stand in the way of the exchange of such personal data across agencies. Compliance with such regulations may therefore require special measures in an MSE study, as will be discussed below at some length.

Previous research has demonstrated that the third assumption of equal inclusion probabilities need to be met by only one of the lists used and not by all (Van der Heijden et al., 2012). In the case of human trafficking victims, national administrations of the police usually qualify as such. Nevertheless, even this relaxed assumption concerning equal inclusion on one of the lists can cause problems for DSE with trafficking data which need to be addressed. Various subcategories of trafficking victims might be heterogeneous in their likelihood to be detected by the police or any other relevant organizations. For example, forced labour in remote areas of a large country might be harder to detect than sexual exploitation in well-known red-light districts in cities.

The fourth fundamental assumption behind DSE~~,~~ is that a victims' probability of being recorded on the list of organization A, e.g. the police, should by itself be independent ~~on~~ of a victim's probability to be recorded on list B of another organization. The fourth assumption logically follows from the third assumption as, when the probability of being recorded on list A is identical for those being on list B or not being on list B, independence holds. In the example of catching fishes, the fact that fishes are caught and marked during the first catch should neither increase nor decrease their probability to be caught again during subsequent catches. The reasons why the probability of members of a hidden population to be recorded by a particular organization can be co-dependent on their placement on the list of another organization are manifold.[12] In the field of human trafficking the placement of victims on lists of service-providers may be co-dependent on their prior placement on a list of law enforcement institutions. In fact, in countries with National Referral Mechanisms law enforcement agencies are supposed to refer identified victims of human trafficking to service providers. Some measure of statistical dependence is therefore to be expected.

We conclude that the last two assumptions are problematic and will not hold in general. There are two solutions to this. For the third assumption, persons having the same probability to be included on each of the lists, a solution is to make use of covariates for background characteristics such as information on the type of exploitation of each victim to the dataset for use in the MSE. When this 'covariate' is added, the assumption of equal probabilities of persons to be included on a list no longer applies to all victims but to categories of victims such as victims of forced labour or

12. Dependency between lists can be based on the fact that subgroups with certain characteristics are more likely to be placed on two lists than others. Such 'apparent dependency' disappears when the said characteristic is included in the analysis as covariate.

of sexual-exploitation.[13] This applies likewise to other relevant covariates such as age and sex. The inclusion of key covariates in the analysis has the additional advantage that the estimation results can be broken down by type of exploitation, age and sex as required for monitoring progress in implementing SDG 16.2.

Despite the usefulness of including covariates to make the assumptions of the DSE model more realistic, the problematic assumption of independence of inclusion in the two lists persists. Fortunately, this problem is solved in MSE, with the use of three or more lists. When three lists or more are used, it is possible to determine for all persons on list 3, whether and to what extent, the likelihood of their placement on list 2 (e.g. of shelter homes) is co-dependent on their presence on list 1 (e.g. the police). By including such interactions between lists in the estimation model, it is possible to correct for known dependencies between certain lists in the estimation. In section three we will demonstrate how this worked out in the MSE using Irish trafficking data divided over three lists.[14]

The independence assumption in DSE is seen as a major obstacle, as violation of this assumption can lead to seriously biased estimates. For this reason, in this handbook we take the position that we advise only to apply MSE, i.e. make use of more than two lists. We also advise to include covariates, when they are available.

The 'official status' of MSE-based estimates

As explained above, MSE ideally uses data from three or more official, administrative lists of victims to estimate the total size of a hidden population. By using existing administrative data for estimating the 'hidden figure' of unobserved victims, the researcher adopts the legal and operational definitions used by the key institutions involved. This is especially the case if the administrative lists form part of a National Referral Mechanism or another set of arrangements for inter-agency cooperation in the rescue and support of victims. Typically, such cooperation mechanisms provide guidance to practitioners from all institutions involved concerning the

13. In the example of the capture and recapture of fishes, the underlying assumption of the estimation was, as said, that the unmarked fishes in the pond possess roughly the same probability to be caught as the marked fishes. In most circumstances this cannot be taken for granted since there might be various species of fish swimming in the pond with different likelihoods to be caught by fishermen. By including the covariate 'type of exploitation', the estimation procedure can be applied to different types of fish separately in order to arrive at a better estimation of the true numbers overall.

14. For an investigation of the impact of referrals between lists on MSE, see also Jones et al. 2014.

operational definitions of human trafficking, in the form of lists of 'signals' of a possible case of human trafficking. A prime example of such lists is the international list of indicators designed by a joint working group of the European Commission and ILO (ILO, 2009).

From a policy perspective, the use of officially approved definitions and operational indicators of human trafficking victims is one of MSE's strengths. Unlike survey-based estimates, typically based on operational definitions designed by researchers, and as interpreted by individual respondents themselves, MSE-based estimates reflect the state-approved operational definitions of each country as interpreted by specialised public officials or staff working for dedicated NGO's or researchers adopting prevailing definitions. The MSE estimates show the numbers of victims which would be recorded if all institutions involved in the fight against human trafficking in the country were optimally enabled to implement their existing mandates to fight human trafficking as officially defined in the country.

Experiences over the past few years have taught that MSE based estimates were often two to four times higher than the numbers of victims currently recorded (for an overview of results see table 4 in section 4). Such estimates allow policymakers to set targets for improved action that can realistically be met in the foreseeable future. In addition to providing a realistic assessment of what improved national trafficking policies could achieve overall, the results provide insight into the categories of trafficking victims which have in recent years remained the most hidden and should therefore be prioritized. In the first Dutch study, using covariates, victims of forced labour and child victims were found to be relatively less likely to be detected/ recorded than other categories of victims (UNODC, 2017. Such results provide fact-based guidance for policymakers concerning priorities in detection or outreach of service-providers.[15] Taken together the results of MSE studies allow policy-makers to make their anti-trafficking policies more data-driven, including by suggesting achievable, quantitative performance targets per type of exploitation or category of victims for upcoming years.

15. In the Netherlands a new policy plan issued in 2019 gave higher priority to detecting chold victims and victims of forced labour (See Ministries of Justice and Security, Social Affairs and Employment, Health, Welfare and Sports, and Foreign Affairs, Government of the Netherlands, *Together against human trafficking: An integrated program to tackling sexual exploitation, labor exploitation and criminal exploitation,* February 2019 (in Dutch))

Data protection issues

The conduct of MSE relies on the availability of three or more suitable administrative lists concerning victims of trafficking and the possibility to determine which appear on more than one list (matching). Since the status of a trafficking victim constitutes sensitive personal data, for example in the case of victims with irregular status and/or those involved in illegal prostitution or other criminal activity, the exchange of personal data on victims from various institutions requires due attention to data protection concerns. While contexts and legislation vary, a clear legal basis for the exchange and collection of data must be established. A common and often preferred legal basis for the collection and exchange of personal data of trafficking victims in assistance settings is the informed consent of the registered person, or in the case of minors, of their guardian. In practice, such consent can sometimes be difficult to obtain. Victims or guardians may have trouble understanding its implications. Additionally, it may be difficult or even inadvisable to contact victims to obtain consent to share their data for research after they have moved on from contact with NGOs or others, either due to missing address information or because of the danger that the request may fall into the hands of a third party. Where data are collected and shared on other grounds than consent, it should be considered whether there are ethical implications, with a particular focus on not undermining trust and confidence between trafficking victims and service providers/authorities.

In countries where comprehensive statistics on human trafficking victims are produced on a regular basis, data is collated from different public and private institutions. To prevent double counting of the persons recorded by more than one institution, the dataset usually includes either full names or unique identifiers such as birth data and first initials. To address data protection concerns, several countries have set up dedicated state institutions or NGO's as clearing houses of data on victims of human trafficking from different sources, operating under strict regimes of data security and protection. In some countries, for example Rumania, Serbia, Slovakia and Portugal, such agencies have been put on a statutory footing. In countries where such agencies are operational, existing multi-source datasets can be de-identified after matching and shared with researchers for the purpose of MSE. Where no such agencies are operating, trusted research teams must be tasked to collect and process data from various sources using unique identifiers such as birth date and initials to allow matching. This procedure has been followed in Australia (Lyneham et al., 2019) and in some local studies in the USA (Farrell et al, 2019; Anderson et al., 2019). To comply with data protection requirements, the identifiers can also be encrypted whereby the 'key' is known to the contributing agency and the research team only. Another option is to involve a trusted third party such a Statistical Authority or public notary in the collection and matching of data on victims recorded on different lists [16]. In that scenario, the unique identifiers can be deleted after the matching is done and the researchers will receive a fully anonymized dataset for analysis. In the next section measures to assure compliance with data protection regulations will be discussed in more detail.

SUMMARY POINTS

- Since statistics on recorded cases of trafficking victims largely reflect detection and recording efforts of relevant institutions, they are flawed as indicator of the true magnitude of the problem in a country, and cannot be reliably used for national planning purposes, target setting or reporting on progress in the implementation of SDG 16.2 (eradication of all forms of human trafficking) to the United Nations.
- A tested option to estimate true prevalence is the conduct among the general population of surveys on personal experiences with human trafficking/exploitation. Such studies are costs intensive, especially when repeated periodically to monitor trends and tend to produce prevalence estimates that go far beyond recorded numbers.
- As an alternative it seems worth considering the application of Multiple Systems Estimation (MSE) to (potentially) available datasets on recorded victims collected by law enforcement agencies, prosecutors and private sector organizations providing services to victims. Such estimates can be made against modest additional costs. They provide an estimate of the numbers of victims of human trafficking as officially defined which can realistically be rescued and supported when institutional capabilities are improved and/or priorities reset.
- Although estimation is possible using only two separate lists of victims, leading to dual systems estimation, we advise MSE using a minimum of three separate lists of victims, for example those of police forces, labour inspectorates, victim support organizations and others (e.g. child care services, law firms, international organizations). The reason for this advice is that the assumptions for MSE are more easily fulfilled than for dual systems estimation. There is no need for full harmonization of the

16. Encryption was applied in local studies in the USA (Farrell et al., 2019). In the Netherlands local institutions from two cities, Utrecht and Ede, were invited by a research team to upload data on human trafficking cases directly to a secured website maintained by the Central Bureau of Statistics of the Netherlands. After the linking of the datasets provided by these institutions, the CBS supplied an anonymized integrated dataset to the research institute Regioplan for further analysis (Buimer et al., 2021).

inclusion criteria of these lists as long as each reflects prevailing operational definitions of trafficking in the country.

- Most administrative databases make distinctions between type of exploitation, gender, age, and nationality. To the extent possible such covariates should be included in the analysis, since this will make estimates of victims technically more robust. Because the estimates can then be differentiated across categories of victims, they are also more useful for policy purposes, for example, for setting new targets for categories of the most hidden victims (e.g. minors).
- If multi-institutional arrangements are in place to detect and support trafficking victims and collect multi-source statistics, MSE is usually feasible with minimal cost implications besides hiring a qualified analyst (see UNODC, 2020). Eventually, MSE could develop into a regular reporting task for the national agency responsible for statistical reporting on human trafficking.
- If no National Referral Mechanism or similar arrangement for detecting and supporting trafficking victims is in place, MSE requires the involvement of a research team to approach the institutions involved and extract data on presumed victims from existing administrative registers applying operational definitions of human trafficking in line with national legislation and case law (for an example of such studies in the USA, see Farrell, 2019).
- For the purpose of MSE persons placed on different lists must be matched against each other using unique identifiers such as birthdate and initials. In some countries such matching can be done by existing clearing houses mandated to collate human trafficking data for statistical purposes. In other countries data protection issues around the linkage of lists must be addressed, for example by using encryption measures and/or the involvement of trusted third parties such as Statistical Authorities Examples of this can be found in studies conducted in Australia (Lyneham et al., 2019), the USA (Farrell et al., 2019; Anderson et al., 2019) and the Netherlands (Buimer et al., 2021).

SECTION 2

Identifying, collecting and organizing data suitable for MSE

Obviously, is not possible to apply the statistical MSE methodology without suitable data sets. Once a country has decided to carry out a national study to estimate the number of trafficking victims with the MSE methodology, the first task is to find out whether there are sufficient suitable administrative data available. As said, most countries with a decisive response to trafficking already collect, store and publish data on victims of trafficking from different sources at the national or local level, although the dataset itself might not be readily accessible to researchers. This part of the manual will tackle the technical data requirements and the steps to take to collect and prepare the data for the statistical estimations.

Clarification of terms

In the context of MSE and for the purpose of this manual, a "victim" of trafficking is a person recorded as such by competent entities/national institutions as either a presumed or a definitely identified victim of human trafficking. This implies a 'victim' is a person who has suffered trafficking victimization, and because of this, has somehow come into contact with an entity that is part of a country's response to human trafficking. Moreover, the person has been recorded as such by the entity as meeting some commonly agreed upon criteria.[17]

A *'list'* is a register of individual trafficking victims maintained by the entities that come into direct contact with victims. Lists may vary significantly in terms of quality and quantity of information about each victim. However, most lists will include, at a minimum, some basic individual information (for example, name, case number, date of birth, date of entry into the organization's registry) as well as some profile

17. The victims may not have formal status as identified trafficking victims but may be classified as, inter alia, 'presumed', 'suspected', 'possible' or 'probable' victims.

information (sex, nationality and form of exploitation) about each victim that the organization has encountered.

'Contributing entities' are those organizations that may provide victim data to be used for the study, as a result of their role in the country's trafficking response. In most countries, the most relevant contributing entities will include law enforcement, Non-Governmental Entities (NGO's) assisting victims of trafficking, labour inspectorates and migration authorities. In many countries there are often numerous other possible contributors, including anti-trafficking coordinating bodies, local/regional assistance providers, providers of services to migrant workers, prosecutors, international organizations such as IOM, and public hotlines, to mention a few. If the numbers are small, these secondary providers can be grouped together under a list of Others (as was done in the study on Ireland presented in section three).

Data collection

Countries with a formalized National Referral Mechanism usually encourage a wide range of public and private institutions to assign the status of presumed victim to persons meeting a threshold of officially established operational indicators. This status offers such persons certain entitlements such as accommodation and a temporary residence permit in case they are at risk to be expelled. In most of these countries the authority to assign the definite status of an *identified* victim receiving a wider and more permanent set of entitlements is vested in the national police. The United Kingdom provides an example of a fully institutionalised National Referral Mechanism which assigns both statuses. In this country around a third of all presumed victims are eventually recognized as definitely identified victims.

A preliminary issue to consider when planning a national study is whether to use data pertaining to *presumed* victims or only to *formally identified* victims. In most countries victims who have been recorded solely by NGOs are regarded as presumed victims. If the MSE methodology is applied to lists of identified victims only, victims who only come into contact with NGOs - for instance because they are unwilling to cooperate with the police - are excluded from the analysis, thereby severely restricting the scope of the estimate. Moreover, using only data from formally identified victims would reduce the number of countries where MSE can be applied, as there will often be insufficient data from identified victims alone. For this reason, MSE studies in the domain of human trafficking have so far always used integrated, multi-source datasets on *presumed* victims (of which only some have ultimately been formally identified).[18]

The selection of suitable lists

The primary challenge for the implementation of a national study with the MSE methodology is to obtain roughly compatible victim lists from at least three different entities.

To ensure sufficient reliability of the estimates, information on a critical mass of victims is required. While there is, from a methodological perspective, no strict lower limit, analyses of small datasets will produce estimates with large margins of error, of limited use for policy purposes. The MSE studies conducted so far have been carried out with datasets containing information on at least 150 victims in total for a single-year study, or, with data for multiple years, a minimum of 50 victims per year. Although minimum numbers cannot be given, there should be at least some overlap between the lists used for the analysis. One option to increase overlap is the collation of smaller list into larger ones.[19] Researchers are currently studying the application of the MSE methodology to scarce datasets, so this recommendation may become subject to change in the future.[20]

Required as a minimum are micro data concerning uniquely identifiable persons containing information on the contributing entities which have reported them in the course of a year and, preferably, a set of covariates. As central agencies collating such data seek to correct their aggregated totals of victims for 'double counting' by reporting institutions, information on recording by different entities of the same victim is usually available in their regular datasets. In some cases, however, only the first recording entity is entered into the dataset and information on further recording has to be derived from case files and added to the regular dataset retroactively.

In countries where a central entity with a national mandate to collect such data exists– a so called trafficking data 'clearing house' - this entity is obviously best placed to lead the data collection work, in liaison with the researchers in charge of the MSE study. Preferably, the MSE analysts liaise with a counterpart at the agency working

18. Theoretically, it seems possible to carry out MSE on an integrated dataset distinguishing between presumed and identified victims but this has, to our knowledge, not yet been done (Overstall et al., 2014).

19. See Sharifi Far et al., 2020.

20. See, for example, Chan, Silverman and Vincent, 2021.

as clearing house who can explain how the central agency and each contributing entity assembles and processes the data.

As explained in the first section, MSE does not necessarily require equal inclusion chances on all lists. If all victims, or categories of victims, have an equal chance to be placed on at least one of the lists, for example the one from the national police, other lists may apply different inclusion criteria, for example focussing on victims of forced labour only.

Although there is no requirement either that all lists apply exactly the same definitions of human trafficking, it is advisable that all lists broadly adhere to the official definitions of human trafficking in the country. To the extent that this is the case, the estimate of the true numbers of victims will be more convincing for policymakers since it conforms to their own, official definition of trafficking.

As known, in almost all countries the Palermo definition of human trafficking has been incorporated into national law. This definition has been operationalized into lists of 'signals' which can be recognized by frontline workers and used to identify cases of presumed victimization by human trafficking. A prime example of such list was developed by the ILO, in collaboration with a working group of the EU Anti-Trafficking Coordinator (ILO, 2009). Similar lists of detailed indicators or 'signals' have been published by an EU-funded European consortium led by the French Ministry of Foreign Affairs (2013) and, more recently, by the Croatian Red Cross (2019). Across Europe versions of this list, adapted to local situations, are in use among entities contributing data to centralised national databases.

An example of how an entity contributing data on presumed victims to a central clearing house identifies presumed victims with a list of operational indicators is provided by the Dutch NGO Fair Work, assisting migrant workers exposed to bad working conditions. Staff and volunteers are informed of possible cases of human trafficking by a wide array of institutions in the country and by the victims themselves. They are instructed to enter cases as presumed victims of forced labour/human trafficking into Fair Work's database if the case ticks three or more boxes of the list of 'bad working conditions' or at least one box of 'bad working conditions' and at least one box of any of the other four lists (multiple dependency, limitation of freedom, violation of integrity, organized structure of exploitation.). Figure 2 presents an overview of the five list of indicators/boxes.

Figure 2 Operational indicators of forced labour used by FairWork Netherlands

0 ☐	**1. Multiple dependency (=signal)**
☐ yes	*The employer/exploiter also arranges housing, clothing, transport etc.*
	☐ Didn't arrange the traveling, visa etc themselves
	☐ Recruited in country of origin
	☐ Working on a false or forged passport
	☐ Lack of own living space in The Netherlands
	☐ Housing linked to employment (obliged use of housing offered by employment agency, forced to move out after being dismissed)
	☐ Sleeping in workplace
	☐ Unfamiliarity with own work address
☐ yes	*The client is in a vulnerable position*
	☐ Limited mental capacity
	☐ Bad economic situation in country of origin
	☐ Illegally staying/working in The Netherlands
	☐ Client does not speak Dutch (or English)
☐ yes	*The client is in debt*
	☐ Debts to someone else than the employer
	☐ Employer has paid an amount for transfer of employee
☐ yes	*Not being able to control (a part of) own income*
	☐ No possession of own bank account/cash card
	☐ Employer forces employee to give access to his bank account
	☐ Obliged to buy transport, food, and/or other services at just one supplier
☐ yes	*Other, namely:* ☐

☐ **TOTAL**	**2. A strong limitation of basic freedom of the person concerned (=signal)**
☐ yes	*The victim can- or may not, be in touch with the outer world.*
☐ yes	*The victim doesn't have independent freedom of movement*
☐ yes	*The refusal of medical aid*
☐ yes	*The victim doesn't possess own identity papers*
☐ yes	*Employee is tied to the employer through a debt*
☐ yes	*Forced criminal activities*
☐ yes	*Other, namely:*

☐ Total	**3. Working under bad conditions (=signal)**
☐ yes	*Contract is not of adequate quality*
	☐ No labour-contract
	☐ No copy of labour-contract
	☐ Labour-contract is not in the language of migrant
	☐ Only a verbal labour-contract
☐ yes	*Work differs from the employers' promises*
	☐ No or less work: working based upon a zero-hours contract, functioning as stand-by
	☐ Lower wage than promised
☐ yes	*Labour rights are not respected*
	☐ Many overtime hours at work (per day or week)
	☐ Overtime hours are not paid
	☐ Cheating with hour registration
	☐ No (regular/holiday) days off
	☐ Dismissal without reason
	☐ Wage under the official minimum wage
	☐ Wage paid in cash
	☐ Only received money in advance
	☐ No paychecks
	☐ Handwritten paychecks
	☐ No annual statement
	☐ No payment of holiday bonuses
	☐ No payment in case of illness
	☐ Forced to work in case of sickness, or non-adapted work during pregnancy
	☐ No payment of last period prior to dismissal
	☐ Payment of juvenile wages when employee is older than 23 years old.
	☐ Payment based on incentive wages
	☐ No premiums for overtime, holidays or irregular working hours
☐ yes	*Restrained wages*
	☐ Fines (in case of dismissal, parking etc.)
	☐ For schooling
	☐ For BSN (identity number)
	☐ For work outfits/equipment/trainings etc.
	☐ For opening of bank accounts
	☐ Without clear reasons
	☐ For transport from/to country of origin
☐ yes	*Work under dangerous or unhealthy circumstances*
	☐ Physical or mental damage caused by labour

	☐ Works with dangerous substances or old/poor materials
	☐ Has to execute (dangerous) activities that are not agreed upon
	☐ No instructions
	☐ No protective measures in case of dangerous or unhealthy circumstances
☐ yes	*Bad housing*
	☐ Quality of housing doesn't suffice the official norms
	☐ Very high renting prices
	☐ The house is too small for the amount of people that live in it
	☐ Refusal of privacy
	☐ No heating system
	☐ insufficient sanitary facilities
☐ yes	*Unequal treatment*
	☐ Low wages compared to collegues
	☐ CAO is not applied (collective employment agreement)
	☐ Working under different (worse) conditions than Dutch employees
☐ yes	*Restrained national insurance premiums*
	☐ In case of disability not registered by the UWV or immediate dismissal after declaration of disability/workplace accidents
	☐ No distribution of national insurance premiums
	☐ No healthcare insurance, or no information about the obligation to insure
	☐ Is paying premiums to employer but is not insured
☐ yes	*Extreme forms of protection/control*
	☐ Buildings with cameras (also inside), bodyguards etc.
☐ yes	*Other, namely:* ☐

☐ **Total**	**4. Violation of physical or mental integrity of the person concerned (=signal)**
☐ yes	*Sexual intimidation*
☐ yes	*Execute sexual acts as condition to obtain a job*
☐ yes	*Threatened or confronted with violence*
☐ yes	*Fear for colleagues and/or manager, bullying etc.*
☐ yes	*Discrimination, namely....*
☐ yes	*Other, namely:* ☐

☐ **Total**	**5. The exploitation is not incidental but there is a pattern or some kind of organised structure (=signal)**
☐ yes	*In organized structure with employment, labour or travel agencies*
☐ yes	*Exploitation is systemic, new people are recruited over and over again*
☐ yes	*The duration of the exploitation was long (more than one year)*
☐ yes	*Other, namely:* ☐

Within Europe the establishment of centralised multi-source databases on victims of trafficking is actively promoted by the Council of Europe (with 40 Member States) and the European Union (27 Member States). This means that opportunities for carrying out MSE are gradually increasing in this region. In most countries outside Europe, consolidated lists of victims maintained by a central 'clearing house' do often not yet exist, least of all at national or federal level. In many countries, some data on victims are collected by various organizations such as the police and NGO's but not collated into a single dataset, maintained by a dedicated 'clearing house'. Integration of data on trafficking victims requires high levels of cooperation, harmonization and trust which can be especially difficult to achieve in countries with federal state structures in North and South America. In such countries, MSE studies are still feasible but require a preparatory stage wherein existing data from various entities are collected, standardized and consolidated by research teams. Steps towards data sharing on trafficking victims on a regular basis have, inter alia, been taken in Brazil (the Human Trafficking Data Lab) and in Chili (Quinteros Rojas, Dufraix Tapia & Ramos Rodríguez, 2019).

When data on recorded victims must be collected for the purpose of MSE, it is vital that all contributing institutions use the same unique identifiers. Such studies therefore need to be based on agreements between the contributing entities to share personal data with a research team, respecting national and international standards of data protection. Such multi-staged MSE studies have so far been conducted in three unnamed locations in the United States (Farrell et al., 2019) and, concerning child victims of sexual exploitation, in the state of Ohio (Anderson et al., 2019)[21]. In Australia an MSE study whereby the data had first to be collected from various agencies by a dedicated research team was conducted at the federal level (Lyneham, Dowling and Bricknell, 2019).

To persuade different entities to share their sensitive victim data may require a leadership role by senior officials or politicians. In Australia, for example, the Minister of Justice tasked the Australian Institute of Criminology with conducting an MSE study on human trafficking prevalence and to collect the necessary data for that purpose.

As discussed, entities recording data on presumed victims usually apply checklists of operational indicators of the commonly shared Palermo definition of human trafficking. Experience has taught that experienced practitioners working in law enforcement, victim support or otherwise share a basic understanding of the operational indicators of what a presumed trafficking victim is. Yet, the operational indicators of the various institutions may differ in some respects, reflecting their unique

21. In the Ohio study a total count of recorded victims controlling for double counting was made but no MSE was carried out.

professional contexts and perspectives. Law enforcement officials, for example, tend to focus somewhat more on indicators of a possible successful criminal investigation of the traffickers and care providers on symptoms among the victims of the impact of the crime. Also, officials of the border police will, of necessity, look at indicators visible during the brief time they can interview non-nationals entering the country. Their situation differs from that of a police officer, interviewing victims at a police station, or a social worker assisting a victim over a longer period of time. As a consequence, entities may sometimes 'miss' a presumed victim recorded by another because of different criteria or of missing information in individual cases. Such incongruities between some of the lists do not, as said, preclude the conduct of a valid MSE.

In principle, the lists can come from any entity detecting and recording victims of trafficking, and any combination of sources. For methodological purposes, it does not matter whether a country uses, for example, a law enforcement list, a labour inspectorate list and a public hotline list, or, alternatively, three lists from different victim assistance NGOs, or any other combination of lists.

Although a minimum of three lists is vital for various technical reasons, to be discussed below, there is no maximum number of lists. When datasets are available containing more than a few hundred entries, the use of more than three lists is in fact recommendable since this will make the estimates more robust. As a case in point, the UK study used five lists and the comprehensive MSE studies in the Netherlands of both 2017 and 2021 six.[22] When more than six lists are available the model selection might become too complicated. To avoid this, some of the available smaller lists may be combined into larger ones. In general, where there are several lists with few victims, these may be combined into one larger list, provided that the total number of lists to be used for the study will remain at least three.[23]

In some countries, one entity may maintain a comprehensive list of victims, and such lists usually also contain indications of which institution originally came into contact with the individual victims listed. On the basis of this information, the list can be subdivided into different contributing entities.[24]

22. These were lists from national police, border police, labour inspectors, regional trafficking coordinators, residential treatment centers and shelters, and an 'others' list that included, for example, victims that had come into contact with non-residential care centers, organizations providing legal advice, and the Dutch Immigration and Naturalization service.

23. For an investigation into the robustness of combining and omitting lists, see Sharifi Far et al.,2021.

24. For example, in Serbia, the Centre for Victims of Human Trafficking maintains a comprehensive list of all victims recorded by one or more institution. By going back to the available individual victim files, the Centre succeeded in disaggregating the list into three separate list of institutions acting as either primary or secondary contact points.

The requirement of separate lists

An assumption behind dual systems estimation methodology, i.e. multiple systems estimation of only two lists is, as explained, that inclusion in list 1 is independent of inclusion in list 2. However, as discussed above, in the case of human trafficking victims, lists of victims cannot generally be assumed to be independent, as appearing on one list often increases the victim's probability of appearing on another. This is because victims are, for example, first detected by law enforcement and subsequently referred to NGOs for support, or they approach NGOs for assistance and are then referred to government agencies. It is also possible that victims placed on one list are less likely to be registered on another list (negative dependence). The requirement of including a third list (or more) in the statistical model is a strategy to deal with such dependencies between lists.[25] When a third list is included, the models used in MSE allow to include pairwise dependencies between all three lists. Thus, possible dependence can be corrected for in the estimation procedure.

Although the assumption of independence between lists can be relaxed by using at least three lists, this does not imply that dependence between lists is of no consequence for MSE. A pilot study with data from victim lists fully based on referrals have shown that such data are not suitable for MSE.[26] The key lesson is that institutions that provide victim lists may accept referrals from each other, but each must also act as a separate, primary contact point for victims of trafficking.

Matching cases through unique identifiers

The crux of the MSE methodology is to establish the overlaps between the different victim lists and use that information to estimate the number of victims that do not appear on any lists. In order to do so, analysts need to determine on how many of the different lists the same individual victim appears. This is done by a process of 'matching', or individual record linking, which clarifies - victim by victim - the number of lists on which each victim appears.

25. This is further explained in Section 3, Statistical modelling.

26. The test involved using data from a database on assisted victims maintained by the International Organization for Migration (IOM) in Belarus which contained information on the various institutions that had referred victims to the IOM programme of assistance. Through a process of list separation, victims were identified on the basis of the institutions that had referred them to IOM, including those who had been referred by more than one institution. Although the final model produced a reasonably stable estimate, the used dataset was felt to insufficiently meet the assumption of list independence, and the results of the estimation were not published.

For matching to be carried out accurately, a unique identifier is required for each individual victim appearing on any of the lists. Unique identifiers like birth date or name are often present in the datasets of the institutions involved. In several European countries statutory-based or consensual arrangements have been set up for the pooling of data on identified victims. In such arrangements participating state and non-state agencies share data on identified victims with a repository. For example, in the Netherlands state and non-state agencies share such data with a dedicated and state-funded Non-Governmental Organization, CoMensha. In Portugal, Rumania, Serbia and Slovakia, inter alia, such data sharing with a dedicated state institution, usually an anti-trafficking unit, is, as said, statutory-based. [27]

In many countries state agencies and non-state agencies are not allowed or unwilling to share personal data of a sensitive nature such as the status of a trafficking victim. Under these circumstances secure identity keys can be constructed by concatenating identifying elements. A possible formula is the use of the first letters of first name and family name and birth date. In a study in the United States, researchers collating data from different institutions used the first three letters of an individuals' first name, the first two letters of their last name, their month and year of birth, and their state of residence (Farrell, et al., 2019). A more advanced option is to scramble details from victims' identifiers. For example, Jane Doe; date of Birth: 01/01/1990; social security number: 123-45-6789 could be scrambled and recoded as 19J01D01678990 to ensure anonymity. Such encryption is recommended by American researchers when using record data on human trafficking for prevalence estimation (Anderson, Kulig & Sullivan, 2019b; O'Connell et al., 2020). In this scenario, access to the decryption keys must of course be strictly reserved to the contributing agencies and the repository.

If no unique identifiers can be employed, other information – location, length and form of exploitation, sex, age, country of origin, residence status, recruitment method, information about the trafficker, place of identification, and so on – can be collated into a comprehensive victim profile and used for matching,[28] even though this is laborious in cases of large datasets. Mismatches and missed matches may occur, which could affect the estimates. However, in the MSE studies undertaken to

27. For example, in Portugal the Observatory on Human Trafficking has been mandated by Decree/Law No 229/2008 to collect data on human trafficking from various sources.

28. The exact data fields to be used depend on what data is consistently captured across all victim lists. It is important to use enough fields to distinguish between similar profiles; for example, a combination of sex, age, citizenship and form of exploitation alone is unlikely to be sufficient.

date, the impact of matching errors has been marginal because unique identifiers could be used. [29]

Matching may be carried out by either the data 'clearing house' entity or the researchers in charge of the MSE study. The matching can also, as said, be done by a trusted third party, such as a National Statistical Office. It is important to note that while it is inevitable that the entity that will carry out the matching knows the identities of individual victims, this information can be deleted after the matching is done. The analysts carrying out MSE will normally work with a fully anonymized dataset (for an example, see Table 1 hereunder presenting data shared by the Irish Anti-Human Trafficking Unit with Utrecht University).

The datasets used in the Irish and British studies

To illustrate the type of dataset required for the statistical MSE calculations, Table 1 below shows part of the consolidated list of individual victims collected by the the Irish Department of Justice and Equality's Anti-Human Trafficking Unit (AHTU). This dataset relates to victims identified in 2014 by three different groupings of organizations: national police (An Garda Siochana), migrant organizations/IOM, and NGOs providing services to victims. The second and third lists are both combined lists, coalescing available smaller lists. In this case, AHTU obtained lists with unique identifiers from each of these organizations which permitted matching across lists. The dataset shared with the analysts was anonymized and only included the reference number of the individual victim. It also included the covariates sex, citizenship, age (adult/minor) and type of exploitation. The last covariate was made dichotomous by distinguishing between sexual exploitation and Other only. The three columns on the right side present the results of the matching, as they indicate which organization/s has/have identified the individual victims.

29. For an investigation of the use of datasets with incomplete matching for MSE see Sutherland and Schwarz, 2005.

Table 1 *Sample from the consolidated list of presumed victims identified by the Irish national police (An Garda Siochana), migrant organizations/IOM and dedicated NGOs*

Reference	Gender	Irish / Non-Irish	Age	Exploitation	An Garda Síochána	MRCI, ICI, IOM	Ruhama, SVCC, DL
2014-040	Male	Non-Irish	Adult	Labour	1	1	0
2014-041	Male	Non-Irish	Adult	Labour	1	1	0
2014-042	Male	Non-Irish	Adult	Labour	1	0	0
2014-043	Male	Non-Irish	Adult	Labour	1	0	0
2014-044	Female	Non-Irish	Adult	Sexual & Labour	1	0	0
2014-045	Female	Non-Irish	Adult	Labour	1	0	1
2014-046	Female	Non-Irish	Adult	Sexual	1	0	0
2014-047	Female	Non-Irish	Adult	Forced criminality	0	0	1
2014-048	Female	Non-Irish	Adult	Sexual	0	0	1
2014-049	Female	Non-Irish	Adult	Sexual	0	0	1
2014-050	Male	Non-Irish	Adult	Forced criminality	0	1	0
2014-051	Female	Non-Irish	Adult	Forced marriage	0	0	1
2014-052	Female	Non-Irish	Adult	Sexual	0	0	1
2014-053	Female	Non-Irish	Adult	Sexual	0	0	1
2014-054	Female	Non-Irish	Adult	Sexual	0	0	1
2014-055	Female	Non-Irish	Adult	Sexual	0	0	1
2014-056	Female	Non-Irish	Adult	Sexual	0	0	1
2014-057	Female	Non-Irish	Adult	Sexual	0	0	1
2014-058	Female	Non-Irish	Adult	Sexual	0	1	1
2014-059	Female	Non-Irish	Adult	Sexual	0	0	1
2014-060	Female	Non-Irish	Adult	Sexual	0	0	1
2014-061	Female	Non-Irish	Adult	Sexual	0	0	1
2014-062	Female	Non-Irish	Adult	Sexual	0	0	1
2014-063	Female	Non-Irish	Adult	Sexual	0	1	1
2014-064	Female	Non-Irish	Minor	Sexual	0	1	0
2014-065	Female	Non-Irish	Adult	Forced begging	0	0	1
2014-066	Female	Non-Irish	Adult	Sexual	0	1	1
2014-067	Female	Non-Irish	Adult	Labour	0	1	0
2014-068	Female	Non-Irish	Adult	Sexual	0	0	1
2014-069	Male	Non-Irish	Adult	Forced criminality	0	1	0

Source: Anti-Human Trafficking Unit, Department of Justice and Equality, Ireland.

The Irish dataset of which a sample is shown in Table 1 shows that several victims were identified by just one organization, but some were identified by two (and in other parts of the dataset there were some victims identified by all three). For the purpose of the MSE calculations, the data are expressed in a contingency table where the overlaps between the different lists are quantified. For an MSE with three lists, named A, B and C for illustration purposes, such a contingency table would include counts of individuals in list A only, list B only, list C only, lists A and B, lists A and C, lists B and C and lists A, B and C. Yet, there are no individuals observed that are not in all three lists, and this count is estimated in MSE, cross-classified by the covariate combinations if such covariates are available.

In the dataset used for an MSE study in the United Kingdom, data from several small lists were grouped together under five principal lists, detailed below:

- LA: Local authority
- NG: Non-governmental organization
- PF: Police force/National Crime Agency
- GO: Government organization (mostly Home Office agencies such as the UK Border Force and the Gangmasters Licensing Authority)
- GP: The general public, through various routes.

Of the 2,744 victims included in the 2013 database, some appeared on two, and a few on three or four of the five lists. The contingency table for this study is presented below, showing the summarized distribution of the presumed victims over the five lists (Table 2).

Table 2 *Contingency table of presumed trafficking victims in the United Kingdom, 2014*

LA	X					X	X	X								X	X	X
NG		X				X			X	X	X				X	X	X	X
PF			X				X		X			X	X		X	X		X
GO				X				X		X		X		X	X		X	X
GP					X						X		X	X				
number	**54**	**453**	**995**	**695**	**316**	**15**	**19**	**3**	**62**	**19**	**1**	**76**	**11**	**8**	**4**	**1**	**1**	**1**

Source: National Crime Agency/Bales, Hesketh & Silverman, 2015)

Using covariates to enhance robustness and policy relevance of results

As explained above, the inclusion of covariates - such as sex, age and form of exploitation - in the study is highly recommended, for both technical and policy reasons. In the absence of covariates, MSE allows for dependence between lists, where these list dependencies are equal for different segments of the population. Methodologically, once the study includes information on the victims' sex, age and form of exploitation, the assumption of equal list dependencies for all victims can be replaced by the less demanding assumption of equal list dependencies within the cross-classification of all covariates. For example, a study that includes the covariate sex would permit males and females to have different list dependencies. These differences can be recognized in the estimation procedure and be corrected for in the total estimate.

The Dutch experience has confirmed that the inclusion of covariates is indeed advisable. In addition to making the total estimate more robust - an analysis with the use of covariates produced significantly different, in this case lower estimates than an older study not using covariates - the inclusion of covariate information made it possible to calculate separate estimates of the 'hidden figures' of, for example, cross classifications of age (minor victims, adult victims), gender (female victims, male victims) and type of exploitation (victims of sexual and non-sexual exploitation (all other categories)).[30]

SDG indicator 16.2.2 calls specifically for victimization rates by sex, age and form of exploitation. The inclusion of sex and age as covariates reflects the special focus of the Palermo-protocol on the protection of women and children. The most important forms of exploitation to be distinguished can differ per country. In many countries the most common forms are sexual exploitation and forced labour. In most studies the covariate type of exploitation distinguished between sexual exploitation and other forms (such as forced labour, forced begging, forced criminality and organ removal). In studies using data from Rumania, Serbia and Slovakia begging/criminal activity was included as a separate, third form of exploitation because of its prevalence in the dataset.

30. See Ministries of Justice and Security, Social Affairs and Employment, Health, Welfare and Sports, and Foreign Affairs, Government of the Netherlands, *Together against human trafficking: An integrated program to tackling sexual exploitation, labor exploitation and criminal exploitation,* February 2019 (In Dutch).

Depending on the nature of trafficking in a country, other covariates may also be included in an MSE study. Including additional covariates boosts, as explained, the robustness of the results, and also enhances the policy relevance of the results. For example, in many countries, detected trafficking victims include both local citizens and foreigners. This is especially the case of typical destination countries for trafficking where foreigners are transferred for the purpose of exploitation (in addition to domestic trafficking of local citizens). In these countries, including nationality as a covariate is recommended as local citizens and foreigners may experience different forms and levels of trafficking and different likelihoods to be detected as well.

Moreover, in countries which are largely countries of origin, detected victims may have been exploited outside their home country, and possibly may have been repatriated afterwards. When working with datasets from such countries it is vital to include the territory of exploitation in the dataset in order to be able to estimate the numbers exploited on the country's own territory and the number of nationals exploited elsewhere separately.[31] If prevalence rates in countries that serve primarily as source countries for trafficking victims are to be compared with those of typical destination countries, they should represent victims who have been exploited in their home countries. We will return to this issue in the final section.

SUMMARY POINTS

- In countries with National Referral Mechanisms, data on presumed victims of human trafficking are often collected by a central Anti-Human Trafficking Office or dedicated NGO. If such clearing house exists, its available datasets will usually meet the key requirements for MSE listed hereunder or can be reconstructed to do so.
- If no centralized, multi-source dataset exists, data must be collected from institutions collecting such data in the framework of an ad hoc research project to develop an integrated database for the purpose of conducting MSE.
- Data must be available on persons who meet operational indicators of presumed victims in line with national legislation from a minimum of three different entities that come into contact with trafficking victims, for example, police, NGO's, immigration authorities, labour inspectorates and relevant international organizations.

31. In publications of the ILO and the Minderoo Foundation/Walk Free on human trafficking regional and national rates of victims relate to victims exploited on national territories of regions/countries.

- If data from different entities need to be consolidated, a set of definitions and operational indicators of a presumed victim may need to be applied retroactively by the research team to extract relevant cases from existing administrative datasets.
- The institutions that provide victim lists must each operate as independent contact points for trafficking victims, besides accepting referrals from each other. For example, where police refer identified victims to an NGO for the provision of support, this NGO should also directly come into contact with victims in other ways.
- Each victim that appears on any list needs a unique identifier such as birthdate and initials or a coded or encrypted identity key in order to permit matching across lists to determine overlaps (whether a victim appears on one, two or more lists). For reasons of data security, the matching can also be conducted by a trusted third party which deletes the identifiers thereafter.
- To arrive at sufficiently reliable estimates, the total number of trafficking victims across all lists should be at least 150, or when multi-year datasets are used, a minimum of 50 victims per year; there should be a reasonable amount of overlap between the various lists.
- Ideally, the dataset should include data about basic characteristics of the victims such as age (adult or child), sex and type of exploitation (sexual, forced labour, forced begging/criminal exploitation). Other relevant covariates may include nationality and location of exploitation (in home country or abroad in a destination country).
- Since the use of multiple lists and covariates requires, as will be demonstrated in the next section, an advanced statistical modelling exercise, it is advisable to involve an analyst with prior experience with MSE, preferably with human trafficking data. The Crime Research Section of UNODC stands ready to offer advice to governments interested in conducting MSE to monitor progress in attaining SDG 16.2.

SECTION 3

Statistical modelling for Multiple System Estimation

In this section we first provide the theory behind Multiple Systems Estimation. MSE is also known under a variety of other names, most prominently Capture-Recapture. In our exposition we follow a review by the International Working Group on Disease Monitoring and Forecasting (1995).[32] We supplement the theory with *R-code*. An MSE using data collected in Ireland will serve as an illustrating example.[33]

Theory

While MSE application for estimating the number of unobserved victims of trafficking in persons should always, as discussed, be based on data from at least three lists, the discussion will start with a two-list example to illustrate how the method works.

In MSE multiple lists of individuals are linked. A list refers to a list of individuals that are part of a common population (in the context of this manual: the population of victims of human trafficking). A simple situation is that of two lists R1 and R2, both with level 1 when observed in the list, and 0, when not observed in the list.

32. Other key references are Van der Heijden et al. (2012), Bird & King (2018) and Boehning, van der Heijden & Bunge (2018).

33. In this elaboration we explain the stepwise choice of the best-fitting model accounting for and specifying interactions between lists and covariates. For alternative approaches to modelling whereby the estimates of various models are averaged, see Burnham and Anderson (2002) and Silverman (2020).

Table 3 *Contingency table with two lists*

		R2	
		1	0
R1	1	n_{11}	n_{10}
	0	n_{01}	n_{00}

The table above shows the cross tabulation of two lists, R1 and R2. We denote expected values using '*m*' and observed values by '*n*'. After linking the two lists, there are n_{11} individuals that are in both R1 and R2, there are n_{10} individuals that are in R1 but not in R2, and there are n_{01} individuals that are in R2 but not in R1. There are n_{00} individuals that are in neither of the lists. This number is unknown and to be estimated. The observed part of the population (documented victims of trafficking) is denoted by $n = n_{11} + n_{10} + n_{01}$. The statistical problem is to estimate the size of the total population, that we denote by N. If we would know n_{00}, this would give us N as $N = n + n_{00}$. Hence, we aim to find an estimate for n_{00}, so that we have an estimate of N.

We can find an estimate for n_{00} by making a number of assumptions. We assume that it is possible to match the individuals in R1 to those in R2 without making errors. Second, we assume that the probability of inclusion in R1 is statistically independent from the probability of inclusion in R2. Related to this is the assumption that either the inclusion probabilities for R1, or the inclusion probabilities for R2, are homogeneous over the individuals.

The assumption of independence allows us to find an estimate for n_{00}. Under independence we have that the odds of the values in the rows are identical to the odds of the values in the columns. For example, for the rows, $n_{11} / n_{10} = n_{01} / n_{00}$. In other words, the odds ratio is 1. If we denote the expected value estimate for cell (0,0) as m_{00}, and an estimate by adding a hat '^', then it follows that:

$$\hat{m}_{00} = \frac{n_{10} n_{01}}{n_{11}}$$

For two lists this is very simple, but it becomes more complicated in the situation that there are more than two lists and covariates. Therefore, we switch to the more general loglinear model, that has the independence model as a special case. The loglinear model has the advantage that it can be easily extended to more lists and covariates. In general, if we index the levels of R1 and R2 by i and j respectively, then the expected value for cell (i, j) is given by

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)}$$

where ‘log’ stands for the natural logarithm, and we have to impose restrictions to identify the model. As there are only three observed counts, only three parameters can be identified. For a general 2 x 2 table where a cell (0,0) is observed, a loglinear model will have an interaction term $u_{12(ij)}$, but in our context where the value for the cell (0,0) is unobserved, this term has to be set to zero as it cannot be identified. Also, to identify the parameters we set all parameters equal to 0 if an index is 0, thus we set $u_{1(0)} = u_{2(0)} = 0$. The result is that

$$\log m_{00} = u.$$

In other words, we fit a model for the three observed cells and project it on the fourth cell (0,0). Once we have an estimate for u, then we can plug this estimate into the equation and we get

$$\hat{m}_{00} = \exp \hat{u},$$

where exp (x) is the exponential function of x. This estimate is used as an estimate for n_{00} to find an estimate for N, the population size.

Covariates

When there are two lists, R1 and R2, statistical dependence of the inclusion probabilities may arise which would violate a key assumption. It may be the case that some victims are more likely to be included in R1 as well as in R2, than others. This may be due to their visibility. Assume, as an example, that sex is related to this visibility, for example because there is an active policy to identify women trafficked for sexual exploitation, and these are therefore more easily found. In this case, not including sex would lead to a violation of the model assumption. However, if we would include sex as a covariate and make separate estimates for females and males, the violation would be less severe.

So, let us assume that we have two lists and one covariate, say sex, denoted by 4, with levels $x = 0, 1$ for females and males respectively. Then the loglinear model that can be fit is

$$\log m_{ijx} = u + u_{1(i)} + u_{2(j)} + u_{4(x)} + u_{14(ix)} + u_{24(jx)}$$

This model assumes that for both males and females, inclusion in list R1 and in R2 are independent, but that the inclusion probabilities for males may differ from those for females. This model would be denoted by the highest fitted margins in square brackets, i.e. [14][24].

In this context we have to estimate counts for two cells, the cell for females (0,0,0) and the cell for males (0,0,1). We again identify the parameters by setting them to 0 if an index is 0, and estimate

$$\hat{m}_{000} = log(\hat{u})$$

for females and

$$\hat{m}_{001} = log\left(\hat{u} + \hat{u}_{4(1)}\right)$$

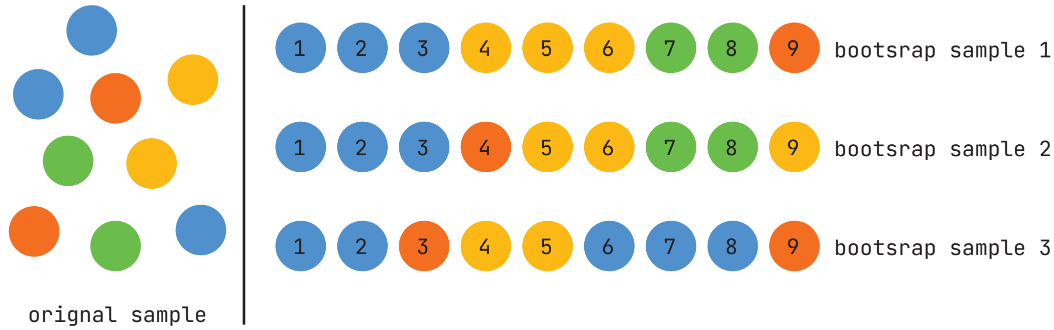
for males.

Confidence interval for estimate of N

We describe how to find a confidence interval for the estimate of *N* in the context of two lists, as this generalizes to more than two lists in a straightforward way. We use the parametric bootstrap (compare International Working Group on Disease Monitoring and Forecasting, 1995). We illustrate how that works with the figure with the nine colored balls below (Figure 3).

We have three observed counts symbolized by the orange, green and yellow balls, and we have an estimated count of the cell (0,0) symbolized by the blue balls. The sum of these four counts gives the estimated population size. These four counts are used to calculate four probabilities of a multinomial distribution. In step 1 we draw a first bootstrap sample with replacement using the four probabilities of the multinomial distribution and the rounded estimate of *N*. In Figure 3 below we have three blue balls, so at each draw the probability of drawing a blue ball is 3/9, and for each of the other colors the probability is 2/9. In step 2 we eliminate the observations that fell in cell (0,0), i.e. the blue balls in the bootstrap samples. In step 3 we estimate *N* for the first bootstrap sample, i.e. we estimate the number of blue balls given the number of orange green and yellow balls. This ends the procedure for the first bootstrap sample. We do this, say, 1,000 times, and this provides us with 1,000 bootstrap sample estimates of *N*. We order these into a distribution of estimates. By taking the 2.5 and 97.5 percentile value this yields a 95 per cent percentile confidence interval for the estimated N found in the original sample (see Figure 3).

Figure 3 Finding confidence intervals through bootstrapping; an illustration



This confidence interval has the advantage over analytic approaches that it is not assumed a priori that the confidence interval is symmetric. Also, by including this complicated procedure for cell (0,0), by sampling it and then eliminating the count, we allow for fluctuation in the number of observations n.

Three lists

In the section for two lists, we saw that there are only three counts observed and therefore we can only fit a model with three parameters. However, when there are three lists, the number of counts observed is 2 × 2 × 2 - 1 = 7, where the '-1' stands for the count that has to be estimated. Hence a model with 7 parameters can be fit. We denote the third list as R3 with index k ($k = 0, 1$). The most general loglinear model that we can fit to the data, i.e. the seven counts, is

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$$

This shows that the big advantage of including a third list is that we do not assume statistical independence between the lists anymore as the model has three two-factor interaction parameters, namely $u_{12(ij)}$, $u_{13(ik)}$ and $u_{23(jk)}$. In practical situations this allows for a more realistic model that can be fit. There is only the restriction that there is no three-factor-interaction, meaning that an interaction between two lists is identical over the levels of the third list, or in other words, that the odds of inclusion in two lists does not depend on the in- or exclusion of a third list.

The model is fit in a similar way as for two lists: we estimate the count for cell (0,0,0) using

$$\hat{m}_{000} = log(\hat{u})$$

This estimate is used as an estimate for n_{000} to find an estimate for N, the population size.

It is standard practice to fit only so-called hierarchical loglinear models (Agresti, 2013). This means that, if a higher-order term (here: a two-factor interaction) is in the model, the lower-order terms (here: the marginal terms) are in the model as well. (If such marginal terms are lacking, the model would become hard to interpret.) Therefore, there is an efficient way to denote loglinear models, namely by placing the highest terms between square brackets. For example, [12][23] is the loglinear model where $u_{13(ik)} = 0$.

Model selection

While the inclusion of as many lists/registers and covariates as possible is advantageous for meeting the independence and equal inclusion probabilities assumptions, there is a downside. Firstly, the number of potential models increases very rapidly with the number of variables. For example, with two lists there is one, with two lists and one covariate there are four, and with three lists there are eight potential models to choose from. So, when the number of variables increases, the problem of finding the best model will become more and more difficult. Secondly, with an increasing number of variables, the frequencies in the cells of the contingency table become smaller and smaller, which may lead to unstable population size estimates. Therefore, the quality of the estimates depends to a great extent on striking the right balance between sample size, number of lists and covariates, and model complexity.

It is standard practice in loglinear modelling to fit a series of models with the aim to find the most restrictive model that fits the data adequately. This most restrictive model has the property that, in general, the confidence interval of the estimates under this restrictive model will be smaller than the confidence interval for the saturated model (Agresti, 2013).

We fit a series of models using the so-called STEP procedure programmed in *R*. We start from the model only having an intercept, and then start adding loglinear marginal and interaction terms in consecutive steps, where in each step it is checked whether, when a term is added, other terms become redundant. In the context of population size estimation two approaches are popular for deciding when to stop and choose a final model. One is the AIC[34] and the other is the BIC[35] criterion. Both

34. Akaike information criterion.

35. Bayesian information criterion.

are functions of the deviance of a model, where in the AIC there is a penalty for adding parameters, and in the BIC for adding parameters that involves n. There is a tendency for the BIC to find a more restrictive model than found in the AIC if the sample size is large.

A general warning is in place here. It is possible to end up with a model that is too restrictive, and hence has a confidence interval for N that is too small. The confidence interval is optimal under the model one ends up with but does not take the model fitting procedure into account. This problem was already noticed in the International Working Group on Disease Monitoring and Forecasting (1995), and is still a subject of active research.

Checking the data and preliminary steps

In this section the analysis of Irish data will serve as an illustration. This relatively small dataset includes 302 individuals, and the original Excel data file has one record for each individual (see Table 4 for details).

The first step is to convert the Excel file into a data frame called 'd' for use in R. The three lists are denoted by the letter R and the numbers 1 to 3, and the covariates are denoted with a single letter, including the variable years (see Table 4).

Table 4 *Denotation of the Irish dataset*

Variable	Description	Measurement level	Levels/categories
R1	An Garda Síochána (police)	Numeric	0 = not observed in list, 1 = observed in list
R2	MRCI, ICI, IOM (migrant aid)	Numeric	0 = not observed in list, 1 = observed in list
R3	Ruhama, SVCC, DL (NGO's)	Numeric	0 = not observed in list, 1 = observed in list
S	Sex	Factor	F = female, M = male
A	Age	Factor	A = adult, M = minor
N	Nationality	Factor	I = Irish, N = Non-Irish
E	Exploitation (Type of)	Factor	S = sexual, O = other
Y	Year of observation	Factor	2014, 2015, 2016

We now provide a checklist for data quality, including instructions about going back to those gathering the data for clarifications or requests for reordering of data.

On this data file we perform the following checks:

1. Distribution of the lists over the years
   a. check for extreme fluctuations over the years
   b. check for zero frequencies

2. Overlap between pairs of lists
   a. check for sufficient overlap (good)
   b. check for complete overlap (not good)

3. Frequencies of the covariates check for sufficient spread over the years

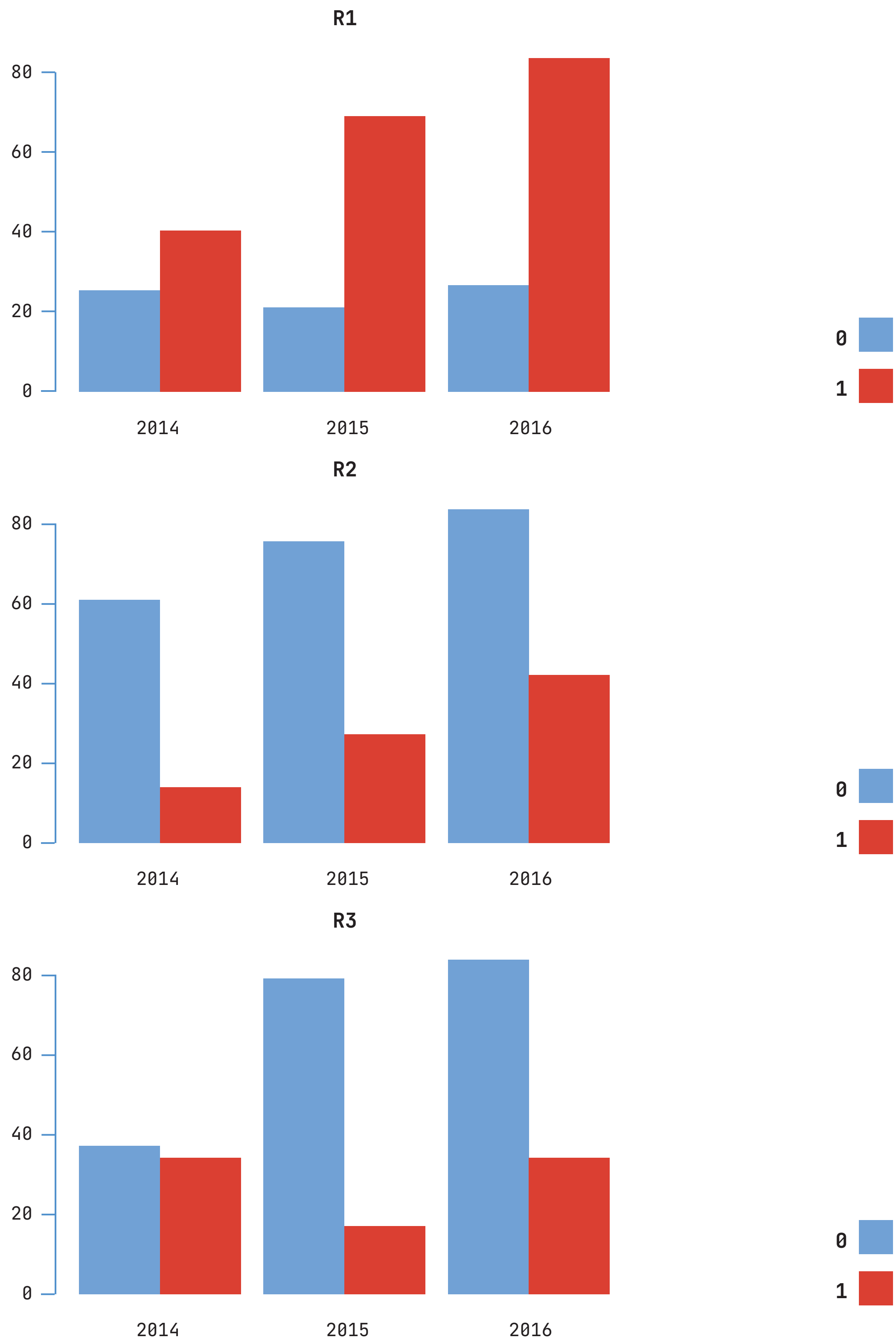
If there is insufficient overlap between pairs of lists – e.g. there are no or very few observations that appear in both lists - and/or spread of the covariates – e.g. some covariate values a frequency of 0 or close to zero in a particular year – numerical problems may occur with the estimation of the interaction effects between these lists or between the covariates and the years. Typically, the corresponding parameters estimates take large values and/or have large standard errors. In itself, this need not lead to unrealistic population size estimates, but it can also occur that the population size estimate explodes. Chan, Silverman and Vincent (2020) discuss the detection and treatment of cells with sampling zeros in more detail.

Distribution of the lists over the years

We make the following graphical displays (bar charts) with the observed frequencies in the Irish dataset on each of the three lists (Figure 4)[36].

36. R code for making the 1st plot: `barplot(xtabs(~ R1 + Y, data = Ireland),beside = TRUE)`

Figure 4 Bar chart of observed frequencies on lists R1, R2 and R3



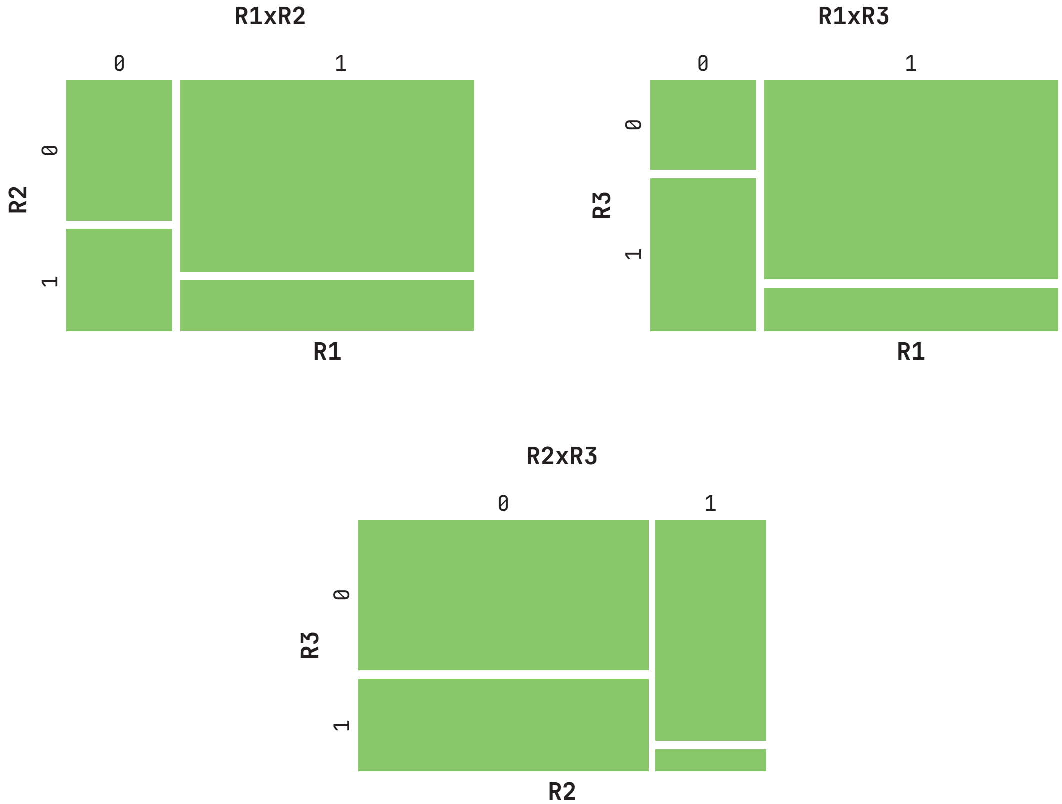
There is an increase of observations over the years. This increase seems to be largely due to an increasing number of observations by the An Garda (R1).

Next, we make mosaic plots of all pairs of lists to check whether there is sufficient overlap. The surfaces of the squares in the plots represent the frequencies of being observed in one of the two lists, both lists, or not being observed in either list. We would like the surfaces of all these combinations to be present. Absence of a surface means that there is a combination missing, which – depending on the combination that is missing - may be cause for alarm:

- if 00 combination is missing: the two lists cover all observations in the data, so the remaining lists do not contribute any new observations. This may be an indication that one of these two lists is a combination of the remaining lists. This needs to be checked with the list-providing entity.
- If 01 or 10 is missing: this is an indication that one of the lists is a combination of other lists. This needs to be checked with the list-providing entity.
- If 11 is missing: this is no cause for alarm. It most likely indicates that the two lists are negatively associated; the inclusion in one list reduces the probability of inclusion in the other list. The effect of a zero frequency of 00 is that the parameters of the model will be inflated, but the population size estimates are not.

In each of the three plots in Figure 5 below we see 4 rectangles, whose surface reflects the proportion of observations in the cells 00, 10, 01, and 11 respectively. If there would have been a cell with no observations, one of the rectangles would be missing, i.e. would have a zero surface.

Figure 5 *Graphic representation of the distribution of cases over three lists*

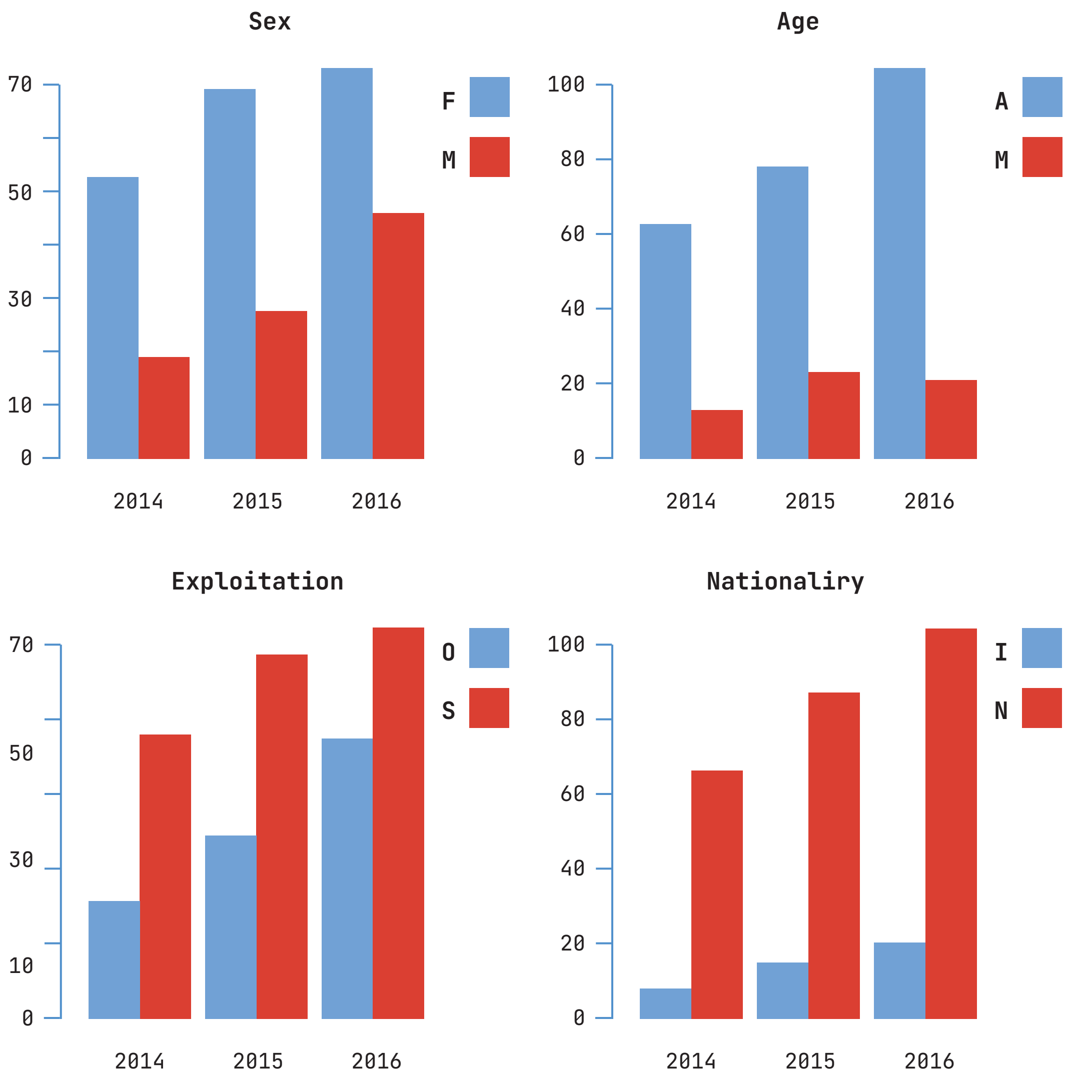


We see that the surface in the third plot of R2 = 1, R3 = 1 is the smallest. This cell still has 7 observations nevertheless. So, none of the list combinations have near-zero frequencies, and there is sufficient overlap between the pairs of lists for reliable parameter estimation.[37]

We also check the distribution of the covariates over the years (see figure 6).

37. R code for making the 1st plot: `mosaicplot(xtabs(~ R1 + R2, data = Ireland))`

Figure 6 Distribution of four covariates in bar charts per year



The increase of observations over the years does not seem to depend on any of the covariates in particular; the levels of all four covariates seem to show a relatively steady increase over the years, except maybe for the small decrease in the frequency of observed minors from 2015 to 2016. This means that there is little to no interaction between year and each of the covariates: otherwise the increase in frequencies over the years would be different for the levels within the covariates.

The R package 'mse'

The R package 'mse' has been especially developed for this manual by author Maarten Cruyff to be used for multiple systems estimations of victims of human trafficking. The package will be available on CRAN, but can also be installed from GitHub with the R command:

```
devtools::install_github("MaartenCruyff/mse")
```

The package contains two functions, one for model selection and one for determining confidence intervals. The function to perform a search for the best fitting model is:

```
model_search(x, lists = NULL, max_model = Freq ~ .^2,
             year = NULL, degree_year = NULL)
```

where `x` denotes the data (a data frame with one record per individual observation, or a contingency table), lists the column numbers with the lists, `max_model`, the maximal model that can enter the search path, `year`, the name of the variable with the year of the observation (if observations are collected over multiple years), and `degree_year`, the degree of the polynomial for year (if left unspecified, dummy variables will be created).

The minimal model in the model search is specified by the argument `min_model`, which by default is the main-effects model Freq ~ . The argument `start_model` offers the option to start the search with a user-defined model, and if left unspecified the model search starts with the model specified in `min_model`.

The function that performs a parametric bootstrap on the best fitting model is

```
boot_mse(object, modelnr = NULL, rep = 1000, seed = 1)
```

where object is the object created with `model_search()`, `modelnr` is the number of the model in the search path (the default NULL selects the model with the lowest BIC), `rep` is the number of bootstrap samples, and `seed` is for reproducibility of the results, and returns tables and plots of the population size estimates, stratified over the years (if more than one) and over covariates and pairs of covariates (if present).

This next section explains the principles of multiple systems estimation again, now in the context of the Irish data, and illustrates the working of the mse package on these data. The section starts with the simplest case of two lists without covariates, and works from there to the simultaneous analysis of all variables available in the Irish data.

Two lists, no covariates, no years

We now apply the theoretical principles discussed above. For the lists R1 and R2, the contingency table for the year 2016 years looks like the one in Figure 7.

Table 4 Contingency table of two lists for year 2016 in Irish dataset

```
  R1 R2 Freq
1  0  0    0
2  1  0   64
3  0  1   11
4  1  1   31
```

The first row of this contingency table refers to the persons who were not observed on any of the two lists, and consequently has frequencies of zero.

We start with discussing the assumptions and violations of assumptions for this contingency table. The independence assumption may need some clarification. It means that the probability of inclusion in R1 is independent of probability of inclusion in R2. Sometimes this is misunderstood, when people assume that when the lists are built independently from each other, this assumption is fulfilled. However, a key violation occurs when some victims have a much higher probability to be included in both lists than other victims, and this may happen even though the lists are built independently from each other, e.g. by cross-referrals between the entities maintaining the lists.

Earlier it was indicated that statistical independence of the inclusion probabilities means that the odds ratio equals 1, and hence

$$\hat{m}_{00} = \frac{n_{10}n_{01}}{n_{11}} \frac{64\cdot 11}{31} = 22.7$$

This makes the total population size estimate 64 + 11 + 31 + 22.7 = 128.7.

Using the mse package the model search produces the following model, log linear coefficients and 95% confidence interval:

```
R1R2 <- model_search(x = d[d$Y == 2016, 1:2], lists = 1:2)
```

```
  Step Df Deviance Resid..Df Resid..Dev AIC AICc BIC  Nhat
1      NA       NA         0          0   0    0   0 128.7
```

```
R1R2$coefs
```

```
[[1]]
(Intercept)         R11         R21 
  3.1227912   1.0360919  -0.7248959 
```

```
boot_mse(R1R2)
```

```
$x
    nobs Nhat min95 max95
all  254  379   326   447
```

Note the data specification `d[d$Y = 2016, 1:2]` selects observations made in 2016, and the variables R1 and R2, and that with these two variables the only model that can be fitted is `Freq ~ R1 + R2`. The population size estimate is 379 with a 95% confidence interval ranging from 326 to 447.

Two lists, one covariate, no year

We now include the covariate sex (S). We refer earlier sections for an explanation of the importance of including covariates, when they are available. The contingency table for these data is given in Table 8.

Table 5 Contingency table for two lists and the covariate sex (F/M) for year 2016 only

```
  R1 R2 S Freq
1  0  0 F    0
2  1  0 F  114
3  0  1 F   27
4  1  1 F   20
5  0  0 M    0
6  1  0 M   57
7  0  1 M    8
8  1  1 M   28
```

The stepwise model search produced three models: a model with no interaction, an interaction between Sex and list 1 and with an interaction between Sex and list 2 respectively. The program also shows the Nhat (estimated numbers) of the models and the values for the best fit measures (AIC, AICc and BIC). At the bottom the results of the bootstrapping (minimum and maximum estimates) are presented.

```
R1R2S <- model_search(x = d[d$Y == 2016, c(1, 2, 4)], lists = 1:2)
```

```
    Step Df Deviance Resid..Df Resid..Dev  AIC AICc  BIC  Nhat
1        NA       NA         2       21.0 17.0 16.7 11.3 128.7
2 + R2:S -1      8.8         1       12.3 10.3 10.1  7.4 128.7
3 + R1:S -1     12.3         0        0.0  0.0  0.0  0.0 163.0
```

```
R1R2S$coefs[[3]]
```

```
(Intercept)         R11         R21          SM      R21:SM      R11:SM
  4.0125145  -0.2513144  -1.8152900  -3.4528988   1.9488214   2.7362211
```

```
boot_mse(R1R2S)
```

```
$x
    nobs Nhat min95 max95
all  125  163   123   275

$S
  nobs Nhat min95 max95
F   77  114    74   225
M   48   49    38    62
```

The fit measures AIC, AICc and BIC are all the lowest for the 3rd model, so this is the preferred model (the model includes an interaction between sex and the second list). The overall population size estimate of the preferred model is 163, with a 95% confidence interval ranging from 123 to 275; for females separately, the estimate is 114 (74, 225) and for males 49 (38, 62).

Two lists, one covariate, and year

We now extend the model by including the variable Year, indicated with Y, as a covariate. The contingency table (with the cells R1 = 0, R2 = 0 removed) is presented in Table 6.

Table 6 Contingency table of two lists, the covariate sex (Females and Males) and year (2014, 2015 and 2016)

```
   R1 R2 S    Y Freq
2   1  0 F 2014   30
3   0  1 F 2014    6
4   1  1 F 2014    1
6   1  0 M 2014   13
7   0  1 M 2014    5
8   1  1 M 2014    2
10  1  0 F 2015   41
11  0  1 F 2015   12
12  1  1 F 2015   12
14  1  0 M 2015   23
15  0  1 M 2015    1
16  1  1 M 2015    2
18  1  0 F 2016   43
19  0  1 F 2016    9
20  1  1 F 2016    7
22  1  0 M 2016   21
23  0  1 M 2016    2
24  1  1 M 2016   24
```

The variable year has 3 levels/categories, which are represented by two dummy variables with the argument `year = "Y"`, which allows the effect of year on the log of the fitted frequencies to be non-linear. The results of the model search are now as follows:

```
R1R2SY2 <- model_search(x = d[, c(1, 2, 4, 8)], lists = 1:2, year = "Y")
```

```
    Step Df Deviance Resid..Df Resid..Dev  AIC AICc  BIC  Nhat
1        NA       NA        12       52.5 17.8 16.6  1.7 378.7
2 + R1:S -1      3.5        11       49.0 16.2 15.2  3.9 378.7
3 + R2:S -1      9.7        10       39.4  8.6  7.7  0.0 424.2
4  + S:Y -2      5.3         8       34.0  7.2  6.6  6.0 424.2
5 + R2:Y -2      3.8         6       30.2  7.4  7.1 13.7 424.2
6 + R1:Y -2     11.5         4       18.7  0.0  0.0 13.6 573.6
```

```
round(R1R2SY2$coefs[[3]], 3)
```

```
(Intercept)        R11        R21         SM      Y2015      Y2016
      3.542     -0.300     -1.740     -2.246      0.468      0.620
     R11:SM     R21:SM
      1.553      1.030
```

The additional argument `degree_year = 1` treats year as a numerical variable, with a single parameter having a linear effect on the log of the fitted frequencies. Results of the model search are slightly different, as shown below.

```
R1R2SY1 <- model_search(x = d[, c(1, 2, 4, 8)], lists = 1:2, year = "Y",
                        degree_year = 1)
```

```
    Step Df Deviance Resid..Df Resid..Dev  AIC AICc BIC  Nhat
1         NA       NA        13       53.9 21.1 20.7 6.3 378.7
2 + R2:Y -1      4.1        12       49.8 19.1 18.7 7.9 378.7
3 + R1:Y -1     11.9        11       38.0  9.2  9.0 1.8 465.5
4 + R1:S -1      3.5        10       34.5  7.7  7.6 4.0 465.5
5 + R2:S -1      9.7         9       24.8  0.0  0.0 0.0 542.7
```

```
round(R1R2SY1$coefs[[3]], 2)
```

```
(Intercept)         R11         R21          SM           Y       R21:Y
       3.58        0.00       -1.58       -0.55       -3.34        3.34
      R11:Y
       4.08
```

The BIC favors the 3rd model with two parameters for year, and the 6th model with a single parameter for year. These two models cannot be compared directly, because within each table the BIC values are scaled such that the minimum value is 0. The unscaled values are obtained with

```
R1R2SY2$minima["BIC"]
```

```
  BIC
156.3
```

```
R1R2SY1$minima["BIC"]
```

```
  BIC
147.4
```

This shows that the model with a single parameter for year is more parsimonious than the model with two parameters for year, and is therefore to be preferred.

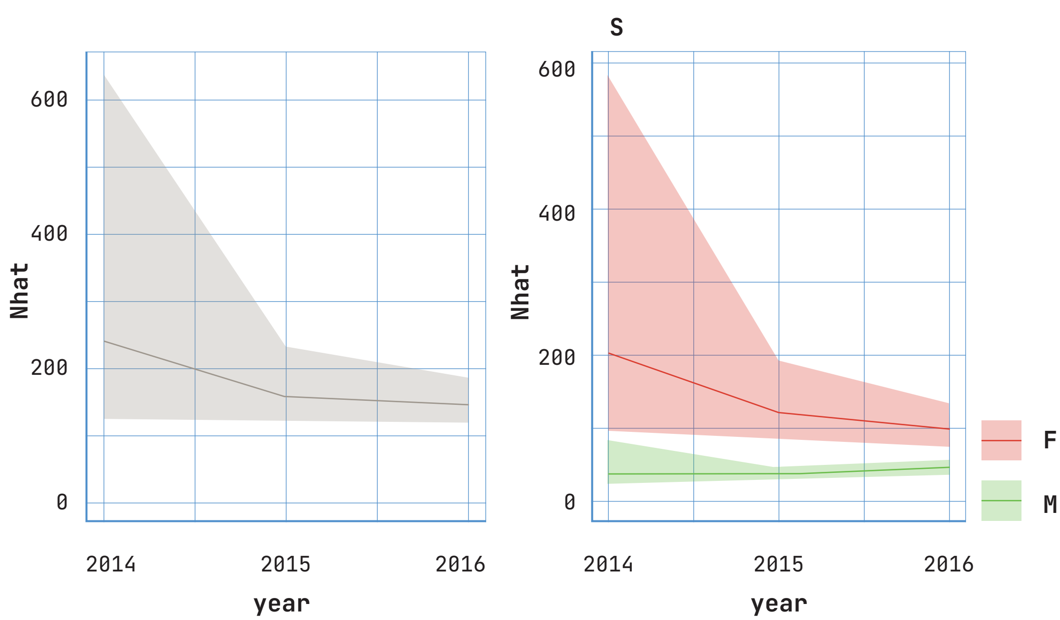
When the argument year is specified, the function `boot_mse()` returns tables as well as plots with the estimates and 95% confidence intervals. The tables obtained are as follows:

```
out <- boot_mse(R1R2SY1)
out$tables
```

```
$x
    nobs Nhat min95 max95
all  302  543   393  1012

$Y
     Year nobs Nhat min95 max95
2014 2014   75  241   124   641
2015 2015  102  157   120   235
2016 2016  125  145   118   185

$S
  Year nobs Nhat min95 max95
F 2014   55  203    97   587
F 2015   73  121    86   195
F 2016   77   98    74   134
M 2014   20   38    23    85
M 2015   29   36    28    47
M 2016   48   47    37    58
```

The plots are stored in the list element `out$plots`. Figure 7 presents the plots showing minimum and maximum estimates, overall and for males and females separately.

Figure 7 Estimated numbers of victims according to the best fit model including 2 lists, the covariate sex and three years.



The tables and plots show wide confidence intervals for the estimates of 2014. Inclusion of another list and/or more covariates may help to reduce the width of these intervals.

All lists, all covariates and all years

As a final example, we perform a model search on all lists and all covariates. As in the previous example, we start with performing a model search with a linear and non-linear effect for year, and a comparison of the BIC's of both models. In this case the model with two parameters for year performs slightly better (BIC of 473.2 against 475.2). The results of the model search for this model are:

```
ms2 <- model_search(d, lists = 1:3, year = "Y", degree_year = 2)
```

```
      Step Df Deviance Resid..Df Resid..Dev   AIC  AICc   BIC        Nhat
1          NA       NA       326      679.0 505.4 501.0 453.2       464.9
2    + A:N -1    165.9       325      513.1 341.5 337.3 293.0       464.9
3    + S:E -1     97.9       324      415.2 245.6 241.6 200.8       464.9
4   + R3:S -1     50.4       323      364.8 197.2 193.4 156.1       464.9
5    + N:E -1     42.0       322      322.9 157.3 153.6 119.9       464.9
6   + R3:A -1     41.6       321      281.2 117.6 114.2  84.0       464.9
7   + R2:N -1     35.3       320      245.9  84.3  81.1  54.4       474.9
8    + S:N -1     24.1       319      221.7  62.1  59.2  35.9       475.1
9   + R2:E -1     18.6       318      203.2  45.6  42.9  23.0       464.5
10  + R3:Y -2     18.9       316      184.2  30.6  28.5  15.5       464.5
11  + R1:S -1     12.1       315      172.2  20.6  18.8   9.2       459.1
12 + R2:R3 -1     10.6       314      161.6  12.0  10.5   4.3       432.3
13  + R2:A -1      7.9       313      153.7   6.1   4.9   2.1       432.4
14  - R2:N  1      4.0       314      157.7   8.1   6.6   0.4       432.9
15  + R3:E -1      6.1       313      151.6   4.0   2.8   0.0       429.6
16  + R2:N -1      4.1       312      147.4   1.8   1.0   1.6       429.0
17   + S:Y -2      5.8       310      141.6   0.0   0.0   7.2       429.0
18  + R2:Y -2      3.6       308      138.0   0.4   1.2  15.0       429.7
19  + R1:Y -2      3.6       306      134.3   0.7   2.5  22.8       438.0
20  + R3:N -1      1.8       305      132.5   0.9   3.1  26.6       437.9
21  + R1:N -1      1.2       304      131.3   1.7   4.4  31.1 285687136.5
22   + N:Y -2      2.1       302      129.1   3.5   7.4  40.4 282461493.4
23   + A:Y -2      6.2       300      122.9   1.3   6.3  45.6 279138606.6
```

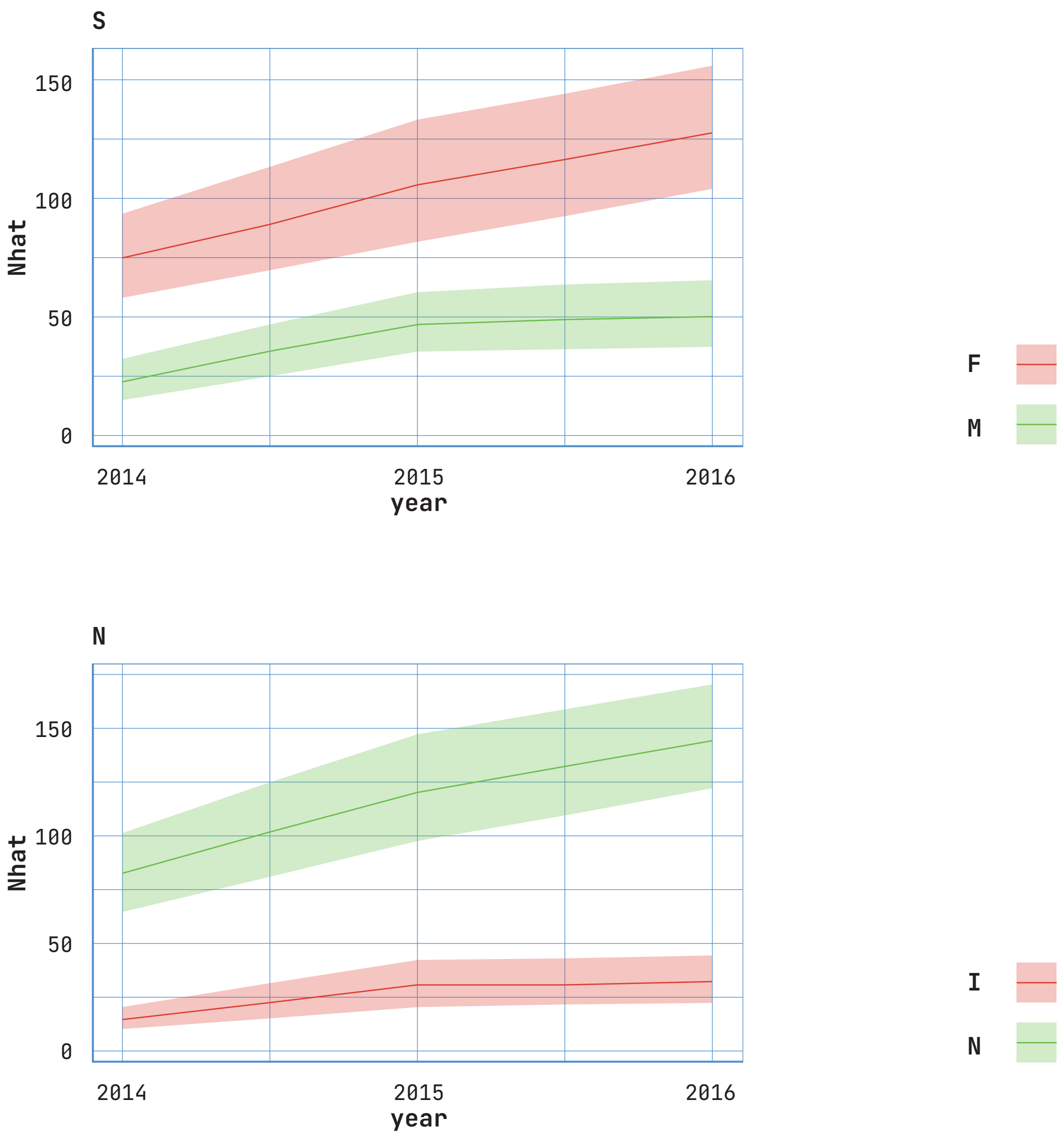
The BIC favors model 15 and the AIC and AICc model 17, but both models result in the same population size estimate (430 and 429 respectively). The last three models show what can happen if the data are overfitted; the population size estimates explode to unrealistically high numbers.

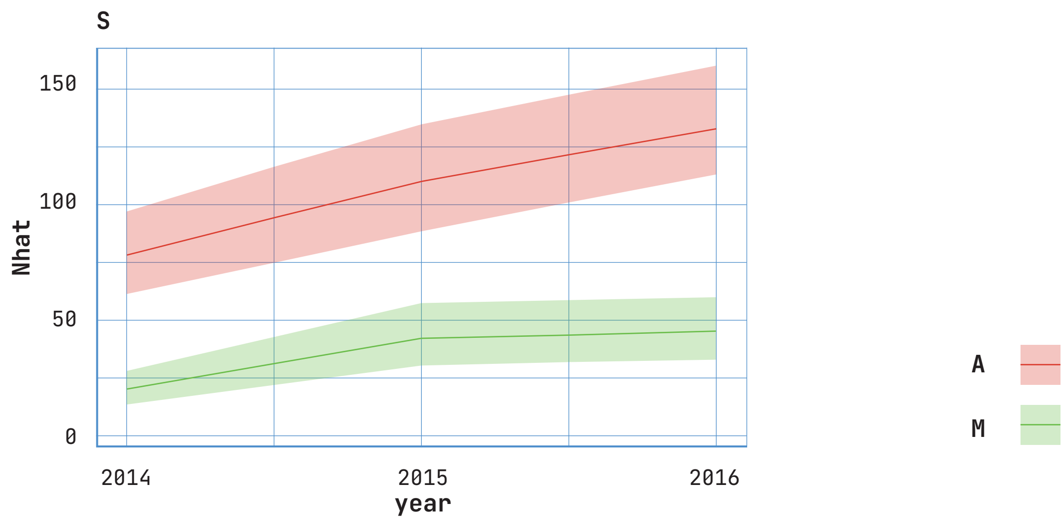
The model search includes one interaction between lists, namely between R2 and R3 (see model 12). The interpretation of this interaction can be derived from the parameter estimates. The parameter estimates of model 15 (the best model according to the BIC, presented in Annex A (last page of document) shows a negative parameter estimate (– 1.3) for the interaction between R2 and R3. This means that persons observed in R2 are less likely to be observed in R3, and vice versa, than would be expected under (conditional) independence of the lists.

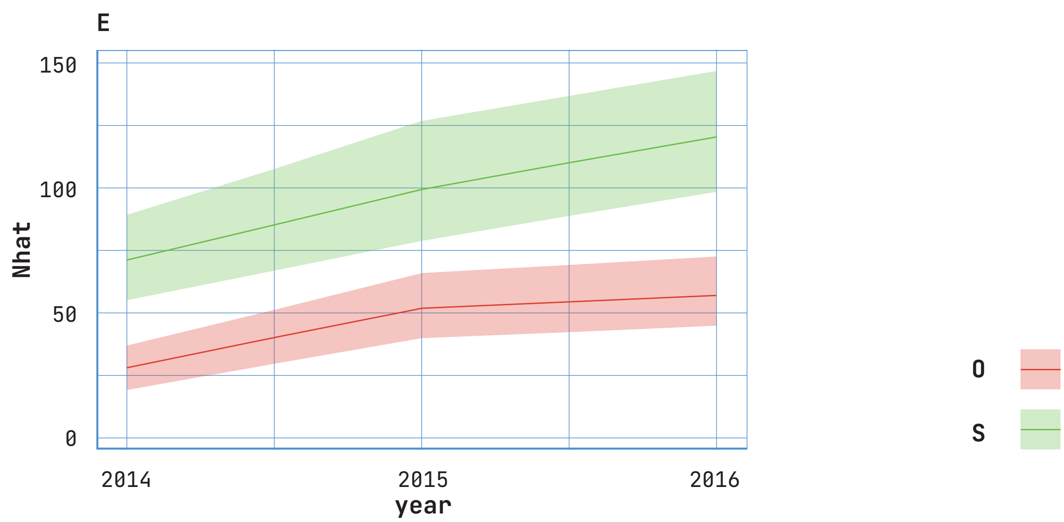
The next pages show the tables and plots created by applying the `boot_mse()`function to the model `ms2` (model 15), overall and for the levels of the covariates separately (Sex: Females/Males; Age (Adult/Minor; Nationality (Irish/Non Irish) and Exploitation (Sexual and Non Sexual).

```
$x
    nobs Nhat min95 max95
all  302  430   386   485

$Y
     Year nobs Nhat min95 max95
2014 2014   75   98    76   121
2015 2015  102  153   123   186
2016 2016  125  179   151   212

$S
  Year nobs Nhat min95 max95
F 2014   55   75    58    94
F 2015   73  106    82   134
F 2016   77  128   104   156
M 2014   20   23    15    32
M 2015   29   47    35    61
M 2016   48   51    38    66

$A
  Year nobs Nhat min95 max95
A 2014   62   78    61    97
A 2015   78  110    88   135
A 2016  104  133   113   160
M 2014   13   20    13    28
M 2015   24   42    30    57
M 2016   21   45    33    60

$N
  Year nobs Nhat min95 max95
I 2014    8   15    10    21
I 2015   15   31    21    43
I 2016   20   33    23    45
N 2014   67   83    65   102
N 2015   87  121    98   148
N 2016  105  145   123   172

$E
  Year nobs Nhat min95 max95
O 2014   22   28    19    37
O 2015   34   52    40    66
O 2016   52   57    45    73
S 2014   53   71    55    89
S 2015   68  100    79   127
S 2016   73  121    99   148
```

Figure 8 *Estimated numbers of victims according to the best fit model including 3 lists, four covariates (S, A, N and E) and three years.*







The results presented in the tables and plots above show that the confidence interval of the estimates for 2014 has been significantly reduced by the inclusion of a third list and more covariates (compare figure 8 with figure 7 above).

Results for combinations of covariates are presented in Annex A.

SECTION 4

Overview of results of MSE studies and discussion

Being able to reliably measure a phenomenon is the first step towards a concerted and sustainable response. The report of the Australian Institute of Criminology on its MSE-based study draws the following conclusion: “Quantifying the extent of human trafficking and slavery victimization enhances our understanding of the problem and provides a sound evidence base for informing policy development, the provision of victim support and law enforcement activities” (Lyneham, Dowling & Bricknell, 2019, p.1). The value of the MSE methodology is that it provides a sound and cost-effective route for countries to estimate the true scale and nature of their trafficking problem.

A national study using the MSE methodology will produce an estimate of the total number of victims of human trafficking in the country. This victim estimate is a headline statistic that, within modest statistical margins of uncertainty, indicates the true level of trafficking victimization in the country as currently defined by the authorities. It also shows how much of this total estimate has come to the attention of law enforcement and/or dedicated NGOs and how much hasn’t.

If a national study is carried out with data on the sex, age and form of exploitation of victims, the results can also be used to highlight which types of human trafficking are more common – for example, for sexual exploitation or for non-sexual exploitation - and which groups of the population are most at risk. This can provide invaluable information for the planning of prevention programs and the setting of priorities in investigation or victim support. More specifically, MSE using covariates, provides insight in which groups of victims are more or less likely to be detected by either law enforcement or support organization. If data for multiple years are used to produce an MSE study, it is possible to produce analyses of changes over time in victim detection. The results can demonstrate both trends in the prevalence of human trafficking and whether certain victims have become more or less ‘visible’ to the authorities or NGO’s over time.

Some key results of MSE worldwide

The key results of MSE studies are estimates of the total numbers of persons who have for some time been trafficked/exploited in a country in the course of a year.[38] For the purpose of international comparison, these numbers should be expressed as rates per 100,000 population as required for SDG indicator 16.2.2. Another interesting comparative measure is the ratio between recorded and estimated numbers per country since gives insight in the effectiveness of existing programs to detect victims.

Although the results of MSE studies in different countries are not strictly comparable, as will be discussed extensively below, we will present some international findings by way of illustration. Figure 9 presents the key results of ten MSE studies conducted in recent years for further scrutiny.

Figure 9 Recorded and estimated numbers of trafficking victims, confidence intervals of the estimates, ratios between recorded and estimated numbers and estimated numbers per 100.000 inhabitants, plus notes, from ten recent MSE studies (Durgana & Van Dijk, 2021)

Country/ Location	Year	Recorded	Estimated	Min/Max	Ratio Rec/Est	Victims Per 100.000	Notes
Australia	2016/ 2017	414	1567	1300-1900	1:4	3	adults only
Ireland	2016	125	197	148-213	1:2	4	
The Netherlands	2014/ 2015	1450	6500	6250-6750	1:4	38	
Romania	2015	880	1328	1062-2083	1:2	7	incl abroad
Slovakia	2016	45	87	59-164	1:2	1.5	incl abroad
Serbia	2016	146	970	670-1560	1:7	14	incl abroad

38. This rate provides a so-called flow estimate, to be distinguished from the rate of persons being exploited in a country at any time (a so-called stock estimate). Stock estimates can be calculated by multiplying flow estimate with the average duration of the exploitation. For example, if the average duration of exploitation is three months, a flow estimate of 1000 victims per year yields a stock estimate of 250 victims at any given time.

Country/ Location	Year	Recorded	Estimated	Min/Max	Ratio Rec/Est	Victims Per 100.000	Notes
United Kingdom	2014	2744	11000	10000-13000	1:4	17	
USA 1 (East)	2016	290	657	549-806	1:2	109	minors only
USA 2 (West)	2016	345	2235	1,606-3,609	1:6	972	
USA 3 (New Orleans)	2016	185	1000	650-1600	1:5	69	

From a comparative perspective, the rates of estimated numbers per 100,000 inhabitants seem the most pertinent statistic. The first MSE study to be carried out on human trafficking data, done in the United Kingdom, produced a victimization rate of 17 per 100,000 population for 2014 (Bales, Hesketh & Silverman, 2015). The UNODC/Dutch National Rapporteur study on Dutch data, using co-variates, found a rate of 38 per 100,000 population in the Netherlands for 2014/2015. The Australian MSE, not using covariates, study reported an estimated rate of 3.3, mainly consisting of adults (Lyneham, Dowling & Bricknell, 2019). Other MSE studies – all using covariates – produced rates of 12 for Serbia, 6 for Romania and 3 for both Ireland and Slovakia. In the three local studies in the United States, much higher prevalence rates were estimated than in the national studies in Europe and Australia.

Possible explanations for these differences can first of all be found in the methodologies of the studies. Some studies included covariates and others did not. Although the use of covariates can lead to both higher and lower estimations, they often affect the estimates and thus compromise comparability. The divergent results from the American studies might also be related to the methodology applied. First, these were results from local studies. Local prevalence of crime may deviate strongly from a national rate as has been found in many criminological studies. Perhaps the studies were conducted in places with high concentrations of human trafficking. Another factor possibly explaining high prevalence in the American studies is that MSE was conducted on lists of presumed victims as defined by the researchers, rather than on administrative lists produced by police officers and other practitioners applying official definitions of human trafficking.

International comparability

When interpreting the variation in victimization rates, several other factors compromising comparability must be considered. A first factor is the definition of human trafficking in national law. In terms of legislation, the UN Trafficking in Persons Protocol has served as an inspiration for many national laws, as the vast majority of countries by now have legislation that is broadly in line with the provisions outlined in the UN Trafficking in Persons Protocol.[39] Nevertheless, there is legal ambiguity and room for interpretation in the UN Trafficking in Persons Protocol 's definition of trafficking in persons. This ambiguity is reflected in the different ways that countries have chosen to criminalize trafficking, and how the definition is operationalized by involved agencies and in enforcement prioritization. In some countries, the detection of human trafficking is for example largely focussed on trafficking of females for sexual exploitation, while in others, cases of domestic trafficking are not included. These factors have a strong influence on the types of victims that appear on lists used for MSE and can compromise the international comparability of results[40].

A second factor hampering international comparability, is the concept of a presumed victim. Neither in the UN Protocol nor in any other international law this concept has been narrowly defined. The concept of a presumed victim often forms part of a National Referral Mechanism. In the United Kingdom, for example, presumed victims are strictly defined as persons who pass a 'reasonable grounds test' that they have been trafficked (an expert of the Competent Authority 'suspects but cannot prove' that a person is more likely to be a victim of trafficking than not). In the Netherlands, the concept of an identified victim does not formally exist. For practical purposes, law enforcement agencies and NGO's are instructed to see as a 'presumed victim' any person who, in the eyes of first responders, '*shows any sign that they might have been trafficked*.' The Dutch definition seems broader and therefore to put the bar a bit lower for including persons as victims. Definitions used in most other countries resemble the one used in the UK.

Thirdly, the universe of presumed trafficking victims differs fundamentally between typical source countries of human trafficking and countries of destination. Of the ten countries where MSE has been applied to victim data, Serbia, Romania and Slovakia are predominantly countries of origin of trafficking. Presumed victims have mostly been victimized abroad and repatriated to their home countries. The numbers of recorded victims, and of estimated totals of such victims, provide important information to policymakers about the numbers of victims who should

39. 169 of the 181 countries assessed for the 2020 *Global Report on Trafficking in Persons.*

40. Within Europe, legislation and policies on human trafficking have been further harmonized by the EU Directive of 2012 and the Council of Europe Convention on Action against Human Trafficking of 2009. Both require the incorporation of the Palermo definition in domestic law and encourage the collection of multi-source statistics on victims so defined.

be provided with protection and support in the country. However, these persons have not all been exploited on the national territory of the country, and the rates for these countries are therefore not strictly comparable with those of destination countries such as the UK, The Netherlands, Australia and Ireland. To calculate rates of persons exploited on the territory of a country, the Slovak, Serbian and Romanian rates should be broken down into rates for those exploited abroad and those exploited on the national territory. The latter rates are considerably lower than the total rates presented in figure 9.

Estimated numbers by covariates

As said, if covariates are used in the calculations, the overall estimates are more robust and separate trafficking victimization rates can be calculated for different population groups, as requested by SDG indicator 16.2.2.

In the Dutch study, four covariates were included: sex (male/female), age (minor/adult), type of exploitation (sexual/non-sexual) and citizenship (Dutch/non-Dutch). The victimization rates per 100,000 population for sexual and non-sexual exploitation were 25 and 13 respectively, indicating that in the Netherlands, trafficking for sexual exploitation was the predominant form of trafficking in that period. However, a repeat of the study using data from 2016-2019 shows that the estimated totals of cases of sexual exploitation and for forced labour have converged and are now almost equal (Van Dijk, Cruyff & Van der Heijden, 2021). Such shift in the composition of the estimated totals of trafficking victims has obvious implications for anti-trafficking policies.

MSE studies using covariates can produce victimization rates per population subgroup. These rates have been found to vary greatly. In the Netherlands, the rates per year for 2015 were 32 victims per 100,000 adult Dutch females and 1 per 100,000 adult Dutch males. The reason for this difference is that males are only rarely sexually exploited and are somewhat less likely to be subjected to forced labour as well. Assuming that child victims of human trafficking in the Netherlands are usually between 12 and 17 years old, the victimization rates of minors are 235 per 100,000 girls (between 12-17 years old) and 14 per 100,000 boys (between 12-17 years old). The rate of Dutch nationals is estimated as 22. The rate among non-Dutch (legal or irregular) residents is estimated at 326 victims per 100,000. Some of these migrants have no doubt been brought into the country for the very purpose of trafficking exploitation.

Of the ten studies, six have used covariates. If in the future results of more countries become available, this will allow more focussed international comparisons of estimated numbers per type of exploitation, sex and age. With such refinements, the comparability of the results will be considerably increased.

Ratios between recorded/observed and estimated cases

In Figure 9 international findings are presented regarding the ratios between recorded and estimated numbers. The ratios per country differ within a limited range of 1:2 and 1:7. The typical ratio in the larger destination countries – Australia, Netherlands, United Kingdom and USA (New Orleans) is 1:4 or 1:5. In all these countries recording entities such as police and NGO's together detect a quarter of all presumed victims or less. With extra efforts, they could easily double or even triple their recorded numbers and thereby the numbers of victims rescued and supported. The scope for improvements in detection was found to be considerably lower in Serbia (3), Romania (2) and Ireland (1,5).

Although the limited number of participating countries does not allow firm conclusions, the results suggest that the ratios between recorded and estimated victims are smaller in typical countries of origin than in countries of destination, although the relatively low ratio in Ireland is a notable exception. A possible explanation for this pattern might be that in countries of origin, the majority of presumed or identified victims are repatriated nationals victimized, and often primarily detected and identified, abroad. Persons previously identified by institutions and/or by themselves as victims in a destination country might be more visible for relevant institutions in their home country upon repatriation than victims in destination countries. In this context it should also be acknowledged that for the identification of victims of trafficking, institutions in countries of origin largely depend on detection efforts by their counterparts in countries of destination. In typical countries of origin, victim identification is co-dependent on law enforcement efforts in countries where their nationals are exploited and/or international referrals from service-providing NGO's from those countries. The border-crossing nature of human trafficking complicates the policy implications for source countries of MSE-based results concerning victims exploited abroad. Such results primarily shed light on the efficacy of national arrangements to deal with referrals of victims from abroad.

Final observations and discussion

From the MSE studies that have been carried out so far, several lessons can be learned.

Experience has taught that to make existing datasets suitable for MSE, they often have to be somewhat adapted, for example by combining smaller lists of similar institutions into larger ones. Close consultation between the operational handlers of the original data system and the researchers is key to a successful MSE study.

For governments, an advantage of MSE is that it takes as its starting point the governments' own administrative data, recorded by officially dedicated institutions working within the legal and operational infrastructure of the country. For this reason, MSE-based estimates provide pertinent feedback to policymakers on the maximum number of victims which could be recorded if all institutions involved in detecting and supporting victims were well established and optimally resourced. The knowledge that, for example, three or four times more victims could potentially be reached if detection efforts were stepped up, allows policymakers to set realistic, evidence-based targets for their country's anti-trafficking policies within a chosen timeframe.

A clear advantage of MSE studies compared to other methods to measure hidden figures such as large-scale population surveys is their cost-effectiveness. In more and more countries, comprehensive national referral systems are introduced to improve detection and support of trafficking victims. A positive side-effect of this development is that in an increasing number of countries digitalized, integrated datasets on victims will become available in appropriately secured environments. If these datasets meet the appropriate specifications, MSE-based estimates of the true numbers can be achieved at minimal costs.

The Dutch experience has shown that repeated MSE studies can provide a fact-based insight in long term trends in different types of human trafficking, in this case revealing a gradual shift in the volume of trafficking victims from victims of trafficking for sexual exploitation to victims of trafficking for labour exploitation between 2010 and 2019.

The use of administrative data for MSE also comes with certain limitations. Some categories of victims may be so deeply hidden in society that none of them have ever been recorded at all. In the terminology of marine biology - where capture-recapture methodology was first applied by capturing, marking and recapturing fish in a pond - such segments are called 'ground feeders', meaning fish that never appear in any normal catch of fish. Such segments that do not appear on any of the available lists will remain outside the reach of the multiple systems estimation. Examples in the field of human trafficking are especially deeply hidden subgroups such as irregular immigrants engaging in commercial sex or laboring in isolated sites in the countryside or on ships. Another example are newly emerging groups of victims which have not yet come to the attention of law enforcement or service providers.

A pertinent example from the United Kingdom are minors working as drug couriers for so-called 'county lines' in different parts of the country. By now, these minors make up a significant part of identified victims of trafficking, but they have largely been missed in the MSE study that used data from 2013 since no or almost no such victims had been recorded at the time. The possibility that certain deeply hidden subgroups or groups of victims that have not yet been recognized as such, are not recorded on any list and are therefore missed in the estimates, is another reason why MSE studies should be regularly repeated, using the latest available lists available. It is also a reason to supplement MSE-based estimation with field studies on self-reported victimization among vulnerable groups to detect newly emerging forms of human trafficking and exploitation.

References

Agresti, A., *Categorical Data Analysis,* 3rd ed., Wiley & Sons, New Jersey, 2013.

Anderson, V.R., T.C. Kulig & Chr. J. Sullivan, *Estimating the Prevalence of Human Trafficking in Ohio*, School of Criminal Justice at the University of Cincinnati under grant number 2016-JG-HTP-6096 for the Ohio Office of Criminal Justice, February 2019a.

Anderson, V.R., T.C. Kulig & Chr. J. Sullivan, Agency Records as a Method for Examining Human Trafficking, In M. Delen & D.M.D Siva (Eds), *Methods of Criminology and Criminal Justice*, Emerald Group of Publishing, pp. 165-178, 2019b.

Bales, K.O., O. Hesketh & B. W. Silverman, 'Modern Slavery in the UK; how many victims', *Significance,* pp. 16-21, Vol 12 June 2015.

Bales, K.O., L.T. Murphy & B. W. Silverman, How many trafficked people are there in Greater New Orleans; Lessons in Measurement, *Journal of Human Trafficking*, Vol 6, Issue 4, pp. 375-385, 2020.

Barrick, K. & Pfeffer, R. *Human Trafficking Policy and Research Analyses Project: Comprehensive Review of Prior Prevalence Studies and Recommendations for Field Testing in the United States.* OPRE Report No. 2021-XXX. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2021.

Buimer, L., A. Mack, M. van Meeteren & Y. Bleeker, *Mensenhandel Herkennen en Begrijpen; Lessen van Utrecht en Ede*, Regioplan, (Recognizing and Understanding Human Trafficking; Lessons from Utrecht and Ede), December 2021.

Bird, S. M. & R. King, 'Multiple Systems Estimation (or Capture-Recapture Estimation) to Inform Public Policy,' *Annual Review of Statistics and Its Application*, vol. 5, pp. 95–118, March 2018.

Böhning, D., P.G.M. van der Heijden & J. Bunge, *Capture-Recapture Methods for the Social and Medical Sciences*, CRC Press, Boca Raton, 2018.

Burnham, K. and Anderson, D., *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer, 2002

Chan, L., Silverman, B.W. & Vincent, K., Multiple Systems Estimation for Sparse Capture Data: Inferential Challenges When There Are Nonoverlapping Lists, *Journal of the American Statistical Association*, *116:535*, 1297-1306, 2021.

CoMensha, *Slachtoffermonitor Mensenhandel 2014-2018 uitgebracht: reactie CoMensha op de vijf aanbevelingen,* 18 October 2019 (available online: https://www.comensha.nl/actualiteiten/item/slachtoffermonitor-mensenhandel-2014-2018-vijf-aanbevelingen).

Council of Europe, *9th General Report on GRETA's activities*, 2019 (available at: htpps://rm.coe.int).

Croatian Red Cross, *The identification of victims of human trafficking in transit and destination countries in Europe; A practical guideline for frontline workers,* STEP project, 2019.

Cruyff, M., van Dijk J. & van der Heijden P., The Challenge of Counting Victims of Human Trafficking: Not on the record: A multiple systems estimation of the numbers of human trafficking victims in the Netherlands in 2010–2015 by year, age, gender, and type of exploitation. *CHANCE* 30(3), pp. 41–49, 2017.

De Vries, I. & C. Dettmeijer-Vermeulen, Extremely wanted: human trafficking statistics- what to do with the hodgepodge of numbers? In Kangaspunta, K. (ed.) *Forum on Crime and Society, Special Issue: Researching Hidden Populations: Approaches to and Methodologies for Generating Data on Trafficking in Persons*, 8, pp. 15-37, 2015.

Durgana, D. & J.J.M. van Dijk, Measuring the Hidden Numbers of Human Trafficking Through Multiple Systems Estimation: Lessons Learned and Challenges Outstanding, *J. of Crime and Delinquency,* first published January, 2021.

Farrell, A., M. Dank, M. Kafafian, S. Lockwood, R. Pfeffer, A. Hughes & K. Vincent, *Capturing Human Trafficking Victimization Through Crime Reporting,* National Institute of Justice, January 2019.

French Ministry of Foreign Affairs, *Guidelines for the first level identification of victims of trafficking in Europe,* Euro TrafGuID, June 2013.

Goudie I.B.J & M. Goudie, Who captures the marks for the Petersen estimator, *J. R. Statist. Soc. A 170*, Part 3, pp. 825–839, 2007.

International Labour Office, *Operational indicators of trafficking in human beings, results from a Delphi Study*, Geneva: ILO, 2009.

International Labour Office, Special Action Programme to Combat Forced Labour, *ILO Global Estimate of Forced Labour: Results and Methodology*, 2012.

International Labour Office, Department of Statistics, ICLS/20/2018/Guidelines, *Guidelines concerning the measurement of forced labour,* 2018.

International Working Group for Disease Monitoring and Forecasting (IWGDMF), Capture-recapture and multiple-record systems estimation I: History and theoretical development. *American Journal of Epidemiology*, Nov 15;142(10), pp1047-58, 1995.

Jones, H. E., Hickman, M., Welton, N. J., De Angelis, D., Harris, R. J., & Ades, A. E., Recapture or precapture? Fallibility of standard capture-recapture methods in the presence of referrals between sources. *American Journal of Epidemiology*, *179*(11), 1383–1393, 2014.

Lyneham, S, C. Dowling & S. Bricknell, Estimating the dark figure of human trafficking and slavery victimisation in Australia, *Australian Institute of Criminology: Statistical Bulletin* 16, February 2019.

National Rapporteur on Trafficking in Human Beings and Sexual Violence against Children (Netherlands), *Victims of Human Trafficking. Periodical Report 2012-2016*. *Summary*, 2017.

O'Connell, D, R. Brittingham, D. Borton & L. Mutis, *Human Trafficking in Delaware*, Center for Drug and Health Studies University of Delaware, September, 2020.

OSCE/ODHIR, *National Referral Mechanisms; Joining efforts to Protect the Rights of Trafficked Persons, A Practical Handbook,* 2004.

Overstall, A.M., King R., Bird S.M., Hutchinson S.J., Hay G. Incomplete contingency tables with censored cells with application to estimating the number of people who inject drugs in Scotland. *Stat Med.* Apr 30; 33(9):1564-79, 2014.

Phillips, A.R. *Commercial sexual exploitation; an analysis of prostitution in Kansas City,* Thesis University of Missouri-Kansas City, 2017.

Quinteros Rojas., D. R. Dufraix Tapia, R. Ramos Rodríguez, Human Trafficking Cases in Chile: Challenges for Reducing the “Dark Figure” *The Palgrave International Handbook of Human Trafficking*, pp.1151-1164, 2019.

Sharifi Far, S., R. King, S. Bird, A. Overstall, H. Worthington & N. Jewell, Multiple Systems Estimation for Modern Slavery: Robustness of List Omission and Combination, *Crime & Delinquency*, vol. 67, 13-14: pp. 2213-2236, First Published August 22, 2020.

Silverman, B.W. *Modern slavery: an application of multiple systems estimation*. Gov. UK. 2014.

Silverman, B.W., Multiple-systems analysis for the quantification of modern slavery: classical and Bayesian approaches, *Statistics in Society, Series A*, Volume 183, Issue 3, pp 689-1326, 2020.

Silverman B.W. and K. Vincent, ‘Multiple Systems Estimation for Sparse Capture Data: Inferential Challenges when there are Non-Overlapping Lists’, *Journal of the American Statistical Association,* vol. 00, no. 0, pp. 1-10, February 2020.

Sutherland, J. & C. J. Schwarz, Multi-List Methods Using Incomplete Lists in Closed Populations, *Biometrics. J of the Biometric Society*, Vol 61, 1, pp. 134-140, 2005.

UNODC/National Rapporteur on Trafficking in Human Beings and Sexual Violence against Children (Netherlands), *Monitoring Target 16.2 of the United Nations’ Sustainable Development Goals; a multiple systems estimation of the numbers of presumed human trafficking victims in the Netherlands in 2010-2015 by year, age, gender, form of exploitation and nationality*, 2017.

UNODC in collaboration with Walk Free Foundation, *Research Briefs: Monitoring Target 16.2 of the United Nations Sustainable Development Goals: multiple systems estimation of the numbers of presumed victims of trafficking in persons, Ireland, Romania and Serbia,* 2018 (available at: https://www.unodc.org/unodc/en/data-and-analysis/tip.html).

UNODC, *Research Brief: Multiple Systems Estimation for estimating the number of victims of human trafficking across the world,* 2015.

UNODC, *UNODC helps countries uncover real magnitude of trafficking in persons and report on SDG 16*, 2018 (see www.unodc.org).

UNODC, *Global Report on Trafficking in Persons,* 2020 (United Nations publication, Sales No. E.20.IV.3), p. 61, 2020

Van der Heijden, P.G.M., J. Whittaker, M. Cruyff, B. Bakker & R. Van der Vliet, 'People born in the Middle East but residing in the Netherlands: invariant population size estimates and the role of active and passive covariates,' *The Annals of Applied Statistics*, Vol. 6, No. 3, pp. 831–852, 2012.

Van der Heijden, P.G.M., de Vries, I., Böhning, D., & Cruyff, M. Estimating the size of hard-to-reach populations using capture-recapture methodology, with a discussion of the International Labour Office's global estimate of forced labour, In Kangaspunta, K. (ed.) *Forum on Crime and Society, Special Issue: Researching Hidden Populations: Approaches to and Methodologies for Generating Data on Trafficking in Persons*, 8, pp.109-136, 2015

Van Dijk, J.J.M. & P.G.M. van der Heijden *Research Brief. Multiple Systems Estimation for estimating the number of victims of human trafficking across the world*, Vienna: UNODC, 2016

Van Dijk, J. & P.G. M. van der Heijden (2016). *On the potential of Multiple Systems Estimation for estimating the number of victims of human trafficking across the world.* Written for the UNODC, where it appeared as Research Brief. (available at: https://www.unodc.org/documents/data-and-analysis/tip/TiPMSE.pdf.

Van Dijk, J.J.M., Cruyff, M., Van der Heijden, P.G.M.& Kragten-Heerdink, S.L.J, *Monitoring Target 16.2 of the United Nations sustainable development goals: a multiple systems estimation of the numbers of presumed human trafficking victims in the Netherlands in 2010-2015 by year, age, gender, form of exploitation and nationality,* Vienna: UNODC, pp 51, 2018.

Van Dijk, J.J.M., Cruyff, M. & Van der Heijden, P.G.M.*, A Multiple Systems Estimation of presumed victims of human trafficking in the Netherlands during 2016-2019,* The Hague: Ministry of Justice and Security, November 2021.

Walk Free Foundation, *The Global Slavery Index,* 2018.

Walk Free Foundation, Van Dijk, J.J.M., Cruyff, M.& Van der Heijden, P.G.M.*, Monitoring Target 16.2 of the United Nations' Sustainable Development Goals; a multiple systems estimation of the numbers of presumed human trafficking victims in Slovakia in 2016-2018 by year, age, gender and form of exploitation,2019.*

Zhang, S. X., *Trafficking of Migrant Laborers in San Diego County: Looking for a Hidden Population*, San Diego, CA: San Diego State University, 2012.

Zhang, S.X. & J. Loudo-Larsen, Estimating the Size of the Human Trafficking Problem: MSE and Other Strategies, *Crime and Delinquency*, Vol 67, Issue 13-14, pp 1-9, 2021

Annex A

Tables and plots of estimates of human trafficking victims in Ireland with confidence intervals; results of best fit model including three lists, by pairwise combinations of covariates (S and A; S and N; S and E; A and N; A and E; N and E).

$SxA

	Year	nobs	Nhat	min95	max95
F:A	2014	49	61	47	77
F:A	2015	55	77	59	96
F:A	2016	65	97	79	118
F:M	2014	6	14	9	21
F:M	2015	18	29	19	42
F:M	2016	12	31	21	44
M:A	2014	13	17	11	24
M:A	2015	23	33	25	43
M:A	2016	39	36	27	48
M:M	2014	7	6	4	10
M:M	2015	6	13	8	20
M:M	2016	9	14	9	21

$SxN

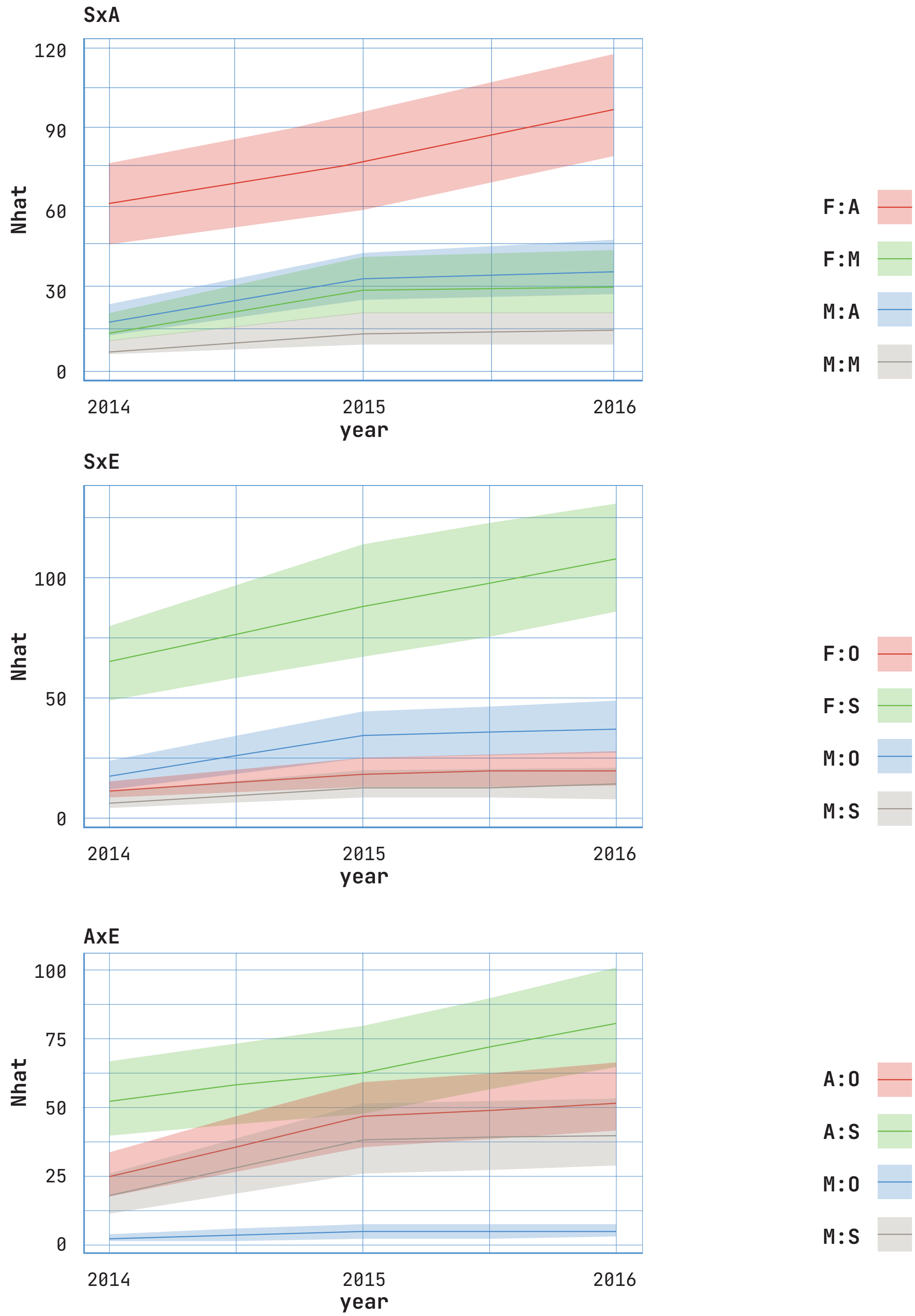
	Year	nobs	Nhat	min95	max95
F:I	2014	3	10	6	16
F:I	2015	11	21	13	32
F:I	2016	11	23	14	34
F:N	2014	52	65	49	81
F:N	2015	62	85	66	106
F:N	2016	66	105	86	129
M:I	2014	5	5	2	8
M:I	2015	4	10	5	16
M:I	2016	9	11	6	17
M:N	2014	15	19	12	26
M:N	2015	25	37	27	48
M:N	2016	39	40	29	52

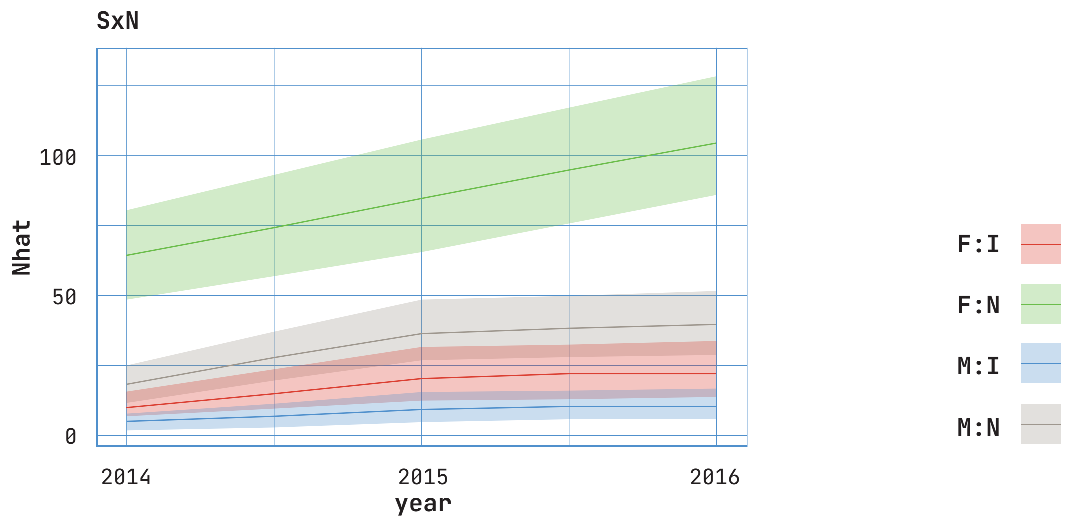
$SxE

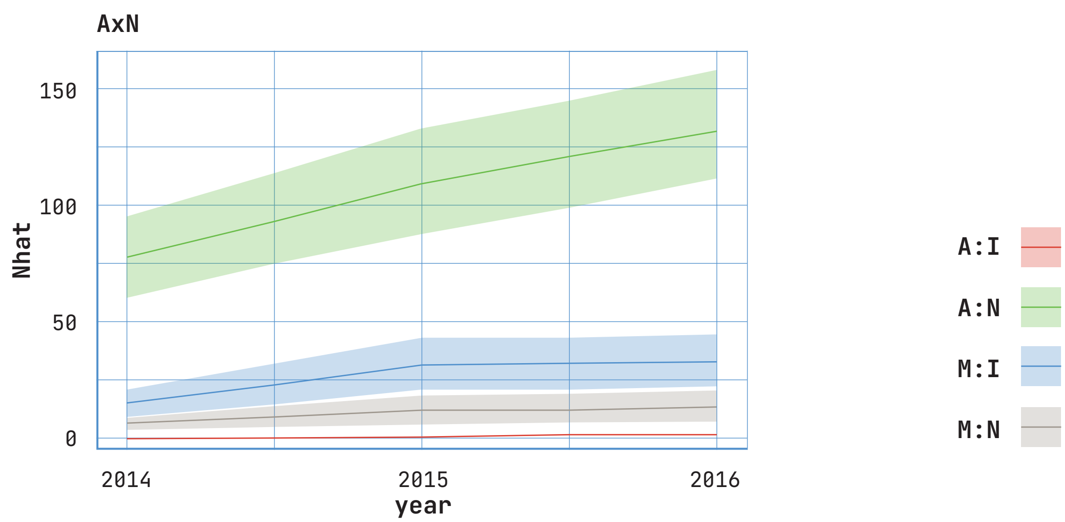
	Year	nobs	Nhat	min95	max95
F:O	2014	8	11	7	15
F:O	2015	13	18	12	25
F:O	2016	14	20	14	28
F:S	2014	47	65	49	81
F:S	2015	60	88	67	114
F:S	2016	63	108	86	132
M:O	2014	14	17	11	24
M:O	2015	21	34	25	44
M:O	2016	38	37	27	49
M:S	2014	6	6	4	10
M:S	2015	8	13	8	20
M:S	2016	10	14	8	21

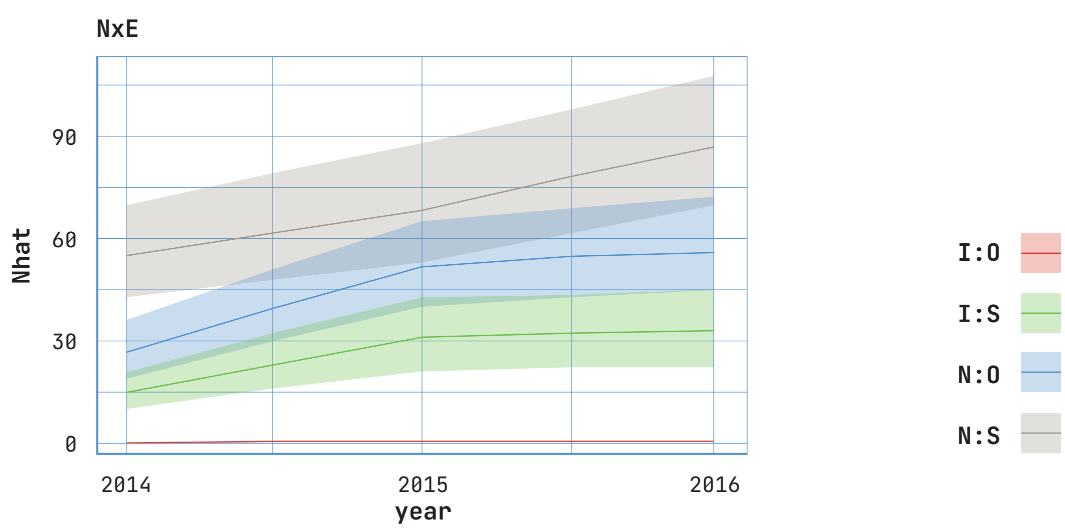
```
$AxN
    Year nobs Nhat min95 max95
A:I 2014    0    0     0     1
A:I 2015    0    0     0     1
A:I 2016    1    1     0     2
A:N 2014   62   78    61    96
A:N 2015   78  110    88   134
A:N 2016  103  133   112   159
M:I 2014    8   15     9    21
M:I 2015   15   31    20    43
M:I 2016   19   33    22    45
M:N 2014    5    6     3     9
M:N 2015    9   12     6    18
M:N 2016    2   13     7    20

$AxE
    Year nobs Nhat min95 max95
A:O 2014   19   25    18    34
A:O 2015   31   47    36    60
A:O 2016   52   52    42    66
A:S 2014   43   53    40    67
A:S 2015   47   63    48    80
A:S 2016   52   81    65   101
M:O 2014    3    2     1     4
M:O 2015    3    5     2     8
M:O 2016    0    5     3     8
M:S 2014   10   18    12    25
M:S 2015   21   38    26    51
M:S 2016   21   40    29    54

$NxE
    Year nobs Nhat min95 max95
I:O 2014    0    0     0     0
I:O 2015    0    0     0     0
I:O 2016    0    0     0     0
I:S 2014    8   15    10    21
I:S 2015   15   31    21    43
I:S 2016   20   33    23    45
N:O 2014   22   28    19    37
N:O 2015   34   52    40    66
N:O 2016   52   57    45    73
N:S 2014   45   56    43    71
N:S 2015   53   69    53    89
N:S 2016   53   88    71   109
```

Figure 9 Estimated numbers of victims in Ireland according to the best fit model including 3 lists, four covariates (S, A, N and E) and three years by pairwise combinations of covariates.




SxN
Nhat
100
50
0
2014
2015
2016
year
F:I
F:N
M:I
M:N


AxN
Nhat
150
100
50
0
2014
2015
2016
year
A:I
A:N
M:I
M:N


NxE
Nhat
90
60
30
0
2014
2015
2016
year
I:O
I:S
N:O
N:S

Parameter estimates model 15

	est
(Intercept)	-24.3
R11	-0.3
R21	-0.2
R31	-1.3
SM	2.3
AM	4.7
NN	25.6
ES	21.9
Y1	6.6
Y2	-3.3
AM:NN	-6.4
SM:ES	-3.7
R31:SM	-1.3
NN:ES	-20.4
R31:AM	-3.4
SM:NN	-2.2
R21:ES	-1.2
R31:Y1	-6.6
R31:Y2	9.7
R11:SM	1.1
R21:R31	-1.3
R21:AM	-2.8
R31:ES	1.0

UNODC
United Nations Office on Drugs and Crime