

TITLE: INTERNATIONAL MIGRATION DATA: CHALLENGES, IMPROVEMENTS, AND WAYS FORWARD

Ahmad Wali Ahmad Yar

Executive summary

Human migration has gained substantial attention in political and academic spheres worldwide in recent decades due to the escalating number of global migrants. Although the increase is not significantly high compared to world population growth, the figures have risen considerably from 161 million in 1995 to 281 million in 2020, a surge of around 27.15% in less than three decades. This rapid growth has brought forth complex challenges related to migration patterns, stocks, flows, and their socio-demographic dimensions. Global remittances and refugee numbers have also seen substantial increases.

This surge in migration poses challenges for policy formulation, migration management, and optimising the benefits of migration. To effectively address these challenges, evidence-based findings play a crucial role. However, despite efforts to collect migration data, significant gaps remain, leading to uncertainty for stakeholders involved in migration management. The disconnect between available migration data and relevant theories exacerbates these challenges, further complicating the situation. Bridging the gap between traditional migration theories and current realities has been a topic of discussion for decades. Migration statistics and theories are intertwined, with data forming the foundation for theoretical development, testing, and validation. The interplay between empirical evidence and theoretical frameworks enhances our understanding of migration complexities, contributing to informed policymaking and better migration management. Migration theories often aim to explain migration patterns, and migration statistics are crucial in supporting or challenging these explanations. To that end, challenges in migration statistics need to be addressed. Inconsistent definitions across countries, unequal coverage of variables, and difficulties in collecting data on irregular migration pose significant issues. Efforts to improve data collection have been made, with international organisations and NGOs involved, but improvements have not been uniform across regions. Moreover, the emergence of big data and digital technology offers potential solutions to complement traditional data sources. Big data provides insights into migration aspects where traditional data is lacking, but its integration into migration research remains relatively unexplored. Integrating perspectives from data-collecting institutions such as National Statistical Institutions (NSIs) is essential for enhancing data quality and accuracy.

In light of these challenges, the ultimate objectives of the present thesis encompass a comprehensive assessment of the quality and relevance of migration statistics in the context of contemporary migration theories and challenges. Through a thorough examination of existing data gaps, this research aims to unravel the intricacies that hinder the acquisition of accurate migration data. Additionally, it seeks to understand and address under and over-coverage issues within irregular migration statistics, proposing strategies to enhance data precision. Furthermore, by integrating perspectives from data-collecting institutions like National Statistical Institutions (NSIs), this research aims to foster collaboration and improve data quality. Lastly, this thesis explores the potential of big data to complement traditional sources, uncovering its capacity to address uncertainties in migration research and its synergy with emerging migration theories. Ultimately, the findings aspire to inform evidence-based policymaking, bolster migration governance, and contribute to a more nuanced understanding of global migration dynamics.

This thesis employs a robust and multifaceted methodology to comprehensively address its research objectives centred around understanding and bridging gaps in migration data. The chosen multimethod approach draws inspiration from Brewer & Hunter (1989), Goertz (2017), Mello (2021), and Seawright

(2016), aiming to provide a comprehensive exploration of the issue by integrating various methodological techniques within a single investigation. To conceptualise migration data gaps, a twofold approach is employed. The study commences with a systematic literature review to conceptualise existing gaps, serving as a basis to develop a method for systematically reviewing data sources. This unique extension of the systematic review methodology offers a clear framework for evaluating statistical data sources and data portals, enhancing its potential for future research. To obtain official perspectives from National Statistical Institutions (NSIs), the study innovatively combines elements from the 'problem-centred expert interview' and the 'expert opinion questionnaire' methodologies, creating a 'problem-centred institution questionnaire'. This hybrid approach allows comprehensive data collection from NSIs, overcoming challenges associated with obtaining opinions from entire institutions. Exploring the potential of big data in migration research involves expert interviews, engaging field specialists to gather insights into the role of big data in complementing traditional sources. Lastly, coverage issues in irregular migration data are examined through a constructivist approach grounded in existing literature, augmented by insights from case studies. By strategically blending these methodologies, the research provides a nuanced perspective on migration data gaps, their relevance to migration theories, and avenues for improvement. These pioneering methodologies contribute not only to research findings but also hold wider applications across various domains. With dedicated chapters for each research objective, the study offers a coherent exploration of migration data gaps and their implications, advancing our understanding of this vital aspect of global migration dynamics.

The results indicate that persisting gaps in migration data have remained a challenge for over a century and that the gaps could be categorised under definitions and measures, drivers or reasons behind migration, geographic coverage, gaps in demographic characteristics and the time lag in the availability of data. The reviewed sources suffer from the gaps which are not mutually exclusive (they are interlinked): the quality and availability of both migration flows, and stocks data vary across regions and countries, and migration statistics highly rely on the arrival of the immigrants.

Examining under and over-coverage in irregular migration, the study reveals complexities due to invisibility and multiple registrations. The thesis elaborates on the existing ambiguities and shortcomings of the statistical coverage and the flaws in the registry systems, particularly for irregular migration. More importantly, it emphasises the potential explanations (and hence the possible implications) of conceptual uncertainties. The study further highlights that NSIs play a crucial role in identifying gaps in migration data collection related to flows and stocks. These gaps are primarily due to differences in data sources and collection methods among countries. Harmonizing definitions and sources would standardize data and enable timely updates. Many countries are transitioning from surveys to administrative sources for better data quality. Gaps in both flow and stock data include key socio-economic characteristics of migrants, impacting international data adequacy. Demographic disaggregation is prioritized for the EU and Sustainable Development Goals, yet variables like education level and reasons for migration are challenging. NSIs use innovative methods, such as side surveys and data augmentation, to fill these gaps.

NSIs' engagement with big data reflects transformative potential, but challenges include legal, ethical, and technological constraints. Lack of expertise and experience hinders full utilisation. The findings emphasise legal structures, ethical guidelines, and international collaboration for big data's effective integration. The study's insights underscore the importance of data for migration theories, policy-making, and practical implementation, addressing gaps in migration trends, theory validation, policy relevance, and societal impacts. Practical implications include innovative data extraction solutions and nuanced discussions about alternative sources. Though limitations exist, the research paves the way for future cross-disciplinary research and more inclusive methodologies. Overall, the study navigates complex challenges, shedding light on migration data intricacies and providing recommendations to inform evidence-based policy formulation and management. In sum, this research contributes to a comprehensive understanding of migration dynamics and advocates for a future where migration is managed and facilitated with wisdom and compassion.

Samenvatting

Menselijke migratie heeft de afgelopen decennia wereldwijd veel aandacht gekregen in politieke en academische kringen vanwege het toenemende aantal migranten wereldwijd. Hoewel de stijging niet significant hoog is in vergelijking met de groei van de wereldbevolking, zijn de cijfers aanzienlijk gestegen van 161 miljoen in 1995 tot 281 miljoen in 2020, een stijging van ongeveer 27,15% in minder dan drie decennia. Deze snelle groei heeft complexe uitdagingen met zich meegebracht die te maken hebben met migratiepatronen, -stocks, -stromen en hun socio-demografische dimensies. Ook het aantal geldtransfers naar het buitenland en het aantal vluchtelingen wereldwijd zijn aanzienlijk gestegen.

Deze stijging in migratie brengt uitdagingen met zich mee voor het vormgeven van beleid en het beheer van migratie en het optimaliseren van de voordelen ervan. Om deze uitdagingen effectief aan te pakken, spelen empirisch onderbouwde vaststellingen een cruciale rol. Ondanks de inspanningen om migratiedata te verzamelen, blijven er aanzienlijke hiaten bestaan, wat leidt tot onzekerheid voor stakeholders die betrokken zijn bij migratiebeheer. De discrepantie tussen beschikbare migratiedata en relevante theorieën versterkt deze uitdagingen, waardoor de situatie nog gecompliceerder wordt. Het dichten van de kloof tussen traditionele migratietheorieën en de huidige realiteit is al tientallen jaren een onderwerp van discussie. Migratiestatistieken en -theorieën zijn onlosmakelijk met elkaar verbonden, waarbij de migratiedata de basis vormen voor het ontwikkelen, testen en valideren van theorieën. De wisselwerking tussen empirisch bewijs en theoretische kaders vergroot ons begrip van de complexe migratieproblematiek en draagt bij aan een empirisch onderbouwde beleidsvorming en een beter migratiebeheer. Migratietheorieën zijn vaak gericht op het verklaren van migratiepatronen, en migratiestatistieken zijn cruciaal voor het ondersteunen of in twijfel trekken van deze verklaringen. Met dat doel moeten uitdagingen inzake migratiestatistieken worden aangepakt. Inconsistentie in definities tussen landen, onder- en overrapportage van statistieken en moeilijkheden bij het verzamelen van data over onregelmatige migratie vormen aanzienlijke problemen. Er zijn reeds inspanningen geleverd om de dataverzameling te verbeteren, met betrokkenheid van internationale organisaties en ngo's, maar de vorderingen zijn niet in alle regio's gelijk. Bovendien biedt de opkomst van big data en digitale technologie potentiële oplossingen als aanvulling op traditionele databronnen. Big data bieden inzicht in migratieaspecten waar traditionele data ontbreken, maar de integratie ervan in migratieonderzoek blijft relatief onontgonnen terrein. Het integreren van perspectieven van data verzamelende instellingen zoals Nationale Statistische Instituten (NSI's) is essentieel voor het verbeteren van de kwaliteit en nauwkeurigheid van de data.

Met oog op deze uitdagingen omvatten de uiteindelijke doelstellingen van dit proefschrift een uitgebreide beoordeling van de kwaliteit en relevantie van migratiestatistieken in de context van hedendaagse migratietheorieën en -uitdagingen. Door middel van een grondige analyse van bestaande data hiaten, beoogt dit onderzoek de complexiteiten te ontrafelen die de verwerving van accurate migratiedata belemmeren. Daarnaast tracht het de ongelijkmatige rapportage inzake onregelmatige migratiestatistieken te begrijpen en aan te pakken, en stelt het strategieën voor om de nauwkeurigheid van de data te verbeteren. Dit onderzoek beoogt bovendien de samenwerking met dataverzamelende instellingen zoals Nationale Statistische Instituten (NSI's) te bevorderen, en de datakwaliteit te verbeteren, door het integreren van de perspectieven van deze instellingen. Tot slot onderzoekt dit proefschrift het potentieel van big data als aanvulling op traditionele bronnen, waarbij zowel het vermogen om onzekerheden in migratieonderzoek aan te pakken als de synergie met opkomende migratietheorieën worden blootgelegd. Het doel is dat de bevindingen empirisch onderbouwde beleidsvorming informeren, migratiebeheer versterken en bijdragen aan een genuanceerder begrip van de dynamiek van wereldwijde migratie.

Dit proefschrift hanteert een robuuste en gevarieerde methodologie om de onderzoeksdoelstellingen, gericht op het begrijpen en overbruggen van de hiaten in migratiedata, uitgebreid aan te pakken. De gekozen multimethodische aanpak is geïnspireerd op Brewer & Hunter (1989), Goertz (2017), Mello (2021) en Seawright (2016), en heeft als doel een uitgebreide verkenning van het onderwerp te bieden door verschillende methodologische technieken te integreren in één onderzoek. Om de hiaten in de migratiedata te conceptualiseren, wordt een tweeledige aanpak gehanteerd. Het onderzoek begint met een systematische literatuurstudie om bestaande hiaten te conceptualiseren, die als basis dienen om een methode te ontwikkelen voor het systematisch onderzoeken van databronnen. Deze unieke uitbreiding van de systematische literatuurstudie biedt een duidelijk kader voor het evalueren van statistische databronnen en dataportalen, waardoor het potentieel voor toekomstig onderzoek wordt vergroot. Om officiële perspectieven van Nationale Statistische Instituten (NSI's) te verkrijgen, combineert het

onderzoek op innovatieve wijze elementen uit de 'probleemgerichte expertinterviews' en de 'vragenlijst voor expertopinions' methodologieën, tot een 'probleemgerichte vragenlijst voor instellingen'. Deze hybride aanpak maakt een uitgebreide gegevensverzameling mogelijk van NSI's en voorkomt uitdagingen die gepaard gaan met het verkrijgen van opinies van gehele instellingen. Het verkennen van het potentieel van big data in migratieonderzoek omvat expertinterviews, waarbij vakdeskundigen betrokken worden om inzicht te krijgen in de rol van big data ter aanvulling van traditionele bronnen. Tot slot wordt de onder- en overrapportage van onregelmatige migratie onderzocht via een constructivistische benadering gebaseerd op bestaande literatuur, aangevuld met bevindingen van casestudies. Door deze methodologieën strategisch te combineren, biedt het onderzoek een genuanceerd perspectief op de hiaten in migratiedata, hun relevantie voor migratietheorieën en mogelijkheden voor verbetering. Deze vooruitstrevende methodologieën dragen niet alleen bij aan de onderzoeksresultaten, maar hebben ook bredere toepassingen in verschillende domeinen. Met specifieke hoofdstukken voor elk onderzoeksdoel biedt deze studie een coherente verkenning van de hiaten in migratiedata en hun implicaties, waardoor het begrip van dit vitale aspect van de wereldwijde migratiedynamiek wordt verbeterd.

Uit de resultaten blijkt dat migratiedata al meer dan een eeuw lang hiaten vertonen en dat deze hiaten kunnen worden onderverdeeld in definities en maatregelen, drijfveren of redenen van migratie, geografische spreiding, hiaten in demografische kenmerken en vertragingen in de beschikbaarheid van gegevens. De bestudeerde bronnen vertonen ook hiaten die elkaar niet uitsluiten (ze zijn met elkaar verbonden): de kwaliteit en beschikbaarheid van migratiedata en gegevens over migratiestromen en stocks verschillen per regio en land, en migratiestatistieken zijn sterk afhankelijk van de aankomst van immigranten.

Deze studie onderzoekt onder- en overrapportage van onregelmatige migratie en onthult de complexiteiten als gevolg van onzichtbaarheid en meervoudige registraties. Het proefschrift gaat in op de bestaande ambiguïteiten en tekortkomingen van de statistische dekking en de gebreken in de registratiesystemen, met name voor onregelmatige migratie. Bovendien wordt de nadruk gelegd op de mogelijke verklaringen (en daarmee de mogelijke implicaties) van conceptuele onzekerheden. Het onderzoek benadrukt verder dat NSI's een cruciale rol spelen bij het identificeren van de hiaten in dataverzameling met betrekking tot de migratiestromen en -stocks. Deze hiaten zijn voornamelijk te wijten aan verschillen in databronnen en verzamelmethode tussen landen. Het stroomlijnen van definities en bronnen zou de data standaardiseren en actuele updates mogelijk maken. Veel landen stappen over van enquêtes naar administratieve bronnen om de kwaliteit van de data te verbeteren. De hiaten in zowel de stroom- als stockdata omvatten belangrijke sociaaleconomische kenmerken van migranten, wat een impact heeft op de internationale bruikbaarheid van de gegevens. Demografische disaggregatie is een prioriteit voor de EU en de Sustainable Development Goals, maar variabelen zoals opleidingsniveau en redenen voor migratie vormen een uitdaging. NSI's gebruiken innovatieve methoden, zoals nevenenquêtes en dataverrijking, om deze hiaten op te vullen.

De betrokkenheid van NSI's bij big data weerspiegelt het transformerende potentieel ervan, maar er zijn ook uitdagingen, zoals wettelijke, ethische en technologische beperkingen. Een gebrek aan expertise en ervaring staan een optimaal gebruik in de weg. De conclusies van dit proefschrift tonen het belang aan van juridische structuren, ethische richtlijnen en internationale samenwerking voor een effectieve implementatie van big data. De bevindingen van het onderzoek onderstrepen ook het belang van migratiedata voor migratietheorieën, beleidsvorming en praktische implementatie, waarbij hiaten in migratietrends, de validatie van theorieën, beleidsrelevantie en maatschappelijke effecten worden aangepakt. Praktische implicaties omvatten innovatieve oplossingen voor data-extractie en genuanceerde discussies over alternatieve bronnen. Hoewel er beperkingen zijn, effent dit onderzoek de weg voor toekomstig interdisciplinair onderzoek en meer inclusieve methodologieën. Samenvattend kan worden gesteld dat dit onderzoek meer inzicht biedt in de complexiteit van migratiedata, en aanbevelingen geeft om een empirisch onderbouwd beleid te formuleren en beheren. Kortom, dit onderzoek draagt bij aan een uitgebreid begrip van de migratiedynamiek en pleit voor een toekomst waarin migratie met wijsheid en mededogen wordt beheerd en gefaciliteerd.