

Chapter 1

DATA AND METHODOLOGY

The analysis presented in this report draws mainly on two primary sources: the British Burma District Gazetteers and the 2019 GAD Township Reports. The first source, the District Gazetteers, were produced by the British authorities in two series—the 1912 series and the 1924 series. They cover detailed historical information at the district level (e.g. historical background, boundaries, state administration, ethnic composition). Given the enormous influence colonial district borders had on present-day administrative borders, the District Gazetteers serve as an important primary source. The second source, the GAD Township Reports, provides township-level ethnicity data.

This chapter elaborates on the GAD Township Reports with two objectives—(1) to provide an overview of background information on the GAD Township Reports and how they are utilized in the analysis, and (2) to identify the limitations of this primary source and discuss their implications.

DEARTH OF ETHNICITY DATA IN MYANMAR

Data is essential for evidence-based policymaking, and ethnicity data is crucial for crafting policies and strategies, including making constitutional arrangements, in order to mitigate problems associated with ethnic diversity, exclusion and discrimination based on ethnic identity. Information on where groups are concentrated could help address issues related to political representation. Additionally, ethnicity data can help us better understand the extent of differences and inequality between minority and majority groups as well as among minority groups.

However, such data is almost non-existent in Myanmar. The 2014 Census, the first census in three decades, could help address this information gap, but the

‘The lack of disaggregated and reliable data remains a fundamental challenge to political participation in Myanmar, which has made it difficult to track and monitor the levels of participation and representation of disadvantaged groups’ (Ebead and Hirakawa 2022: 8).

ethnicity data in the census has been withheld indefinitely.³ To our knowledge, the only source of fine-grain ethnicity data is the GAD Township Reports.

One reason for the lack of ethnicity data is that, prior to the 2014 Census, no census had been taken for three decades. The only post-independence censuses were conducted in 1973 and 1983.⁴ Moreover, while these censuses included ethnicity data at the country and state/region levels, the more fine-grain township-level data did not appear in the released census reports.

Another reason for the omission of ethnic data at large, and at the township level in particular, is the general sensitivity to ethnicity data in Myanmar. There is also a related aversion to publishing details, on the part of the authorities as well as members of minority communities, academics and human rights activists. Several factors contribute to this aversion. One is the concern that ethnicity statistics might be confusing or invalid because officially recorded ethnicity may differ from how individuals self-identify and/or ethnic groups tend to be miscategorized. As such, the resulting numbers may not accurately capture the population size of the ethnic groups. Another is the concern that ethnicity statistics, however accurate or inaccurate, might exacerbate ethnic tensions on the ground. Mary Callahan discusses this concern as follows: 'When the numbers are released, many *lummyo* (ethnic) groups are going to be disappointed with their absolute and relative statistical representations, while viewing the statistics for other groups as suspiciously large, and perhaps artificially inflated by design' (Callahan 2017: 453). Such disappointment, combined with the knowledge of a 'rival' ethnic group with which one's ethnic group shares a township, might result in vulnerable minority groups being targeted with more violence and discrimination.

These are, of course, valid and significant concerns. At the same time, it is important to not conflate data generation with data accessibility. It is also important to note that policymaking and institution-building without any data, based only on speculative information, could lead to adverse effects as well. Furthermore, given the continued centrality of ethnicity in the context of the constitution-building debate (i.e. the Federal Democracy Charter (FDC) and the policies of the NUG), it is essential to draw from all available sources to better understand the ethnic landscape in Myanmar.

GAD TOWNSHIP REPORTS

The GAD had been the backbone of public administration in Myanmar since 1988. One of the GAD's primary roles had been to support coordination and

³ Many Myanmar observers, commentators and scholars have written on the census controversy. See, for example, Palatino (2014).

⁴ During the parliamentary period (1948–1962), there was an attempt to take the first post-independence census in 1953. The plan was to conduct the enumeration in stages, stretching from 1953 to 1955. However, the plan was not completed due to political instability in the country. In the aftermath of the coup in 1962, the socialist government was established and took up the first post-independence census in 1973. See Myanmar (n.d.a).

communication among the Union Government's ministries and to connect them to every level of subnational administrative units all the way down to the wards and village tracts (W/VT) (Saw and Arnold 2014). The GAD has been unique in that it is the only government department that operates at the W/VT level.⁵ Given its expansive coordination role, the GAD collected and stored vital population information obtained by relevant government agencies, including data on population movement and basic demography (Saw and Arnold 2014).

Using the aforementioned information, the GAD compiled annual reports for each township.⁶ The reports, each totalling each totaling about 90 pages on average, include data disaggregated at the township level concerning the population's characteristics as well as economic and social indicators. The ethnicity data utilized in this report is based on a section of the GAD Township Reports indicating the population size of each ethnic group residing in the respective township.⁷

Every GAD Township Report begins with an introduction stating that the report is based on data collected by various departments which is then sent to the GAD. Beyond this brief statement, there is no additional information about the data source. Presumably, the number of schools, the attendance rate and the matriculation exam pass rate came from the Department of Basic Education. Likewise, information on population size, including ethnic headcount, came from the Department of Population (DOP). However, the process through which each department arrived at the numbers reported to the GAD is not discussed.

Since the GAD Township Reports are available in PDF only, it was necessary to extract the tables indicating the township-level ethnic composition data, and then recompile those tables into a usable format (i.e. Excel files). To do so, we utilized a publicly available optical character recognition (OCR) tool.⁸ With this tool, we automated the process of identifying tables and keywords, which enabled the extraction of the ethnic composition tables. These tables (labels and numbers in Burmese) were then manually translated into English in Excel format.

Next, a team of research assistants manually cross-checked the information in the Excel files against the original GAD Township Reports (PDF files). Half of the townships in each state/region (totalling 162 townships) were randomly

⁵ Below the W/VT GAD administrators, there are 100-household heads and 10-household heads, who informally serve as the point persons between the W/VT administrators and the local residents. Under the previous junta (1990–2010), the local GAD personnel served as the government's surveillance officials as well. Today, in post-coup Myanmar, local residents are required to report guest registrations to their local GAD offices.

⁶ However, no report can be found for five townships in Shan State—Mongla, Mongmao, Narphan, Pangsang and Pangwaun.

⁷ It is unclear when the GAD began producing these township reports. We first came across these reports on the Myanmar Information Management Unit (MIMU) website (<https://themimu.info>) in 2018. A Myanmar expert who studies the GAD indicated that she had seen these reports as early as 2013. It is possible that these reports have existed for decades but became accessible to the public only in the early 2010s. Today, these reports are available on the MIMU website as well as on the GAD's website in Burmese (<http://www.gad.gov.mm/my>).

⁸ OCR enables character extraction from PDF files, making them analysable using statistical software. In the case of the GAD 2019 files, two limitations complicated the process. First, OCR tools have not been optimized for Burmese fonts. Second, the quality of the GAD files varied, causing additional errors in the transcription process. That is why, though the extraction process was mostly automated, there was extensive manual oversight from the beginning.

selected for cross-checking.⁹ In this process, we found and corrected just 13 inaccuracies. Given such a low rate of inaccuracy, we are fairly confident that the data analysed mirrors the information indicated in the GAD Township Reports, though we do not claim that the data is an accurate reflection of reality.

Data quality and limitations

Though we are certain that the data was extracted with very little clerical error, concerns over data quality remain—that is, information presented in the GAD Township Reports may not present sound estimates of the real ethnic landscape. In discussing data quality, it is important to acknowledge that data based on the GAD Township Reports, like other statistical undertakings, is not the truth; it is rather an estimate of the truth. To assess the limitations of the data, we consider, in this section, potential sources of data inaccuracy. We conclude that, while the population size indicated in the GAD Township Reports is likely outdated and/or flawed to a certain extent, the estimated population proportion remains a useful approximation of the ethnic landscape.

Ethnic identification

The ethnic headcount indicated in the GAD Township Reports is not based on self-identified ethnicity. As we discuss below, the information on ethnicity in the GAD Township Reports is most likely based on various government records, including national registration cards (also known as citizenship scrutiny cards) and white cards—the latter being temporary identification documents issued to those without a national registration card.¹⁰ These government-issued documents state the holder's ethnicity (*lumyo*), among other personal information. An individual's stated ethnicity may not match their self-identified ethnicity.

The process through which an individual's stated ethnicity was derived was highly opaque. The stated ethnicity is presumably based on the government-issued identification documents of the individual's parents. However, there are anecdotal cases in which an individual's ethnicity differs from that of their parents. Moreover, the parents' ethnicity stated on their identification documents could also differ from their self-identified ethnicity. Furthermore, media reports and NGO reports indicate cases in which bureaucrats ignored individuals' requests to state their self-identified ethnicity on their documents, resulting in a Muslim of Bamar ethnicity being recorded as 'Pakistani Islam', a Shanni being recorded as 'Bamar', a Ta'ang being recorded as 'Shan' and so on (Norwegian Refugee Council and Institute on Statelessness and Inclusion 2018; Consult-Myanmar 2019).

Despite the limitations of the available data, it is difficult to assess the extent of the discrepancy between it (data based on the GAD Township Reports) and data based on self-identified ethnic information as well as the corresponding

⁹ The research assistants were asked to check and correct whether the population size of a particular ethnic group in a given township was accurate.

¹⁰ Holders of white cards were able to vote in Myanmar's 2010 election, but their voting right was revoked in early 2015. The majority of holders of white cards were Rohingya. See Tun (2019).

implication for estimating the ethnic landscape. Consider the Chin ethnic group, for example. It is possible that some percentage of the population reported as 'Chin' in the GAD Township Reports self-identify not as Chin but rather as part of their regional and/or linguistic group, such as Asho or Zomi. If that percentage is fairly small, the discrepancy between the current ethnic data and the self-identified ethnic data would not shift the overall trends in the population share of Chin or other ethnic groups. If that percentage is fairly large, however, then the corresponding discrepancy could be substantial, resulting in largely inaccurate estimates of the ethnic landscape.

Ethnic categorization

As in many countries with a diverse population, a major concern in producing ethnic data in Myanmar is ethnic categorization. Since the creation of the citizenship hierarchy, per the 1982 Citizenship Law,¹¹ the Myanmar Government has recognized 135 'national races' (*taingyintha*) (136 since 2015 [Thawnhmung and Yadana 2017]). These groups are categorized into eight 'major national races': Bamar, Chin, Kachin, Kayah, Kayin, Mon, Rakhine and Shan. This controversial list and the categorization were developed by the socialist one-party regime before the 1983 Census was conducted (Myint 2013).

The list of recognized ethnic groups has caused a long-standing controversy for several reasons. First, determining which groups are recognized effectively provides the grounds for denying citizenship rights to groups that are unrecognized. As such, some have argued that an ethnic hierarchy effectively undermines ethnic equality (Ko and Ford 2022). Second, categorization into eight major groups has also been controversial because most categories are meaningless (Wansai 2017). Some subgroups disagree that they belong to the larger group they are assigned to, on the basis of ethno-cultural differences as well as historical grievances. Furthermore, other subgroups are neither politically nor culturally salient.

Interestingly, the term *taingyintha* appears in the FDC—'a document written as a precursor for a new constitution to replace the military-drafted 2008 constitution' (Chan and Ford 2021)—though the term is left undefined. The NUG has stated that the Rohingya people would have full citizenship in Myanmar (The Irrawaddy 2022). However, given the use of the term *taingyintha* when attributing rights in the FDC, it remains unclear what the status of the Rohingya people and other unrecognized ethnic groups would be in relation to the *taingyintha*.

Given the controversial nature of ethnic categorization in Myanmar, adopting either of the existing paradigms (8 or 135 groups) could be problematic. Interestingly, just over two dozen ethnic groups appear in the GAD Township Reports (see Table 1.1).¹² This proliferation of listed ethnic groups resulted

¹¹ The 1982 Citizenship Law does not actually list the ethnic groups.

¹² The section on ethnicity information in the GAD Township Reports is titled 'Residing Taingyintha in the Township'. It is important to note that the list provided almost mirrors what could be considered 'politically relevant' in Myanmar. Politically relevant groups are the cultural cleavages that matter for political competition and social conflict (Posner 2005: xv). The list of such groups is often different from, and much

from the disaggregation of the Shan subgroups from the Shan.¹³ As expected, all eight major groups are on the list. Additionally, about two dozen subgroups that are generally understood to be distinct from the eight main groups also appear on the list.¹⁴ For example, the Danu, Kokang, Pa-O, Palaung¹⁵ and Wa are listed separately from the Shan—that is, these groups are not included in the Shan population size. Similarly, the Naga also appear separately from the Chin.

Table 1.1. Ethnic groups mentioned in the GAD Township Reports

Akha	Kadu	Lahu	Mone Wong	Salong (Moken)
Bamar	Kanan	Lishaw	Myaung Zi (Hmong)	Shan
Chin	Kayah	Lisu	Naga	Taung Yoe
Danu	Kayan	Loi	Pa-O	Wa
Innثار	Kayin	Loila	Palaung	Ying
Kachin	Kokang	Mon	Rakhine	

Note: The Dawei people and the Hta Naut people are also mentioned in the report for Taunggyi Township; they are, however, not recorded in any other township in Myanmar.

Source: Information retrieved from the 2019 GAD Township Reports; table compiled by the authors.

While successive pre-coup governments insisted on recognizing 135 groups, they actually utilized a paradigm consisting of just a few dozen groups. This discrepancy is remarkable. It suggests an implicit recognition of the fact that the past regimes' ideological narrative regarding the official list of ethnic groups was very detached from reality—or at least impracticable.

The list of ethnic groups mentioned in the GAD Township Reports may not be agreeable to members of all ethnic communities in Myanmar. This list could certainly be further refined to create a more accurate depiction of the ethnic landscape in Myanmar. At the same time, most ethnic minorities are likely to find this list more acceptable than a list of just eight major groups or a list of 135 groups. Given its deviation from the existing paradigm, the ethnic categorization utilized in the GAD Township Reports is arguably less

shorter than, the list of cultural groups in a given country, because not all cultural cleavages matter for people's political behaviour and attitudes. In the context of Myanmar, politically relevant ethnic groups refer to any ethnic group, whether or not it is officially recognized by the government, that has mobilized a political party or a rebel organization.

¹³ The disaggregation of some of the main groups in the 2019 GAD Township Reports, resulting in more than eight groups, is a notable deviation from the 1983 and 1974 Census reports. The extent to which the disaggregation is selective, however, is unclear. Some groups (e.g. Pa-O) appear in the GAD Township Reports separately from Shan in reports from Shan State and other parts of the country. However, the disaggregation of other groups from the main group is selective. For example, the Lisu are disaggregated from the Kachin in reports from Shan State but not in reports from Kachin State. In a similar vein, the Kayan are disaggregated from the Kayah in reports from Shan State but not in reports from Kayah State. The reason for such selectivity is also not apparent.

¹⁴ The categories 'mixed-Chinese' and 'mixed-Indian' (presumably mixed with one of the *taingyintha* groups) also appeared in the table, labelled *taingyintha* for Mudon Township, Mon State.

¹⁵ The Ta'ang population is referred to as 'Palaung' in the GAD Township Reports and in the 2008 Constitution. However, the ethnic group generally prefers 'Ta'ang'.

controversial than what country observers and ethnic stakeholders would expect.

Beyond ethnic categorization, it is unclear how persons of mixed ancestry were recorded in the data. In addition to the recognized groups, the township reports also listed several 'foreign' ethnicities (*naing ngan char tar lumyo myar*). And only a few townships also included a category labelled 'other', which presumably includes persons of mixed ancestry.¹⁶ This omission raises questions about data sources and data quality because government-issued identification cards in Myanmar record ancestry rather than a single ethnicity. For example, an individual with a mother of Kayin ancestry and a father of Bamar ancestry would be designated as 'Bamar–Kayin'. Given that the GAD Township Reports list single ethnic categories, it is possible that such individuals are counted as Bamar or Kayin only. This phenomenon could result in an inflated population share of certain ethnic groups, particularly the Bamar.

Bureaucratic capacity and data quality

The source of the population data, including the ethnic headcount, in the GAD Township Reports cannot be ascertained. Nevertheless, those familiar with the GAD and DOP have suggested that the data likely came from either of the following: the ward/village GAD offices or the GAD Township Reports from the previous year.¹⁷ If the population data is deduced from the former, then the GAD township staff presumably tallied the numbers. If it is deduced from the latter, the numbers from the preceding years are presumably adjusted to account for new data on birth and death records, the issuance of national registration cards, population movement and other relevant information. As such, data quality could be highly dependent on the extent of bureaucratic reach and bureaucratic capacity.

While the bureaucracy in Myanmar was pervasive, extending down to the villages and wards through the GAD, the bureaucratic reach over the population was incomplete. Government registration and other apparatuses the government utilized to record and administer the population were not accessible to many average Myanmar families, especially in remote areas. Additionally, in conflict-affected areas government offices might not have been fully functional, and many ethnic minorities in such areas might also have feared visiting government offices. Consequently, government records are likely to undercount the population, particularly in conflict-affected areas.

Undercounting is evident when comparing the population data from the GAD Township Reports, which again is most likely based on government administrative records, to those from the 2014 Census report, which was

¹⁶ Mudon Township stands out as an anomaly given its mention of 'Chinese hybrid' and 'Indian hybrid' in the categorization of the table 'Residing Taingyintha in the Township' in the GAD Township Report.

¹⁷ It is necessary to interview civil servants staffing the GAD and DOP in order to better understand the process through which the ethnic headcount was obtained. However, given that this project began after the 2021 coup, we are unable to contact such individuals, who are either in hiding due to their participation in the Civil Disobedience Movement or currently still employed. We have, however, discussed the origins of the GAD Township Reports with an independent researcher who contributed an extensive report on the GAD, a consultant who contributed to the 2014 Census and a minor clerk at the township office of the DOP in 2019. Our understanding of the ethnic headcount process is based on their speculations.

based on enumeration (see Table 1.2). As expected, there was some level of population growth between the 2014 Census report and the 2019 GAD Township Reports in most states/regions. The places with significant undercounting were Kachin (-19.6 per cent), Shan (-14.4 per cent) and Rakhine (-10.5 per cent) states. The population in the Yangon Region appears to have been undercounted; this was most likely due to constant population movement in search of economic opportunities rather than bureaucratic capacity to reach the population (the last column in Table 1.2 suggests that the Yangon Region had the best bureaucratic capacity among all the states/regions).

Second, bureaucratic capacity is a concern because the process of compiling statistics based on government records requires that bureaucrats have some level of clerical and mathematical competence. While we do not have a measure of bureaucratic competence in Myanmar, we examined the extent of age heaping as a proxy. Measuring age heaping indexes (Myer's Index, Whipple's Index and the like), with the aim of assessing census quality, has been a common practice in demographic studies. Demographers use such indexes to capture the extent to which certain numbers are 'preferred' by the population surveyed. Those preferences are captured through peaks in the age distribution, often around numbers that end in 0 or 5. They serve as a signal to demographers about a problem with the data generation process which can be the result of a lack of age awareness or difficulties in collecting information from the population living in hard-to-reach places. Political scientists have only recently employed these indexes to assess the bureaucratic capacity of a state (Lee and Zhang 2017).

The last column in Table 1.2 shows the age heaping index (calculated with Whipple's formula) based on census data.¹⁸ While the index we calculated does not directly indicate the quality of GAD data, it speaks to the bureaucratic capacity at the state/region level, which affects the quality of GAD data. Whipple's Index theoretically provides a number between 100 and 500, where the higher values indicate lower accuracy. Overall, Myanmar's Whipple's Index value tells us that the census data is approximate (between 110 and 125). However, there is a notable difference between ethnic minority areas (i.e. states) and regions.

The places with the lowest Whipple's Index score, suggesting the best bureaucratic capacity, are the Yangon (lowest), Ayeyarwady, Bago and Magway regions. The places with the highest Whipple's Index, suggesting the worst bureaucratic capacity, are Shan (highest), Rakhine, Kachin and Kayin states. While this information does not discard intentional undercounting, it does provide some evidence that undercounting in the GAD Township Reports is partly caused by weak bureaucratic capacities in the ethnic minority areas affected by conflict compared with Bamar-dominated regions. This analysis suggests that data from the ethnic minority areas is likely to be of poorer quality compared with data from elsewhere.

¹⁸ Whipple's Index cannot be calculated based on the 2019 GAD Township Reports because township populations are not disaggregated by age. Calculating Whipple's Index requires knowledge of the proportion of the population whose age ends in 0 or 5.

Table 1.2. 2014 Census, 2019 GAD Township Reports and age heaping

State/Region	2014 Census ^a	GAD 2019 Township Reports	Difference in numbers	Difference in percentage	Age heaping ^b
Chin	478,801	532,727	53,926	10.1%	114.7
Kachin	1,689,441	1,412,157	-277,284	-19.6%	118.3
Kayah	286,627	282,697	-3,930	-1.4%	115.6
Kayin	1,574,079	1,595,210	21,131	1.3%	118.2
Mon	2,054,393	2,149,501	95,108	4.4%	111.2
Rakhine	3,188,807	2,885,835	-302,972	-10.5%	120.5
Shan ^c	5,824,432	5,090,862	-733,570	-14.4%	148.6
<i>Total state</i>	<i>13,890,227</i>	<i>13,948,989</i>	<i>58,762</i>	<i>0.4%</i>	<i>121</i>
Ayeyarwady	6,184,829	6,354,312	169,483	2.7%	108
Bago	4,867,373	4,963,294	95,921	1.9%	108.3
Magway	3,917,055	4,230,048	312,993	7.4%	109.1
Mandalay	6,165,723	5,951,666	-214,057	-3.6%	113.4
Nay Pyi Taw	1,160,242	1,118,989	-41,253	-3.7%	111.4
Sagaing	5,325,347	5,542,269	216,922	3.9%	116.2
Tanintharyi	1,408,401	1,505,161	96,760	6.4%	111
Yangon	7,360,703	6,579,444	-781,259	-11.9%	103.1
<i>Total region</i>	<i>36,389,673</i>	<i>36,245,183</i>	<i>-144,490</i>	<i>-0.4%</i>	<i>110.1</i>
Total Myanmar^d	51,486,253	50,194,172	- 1,292,081	-2.6%	115.5

Notes:

- The numbers reflect the combined total of the enumerated and estimated population. Due to conflicts in Kachin State and Kayin State and the situation of the Rohingya people in Rakhine State, 46,600, 69,753 and 1,090,000 people, respectively, are estimated to have not been counted in those areas.
- Whipple's Index of age heaping is typically interpreted as follows: <105, highly accurate; 105–109.9, fairly accurate; 110–124.9, approximate; 125–174.9, rough; and >175, very rough.
- The GAD Township Reports do not include four townships in Shan State—Mongmao, Namphan, Pangsang and Pangwaun. The population in these townships is estimated to be 388,289, according to the 2014 Census.
- The World Bank's estimates of Myanmar's population are 52.3 million for 2014 and 54 million for 2019. World Bank data are publicly available at: <<https://data.worldbank.org/indicator/SP.POP.TOTL?locations=MM>>.

Source: Data retrieved from the 2014 Myanmar Census and the 2019 GAD Township Reports; table compiled by the authors. Differences and age heaping are authors' own.

Given the discrepancies and concerns discussed above, the absolute numbers indicated in the GAD Township Reports should not be taken at face value. At the same time, the estimated population proportion (percentages) still provides a sound approximation of the actual population figures. As discussed above, factors affecting data quality are more likely due to local bureaucratic capacity than an ethnic phenomenon. In other words, low bureaucratic capacity in a given township affects the entire population living there rather than select groups. Take Kyaikmaraw Township, in Mon State, for example (see Table 1.3). The numbers in the second column (population size) for all groups are most likely lower than the actual numbers, but the percentages in the third column

(proportion) would likely be within a small margin of error if the population size were adjusted to the true values. Each ethnic group's population share, or proportion, is the basis for the analysis presented in the remainder of this report.

Table 1.3. Ethnic landscape in Kyaikmaraw Township, Mon State (GAD 2019 Township Reports)

	Population size	Proportion
Bamar	29,047	13%
Karen	47,230	21%
Mon	109,426	48%
Other	41,522	18%

Source: Data retrieved from the 2019 Kyaikmaraw Township GAD Township Report.