sciendo

Journal of Official Statistics, Vol. 39, No. 3, 2023, pp. 275–284, http://dx.doi.org/10.2478/JOS-2023-0013

Letter to Editor
Quality of 2017 Population Census of Pakistan by Age and Sex

Asif Wazir[1] *and Anne Goujon*[2]

This Letter to Editor is a supplement to the previously published article in the Journal of Official Statistics (Wazir and Goujon 2021).

In 2021, a reconstruction method using demographic analysis for assessing the quality and validity of the 2017 census data has been applied, by critically investigating the demographic changes in the intercensal period at national and provincial levels. However, at the time when the article was written, the age and sex structure of the population from the 2017 census had not yet been published, making it hard to fully appreciate the reconstruction of the national and subnational level populations.

In the meantime, detailed data have become available and offer the possibility to assess the reconstruction's outcome more in detail. Therefore, this letter aims two-fold: (1) to analyze the quality of the age and sex distribution in the 2017 Population census of Pakistan, and (2) to compare the reconstruction by age and sex to the results of the 2017 population census. Our results reveal that the age and sex structure of the population as estimated by the 2017 census suffer from some irregularities. Our analysis by age and sex reinforces the main conclusion of previous article that the next census in Pakistan should increase in quality with an inbuild post-enumeration survey along with post-census demographic analysis.

Key words: Census quality; population reconstruction; demographic analysis; Pakistan, Pakistani provinces.

1. Quality of 2017 Population Census of Pakistan by Age and Sex

The constitution of Pakistan does not require the census to be held every ten years as, for example, the U.S. Constitution does. The ten-year period is an international convention under which Pakistan held the census till 1981. Thereafter Pakistan held only two censuses in the last 40 years, that is, in 1998 and 2017. Like in many settings, the censuses in Pakistan have been subject to various types of non-sampling errors including misreporting, nonresponse, recall bias, etc. These quality concerns have not been properly addressed as the Pakistan Bureau of Statistics did not implement a post-enumeration survey (PES) for the last four censuses (though a PES was conducted for the 1981 census, however, results were not published) which was an essential component to assess the potential coverage and content errors in the census enumerations. The census that was held in 2017 was no exception.

[1] United Nations Population Fund, Maputo, Mozambique. Email: wazir@unfpa.org
[2] Population and Just Societies Program, International Institute for Applied Systems Analysis, Schlossplatz 1, 2361 Laxenburg, Austria. Email: goujon@iiasa.ac.at

Acknowledgments: This study is based on publicly available data and was not funded. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the United Nations Population Fund (UNFPA). The authors declare no conflict of interest.

In 2021, we applied, for the first time in the case of Pakistan, a population reconstruction method for assessing the quality and validity of the 2017 census data, by critically investigating the demographic changes in the intercensal period at national and provincial levels, using a range of intercensal surveys (Wazir and Goujon 2021). We showed that on the one hand, while the reconstructed population estimates at the national level can be seen as broadly valid, as compared to the 2017 census figures, on the other hand, the reconstruction does not perform so well at the sub-national level, notably because of the uncertainty involved around internal migration between 1998 and 2017.

However, at the time when the article was written and was under consideration by the Journal of Official Statistics, the age and sex structure of the population from the 2017 census had not yet been published, making it hard to fully appraise the reconstruction of the national and regional populations. In the meantime, detailed data have become available and offer the possibility to assess the reconstruction's outcome in more detail.

The aim of this letter is two-fold: (1) to analyze the quality of the age and sex distribution in the 2017 Population census of Pakistan, and (2) to compare the reconstruction by age and sex to the results of the 2017 population census.

2. Evaluation of Data Quality of the 2017 Population and Housing Census of Pakistan

As a starting point, the single-year age and sex distribution are examined visually to check for anomalies. Figure 1 shows the single-year age distribution by sex for the 1998 and 2017 censuses.

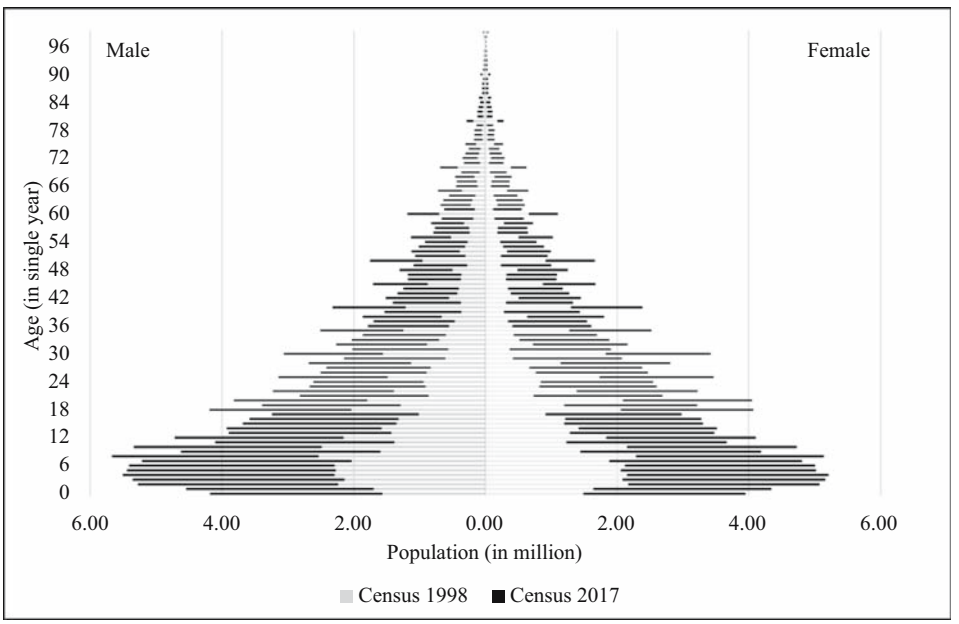

Fig. 1. Single-year age and sex distribution from the 1998 and 2017 censuses in Pakistan.
Source: Authors' calculations based on single-year age-sex data (Pakistan Bureau of Statistics 2022; Minnesota Population Center 2023).

The combined pyramids for the 1998 and 2017 censuses indicate that irregularities in the age distribution are significant for both men and women. The last two censuses show extensive age heaping for males and females alike, with the usual concentration around numbers ending in 0 and 5. Significant under-reporting of children (age $<$ ten years) and age misreporting among the working-age population are visible for both sexes in both censuses. The distortions are less common among older ages. It is worth noting, that their magnitude is less in the 2017 census compared to the 1998 census, suggesting improvements in the accuracy of the age data. This is mainly because the 2017 census used the national identity card (as well as the verified computerized national identity card from the central database) for the census questionnaire enumeration. Therefore, the date of birth was reported more precisely in the 2017 census.

3. Sex Ratio by Age

In the 2017 census, the share of men in the total population was 51.2% at the national level. The proportion was higher in Balochistan (52.6%) and lower in Punjab (50.8%). While the proportion of men in Sindh and KP stood at 52% and 50.6%, respectively.

Further, we also illustrate the sex ratio by single-year age group from the 1998 and 2017 censuses in Figure 2. The age heaping and misreporting in the 1998 census (shown in Figure 1) manifested itself in the distorted pattern of the sex ratio by age, at all ages, decreased in the 2017 census. The sex ratio at birth as well as up to the age of 15 showed a preference for boys in both censuses in Pakistan. Advances in sex-selective abortion technology and persistent discrimination in care practices for girls have resulted in higher female mortality. Between ages 20 and 40, the sex ratio reversed mainly due to slightly higher male mortality at younger ages, and potentially male international emigration for employment, particularly in the gulf countries. Estimates from the Bureau of Emigration and Overseas Employment (2020) show that 5.1 million people migrated overseas for

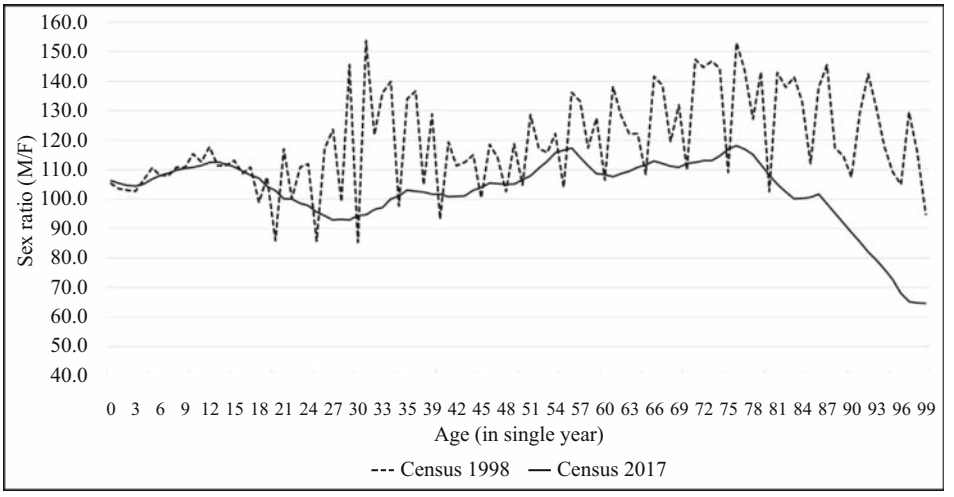

Fig. 2. Sex ratio by single year of age from the 1998 and 2017 censuses in Pakistan.
Source: Authors' calculations based on single-year age-sex data (Pakistan Bureau of Statistics 2022; Minnesota Population Center 2023).

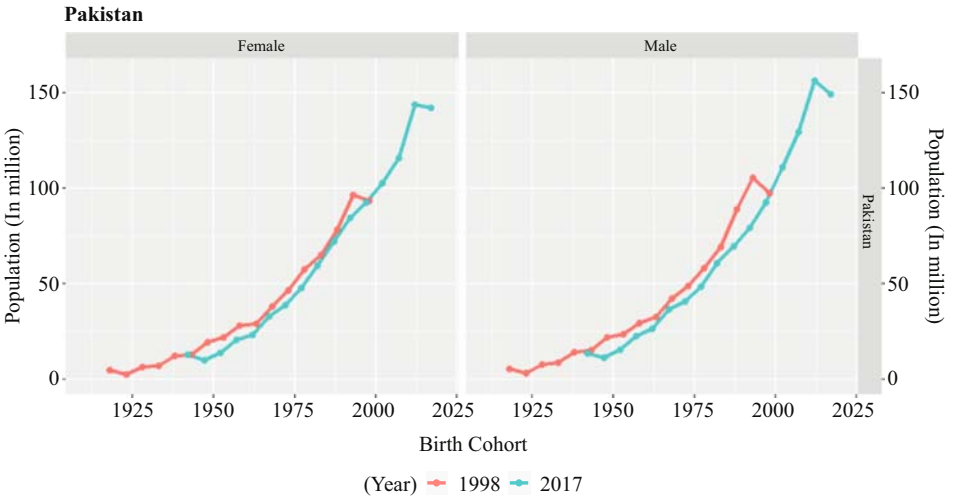

Fig. 3. Birth cohort age-sex distribution from the 1998 and 2017 censuses in Pakistan.
Source: Authors' calculations based on single-year age-sex data (Pakistan Bureau of Statistics 2022; Minnesota Population Center 2023).

employment through formal channels between 2010 and 2017. Finally, at older ages, the sex ratio continuously dropped below 100 due to the female advantage in mortality.

We also performed a graphical birth cohort analysis based on the population data from the last two censuses for men and women in Pakistan depicted in Figure 3. In ideal circumstances, the size of each birth cohort should decline over time (1998 and 2017) due to the mortality, with the assumption that the country did not experience significant migration. In the case of Pakistan, three important patterns emerged from Figure 3: first, the significant under-enumeration of boys and girls under the age of five in the censuses are evident; second, the significant differences in the patterns between the male and female birth cohorts at young ages reinforcing the under-enumeration for both male and female; and third, the almost overlapping of birth cohorts between the two censuses, especially for women. These inconsistencies by age and sex confirmed the presence of irregularities in the enumeration process, particularly in the 2017 census.

Apart from the visual check, we have computed the Whipple's Index (Siegel and Swanson 2004) and the Myers Index (Myers 1954), using the Population Analysis System (PAS) software developed by the U.S. Census Bureau (U.S. Census Bureau 2017). While at the national level and for larger regions like Punjab and Sindh, the 2017 census indices show fair results (in terms of age heaping and/or digit preference for both sexes), small regions have issues of inaccuracy in age reporting, particularly Balochistan and Khyber Pakhtunkhwa (KP).

4. Age Ratio Score

The age ratio score which measures the divergence of one age group from the neighboring two age groups is a valid indicator of errors in age reporting. In the absence of significant changes in fertility, mortality, and migration, the enumerated size of a particular cohort should be approximately equal to the average size of the preceding and following cohorts.

Without irregularities, the age ratio score should be approximately equal to 100. For example, Gerland (2014) utilized the age ratio scores to assess the quality of the census data in India. Figure 4 indicates the age-ratio scores for males (Figure 4a) and females (Figure 4b) from the 1998 and 2017 censuses. The age-ratio scores also reinforced the obvious irregularities in all age groups in both censuses: over-enumeration in the 5–9 age group, and under-enumeration in the 10–14 age group for both boys and girls. For age groups ranging from 15–19 to 45–49, the irregularities were noticeable for men and women in the same way, with higher intensity in the 1998 census. From age 50 onwords, substantial anomalies became evident for both sexes in both censuses, which may be related to age misstatement/exaggeration and the mortality patterns in older persons. Although age-reporting errors have persisted between the last two censuses, the 2017 census shows marked improvements compared with the 1998 census.

Based on the age ratio scores, we have calculated the age-sex accuracy index for five-year age groups (Siegel and Swanson 2004), which is the sum of the age ratio score for

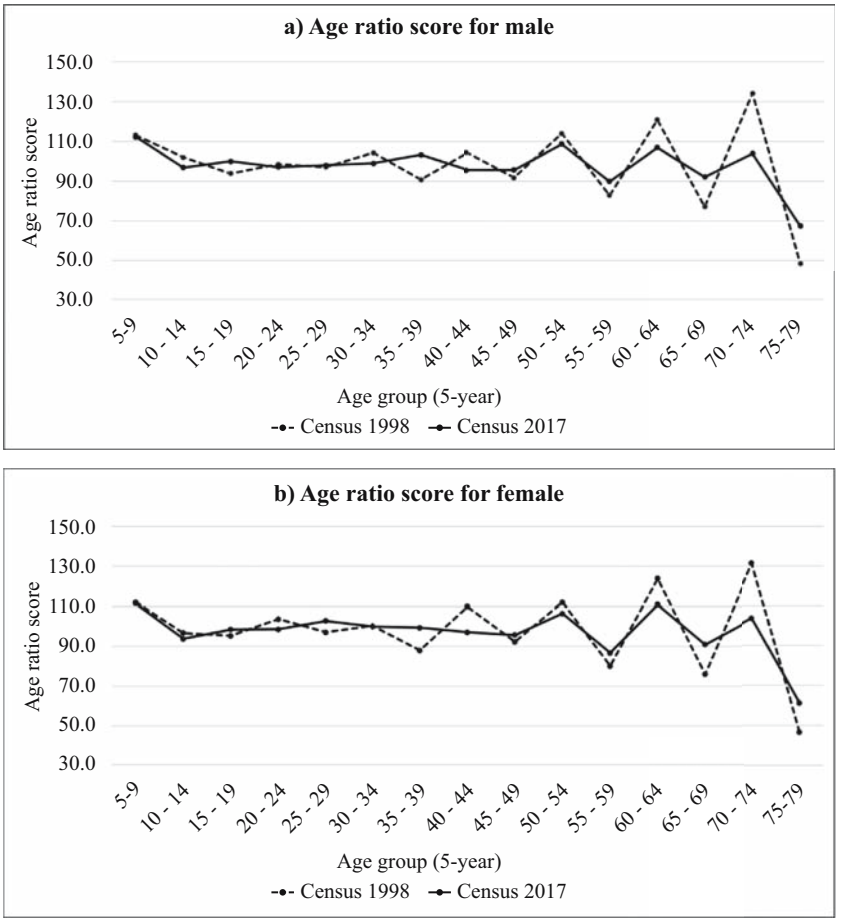

Fig. 4. Age ratio score by age and sex in Pakistan, 1998 and 2017 population censuses.
Source: Authors' calculations based on age-sex data (Pakistan Bureau of Statistics 2022: Minnesota Population Center 2023).

Table 1. Age-sex accuracy index, four provinces, Islamabad, and Pakistan, 2017.

	Index
Balochistan	31.9
KP	25.9
Punjab	22.2
Sindh	25.4
Islamabad	29.5
Pakistan	22.0

Source: Authors' calculations based on 2017 population census data (Pakistan Bureau of Statistics 2022).

males, the age ratio score for females, and three times the sex ratio score. The accuracy scores are divided into three ranges to assess the magnitude of the reporting problem: (1) when the accuracy index is < 20, the census estimates are accurate; (2) between 20 and 40 the estimates are inaccurate, and (3) the estimates are highly inaccurate when the score is > 40. To compute the age-sex accuracy index, we used the "ageSexAccuracy" function from the R package "DemoTools" (Riffe et al. 2019) and applied it to the national level as well as to the four provinces and Islamabad (Table 1).

The age distribution of the 2017 population census at the national level falls under the inaccurate (>20) categories. However, there is a large variation observed at the sub-national level. All provinces were also categorized as inaccurate, the lowest value of the index was found in Punjab (22.2) and the highest in Balochistan (31.9). The analysis shows that the distribution by age and sex resulting from the 2017 census was somehow distorted at the sub-national level.

5. Comparison of the Age-Sex Distribution According to the 2017 Census and the Reconstruction

In 2021, we used demographic analysis to estimate the population size by age and sex based on inter-census (1998–2017) estimates of births, deaths, and internal and international migration (Wazir and Goujon 2021). Here, we compare the reconstructed 2017 population with the official results of the 2017 census population by age and sex using a goodness-of-fit-statistical test computed as the absolute difference (official minus reconstructed population) and the average absolute deviation by age and sex (see Figure 5) along with the 95% confidence interval (CI). An overcount refers to the census population being higher than the reconstructed population, while an undercount refers to the opposite situation.

While our analysis of the total population showed little difference at the national level, we find that there are quite substantial differences by age and sex. We find an overcount of children under age 15 in the census and an undercount for all other age groups (from 15-19 to 75+). There were ten million (95% CI from 3.0 to 15.5) more children under the age of 15 in the census compared to the reconstruction at the national level, with 5.5 million (95% CI from 4.4 to 10.4) boys and 4.5 million (95% CI from 3.9 to 9.3) girls.

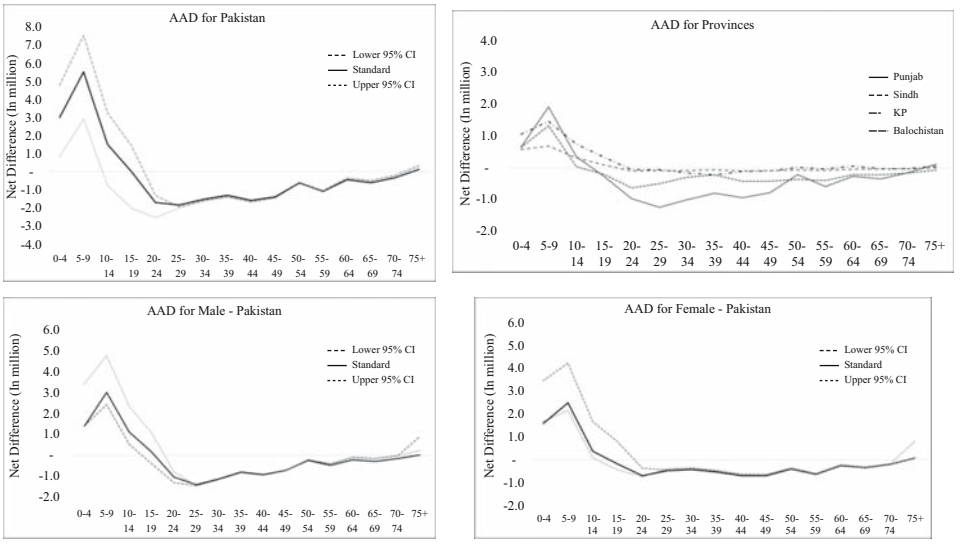

Fig. 5. Average absolute deviation (AAD) by age and sex at the national and provincial level.
Source: Authors' calculations based on 2017 population census data (Pakistan Bureau of Statistics 2022; Wazir and Goujon 2021).

At the sub-national level, the overcount in the 0–14 age group was spread as such: 3 million in Punjab (95% CI from −0.1 to 5.1) and 2.1 million in Sindh (95% CI from −0.4 to 3.4). The overcount was substantial in KP with 3.3 million (95% CI from 1.8 to 4.7) and in Balochistan with 1.6 million (95% CI from 0.9 to 2.2), which are relatively small regions. It is worth noting that the difference was particularly acute for the 5-9 age group compared to the 0–4 and 10–14 ones (see Figure 5 and Table 2).

Table 2. Average absolute deviation by age at the national and provincial level (in millions).

	Punjab	Sindh	KP	Balochistan	Pakistan
0–4	+0.7	+0.7	+1.1	+0.6	+3.0
5–9	+1.9	+1.3	+1.5	+0.7	+5.5
10–14	+0.4	+0.1	+0.8	+0.3	+1.5
15–19	−0.3	−0.2	+0.3	+0.1	+0.0
20–24	−1.0	−0.6	+0.0	−0.1	−1.7
25–29	−1.2	−0.5	+0.0	−0.1	−1.8
30–34	−1.0	−0.3	−0.1	−0.1	−1.5
35–39	−0.8	−0.2	−0.2	+0.0	−1.3
40–44	−0.9	−0.4	−0.1	+0.1	−1.6
45–49	−0.8	−0.4	−0.1	+0.1	−1.4
50–54	−0.2	−0.4	+0.0	−0.1	−0.6
55	−0.6	−0.4	+0.0	−0.1	−1.1
60–64	−0.3	−0.2	+0.1	+0.0	−0.4
65–69	−0.3	−0.2	+0.0	+0.0	−0.6
70–74	−0.1	−0.1	+0.0	+0.0	−0.3
75+	+0.1	−0.1	+0.1	+0.0	+0.1

Source: Authors' calculations based on 2017 population census data (Pakistan Bureau of Statistics 2022; Wazir and Goujon 2021).

The difference in the age structure between the census and the reconstruction can have several causes. Besides the deficiencies of the census as noted in the previous section, it is possible that the fertility and mortality levels used in the reconstruction and originating from the existing demographic surveys underestimate the number of births and deaths, especially at the sub-national level.

6. Comparison of the Sex Ratio According to the 2017 Census and the Reconstruction

Figure 6 shows the sex ratio (number of men to 100 women) for all ages according to three sources: the 2017 population census, model life tables, and the reconstruction. The sex ratio computed from the model life table East Model life table (Coale et al. 1983), is based on the life expectancies at birth for provinces and national level (as calculated by Wazir and Goujon 2021). The model life tables show that sex ratios should be declining with age, which was not the case in the reconstruction. Except for KP, the sex ratios for children up to age 15 from the reconstruction were close to the value computed from model life tables, while the ratios calculated from the 2017 census show an irregular pattern, for all ages. The sex ratios from the reconstruction between ages 15 and 50 were smooth and matched those from the model life tables, except for Sindh and Balochistan. In these two provinces, females were likely to be underreported, leading to systematic excess of males in the younger cohorts.

The sex ratios for older ages (age 50 and above) estimated from the census and the reconstruction were contradictory to those using model life tables for all provinces and the national level. All provinces and national levels experienced a significantly high level of sex ratios – above 100 – for older ages and remained persistent. The plausibility of these distortions in older ages is difficult to explain, whether they were due to higher mortality of women, systematic under-reporting of older cohorts, or other omission issues. Those

Fig. 6. Sex ratio by age in Pakistan from 2017 census, model life table, and reconstructed population.
Source: Authors' computation based on 2017 population census data (Pakistan Bureau of Statistics 2022; Coale et al. 1983; Wazir and Goujon 2021).

results tend to point in the direction of data issues in the 2017 census as well as in the mortality and internal migration indicators estimated and used in the reconstruction.

Finally, the inconsistencies in the pattern of sex ratios in adulthood in the 2017 census suggest that this was a real structural anomaly, potentially representing reporting inaccuracies. Therefore, an in-depth province-specific investigation would be necessary to disentangle these factors.

7. Conclusion

Our results reveal that the age and sex structure of the population as estimated by the 2017 census suffered from some irregularities, as is the case in many low-income countries, both at the national and sub-national levels. These irregularities partly explained the observed differences between the census and the reconstruction. Similarly, the validity of the reconstruction is limited by the data estimates, in terms of fertility, mortality, and internal migration, which become more visible when looking at the age structure. At the sub-national level, the overall irregularities were amplified by issues related mostly to the difficulties in estimating internal migration. This was particularly visible in the small provinces of Balochistan and KP.

The analysis by age and sex reinforces the main conclusion of our article that the government and the international community should ensure that the next census in Pakistan increase in quality with an inbuild post-enumeration survey along with post-census demographic analysis. To resolve the contention about the provincial population size, the census should use both dejure and de facto population count within a short enumeration period (ideally three to five days) and collect the information on the previous place of residence and duration of stay at the current residence (status of internal migration).

8. References

Bureau of Emigration and Overseas Employment. 2020. Labor Migration Report. Islamabad: Bureau of Emigration and Overseas Employment. Available at: https://beoe.gov.pk/reports-and-statistics.

Coale, A., P. Demeny, and B. Vaughan. 1983. Regional Model Life Tables and Stable Populations (Vol. 2nd Edition). New York 10003, NY, USA: Academic Press Inc.

Gerland, P. 2014. UN *Population Division's Methodology in Preparing Base Population for Projections: A case study for India* Asian Population Studies: 10(3): 274–303. DOI: https://doi.org/10.1080/17441730.2014.947059.

Minnesota Population Center 2023. *Integrated Public Use of Microdata* Series (IPUMS), International: Version 6.5 [data set]. Minneapolis, MN: University of Minnesota.

Myers, R.J. 1954. "Accuracy of Age Reporting in the 1950 United States Census." *Journal of the American Statistical Association* 49(268): 826–831. DOI: https://doi.org/10.2307/2281542.

Pakistan Bureau of Statistics. 2022. *Final Results of 6th Population and Housing Census*, Islamabad: Pakistan Bureau of Statistics, Ministry of Statistics, Islamabad, Islamic

Republic of Pakistan. Available at: https://www.pbs.gov.pk/content/final-results-census-2017-0 (accessed January 2023).

Riffe, T., J.M. Aburto, M. Alexander, S. Fennell, l. Kashnitsky, M. Pascariu, and P. Gerland. 2019. *DemoTools: An R Package of Tools for Aggregate Demographic Analysis* (version 01.13.76). Available at: https://github.com/timriffe/DemoTools/ (accessed January 2023).

Siegel, J., and D. Swanson. 2004. *The Methods and Material of Demography*. San Diego, California, USA: Elsevier Academic Press. Available at: https://books.emeraldinsight.-com/page/detail/the-methods-and-materials-of-demography/?k=9780126419559.

U.S. Census Bureau. 2017. *Population Analysis System (PAS) Software*. Available at: https://www.census.gov/data/software/pas.html (accessed January 2023).

Wazir, A., and A. Goujon. 2021. "Exploratory Assessment of the Census of Pakistan Using Demographic Analysis." *Journal of Official Statistics* 37(3): 719–750. DOI: https://doi.org/10.2478/jos-2021-0032.

Received July 2022
Revised November 2022
Accepted February 2023