The National Survey on Population and Family Health (NSPFH), Morocco-2018: a Data Quality Assessment

Abdelghani Asraoui1*, Chems Eddouha Khassouani2 and Abdelmajid Soulaymani3

1 3 Center for Doctoral Studies, Ibn Tobail University, Kénitra-Morocco
2 Provincial Hospital Center, Ministry of Health, Ifrane-Morocco

Abstract. The National Survey on Population and Family Health (NSPFH) is an important source of data in Morocco. Its objective is to assess, periodically, the state of population health and the impact of policies and programs put in place, by updating the main demographic and health indicators. The report of the sixth NSPFH, conducted during the period October 2017-January 2018, illustrates updated socio-demographic data and the new prevalence of several diseases. As the NSPFH results are a reference for decision-makers, researchers and professionals, it seems necessary to promote them and give them more credibility. The objective of this paper is to verify the reliability of the NSPFH data and results using demographic data quality assessment methods (graphic methods: age-gender pyramid and distribution by gender and area of residency, and statistical methods: non-response rate and age accuracy indexes) to ensure if they are of good quality. This study showed that the NSPFH-2018 data were of good quality. Indeed, the non-response rate did not exceed 1.1% for all questionnaires (household 1.1%, woman 0.5% and elderly 1.1%). The age-gender pyramid confirmed the demographic transition towards ageing and the downward trend of fertility in Morocco. The distribution by area of residency confirmed the trend towards urbanization of the country (61% urban and 39% rural). The Whipple (1.05), Myers (4.73), Bachi (2.31) indexes and the United Nations Combined Index (31.21) were all within the standards for a good quality of age declaration. In addition, the results were consistent with each other and in line with the national and international health context.

Keywords: Assessment, Data Quality, NSPFH Morocco-2018, Age declaration, Consistency.

1 Introduction

For more than thirty years, the Ministry of Health in Morocco has been carrying out demographic and health surveys, according to an internationally approved standard methodology. These large-scale surveys are one of the main sources of population and health data at the national level and a major component of the national health information system [1]. The National Survey on Population and Family Health (NSPFH-2018) consisted of the sixth edition of surveys of this type. It aims to assess the population health status and the impact of policies and programs put in place, by updating the main demographic and health indicators. It also aims to make available, to the Government, civil society and development partners in the country, recent statistics for a better appreciation of the extent of health problems and an assessment of progress in implementing Morocco’s international commitments, particularly in the context of the Sustainable Development Goals (SDGs) [2].

Thus, the NSPFH is an indispensable source for professionals at all levels. It is also the best tool around which the debate on the particular population needs can be engaged. Following our review of literature on the subject, we found that the data quality of this survey was never assessed. Furthermore, just two studies included data from Morocco. The first was conducted in 2007 by Spoorenberg where he applied the modified Whipple Index to the 2004 Morocco census data [3]. In the second, Fajardo-Gozález et al. evaluated, in 2014, the quality of age reporting in the census of several countries including Morocco [4]. However, the data from the NSPFH, like those from any collection operation, cannot be free of errors. Therefore, it was interesting to carry out an analysis to provide qualitative and quantitative indicators of the confidence that can be given to the results from the NSPFH.

In this paper, we assessed the data from the NSPFH-2018 to ensure if they are of good quality. To do this, we have applied demographic methods of data quality assessment and analyzed the consistency of the results issued from the survey. The objective was also to give more credibility to the results of the NSPFH-2018. Furthermore, it was necessary to verify the reliability of the data from this crucial and indispensable source of health information on which decision-makers are based when developing health strategies, policies, plans and programs.
2 Methods

2.1 Data Sources

The data used in this paper are from the results and tables presented in the final reports of the SPFH-2003-04, NSPFH-2011 and NSPFH-2018 available and downloadable on the website of the Ministry of Health [5], DHS Program [6] and population and household projections conducted by the High Commissioner for Planning (HCP) [7].

2.2 Data Quality Assessment

To assess the data quality of the NSPFH-2018, we analyzed: 1) the internal consistency of the NSPFH-2018 results. It involved studying the data quality of the survey and its results using some graphical methods (age-gender pyramid and distribution of the population by age, gender and area of residency) and statistical methods (non-response rate and age accuracy indexes: Whipple, Myers, Bachi and United Nations Combined Index). In addition, we tried to identify several relationships and concordances between the NSPFH-2018 results; 2) the external consistency of the survey’s results. To ensure that the results from the NSPFH-2018 were aligned with expected trends and health policies, plans and programs, we compared them with those of previous surveys, including the SPFH-2003-04, NSPFH-2011 and the demographic projections made by the HCP.

2.2.1 Non-response Rate

To assess the data quality, before testing any other method (graphical or statistical), the non-response rate is the most commonly used indicator, given the considerable bias that can result from non-response in the analysis of the data. If the non-response rate is less than 10%, it is generally considered acceptable [8].

2.2.2 Age-gender Pyramid

According to the National Institute of Statistics and Economic Studies (INSEE) definition, the age-gender pyramid represents the breakdown of the population by gender and age at a given point in time. It consists of two histograms, one for each gender (by convention, men on the left and women on the right) where the numbers are shown horizontally and the ages vertically. The numbers by gender and by age depend on interactions between fertility, mortality and migrations. The shape of the pyramid and its variations over the years depend above all on the variations in fertility [9].

2.2.3 Age Reporting Accuracy Indexes

The errors in the reporting of age have probably been examined more intensively than the reporting errors for any other question in the census/survey [10]. The quality of the distributions by age and gender was estimated by the most commonly used indexes [11]: Whipple [12, 13], Myers [14], Bachi [15] and the United Nations Combined Index [11, 12].

2.2.3.1 Whipple Index (W)

This index is the simplest one. It aims to measure the attraction or repulsion of ages ending in 0 or 5. Its construction is based on the hypothesis of linear evolution of the numbers between the five-year age groups, the young ages (0-22 years) and the older ages (63 and over) are excluded because, in these age ranges, the linearity hypothesis is implausible. The advantage of this index is its simplicity. Its disadvantage is to measure only the preference to numbers 0 and 5. If W<1.05: the data are very accurate; if 1.05 ≤ W ≤1.099: the data are relatively accurate; if 1.10 ≤ W ≤1.249: the data are approximate; if 1.25 ≤ W ≤1.749: the data are rough; if 1.75 ≤ W: the data are very rough [11, 12, 13, 16].

2.2.3.2 Myers Index (M)

Myers Index (M) measures the attraction or repulsion of ages ending in digits between 0 and 9, not just those ending in 0 or 5. It is not possible to directly compare the successive total numbers of individuals reporting ages ending each of the digits from 0 to 9 because of the normal decline in numbers with age. Therefore, Myers proposed to calculate, for each of these digits, an “adjusted number” which, if there were no attraction or repulsion would be equal to 10% of the total adjusted number. M is the sum of the absolute differences in the percentages of each adjusted number with the theoretical 10. If the age declarations are correct, all the adjusted numbers are almost equal and M is around zero. The higher the value of M, the greater the preferences or repulsions for ages ending in particular digits. Its maximum value is reached when there is a preference for all ages ending in a single number and is then worth 180. This index also gives information on the attraction of round ages (ending in 0 or 5). One of the disadvantages of the Myers index is that it is impossible to define precisely the theoretical conditions under which it takes, respectively, values 1 and 0 [11, 12, 13, 17].

2.2.3.3 Bachi Index (B)

One of the disadvantages of the Whipple and Myers indexes is the impossibility of defining precisely the theoretical conditions under which they take, respectively, values 1 and 0. To cope with this, Bachi developed an index that does not present this disadvantage. Its calculation does not consider young and older ages. Based on the 23-72 age group, Bachi found that for populations with a well-declared age, the ratio of 23-72 age group by the total 72 years varies roughly linearly according to the number of units from 0 to 9. Moreover, he showed that the slope of this line does not vary much if we slightly change the age limits. The Bachi index can vary between 0 (no preference or
2.2.3.4 United Nations Combined Index (UNCI)

The UNCI is different from the three indexes described above. Indeed, it is calculated based on the distributions by age groups and no longer by years of age, and it attempts to measure the regularity of the distributions by gender and age. Compared to the methods of Whipple, Myers and Bachi, this method has the advantage that the calculated index reflects changes in the number of omissions according to age groups, intentionally inaccurate age declarations and preferences for ages ending in a given number of units. Therefore, this index better reflects the overall accuracy of the statistics by age. However, when the population size is high, the age distribution is, largely, random and the value of the index is affected. If UNCI < 20: the data are of good quality; if 20 ≤ UNCI <40: the data are of relatively good quality and can be adjusted; if UNCI ≥ 40, the data are of very poor quality [11, 12, 17].

3 Results and Discussions

3.1 Internal consistency

3.1.1 Non-response rate

The NSPFH-2018 is a large-scale survey in which 15,022 households were successfully surveyed (Table 1.), with a non-response rate of 1.1% (1.6% in urban and 0.3% in rural area). The number of women aged 15 to 49 who took part in the survey was 9969. The non-response rate for women was around 0.5% (0.7% in urban and 0.2% in rural area). Similarly, of the 3,575 people aged 60 and above, 3,534 were successfully surveyed, representing a non-response rate of 1.1% (1.5% in urban and 0.7% in rural area).

| Area of residency | Total |
|-------------------|-------|
| Urban             | Rural |       |
| Selected households| 9000  | 6293  | 15293 |
| Surveyed households| 8930  | 6254  | 15184 |
| Household non-response rate | 1.6% | 0.3% | 1.1% |
| Individual survey (Women) |      |       |       |
| Eligible women     | 5568  | 4450  | 10018 |
| Eligible women surveyed | 5528  | 4441  | 9969  |
| Women non-response rate | 0.7% | 0.2% | 0.5% |
| Individual survey (Elderly) |      |       |       |
| Eligible elderly people | 1968  | 1607  | 3575  |
| Eligible elderly people surveyed | 1939  | 1595  | 3534  |
| Elderly non-response rate | 1.5% | 0.7% | 1.1% |

Source: NSPFH-2018, Final report

The non-response rate was well below 10% for the three NSPFH-2018 questionnaires (household 1.1%, women 0.5% and elderly 1.1%). It confirms that the interviews were carried out on almost all statistical units selected. It is an indicator of good data quality [8]. The small value of the non-response rate is explained by the benefit of interviewer-assisted survey methods, where the interview is personalized, questions and concepts can be interpreted, and the interviewer can increase the response rate and data quality overall [18]. However, the response rate is only an indicator of the possibility of non-response bias. Even when the response rate is high, there may be a non-response bias. Therefore, a non-response bias analysis is always recommended [19].

3.1.2 Population Structure according to gender, age and area of residency

The results of the NSPFH-2018 showed that the gender distribution of the population presented equality between men and women. Indeed, of the 67,795 individuals surveyed, 33,926 were women (50%) and 33,869 were men (50%). The sex ratio was almost equal to 100 men per 100 women. Similarly, the sex ratio in urban area was 99 men per 100 women (20664/20871=99.0), while it was just over 101 men per 100 women in rural area (13206/13054=101.2). For the age group 0-4 years (Fig. 1.), the sex ratio was 108 boys per 100 girls. It decreased with age, as the mortality rate of boys is generally higher than that of girls. From the age of 20, women were in majority. Since the age of 50, the sex ratio represented some irregularities. These variations are observed, not just in Morocco, but also in several African countries [17]. Furthermore, it should be noted that, generally speaking, the sex ratio decreases at an advanced age, given the longer life expectancy of women. It is, therefore, largely dependent on the age structure of the population [20].
The analysis of the age-gender pyramid of the surveyed population (Fig. 2.) revealed that the general trends of the Moroccan population structure (including fertility, mortality and ageing) were reflected in the NSPFH-2018. Indeed, a demographic transition towards ageing, which is gradually taking place, has been observed. The shape of the age-gender pyramid also showed that children were slightly more numerous than girls in the 0-4 and 5-9 age groups. A slight shrinkage in the 0-4 age group was also observed. According to Hardy (2002), it is explained by the Fertility rates declining [21]. The results showed that there were 2.38 births for every woman aged 15 to 45 in 2018 compared to 2.6 in 2011 [22]. Overall, the population pyramid showed a symmetry between men and women. However, an increase in numbers of women aged 50-54 at the expense of the younger age group was noted. It was an anomaly, which can be attributed to investigators who, to avoid extra work during the individual survey, tend to transfer, during the household survey, some women of the 45-49 age group to the 50-54 age group. Because after the age of 50, women were no longer eligible for the individual survey [23]. Then, it was an observation error. It came from the risks associated with the observation itself. It depends on the rate of coaching, the quality and training of the interviewers, the clarity of the questionnaire, etc [24]. It can also reflect the accuracy of age declarations among women, particularly in a population where 39.5% [25] of women have no educational level [2] and a good number of individuals are not registered at Civil status [23, 25]. In addition, the decline in malnutrition rates may be a potential explanation, as infant protein malnutrition syndrome was and is (in the poorest economies) a limiting factor in an adult’s cognitive abilities (which may lead to reporting errors by age) [26].

In addition, the results showed that 41,535 lived in urban area (61.3%) and 26,260 in rural area (38.7%) (Fig. 3.). The trend towards urbanization of the Moroccan population was reflected by the high proportion of the urban population at the expense of that of the rural area [27, 28].

### 3.1.3 Age Reporting Quality

The results (Table 2.) showed that the Whipple index was equal to 1.05 (1.04 for men and 1.05 for women). We concluded that, when reporting age, there was no attraction or repulsion for the particular ages ending in 0 or 5. The W value was almost in the same order as 1.04 calculated for the 2016 Iran census by Sadoghi et al. in 2020 [29]. On the other hand, W was less than 1.13 calculated, based on data from the 2004 Morocco census, by Fajardo-Gozález and collaborators in 2014 [4]. The results also showed that the Myers Index value was 4.73 (4.25 for men and 5.20 for women). This value was much lower than the maximum value of 180 (where all ages end in the same number). Thus, the Myers index led to the same results as Whipple’s and generalized it for all ages (ending in 0.1, 2, …9). By comparison, the Myers index was below 4.8 and 4.99 calculated, respectively, by Sadoghi et al. in 2020 [29] and Fajardo-Gozález et al. in 2014 [4]. As for the Bachi index, the calculations resulted in a value of 2.31 (2.20 for men and 2.41 for women). This value was close to 0. Therefore, B confirmed the previous results.

In addition, the UNCI value was 31.21 (between 20 and 40). Therefore, the age and sex structure from the
NPSFH-2018 was of relatively good quality and can be adjusted. It confirmed the irregularities in the age-gender pyramid and the sex ratio observed previously which were due to the observation errors. Similarly, it should be noted that the UNCI does not examine the problem of age attraction as the previous indexes do. In addition, the age group affects this method, as it applies to data classified by five-year or ten-year age groups [10, 11, 12]. Therefore, it is advised to limit the calculation of age group ratios and sex ratios to age groups up to 70 years. Above this age, the series shows significant variations [17].

Based on the 2004 Morocco census, Fajardo-Gozález et al. found that UNCI was of 24.92 [4].

**Table 2. Quality of Age Reporting, NPSFH-2018**

| Index | Men | Women | Total |
|-------|-----|-------|-------|
| W     | 1.04| 1.05  | 1.05  |
| M     | 4.25| 5.20  | 4.73  |
| B     | 2.20| 2.41  | 2.31  |
| UNCI  | 31.21| 31.21| 31.21 |

Source: Authors’ work on NPSFH-2018 data

### 3.1.4 Consistency of Results

The results of the NPSFH-2018 presented several concordances. Indeed, the low proportion of pregnancies among women aged 15-49 (7.1%) is due to the high use of contraception (70.8%), the unmet need for family planning (11.3%) and the high average age at first marriage (31.9 years for men and 25.5 years for women). It explains the downward trend in Fertility rate [30]. It declined by 1.2 births per woman between 2011 and 2018, 2.38 and 2.6 children per woman respectively. The decline in maternal mortality (72.6 maternal deaths per 100,000 live births in 2016 compared to 112 in 2010) is justified by the improvement in the proportion of prenatal consultations conducted by qualified personnel (88.5% in 2018 compared to 77.1% in 2011) and the proportion of deliveries assisted by qualified personnel (86.6% in 2018 versus 73.6% in 2011) as well as the high proportion of caesarean delivery (21.2%) [31, 32, 33]. The downward trend in under-5 mortality (22.2 per 1,000 live births in 2018 versus 28.8 in 2011) is also due to an increase in the proportion of prenatal consultations conducted by qualified personnel, the proportion of deliveries attended by qualified personnel, the proportion of children breastfed (97.1% in 2018 versus 96.6% in 2011) and the proportion of children completely vaccinated (90.6% in 2018 versus 87.7% in 2011) [31, 33].

These concordances, and others, reflected the internal consistency of the NPSFH-2018 results and their alignment with expected overall trends.

### 3.2 External Consistency

When we compare the population pyramid of Morocco in 2018 [7] (Fig. 4.) with that of the NPSFH-2018 [2] (Fig. 2.), it revealed that the general trends in the population structure (including fertility, mortality and demographic transition towards ageing) have been reflected in the NPSFH-2018.

Moreover, our study revealed that the NPSFH-2018 results were in line with general trends and the national and international health context. Indeed, the maternal mortality ratio (Fig. 5.) rose from 112 deaths in 2010 to 72.6 maternal deaths per 100,000 live births in 2016, with a 35% reduction at the national level (39% in urban area and 25% in rural area). These developments were clearly in line with Morocco’s significant progress in improving maternal health and its commitment to the international community concerning the 5th Millennium Development Goal (MDG 5) (83 deaths per 100,000 set for 2015) [34].

Similarly, the results confirmed the downward trend in under-five mortality rate, which declined from 47 deaths per 1,000 live births during the 1999-2003 period to 30.5 between 2004 and 2011 to reach 22.2 deaths per 1,000 live births in 2018.

Concerning non-communicable diseases (NCDs), we found that Morocco, like all countries over the world, is registering an increase in the prevalence of several diseases of the NSDS category [35]. For example, in 2018, the proportion of people with at least one chronic disease (including diabetes and hypertension) was 21%, with an increase of 2.8 points compared to 2011 (18.2%). Diabetes affected 4.8% of the Moroccan population, indicating an increase of 1.5 percentage points compared to 2011 (3.3%) and 7.8% of patients declared they do not regularly follow medical treatment. The same results also revealed that 6.8% of the
Moroccan population contracted hypertension, which increased 1.4 points compared to the 2011 survey (5.4%). Just 84.7% declared they regularly follow medical treatment for the disease. Thus, we provided several examples to illustrate and prove the external consistency. Therefore, we conclude that the results of the NSPFH-2018 were in line with expected trends and the national and international health context.

4 Conclusions

In this paper, we assessed the data quality of the NSPFH-2018. We applied several demographic data quality assessment indexes (Whipple, Myers, Bachi and the United Nations Combined Index) to examine the quality of age reporting. We also analyzed the consistency of the results of this survey to ensure that they were consistent with each other and aligned with the national and international context. The results of our analysis revealed that the NSPFH-2018 data were of good quality. The non-response rate was very low and did not exceed 1.1% for all questionnaires. The age-gender pyramid confirmed the demographic transition of the Moroccan population towards ageing and the downward trend in fertility and mortality. The distribution by area of residency confirmed the trend towards urbanization in Morocco. The Whipple, Myers, Bachi and UNCI indexes provided information on good data quality of age reporting. However, irregularities in the population structure by age groups of five years and sex were observed. Moreover, the survey results were consistent and aligned with general trends and the national and international health context.

As a result, the use of the NSPFH-2018 results is recommended. It also seemed interesting to prepare a data quality report for future surveys. We also recommend sharing the survey database with the scientific community to allow researchers to conduct indepth analysis of health problems.

In terms of limits, we note that we used the weighted data represented in the final reports of the surveys studied. However, to assess the quality of age reporting, it is preferable to use the raw data before weighting [10, 11, 12]. In addition, analysis of the response rate for each variable would be more meaningful than analysis of the overall response rate [19].

References

1. M. Azelmat, M. Ayad, A. Abdelmoneim, Ministry of Public Health (Morocco), Macro International Inc, League of Arab States, Enquête sur la Population et la Santé Familiale (EPSF-2003-2004), xvii-xviii (2005)
2. Ministry of Health (Morocco), Enquête Nationale sur La Population et la Santé Familiale (ENPSF-2018), 6-10 (2019)
3. T. Spoorenberg, C. Dureuilh, Population, 62, 847-859 (2007)
4. J. F. González, L. Attanasio, J. Trang Ha, An Assessment of the Age Reporting in the IPUMS-I Microdata, Paper submitted for presentation at the 2014 Annual Meeting of the Population Association of America (2014)
5. Ministry of Health (Morocco) Web site: https://www.sante.gov.ma/Publications/Estudes_enquete/Pages/default.aspx
6. DHS Program Web site: https://dhsprogram.com/data/
7. HCP (Morocco), Projections de la Population et des Ménages, 2014-2050 (2017)
8. M. M. Seastrom, Taux de réponse comme outil de gestion de la qualité des données, Recueil du symposium de statistique au Canada (2001)
9. National Institute of Statistics and Economic Studies (INSEE) Web site: https://www.insee.fr/en/metadonnees/definition/cl230
10. H. S. Shryock, J. S. Siegel, E. A. Larmon, The methods and materials of demography, 2, 115-119 (1975)
11. F. Gendreau, F. Gubry, L. Lohle-Tart, E. V. De Walle, D. Waltisperger, Manuel de Yaoundé : Estimations indirectes en démographie africaine, 189-206 (1985)
12. G. Roger, D. Waltisperger, C. Corbille-Guittton, Roger, Gilles, Dominique Wal, Les structures par sexe et âge en Afrique. Groupe de démographie africaine (1981)
13. J. A. Nasir, A. Hinde, Pakistan Journal of Statistics, 30, 265-272 (2014)
14. R. J. Myers, Journal of the American Statistical Association, 49, 826-831 (1954)
15. R. Bachi, Measurement of the Tendency to Round Off Age Returns. Bulletin de l’Institut International de Statistique (Proceedings of the 28th Session, Rome, Sept. 6-12,1953), 34, 129-138 (1954)
16. A. Noumbissi, Population (french edition), 47, 1038-1041 (1992)
17. F. Gendreau, La population de l’Afrique: Manuel de démographie (1993)
18. Statistics Canada, Survey Methods and Practices, 37-54 (2010)
19. J. R. Halbesleben, M. V. Whitman, Health Services Research, 48, 913-930 (2013)
20. A. Golini, P. Golini, R. Lombardi, Cahiers Québécois de Démographie, 18, 195-210 (1989)
21. I. C. W. Hardy, Sex ratios: Concepts and research methods (2002)
22. Ministry of Health (Morocco), Enquête Nationale sur la Population et la Santé Familiale (ENPSF-2011) (2012)
23. T. W. Pullum, An assessment of age and date reporting in the demographic and health surveys, 1985-2003, Demographic and Health Surveys Project, ORC Macro, Calverton, Maryland, USA, 5 (2006)
24. A. Kamal, Sondage et Tests Statistiques, 4-44 (1998).
25. D. Crayen, J. Baten, Explorations in Economic
26. K. Haandrikman et al., Time & Society, 13, 339-362, (2004)
27. HCP (Morocco), Dynamique urbaine et développement rural au Maroc, 17-63 (2005)
28. R. Escallier, Méditerranée, 81, 107-112 (1995)
29. Z. Sadoghi, A. Teimouri, A. Pezhhan, Journal of Health System Research, 16, 38-45 (2020)
30. K. Machiyama, J. B. Casterline, J. N. Mumah, F. Akhter Huda, F. Obare, G. Odwe, C. W. Kabiru, S. Yeasmin, J. Cleland, Reproductive Health, 14, 1-11 (2017)
31. J. Villar, E. Valladares, D. Wojdyla et al., The Lancet, 367, 1819-1829 (2006)
32. M. Koblinsky et al., Journal of Health, Population, and Nutrition, 26, 280 (2008)
33. C. P. Benziger, G. A. Roth, A. E. Moran, Global Heart, 11, 393-397 (2016)
34. United Nations, Millennium Development Goals (2000)
35. World Health Organization, Noncommunicable Diseases Country Profiles 2018 (2018)