

Session III: Establishing the base population

8 March 2016

Cheryl Sawyer, Lina Bassarsky

Population Estimates and Projections Section

www.unpopulation.org



Regional Workshop on the Production of Population Projections
Addis Ababa, 7-11 March 2016



Establishing the base population

- An accurate base population is essential to ensure accuracy of the projection
- Steps to establishing the base population:
 - Detecting errors in data
 - Correcting distorted or incomplete data
 - Moving the population to mid-year

Evaluation of age and sex distribution data *refresh*



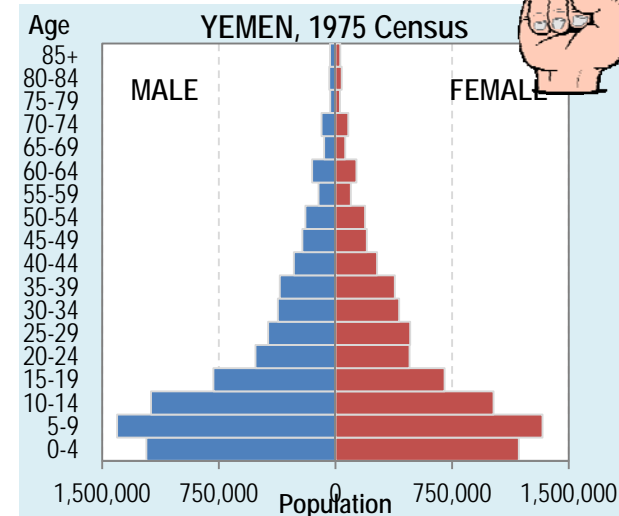
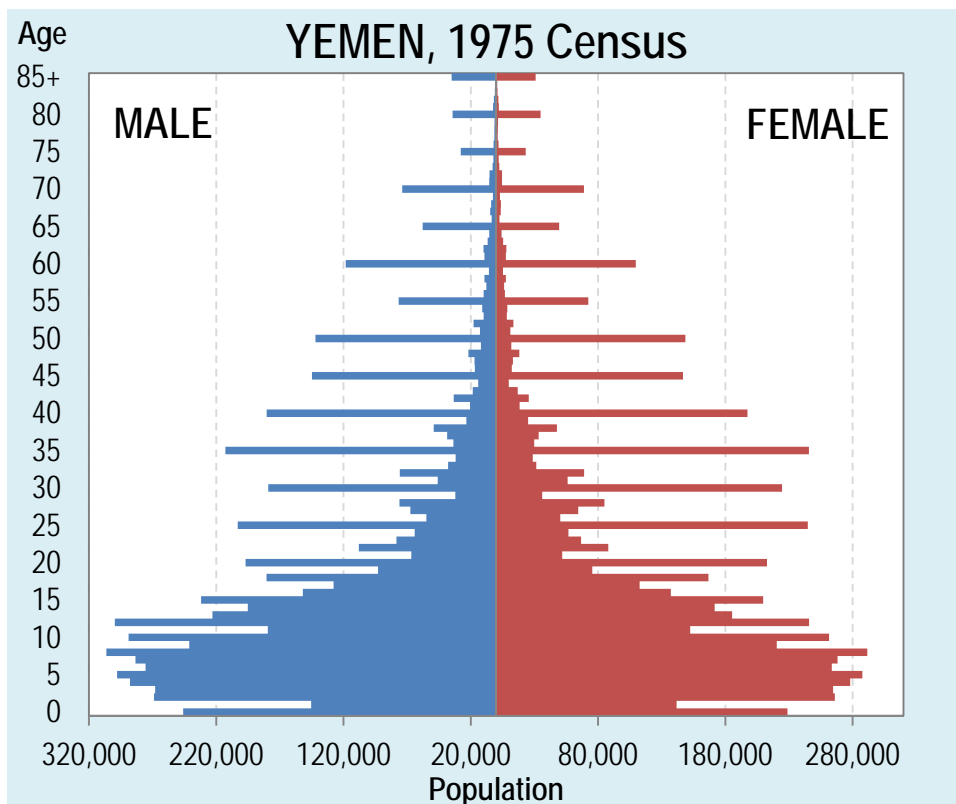
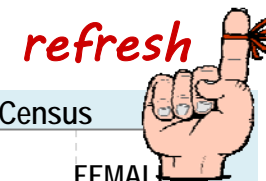
- Graphical analysis
 - Population pyramids
 - Graphical cohort analysis
- Age and sex ratios
- Summary indices of error in age-sex data

What to look for at the evaluation



- Possible data errors, ex.
 - Age misreporting (age heaping and/or age exaggeration)
 - Coverage errors – net underenumeration (by age or sex)
- Significant discrepancies in age-sex structure due to extraordinary events
 - High migration, war, famine, HIV/AIDS epidemic etc.

Population pyramid – Detecting errors



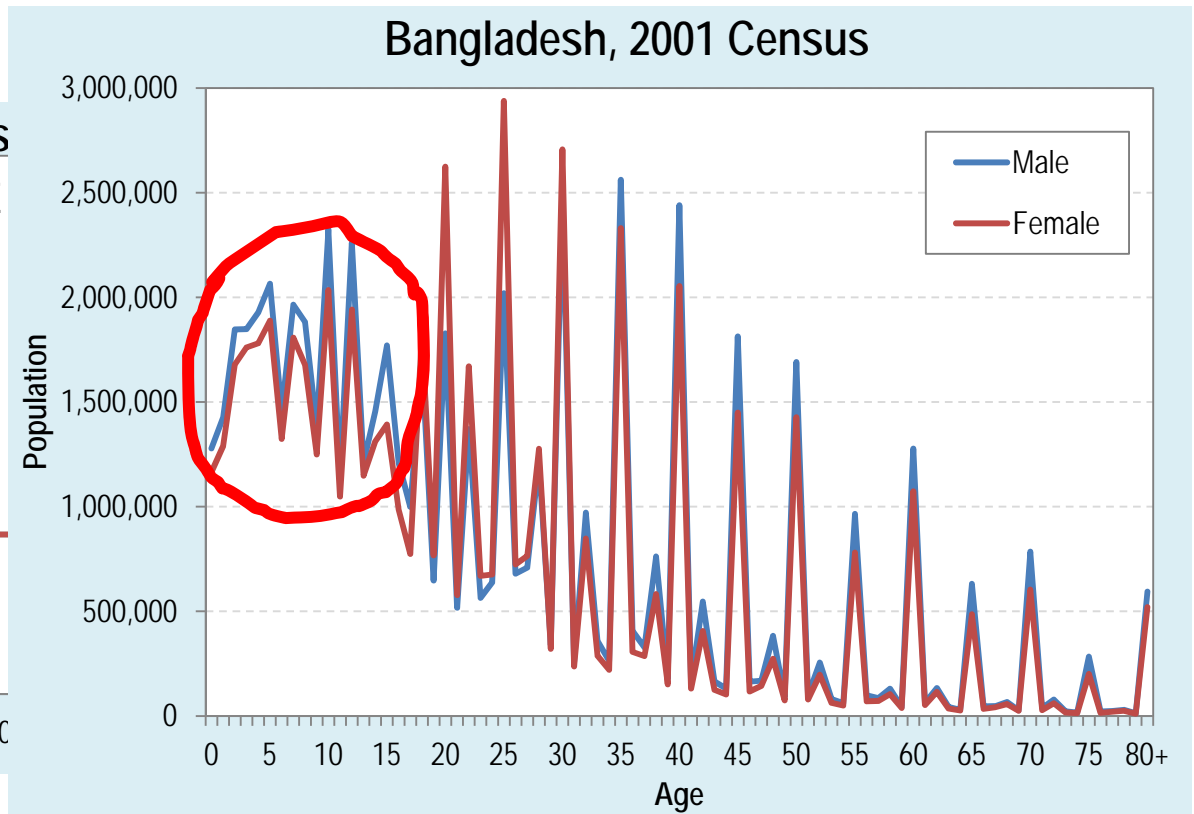
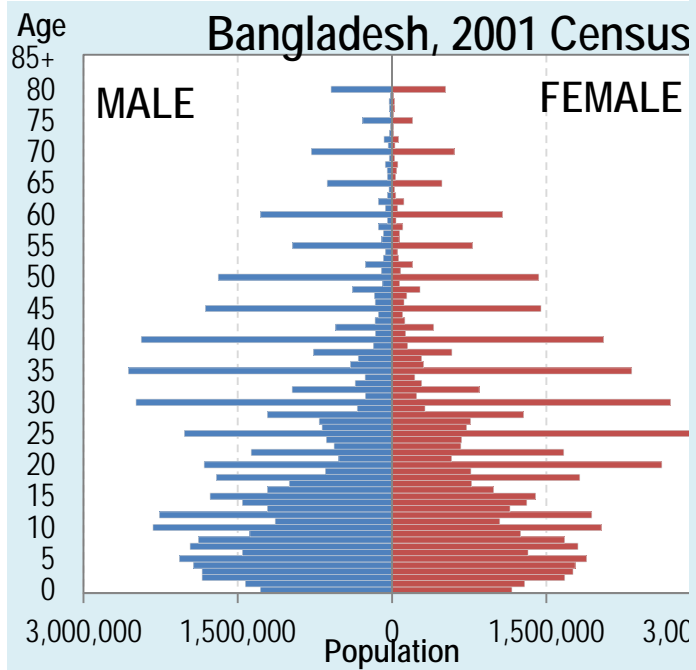
- Age misreporting errors (heaping) among adults
- Under enumeration of young children (< age 2)
- High fertility level
- Smaller population in 20-24 age group
 >> extraordinary events in 1950-55?
- Less men relative to women in 20-44 age group
 >> labor out-migration?

Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Population pyramid – Line instead of bars

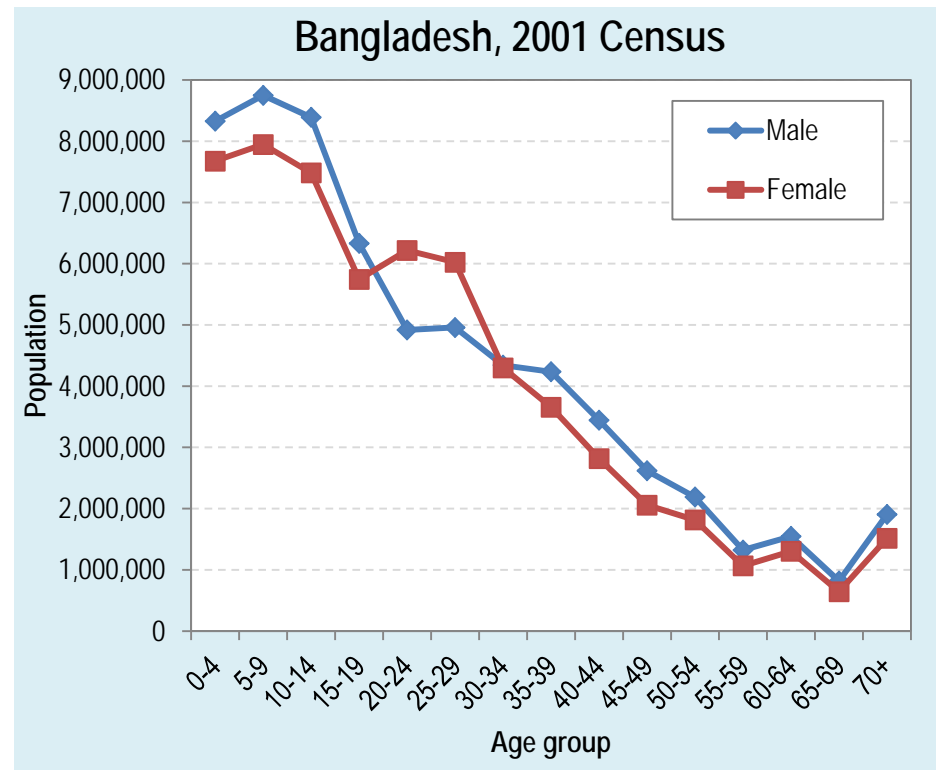
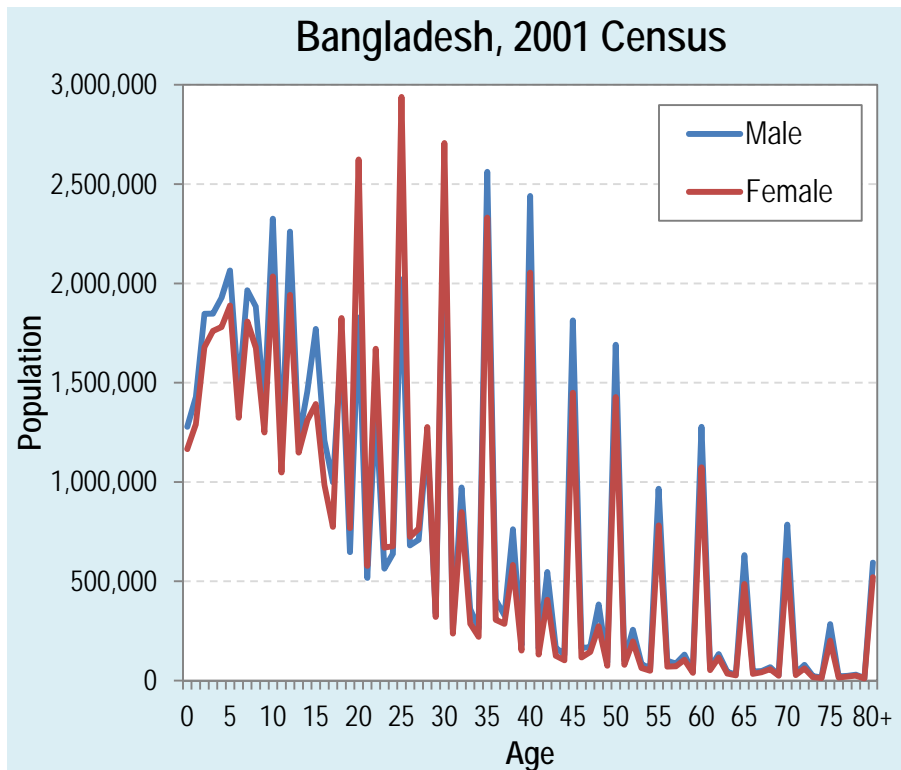


- Population Pyramid (bar chart)
 - >> Not always easy to determine differences by sex
- Use of line chart



Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Population pyramid – Line instead of bars (2)



Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Graphical Cohort Analysis (1)

- Tracking actual cohorts over multiple censuses
- The size of each cohort should decline over each census due to mortality, if no significant international migration
- The age structure (the lines) for censuses should follow the same pattern in the absence of census errors
- An important advantage - possible to evaluate the effects of extraordinary events and other distorting factors by following actual cohorts over time

Graphical Cohort Analysis (2)



Mozambique, 1997 and 2007 Censuses

	1997 Census		2007 Census		Birth Cohort
	Male	Female	Male	Female	
0-4	1,353,206	1,388,350			2002-2007
5-9	1,112,321	1,113,675			1997-2001
10-14	947,236	878,429	1,222,668	1,183,939	1992-1997
15-19	774,327	854,078	925,729	991,323	1987-1992
20-24	637,113	827,614	774,413	986,526	1982-1997
25-29	509,109	654,465	707,603	841,416	1977-1982
30-34	410,148	477,562	583,689	667,865	1972-1977
35-39	373,813	428,395	481,396	556,191	1967-1972
40-44	270,046	303,147	366,518	389,087	1962-1967
45-49	257,070	282,098	321,236	328,660	1957-1962
50-54	178,902	212,060	231,232	283,288	1952-1957
55-59	162,122	174,234	194,011	208,657	1947-1952
60-64	114,335	125,096	140,146	159,557	1942-1947
65-69	100,425	109,288	113,840	127,794	1937-1942
70-74	47,407	50,607	72,288	81,329	1932-1937
75-79	41,529	42,858	55,448	61,012	1927-1932
80-84	15,305	17,326	22,417	28,278	1922-1927
85-89			16,576	19,448	1917-1922
90-94			4,803	5,883	1912-1917

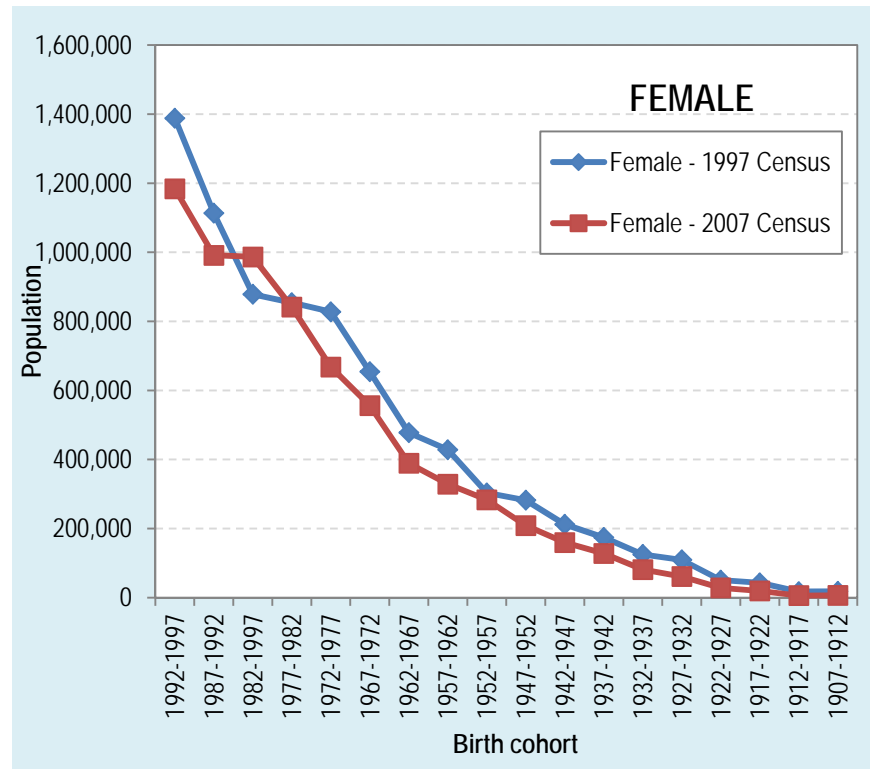
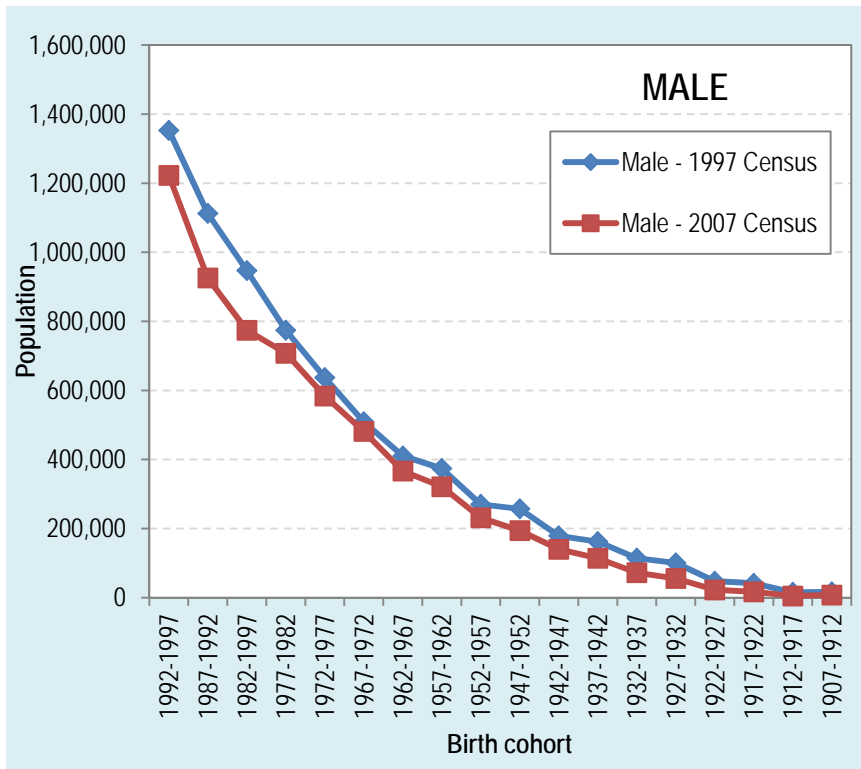
- Data is organized by birth cohort
- Exclude open-ended age category
- People who were born in the same years are compared in the analysis

Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Graphical Cohort Analysis (3)



Mozambique, 1997 and 2007 Censuses



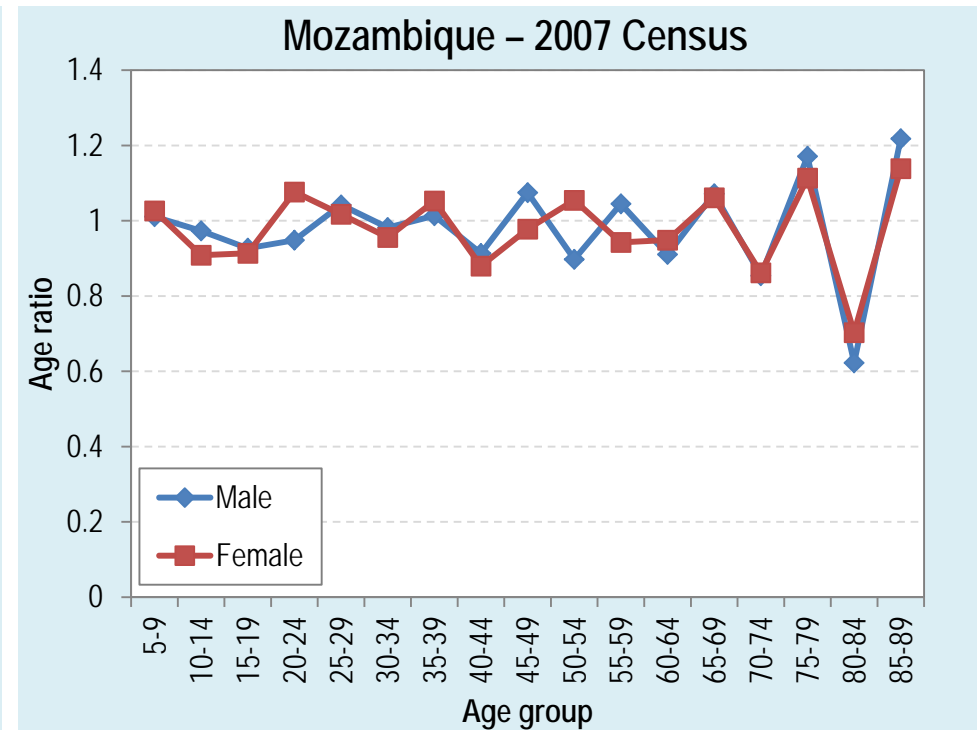
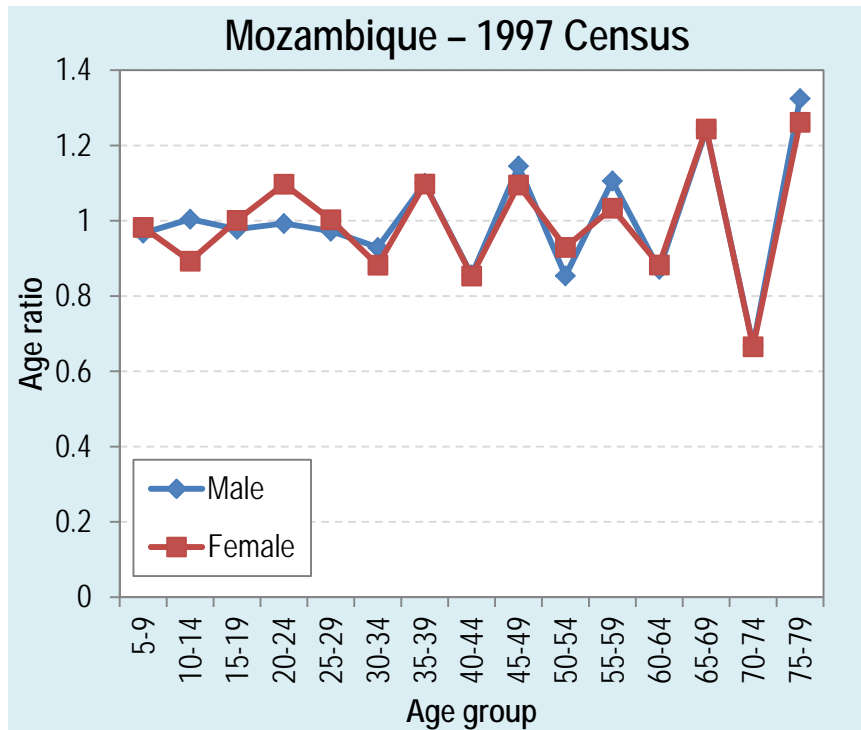
Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Age ratios (1)



- In the absence of sharp changes in fertility or mortality, significant levels of migration or other distorting factors, the enumerated size of a particular cohort should be approximately equal to the average size of the immediately preceding and following cohorts
- The age ratio for a particular cohort to the average of the counts for the adjacent cohorts should be approximately equal to 1 (or 100 if multiplied by a constant of 100)
- Significant departures from this “expected” ratio indicate either the presence of census error in the census enumeration or of other factors

Age ratios (2) – example



Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Sex ratios (1) - calculation

Sex ratio by age group

$$\text{Sex Ratio} = \frac{{}_5P_x^m}{{}_5P_x^f} \quad \text{or} \quad \text{Sex Ratio} = \frac{{}_5P_x^m}{{}_5P_x^f} \cdot 100$$

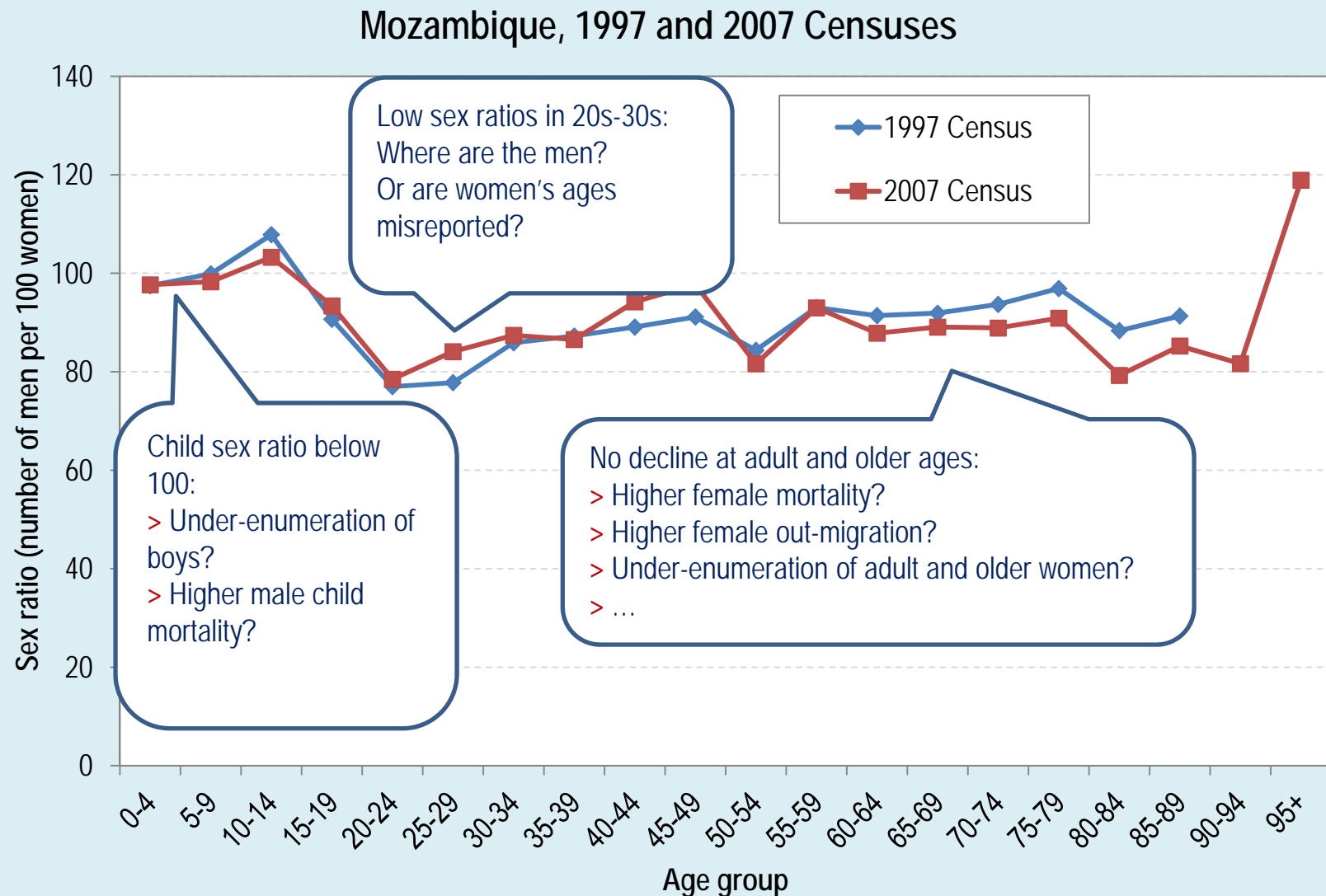
Where

${}_5P_x^m$ = Male population enumerated in a specific age group

${}_5P_x^f$ = Female population enumerated in the same age group

Value of sex ratio	Interpretation
1	Same number of men and women in a given age group
Above 1	More men than women in a given age group
Below 1	Less men than women in a given age group

Sex ratios (2) – Example



Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Tools for evaluation of age/sex data

- PYRAMIDS.xlsx (UNPD)
 - New template by UNPD to produce population pyramids and line graphs by single-year and 5-year age group
- SINGAGE.xls (PASEX)
 - Graphs data by single year of age
 - Calculates indicators of age misreporting (Whipple and Myers' indices)
- AGESEX.xls (PASEX)
 - Calculates age ratios and sex ratios for 5-year age groups
- GRPOP-YB.xls (PASEX)
 - Plots cohorts from 2-3 censuses

Correcting for age misreporting (smoothing)

- It is important to smooth reported age distributions if they are erratic
- PASEX spreadsheet AGESMTH.xls implements 5 methods, of 3 types
- 1. Do not modify the total population - accepting population in each 10-year age group, then divide into 5-year age groups
 - Carrier-Farrag
 - Karup-King-Newton
 - Arriaga's formula (also the first and last group)

Age	Population
20-29	a
30-39	b
40-49	c

$$\text{Pop (35-39)} = f(a, b, c)$$

Correcting for age misreporting (smoothing)

2. Slightly modifying total population - smoothing the 5-year age groups
 - The United Nations Method

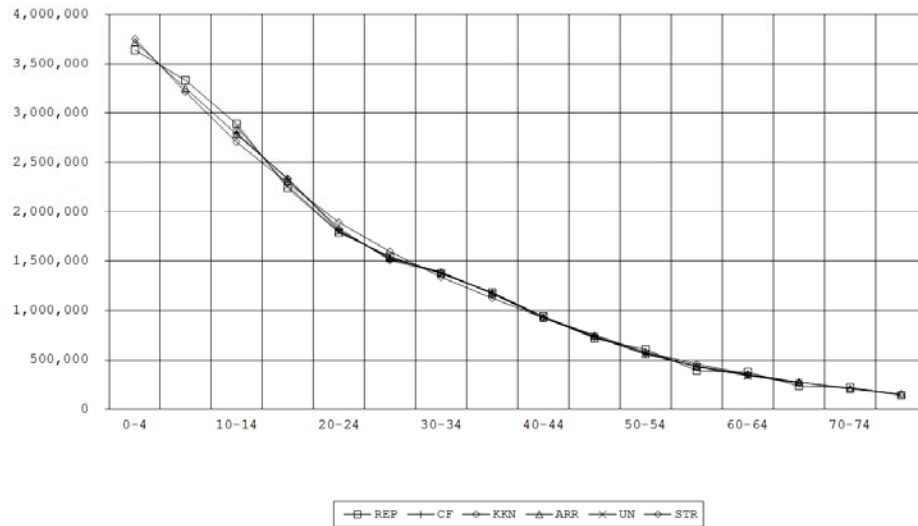
3. Strong smoothing – modifying totals based on consecutive 10-year age groups, then using Arriaga's for the 5-year population

Smoothing Example

PASEX spreadsheet – AGESMTH.xls (using workshop sample data)

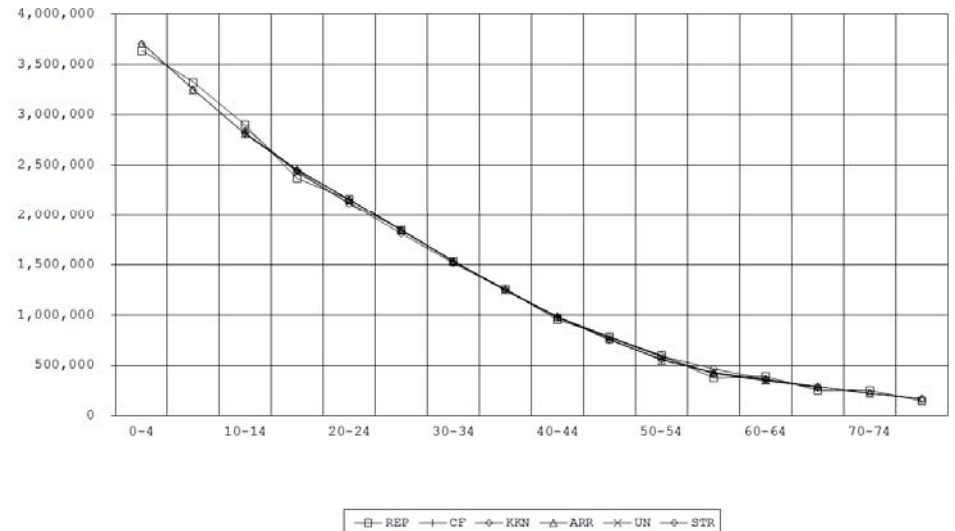
EASTLANDS: 2012

1. Male Population by Age



EASTLANDS: 2012

2. Female Population by Age



Smoothing Example

PASEX spreadsheet – AGESMTH.xls (using workshop sample data)

B. Summary of Indices Measuring the Accuracy of Data

Index	Reported	Smoothed				
		Carrier Farrag	K.-King Newton	Arriaga	United Nations	Strong
Sex ratio score	5.49	3.81	4.16	3.83	4.27	1.79
Male age ratio score	7.35	2.95	3.06	3.12	2.74	1.75
Female age ratio score	6.87	1.96	2.15	1.95	2.32	1.76
Accuracy index	30.68	16.36	17.70	16.57	17.86	8.88

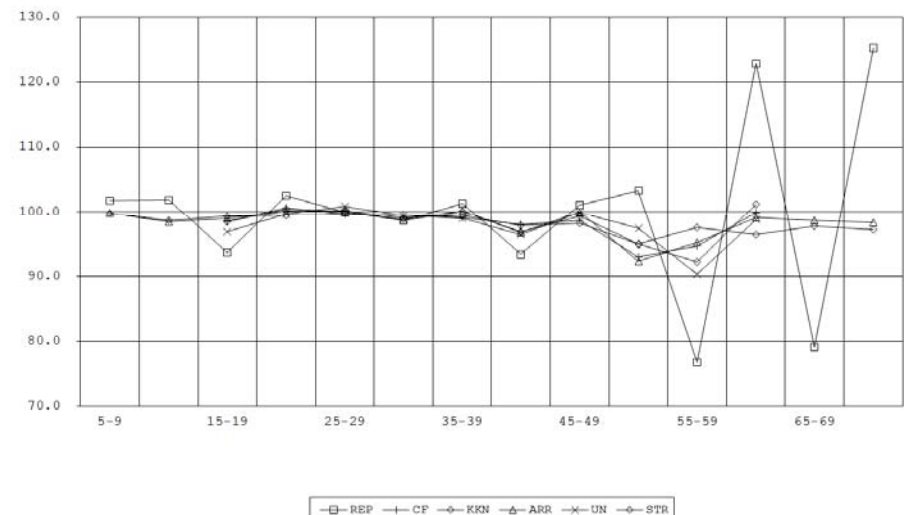
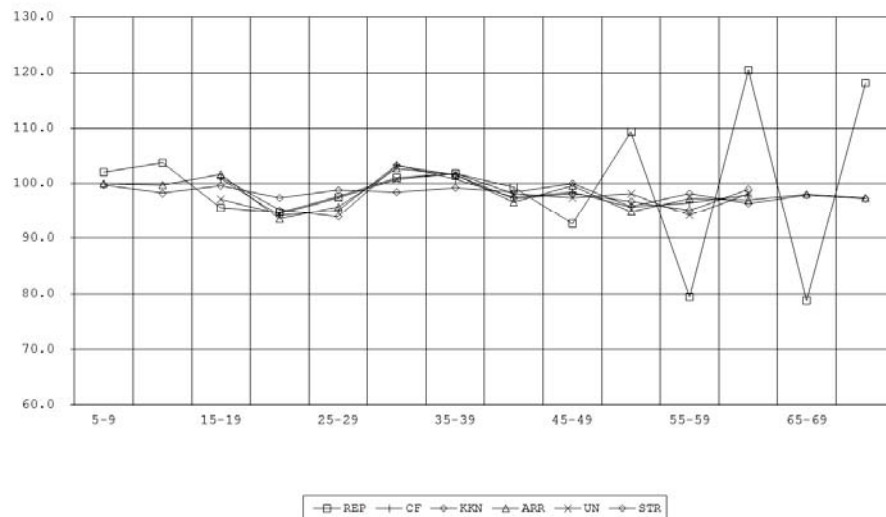
Note: The accuracy index is the sum of the male and female age ratio scores plus three times the sex ratio score, all calculated using data for ages

EASTLANDS: 2012

EASTLANDS: 2012

3. Male Age Ratios by Age

4. Female Age Ratios by Age



Smoothing age patterns – Caution needed



- No generalized solution for all populations
- Methods produce similar results
- Technique used depends on errors in age-sex distribution
- Be cautious in using strong smoothing
- If only part of population distribution problematic, no need for smoothing on entire age distribution

Smoothing with correction for underreporting

- Smoothing methods above do not make any adjustment for underreporting
- Need to adjust the census population for projection input
- PASEX spreadsheet BASEPOP.xls
 - Implements Arriaga smoothing methods for ages 10+
 - Adjusts age groups 0-4 and 5-9 based on recent fertility and mortality
 - Can apply an overall adjustment (based on post-enumeration survey, etc.)

Smoothing with correction for underreporting (BASEPOP)- example

EASTLANDS: 2012

Base Population Estimation

A. Reported Population by Age and Sex, Desired
Population by Sex, and Sex Ratio at Birth

Item and age	Male	Female	Both sexes
Reference date			2012.65
REPORTED POPULATION			
All ages	21,869,990	23,058,933	
Under 1	747,826	751,563	
1-4	2,890,156	2,884,287	
0-4	3,637,982	3,635,850	
5-9	3,333,835	3,325,202	
10-14	2,892,583	2,900,004	
15-19	2,238,976	2,369,860	
20-24	1,793,053	2,160,986	
25-29	1,548,762	1,843,732	
30-34	1,380,068	1,529,610	
35-39	1,182,651	1,258,301	
40-44	944,033	955,081	
45-49	718,372	787,547	
50-54	607,361	603,666	
55-59	392,104	381,736	
60-64	380,223	390,494	
65-69	238,972	253,864	
70-74	226,484	251,346	
75-79	144,643	147,489	
80+	209,888	264,165	
Unknown	0	0	
DESIRED POPULATION			
All ages	22,526,090	23,750,701	46,276,791
Sex ratio at birth			1.0300

Dates converter.xlsx

Calculation of reference dates in Year, decimals

Enter information in the blue cells		
Country / Survey:	Eastlands	2012 Census
Date of survey starts:	26-Aug-12	
Date of survey ends:	26-Aug-12	
Midpoint	26-Aug-12	
End prev year (for decimal calc)	31-Dec-11	
Reference date:	2012.65	

* assumed 3% adjustment to total population

Smoothing with correction for underreporting (BASEPOP)- example

EASTLANDS: 2012

Base Population Estimation

A. Reported Population by Age and Sex, Desired
Population by Sex, and Sex Ratio at Birth

Item and age	Male	Female	Both sexes
Reference date			2012.65

REPORTED POPULATION

All ages	21,869,990	23,058,933	
Under 1	747,826	751,563	
1-4	2,890,156	2,884,287	
0-4	3,637,982	3,635,850	
5-9	3,333,835	3,325,202	
10-14	2,892,583	2,900,004	
15-19	2,238,976	2,369,860	
20-24	1,793,053	2,160,986	
25-29	1,548,762	1,843,732	
30-34	1,380,068	1,529,610	
35-39	1,182,651	1,258,301	
40-44	944,033	955,081	
45-49	718,372	787,547	
50-54	607,361	603,666	
55-59	392,104	381,736	
60-64	380,223	390,494	
65-69	238,972	253,864	
70-74	226,484	251,346	
75-79	144,643	147,489	
80+	209,888	264,165	
Unknown	0	0	

← Reported population

DESIRED POPULATION

All ages	22,526,090	23,750,701	46,276,791
Sex ratio at birth			1.0300

If there is a basis to adjust the total population, enter desired total by sex here

assumed 3% adjustment to total population

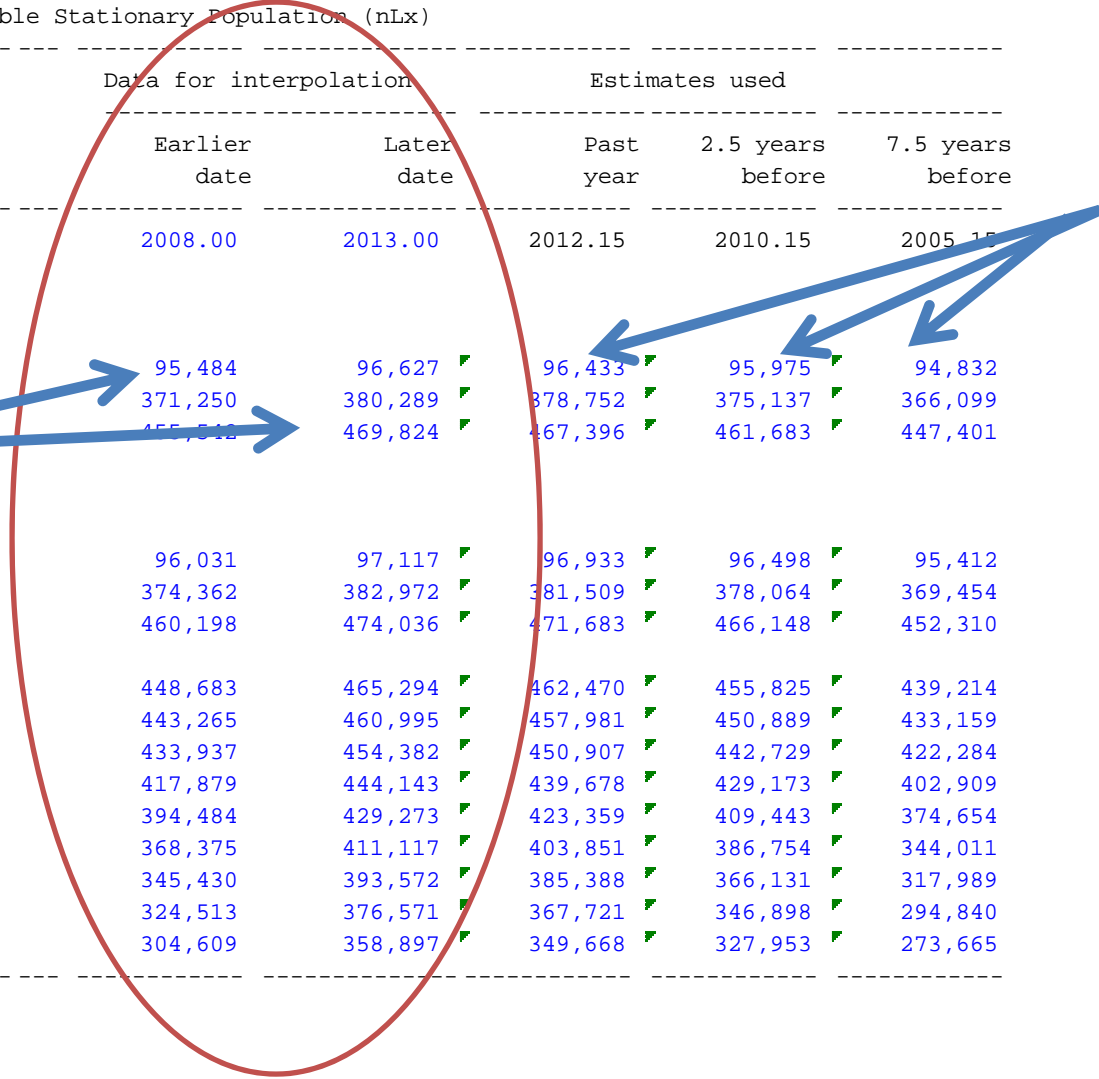
Smoothing with correction for underreporting (BASEPOP)- example

B. Life Table Stationary Population (nLx)

Function and age	Data for interpolation		Estimates used		
	Earlier date	Later date	Past year	2.5 years before	7.5 years before
Dates	2008.00	2013.00	2012.15	2010.15	2005.15
MALE					
1L0	95,484	96,627	96,433	95,975	94,832
4L1	371,250	380,289	378,752	375,137	366,099
5L5	455,512	469,824	467,396	461,683	447,401
FEMALE					
1L0	96,031	97,117	96,933	96,498	95,412
4L1	374,362	382,972	381,509	378,064	369,454
5L5	460,198	474,036	471,683	466,148	452,310
5L15	448,683	465,294	462,470	455,825	439,214
5L20	443,265	460,995	457,981	450,889	433,159
5L25	433,937	454,382	450,907	442,729	422,284
5L30	417,879	444,143	439,678	429,173	402,909
5L35	394,484	429,273	423,359	409,443	374,654
5L40	368,375	411,117	403,851	386,754	344,011
5L45	345,430	393,572	385,388	366,131	317,989
5L50	324,513	376,571	367,721	346,898	294,840
5L55	304,609	358,897	349,668	327,953	273,665

User inputs nLx estimates from life tables for two nearby dates

Workbook interpolates these 3 columns

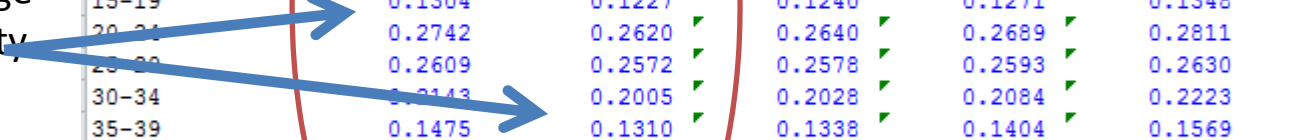
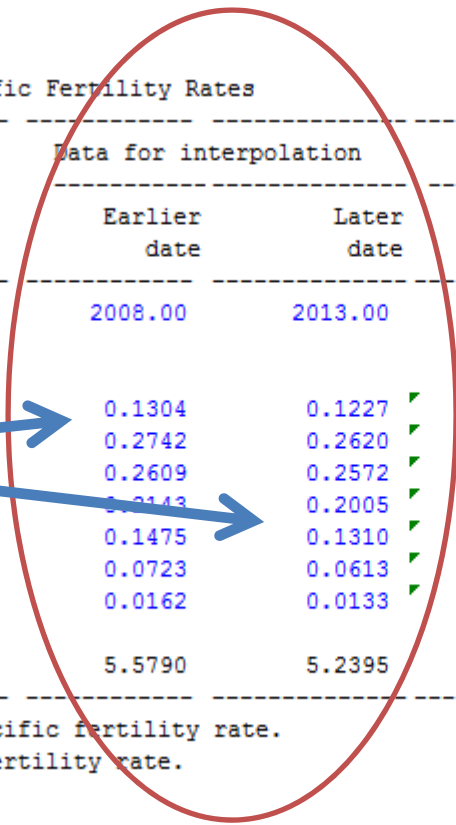


Smoothing with correction for underreporting (BASEPOP)- example

C. Age-Specific Fertility Rates					
Function and age	Data for interpolation		Estimates used		
	Earlier date	Later date	Past year	2.5 years before	7.5 years before
Dates	2008.00	2013.00	2012.15	2010.15	2005.15
ASFR					
15-19	0.1304	0.1227	0.1240	0.1271	0.1348
20-24	0.2742	0.2620	0.2640	0.2689	0.2811
25-29	0.2609	0.2572	0.2578	0.2593	0.2630
30-34	0.2143	0.2005	0.2028	0.2084	0.2223
35-39	0.1475	0.1310	0.1338	0.1404	0.1569
40-44	0.0723	0.0613	0.0632	0.0676	0.0786
45-49	0.0162	0.0133	0.0138	0.0149	0.0178
TFR	5.5790	5.2395	5.2972	5.4330	5.7725

User inputs age-specific fertility estimates for two nearby dates

Workbook interpolates these 3 columns



Smoothing with correction for underreporting (BASEPOP)- Results

D. Reported and Adjusted Populations, by Age and Sex

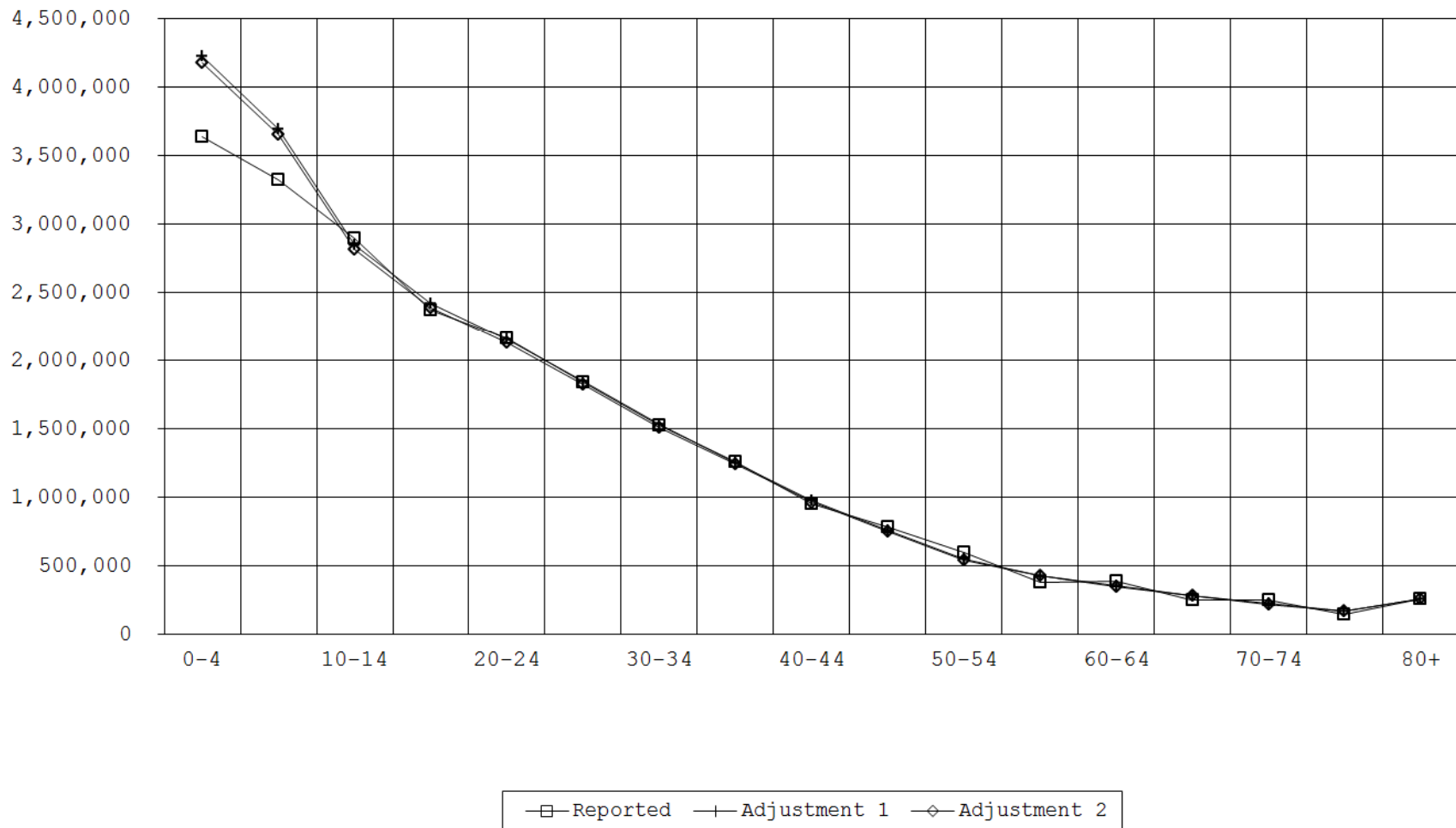
Age	Reported with unknowns distributed		Adjustment 1 (adjusted under 10, smoothed 10+)		Adjustment 2 (desired total using adjusted under 10, smoothed 10+)	
	Male	Female	Male	Female	Male	Female
Adj. factor	1.0000	1.0000	X	X	0.9800	0.9887
All ages	21,869,990	23,058,933	22,986,789	24,021,456	22,526,089	23,750,700
Under 1	747,826	751,563	915,548	893,492	897,198	883,421
1-4	2,890,156	2,884,287	3,405,861	3,332,781	3,337,601	3,295,216
0-4	3,637,982	3,635,850	4,321,409	4,226,273	4,234,799	4,178,637
5-9	3,333,835	3,325,202	3,767,207	3,697,302	3,691,705	3,655,628
10-14	2,892,583	2,900,004	2,838,136	2,850,933	2,781,255	2,818,799
15-19	2,238,976	2,369,860	2,293,423	2,418,931	2,247,458	2,391,666
20-24	1,793,053	2,160,986	1,810,405	2,156,474	1,774,121	2,132,168
25-29	1,548,762	1,843,732	1,531,410	1,848,244	1,500,718	1,827,412
30-34	1,380,068	1,529,610	1,388,848	1,531,763	1,361,012	1,514,497
35-39	1,182,651	1,258,301	1,173,871	1,256,148	1,150,345	1,241,990
40-44	944,033	955,081	923,961	977,970	905,443	966,947
45-49	718,372	787,547	738,444	764,658	723,644	756,039
50-54	607,361	603,666	559,045	552,673	547,841	546,444
55-59	392,104	381,736	440,420	432,729	431,593	427,851
60-64	380,223	390,494	346,115	356,850	339,178	352,828
65-69	238,972	253,864	273,080	287,508	267,607	284,267
70-74	226,484	251,346	211,063	226,127	206,833	223,579
75-79	144,643	147,489	160,064	172,708	156,856	170,761
80+	209,888	264,165	209,888	264,165	205,681	261,187

X Not applicable.

Smoothing with correction for underreporting (BASEPOP)- Results

EASTLANDS: 2012

2 Reported & Adjusted Female Population



Getting to a mid-year estimate

- Projection software needs a mid-year population as base
- PASEX spreadsheet MOVEPOP.xls shifts population from one date to another
- Inputs:
 - Date of census
 - Date desired
 - Population by age and sex
 - Mortality pattern (m_x)
 - Fertility pattern (ASFR)
 - Total net migration

Getting to a mid-year estimate (MOVEPOP)- example

Table
EASTLAND: 2012
A. Population and Mortality Data

Mx from life table

Age	Population		M(x)		ASFR's and TFR	Annual net number of migrants
	Male	Female	Male	Female		
Initial date	2012.65					
Desired date	2012.50					
All ages	22,526,089	23,750,700				10,000
0-1	897,198	883,421	0.03485	0.04136		
1-4	3,337,601	3,295,216	0.00362	0.00399		
5-9	3,691,705	3,655,628	0.00174	0.00213		
10-14	2,781,255	2,818,799	0.00188	0.00207		
15-19	2,247,458	2,391,666	0.00184	0.00269	0.1227	
20-24	1,774,121	2,132,168	0.00215	0.00303	0.2620	
25-29	1,500,718	1,827,412	0.00375	0.00405	0.2572	
30-34	1,361,012	1,514,497	0.00541	0.00579	0.2005	
35-39	1,150,345	1,241,990	0.00816	0.00875	0.1310	
40-44	905,443	966,947	0.00879	0.00994	0.0613	
45-49	723,644	756,039	0.00872	0.01098	0.0133	
50-54	547,841	546,444	0.00909	0.01198		
55-59	431,593	427,851	0.01042	0.01362	5.2395	
60-64	339,178	352,828	0.01499	0.01944		
65-69	267,607	284,267	0.02369	0.02937		
70-74	206,833	223,579	0.03919	0.04586		
75-79	156,856	170,761	0.06522	0.07391		
80+	205,681	261,187	0.10970	0.11934		

Adjusted population from BASEPOP

Age-specific fertility rates estimate

M(x) Age-specific central death rates.
TFR Total fertility rate.

Getting to a mid-year estimate (MOVEPOP)- results

B. Estimated Vital Rates and Population at Desired Date

Item	Value
Crude birth rate	0.04014
Crude death rate	0.00787
Net migration rate	0.00022
Growth rate	0.03249
Population at desired date	2012.50 46,051,833

C. Estimated Population, by Age and Sex: 2012.50

Age	Population			Sex ratio
	Both sexes	Male	Female	
Adjustment factor.....	0.9951			
All ages	46,051,833	22,416,587	23,635,246	0.9484
(0-4)	8,372,538	4,214,213	4,158,325	1.0134
Under 1	1,771,964	892,837	879,127	1.0156
1-4	6,600,574	3,321,376	3,279,198	1.0129
5-9	7,311,616	3,673,759	3,637,857	1.0099
10-14	5,572,833	2,767,736	2,805,097	0.9867
15-19	4,616,572	2,236,532	2,380,040	0.9397
20-24	3,887,300	1,765,497	2,121,803	0.8321
25-29	3,311,952	1,493,423	1,818,529	0.8212
30-34	2,861,531	1,354,396	1,507,135	0.8987
35-39	2,380,705	1,144,753	1,235,952	0.9262
40-44	1,863,289	901,042	962,247	0.9364
45-49	1,472,490	720,126	752,364	0.9572
50-54	1,088,965	545,178	543,787	1.0026
55-59	855,267	429,495	425,772	1.0087
60-64	688,641	337,529	351,112	0.9613
65-69	549,192	266,306	282,886	0.9414
70-74	428,320	205,828	222,492	0.9251
75-79	326,024	156,093	169,931	0.9186
80+	464,598	204,681	259,917	0.7875

Thank you

Questions?

>> until 11 March:



>> After 11 March: sawyerc@un.org
bassarsky@un.org



Regional Workshop on the Production of Population Projections
Addis Ababa, 7-11 March 2016

