Assessing and Improving the Accuracy of Target Population Estimates for Immunization Coverage

World Health Organization

Working Draft

Revised November 2015

Contents

List of Tables

List of Figures

List of Abbreviations

BCG	Bacillus Calmette-Guérin
CBR	Crude Birth Rate
CRVS	Civil Registration and Vital Statistics
DHS	Demographic and Health Survey
DTP	Diptheria-tetanus-pertussis vaccine
EPI	Expanded Programme on Immunization
HepB	Hepatitis B vaccine
HepB3	Third dose of hepatitis B vaccine
Hib	*Haemophilus influenzae* type b vaccine
Hib3	Third dose of *Haemophilus influenzae* type b vaccine
HMIS	Health Management Information System
IMR	Infant Mortality Rate
MCV	Measles containing vaccine
NSO	National Statistical Office
NSS	National Statistical System
Pol3	Third dose of polio vaccine (either oral polio vaccine inactivated polio vaccine)
UN	United Nations
UNPD	United Nations Population Division
UNSD	United Nations Statistics Division
WPP	World Population Prospects
WHO	World Health Organization

Chapter 1. Introduction to the manual

Calculating vaccination coverage from administrative data on numbers of vaccinated persons requires corresponding target population estimates. The accuracy of available target population estimates may be problematic. Evidence of low accuracy includes coverage estimates far in excess of 100%, erratic year-to-year fluctuations, and disease outbreaks in areas with high estimated coverage.

The problem has grown more acute as coverage rates have risen, requiring more accurate target population estimates to monitor changes in coverage.

This manual provides a systematic approach to assessing and improving the accuracy of target population estimates. It describes long-term, best practice methods as well as shorter-term expedients.

1.1 Target audience

The manual is intended for use by national Expanded Programme on Immunization (EPI) managers and their staff; by staff at the World Health Organization (WHO) Headquarters and in WHO regional offices who provide technical support to national EPI programmes; and by other persons concerned with the accuracy of target population estimates, nationally, regionally, and/or globally.

The statistical methods presented in this manual are not particularly difficult, but if some EPI programmes may lack the resources and expertise to implement them. In this situation it is appropriate to seek assistance from the national statistical office or another source of statistical expertise, such as a university or research institute.

The methods presented are illustrated with examples based on actual data received by World Health Organization headquarters in Geneva, anonymized to obscure the identity of individual countries.

1.2 The measurement challenge of higher coverage

As coverage rises, coverage estimates become increasingly sensitive to errors in target population estimates.

This is demonstrated in Table 1.1 and Figure 1.1. The first three columns of the table show true values of target population, vaccinations, and coverage for a hypothetical population, with coverage levels ranging from 10 to 90 percent. The following two columns show the range of estimated target populations corresponding to a ±10% error and the corresponding range of coverage estimates.

If true coverage is 50%, for example, and the estimated target population is 10% low, estimated coverage is 50/90, or 56%. If the estimated target population is 10% high, estimated coverage is 50/110, or 45%. This is how the range 45-56% in the middle row of the table is calculated.

Figure 1.1 plots the coverage levels and error ranges in Table 1.1 as though they referred to the years 1960, 1970, ..., 2010. When coverage rises from 10% to 20%,

Table 1.1
Effect of 10% error in target population estimate on error in immunization coverage estimate at different levels of true coverage

True values			Estimates	
Target Population	Vaccin-ations	Coverage	Target Population	Coverage
100	10	10%	90-110	9-11%
100	20	20%	90-110	18-22%
100	50	50%	90-110	45-56%
100	80	80%	90-110	73-89%
100	90	90%	90-110	82-100%

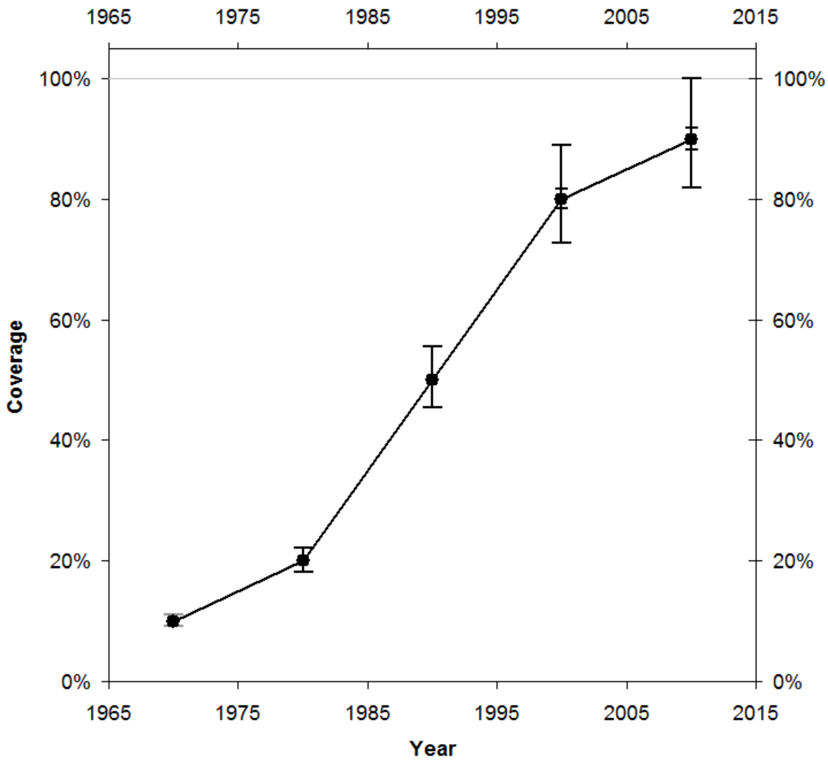

Figure 1.1
Effect of 10% error in target population estimates on estimated coverage

the 10% error in the target population estimate does not obscure the upward trend. The height of the error bars is small compared to the rise in coverage.

The same is true when coverage rises from 20% to 50%, though the height of the error bars increases with increased coverage.

The situation changes when coverage rises further. The height of the error bars for 2000 ($89 - 73 = 16$ percentage points) and 2010 ($100 - 82 = 18$ percentage points) coverage are both larger than the true increase in coverage over this period—10 percentage points).

This has important consequences for estimating change in coverage. If the target population estimate for 1990 is at the lower end of the error range (90 children, giving a coverage of 73%), for example, and the estimate for 2000 is at the upper end of the error range (110 children, giving a coverage of 100%), the estimated change in coverage will be 27%—nearly three times the true increase of 10%.

At the other extreme, if the target population estimate for 1990 is 10% high and the estimate for 2000 is 10% low, the estimated change in coverage will be -7%—a *decline* in coverage.

How accurate must target population estimates be to give reasonable estimates of changing coverage when coverage is 80-90%? Figure 1.1 shows ±2% error bars for 2000 and 2010 only. These ±2% error bars are roughly equivalent, when coverage is 80-90%, to the ±10% bars when coverage is 50-80%.

Higher immunization coverage makes the accuracy of target population estimates more important than it is at lower coverage levels. This is the measurement challenge of higher coverage.

1.3 Target populations: Births and surviving infants

Target populations for the primary EPI vaccines are numbers of births or numbers of surviving infants. Births are the usual target population for BCG vaccination. Surviving infants are the usual target population for other EPI vaccines given in infancy.

Numbers of births are a standard demographic statistic that may be found in official statistical publications. Numbers of surviving infants must usually be estimated from estimated numbers of births and infant mortality rates using the relationship

$$Surviving\ infants = Births \times (1 - Infant\ mortality\ rate).$$

Infant mortality rates, like numbers of births, are a standard demographic indicator that may be found in official statistical publications.

The accuracy of estimates of surviving infants calculated in this way depends on the accuracy of the $Births$ and the $Infant\ mortality\ rate$ estimates. Infant mortality rates are typically less than 100 infant deaths per thousand births, however, so estimates of surviving infants are relatively insensitive to errors in estimated infant mortality rates.

For this reason, Parts I and II focus mainly on assessing and improving the accuracy of estimates of births. If these estimates are sufficiently accurate, the derived estimates of surviving infants will usually be sufficiently accurate as well.

1.4 Assessing accuracy

Part I provides a systematic approach to assessing the accuracy[1] of target population estimates, including annual numbers of national and subnational births and

[1] For definitions, refer to the Glossary on page 75

surviving infants.

1.5 Improving accuracy: Long term goal

Immunization programmes will preferably obtain target population estimates from the country's national statistical system, with no need for primary data collection or estimation by the programme. Few EPI programmes have the expertise and resources needed to produce accurate target population estimates independently.

Target population estimates of accuracy sufficient to monitor coverage changes at high coverage levels will generally be possible only for countries with a well-developed civil registration and vital statistics (CRVS) system. These systems collect information on births and deaths throughout the country as the events occur, making it possible to obtain the most accurate, current, and geographically detailed data possible.

The long term goal will be to promote the development of a civil registration and vital statistics system that will provide target population estimates sufficiently accurate to monitor immunization coverage, especially when coverage is high.

1.6 Improving accuracy: Interim expedients

In the absence of a well-developed CRVS system, population projections based on a national population census will usually provide the most accurate target population estimates. The projections must be based on a census taken during the last 10-15 years, and they must provide annual estimates of births and infant mortality rates. The accuracy of population projection estimates will be lower than the accuracy of estimates from a well-developed CRVS system.

Table 11.1 on page 45 illustrates the kind of estimates population projections may provide. However, not all published population projection estimates will provide this information.

The accuracy of population projection estimates for the country as a whole is likely to be reasonably good for the years following the census, but accuracy tends to decline over time because projections require anticipating future population trends.

The accuracy of subnational population projection estimates tends to be lower than the accuracy of national level estimates, partly because it is more difficult to anticipate future population trends in subnational areas, and partly because of limited information on internal migration.

1.7 Working with the national statistical system

Part II provides a systematic approach to obtaining the most accurate target population estimates possible given the resources provided by the national statistical system. This necessarily involves working with national statistical authorities to identify what resources exist and how they may best be used to produce target population estimates.

The short term goal of this work is to obtain the best possible target population estimates from the existing national statistical system.

An intermediate term goal may be to improve utilization of existing statistical resources. It may be that population projection estimates from a recent census are available, for example, but that they do not provide annual numbers of births and infant mortality rates. The EPI programme may work with the authority responsible for producing the population projections to have these estimates included in published census projection results.

Part I

Assessing Accuracy

Checklist for Assessing Accuracy

STEP 1 Specify the calendar years for which target population estimates are to be assessed. A ten year series ending with the year of the most recent available estimates is preferred. It is advisable to assess estimates for a minimum of five calendar years.

STEP 2 Compare the current national annual estimates of births used by the EPI with estimates from one or more alternative sources (Chapter 3, Comparing Estimates with Alternative Sources). Use the *World Population Prospects* estimates produced by the United Nations Population Division (see Annotated List of Resources) as one alternative source. Flag the result if the difference between the estimates for any year is greater than 10% of the EPI estimate.

STEP 3 Calculate annual growth rates of national births (Chapter 4, Checking Target Population Growth Rates). Flag the result if any growth rate is greater than 10%.

STEP 4 Calculate annual national implied infant mortality rates (IMRs) from national annual numbers of births and surviving infants (Chapter 5 Checking Implied Infant Mortality Rates). Compare these rates with IMRs from one or more alternative sources. Use the *World Population Prospects* estimates (Annotated List of Resources) as one alternative source. Flag the result if the difference between the implied infant mortality rate and the alternative source infant mortality rate for any year is greater than 20% of the implied IMRS.

STEP 5 Compare national numbers of births with sums of numbers of births over all subnational areas (Chapter 6, Checking *Sum of Subnational* Equals *National*). Flag the result if national births does not equal the sum of subnational births for any year.

If any of the checks in Steps 1-5 have resulted in a flag, Steps 6-9 below will help identify the source of the problem, and the accuracy of target population estimates will be improved by the procedures and methods presented in Part II, Improving Accuracy.

STEP 6 Carry out one or more graphical trend analyses as described in Chapter 7, Plotting and Analysing Time Series.

STEP 7 Check documentation (Chapter 8, Checking Documentation to the EPI data on births and surviving infants.

STEP 8 Check missing values (Chapter 9, Checking Missing Values to the EPI data on births and surviving infants.

STEP 9 Check consistency (Chapter 10, Checking Consistency to the EPI data on births and surviving infants.

Chapter 2. Introduction to Assessing Accuracy

This chapter provides an overview of the assessment methods presented in subsequent chapters. It also presents concepts and definitions of terms related to accuracy, error, and data quality.

2.1 Overview of Part I

Chapter 3, Comparing Estimates with Alternative Sources and chapter 4, Checking Target Population Growth Rates provide tests for the accuracy of target population estimates.

Chapter 5, Checking Implied Infant Mortality Rates, provides a test for errors in estimated annual numbers of births and surviving infants.

Chapter 6, Checking *Sum of Subnational* Equals *National*, provides a test for errors in estimated numbers of subnational births.

Chapter 7, Plotting and Analysing Time Series, presents graphical methods that provide insight into inaccuracies in estimates of annual series of target population and immunization coverage.

Chapter 8, Checking Documentation, Chapter 9, Checking Missing Values, and Chapter 10, Checking Consistency, present checks that may be applied to EPI programme data to assess data quality.

2.2 Births and surviving infants

Target populations for the primary EPI vaccines are numbers of births or numbers of surviving infants. Births are the usual target population for BCG vaccination. Surviving infants are the usual target population for other EPI vaccines given in infancy. Surviving infants refers to the number of children reaching their first birthday during a given year.

2.3 Estimates, accuracy, and error

An **estimate** is a quantity calculated from data.

Accuracy refers to the closeness of an estimate to the unknown true value it represents. It is hoped that estimates will be close to true values, but true values remain, strictly speaking, unknown. Using the words “estimate”, “error” and “accuracy” entails accepting that the idea of true values is nonetheless meaningful.

The **error** of an estimate is the difference between the estimate and the true value. It may be assumed without loss of generality that error is calculated as the estimated value minus the true value. A positive error then signifies an estimate that is too large, a negative value an estimate that is too small.

Error is *defined* as the estimated value minus the true value, but this formula cannot be used to *calculate* error. Calculation of error in this way would require knowing the true value. If the true value could be known, consideration of error would be superfluous.

The concept of error is useful for two reasons. First, it is known that most estimates err to some extent, and that estimates may be so wrong as to make them useless. Second, there are methods for assessing the error of an estimate.

The **magnitude** of an error refers its value disregarding the sign. The **direction** of an error refers to whether it is positive or negative. Given a true value of 100, the estimates 95 and 105 have same magnitude, but opposite directions.

Estimates of larger quantities tend to have larger errors than estimates of smaller qualities. The **relative error** of an estimate is the error divided by the true value. In practice, because true values are unknown, relative error may be calculated as error divided by the estimate. The difference between the two calculations is negligible unless the error is large. Relative error is commonly expressed as a percent. Error may be referred to as **absolute error** to distinguish it from relative error.

To illustrate absolute and relative error, consider the 2010 population census of China, which enumerated 1,339,724,852 persons, and the 2011 population census of Botswana, which enumerated 2,024,904 persons.

An absolute error of 1,000,000 persons in relation to the population of China is very small. The relative error is 1,000,000 divided by 1,339,724,852, or 0.07%. In relation to the population of Botswana, the same absolute error is very large. The relative error is 1,000,000 divided by 2,024,904, or 49%.

2.4 Data and data quality

Data may be defined as systematic information on characteristics of statistical units in a defined aggregate. Data consists concretely of **records**, each record referring to a particular statistical unit and providing information on characteristics of this unit, and **documentation** that

1. defines the statistical units represented by the records,
2. defines the aggregate to which the records refer,
3. defines the characteristics of the units and their possible values, and
4. any other information required to understand the information contained on the records.

The records together with the documentation may be referred to as a **dataset**.

Data quality refers to various characteristics of a dataset, including at least the following.

1. The faithfulness of correspondence between the records in the dataset and the statistical units they represent
2. The frequency of missing values for characteristics included on records, taking account of undefined values

3. The accuracy of the information on characteristics provided by the records in the dataset

Chapter 3. Comparing Estimates with Alternative Sources

Annual numbers of births for the country as a whole are the most fundamental target population numbers. The most effective method for assessing their accuracy is to compare them with estimates from one or more alternative sources.

Comparison for a ten year series of annual births ending with the year of the most recent available estimate is preferred. It is advisable to compare estimates for a minimum of five calendar years.

The *World Population Prospects* estimates produced by the United Nations Population Division (see the Annotated List of Resources) could be used as one alternative source.

If a difference between the EPI estimate and an alternative estimate is more than 10% for any year, the method by which the EPI estimates are produced may be investigated and compared with the method by which the alternative estimates are produced to decide which of the two estimates is more accurate.

3.1 Comparing estimates with alternative sources

Estimates of annual national births will usually be available from more than one source. Many countries will have estimates from a civil registration and vital statistics (CRVS) system. Population projection estimates from a recent census may be available. The EPI programme may consult national statistical authorities to determine what national level estimates of annual births are available.

Estimates of annual births and infant mortality rates for all countries with a population of more than 100,000 as of mid-2013 are available from the *World Population Prospects* (Annotated List of Resources).

3.2 Example of Immunization Coverage Data: Country A

Table 3.1 shows illustrative immunization coverage data for Country A. The first panel of the table shows target population numbers, annual numbers of births for BCG and annual numbers of surviving infants for the other vaccines. Note that target populations for non-BCG vaccines are identical with the exception of the denominators for HepB3 for 2003, 2004 and 2005. The note to the table indicates why these target population numbers are different.

Missing value cells are marked by “m”, undefined value cells by “u”. **Undefined** value cells correspond to years in which a vaccine was not provided by the program (section 8.2). **Missing** value cells correspond to years in which the vaccine was provided, but for which no value is available (section 9).

The second panel of the table shows numbers of immunizations. The third and last panel shows coverage, calculated as the number of vaccinations divided by the target population.

Table 3.1
Example of Immunization Coverage Data: Country A

	Vaccine						
Year	BCG	DTP1	DTP3	MCV	Pol3	HepB3	Hib3
			TARGET POPULATION (000)				
2000	883	883	883	883	883	u	u
2001	897	897	897	897	897	u	u
2002	1,063	1,035	1,035	1,035	1,035	u	u
2003	1,078	1,051	1,051	1,051	1,051	86	u
2004	1,099	1,027	1,027	1,027	1,027	178	u
2005	1,118	1,045	1,045	1,045	1,045	772	u
2006	1,138	1,064	1,064	1,064	m	1,064	u
2007	1,158	1,083	1,083	1,083	m	1,083	u
2008	1,179	1,102	1,102	1,102	m	1,102	u
2009	1,068	1,013	1,013	1,013	1,013	1,013	1,013
			VACCINATIONS (000)				
2000	832	984	808	783	795	u	u
2001	907	915	867	845	867	u	u
2002	930	921	914	896	910	u	u
2003	997	981	973	955	971	48	u
2004	1,011	995	982	988	982	105	u
2005	1,013	999	1,000	983	1,000	649	u
2006	1,025	1,013	993	975	m	999	u
2007	1,066	1,047	1,030	1,025	m	1,029	u
2008	1,063	1,048	1,050	1,054	m	1,050	u
2009	1,008	987	972	997	968	971	971
			COVERAGE (%)				
2000	94	111	92	89	90	u	u
2001	101	102	97	94	97	u	u
2002	87	89	88	87	88	u	u
2003	92	93	93	91	92	56	u
2004	92	97	96	96	96	59	u
2005	91	96	96	94	96	84	u
2006	90	95	93	92	m	94	u
2007	92	97	95	95	m	95	u
2008	90	95	95	96	m	95	u
2009	94	97	96	98	96	96	96

Note HepB was introduced in Province A in 2003, extended to Province B in 2004, to Province C in 2005, and throughout the country in 2006. Hib was introduced throughout the country in 2009. Target population for BCG is number of births. Target population for other vaccines is number of surviving infants. u = undefined, m = missing.

The illustrative data in Table 3.1 will be used throughout this manual.

Table 3.2
Comparison of alternative estimates of births and surviving infants: Country A

	Births (000)				Surviving Infants (000)		
Year(s)	EPI	WPP	Ratio	WPP IMR	EPI	WPP	Ratio
2000	883	1,001	0.88	62.2	883	939	0.94
2001	897	991	0.91	59.4	897	932	0.96
2002	1,063	979	1.09	56.7	1,035	924	1.12
2003	1,078	966	1.12	54.1	1,051	914	1.15
2004	1,099	952	1.15	51.6	1,027	902	1.14
2005	1,118	937	1.19	49.2	1,045	891	1.17
2006	1,138	923	1.23	46.9	1,064	880	1.21
2007	1,158	911	1.27	44.6	1,083	870	1.24
2008	1,179	901	1.31	42.2	1,102	863	1.28
2009	1,068	894	1.19	39.9	1,013	858	1.18
2000-09	10,681	9,455	1.13	-	10,200	8,973	1.14

Note EPI estimates are numbers of births and surviving infants reported to WHO headquarters. WPP estimates of births and infant mortality rates (IMR) are from *World Population Prospects: The 2012 Revision.* Both sets of estimates are anomymized by multiplication by a random constant factor. WPP estimate of surviving infants calculated as WPP estimates of births times (1 - IMR) (Chapter 16).

3.3 Example of comparison with alternative source

Table 3.2 compares annual estimates of births and surviving infants for Country A in Table 3.1 with *World Population Prospects* (WPP) estimates.

The first three columns show EPI and WPP estimates of births and the ratio of the WPP estimate to the EPI estimate. The last three columns show the same for surviving infants. The middle column shows the WPP IMR estimates, which are used to compute the WPP estimate of surviving infants from the WPP estimate of births.

The last row of the table shows total births and surviving infants over the 10 year period 2000-2009. The ratios give a summary indication of the difference in level between the EPI and WPP estimates. The EPI estimates of births are 13% higher than the WPP estimates, and the EPI estimates of surviving infants are 14% higher than the WPP estimates.

These differences are greater than 10%. It is therefore appropriate to investigate the method of calculation of the EPI estimates to decide which estimates are more accurate.

The difference in pattern and trend of the EPI and WPP estimates in Figure 3.1 is striking.

The figure shows that

1. the trend of the EPI estimates is up, whereas the trend of the UNPD estimates is down,

2. the EPI estimates jump discontinuously up by about 150,000 births between 2001 and 2002, and discontinuously down by about 100,000 births between 2008 and 2009,
3. a straight line connecting the first two and the last EPI points would have a slope similar to a straight line fitted to the in-between points, and
4. the difference between the EPI and WPP estimates toward the end of the period is large, nearly 300,000 births, or a relative error of 20-30%.

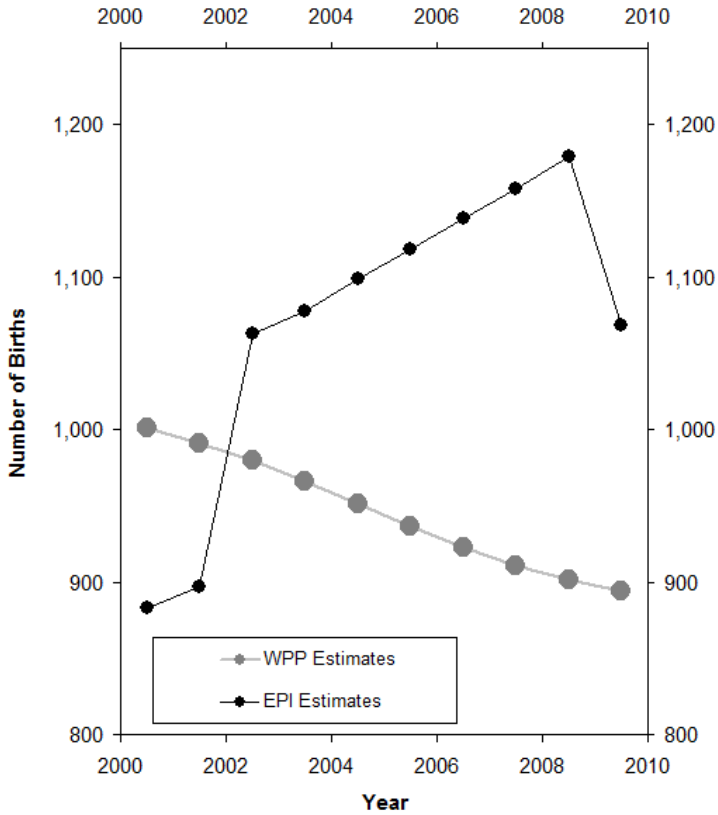

Figure 3.1
EPI and WPP estimates of births compared: Country A

Note The Y axis does not start at 0 to better illustrate the differences. **Source** Table 3.2

The second observation suggests a change in the procedure used to estimate births between 2001 and 2002, the effect of which is to increase the numbers by slightly over 150,000, followed by a reversion between 2008 and 2009 that brings numbers down by about the same amount. The pattern suggests that the second large change might represent a decision that the first change was a mistake to be corrected.

3.4 Interpretation of differences

Comparison of estimates from different sources is a powerful method for evaluating accuracy. For any estimate of births, there is only one true value. If two estimates differ, either or both of them must be incorrect.

If there is no reason to expect that one estimate is more accurate than the other, the magnitude of the difference between the two estimates provides an indication of the magnitude of their error.

If there is reason to expect that one estimate is more accurate than the other, the comparison will provide an indication of the error in the statistic believed to be less accurate.

A large difference between two estimates indicates a large error in either or both of the estimates. A small difference between two estimates may indicate a small error, but the interpretation of small differences is more complicated. A small differences between two estimates may simply indicate that they are the same estimate "in disguise".

Suppose for example that the National Statistical Office produces an estimate of births for a particular year and that this estimate is incorporated into a Health Management Information System (HMIS) maintained by the Ministry of Health.

If the EPI programme receives the estimate from both sources, the NSO and the HMIS, and compares them, there will presumably be no difference between them, but this comparison provides no information on accuracy. The "two" estimates are the same because they are in fact the same estimate from a different source.

The conclusion is obvious by intention in this example. In practice it may be difficult to ascertain how an estimate was produced and from what primary data source.

> It is useful to know that for most national statistical systems there are only two primary sources for annual numbers of births: 1) a civil registration and vital statistics (CRVS) system and 2) population projection estimates from a recent national population census.

Chapter 4. Checking Target Population Growth Rates

Target population size tends to change slowly from one year to the next. Most year-to-year changes are in the range of ±5%. As shown in Section 4.3 below, very few changes lie outside the ±10% range. Growth rates greater than 10% or less than -10% therefore suggest errors in estimated numbers of births.

Changes of this magnitude are difficult to see in the target population numbers themselves. Calculating annual growth rates of target population numbers makes them easy to see.

This chapter explains how to calculate annual growth rates of target population estimates and use them to check for errors.

4.1 Definition of target population growth rate

Growth rates of births are not a standard demographic indicator and may not be found in official statistical populations. They are different from population growth rates, which are a standard demographic indicator and will be found in these publications.

The annual growth rate of births between consecutive years may be defined as

$$\textit{Growth Rate} = \left(\frac{\textit{Births in Year 2}}{\textit{Births in Year 1}}\right) - 1$$

The same formula may be applied to any target population numbers for successive years.

The target population growth rate is a sensitive indicator of year-to-year change in target population numbers. Sharp fluctuations in growth rates of births are likely to indicate errors in the estimates.

4.2 Example of growth rates of births check

Table 4.1 shows annual growth rates of births (BCG column) and surviving infants (remaining columns) for the target population numbers in Table 3.1. Examination of the growth rates in table 4.1 shows the following.

1. Most growth rates in the table are in the range 1.4-1.9%
2. (Calculated) Growth rates "spike up" in 2001-2002
3. The BCG rate for 2003-2004 is 1.9%, but the DTP1, DTP3, MCV and Pol3 rates are all *minus* 2.3%
4. The 2008-2009 growth rates "spike down"
5. The 2003-2006 HebB3 growth rates are extreme positive outliers

Table 4.1
Growth rates of national births: Country A

	Vaccine					
Years	BCG	DTP1	DTP3	MCV	Pol3	HepB3
2000-2001	1.6	1.6	1.6	1.6	1.6	u
2001-2002	18.5	15.4	15.4	15.4	15.4	u
2002-2003	1.4	1.5	1.5	1.5	1.5	u
2003-2004	1.9	-2.3	-2.3	-2.3	-2.3	107.0
2004-2005	1.7	1.8	1.8	1.8	1.8	333.7
2005-2006	1.8	1.8	1.8	1.8	m	37.8
2006-2007	1.8	1.8	1.8	1.8	m	1.8
2007-2008	1.8	1.8	1.8	1.8	m	1.8
2008-2009	-9.4	-8.1	-8.1	-8.1	m	-8.1

Note Growth rates calculated from target population numbers in Table 3.1 using the formula given in section 4.1. As indicated in the note to Table 3.1, HepB was introduced in Province A in 2003, extended to Province B in 2004, to Province C in 2005, and throughout the country in 2006. "u" denotes undefined value. "m" denotes missing value.

The growth rate for 2001-2002 is greater than 10%, and the rate for 2008-2009 is close to, though not less than, -10%. It is therefore appropriate to investigate how the EPI births were estimated to determine whether these extreme rates are real.

The pattern observed in Figure 3.1 suggests a change in method of estimation between 2001 and 2002 that inflated estimated numbers of births and a countervailing change between 2008 and 2009 that deflated estimates numbers of births.

4.3 Variability of annual growth rates of births

For all but the smallest national populations, growth rates of births higher than 10% or less than -10% are rare. This is demonstrated in Figure 4.1, which shows a histogram of 1,332 annual growth rates of births calculated from time series of births for 23 countries with accurate data on numbers of births. The longest series, for Sweden, covers the years 1891-2011. The shortest series, for Slovenia, covers the years 1983-2009.

Only 16 of the 1,332 growth rates are greater than 10%, and only 18 are less than -10%. Overall, over 97% of the growth rates lie within this range.

The variability of growth rates of births may be higher for smaller populations for two reasons. First, smaller populations may experience relatively large numbers of in- and out-migrants. Second, very small populations are subject to random fluctuations in numbers of annual births. The smaller the population, the larger random fluctuations are likely to be.

Figure 4.1
Variability of annual growth rates of births

Note Growth rates calculated for annual series of births for Austria 1951-2010; Bulgaria 1947-2009; Canada 1921-2007; Switzerland 1932-2011; Czech Republic 1950-2011; Germany 1956-2010; Estonia 1959-2010; Finland 1939-2009; France 1946-2010; United Kingdom 1974-2009; Hungary 1950-2009; Japan 1947-2009; Lithuania 1959-2010; Netherlands 1950-2009; Norway 1967-2009; Portugal 1940-2009; Russia 1959-2010; Slovakia 1950-2009; Slovenia 1983-2009; Sweden 1891-2010; United States 1933-2010. Total of 1,332 growth rates of births. Data from the *Human Fertility Database*, www.humanfertility.org, visited 16-Sept-2015). 2 growth rates less than −20 and 4 greater than 20 are excluded from this graph.

Chapter 5. Checking Implied Infant Mortality Rates

Given annual numbers of births and surviving infants it is possible to calculate the infant mortality rate implied by these numbers. These implied infant mortality rates are useful indicators of the quality of the estimates of annual numbers of births and surviving infants.

5.1 Definition of implied infant mortality rate

The infant mortality rate is commonly defined as the number of infant deaths (among children aged <1 year) occurring in a population during a calendar year divided by the number of births occurring to the population during the year,

$$\textit{Infant Mortality Rate} = \frac{\textit{Infant Deaths}}{\textit{Births}}.$$

Infant mortality rates may be expressed as infant deaths per thousand births, so an infant mortality rate of 0.05 may be referred to as a rate of 50 per thousand.

Because

$$\textit{Infant Deaths} = \textit{Births} - \textit{Surviving Infants},$$

the infant mortality rate may be also be calculated as

$$\textit{Infant Mortality Rate} = \frac{\textit{Births} - \textit{Surviving Infants}}{\textit{Births}}$$

When immunization coverage data provides numbers of births and surviving infants, this formula may be used to calculate "implied" infant mortality rates.

These implied rates provide a useful check on the accuracy of the target population because inaccurate target population numbers tend to produce implausibly high or low implied infant mortality rates.

5.2 Example of Immunization Coverage Data: Country B

Table 5.1 shows illustrative coverage data for Country B in the same format as the Table 3.1, with one exception. Table 3.1 shows target population numbers and births in thousands, but Table 5.1 shows them as whole numbers. This difference merits brief discussion.

Estimates of target population and numbers of vaccinations are unlikely to be accurate to more than three significant figures. The 2000 BCG target population in Table 5.1, for example, is 383,586, but the true value might lie anywhere between 383,000 and 384,000, and perhaps outside this interval. The "586" digits are insignificant in this sense: they provide no information about the true target population

Table 5.1
Example of Immunization Coverage Data: Country B

Year	BCG	DTP1	DTP3	MCV	Pol3	HepB3	Hib3
				Vaccine			
			TARGET POPULATION				
2000	383,586	383,586	383,586	506,008	383,586	u	u
2001	398,046	398,046	398,046	398,046	398,046	u	u
2002	457,088	400,867	400,867	400,867	400,867	u	u
2003	435,881	382,234	382,234	382,234	382,234	u	u
2004	449,390	394,081	394,081	394,081	394,081	u	u
2005	461,583	406,308	406,308	406,308	406,308	u	u
2006	451,380	402,174	402,174	402,174	402,174	u	u
2007	594,921	546,735	546,735	546,735	546,735	u	u
2008	618,868	568,706	568,706	568,706	568,706	u	u
2009	632,172	581,344	581,344	581,344	581,344	581,344	581,344
			VACCINATIONS				
2000	136,278	136,697	87,199	174,924	87,199	u	u
2001	195,596	198,471	123,095	208,439	120,339	u	u
2002	213,415	165,833	87,848	193,620	102,119	u	u
2003	280,081	288,725	197,160	243,524	195,681	u	u
2004	322,697	296,194	243,250	290,015	243,209	u	u
2005	429,997	393,596	363,388	336,977	362,962	u	u
2006	487,901	456,595	410,707	416,425	410,387	u	u
2007	495,324	476,143	427,752	367,298	430,726	u	u
2008	480,560	533,744	508,248	456,002	499,851	u	u
2009	605,923	583,070	540,714	506,216	510,412	540,714	540,714
			COVERAGE				
2000	36	36	23	35	23	u	u
2001	49	50	31	52	30	u	u
2002	47	41	22	48	25	u	u
2003	64	76	52	64	51	u	u
2004	72	75	62	74	62	u	u
2005	93	97	89	83	89	u	u
2006	108	114	102	104	102	u	u
2007	83	87	78	67	79	u	u
2008	78	94	89	80	88	u	u
2009	96	100	93	87	88	93	93

Note Target population for BCG is number of births. Target population for other vaccines is number of surviving infants. "u" = undefined value

number. This is the rationale for presenting numbers over 100,000 in thousands rather than as whole numbers.

Presenting whole numbers may be appropriate, however, when working with data for subnational as well as national areas. Numbers for some subnational areas may be so small that rendering them in thousands gives too few significant figures. Presenting numbers in thousands for some areas and whole numbers for other areas invites confusion and is inadvisable for this reason, even if tables contain many insignificant digits.

Whole numbers facilitate "sum checks", which consist of comparing a given total population with the sum of subnational populations. Sum checks are useful for finding errors, which may then be corrected. For whole numbers, the result of a sum check is unambiguous. Numbers in thousands complicate sum checks because rounding errors may introduce differences even when there is no error.

Whole numbers are also important for accurate calculation of implied infant mortality rates, the formula for which involves a relatively small difference—infant deaths—between two relatively large numbers—births and surviving infants.

Coverage data should therefore usually be given in whole numbers, even if these contain substantial numbers of insignificant digits, particularly when subnational as well as national numbers are present.

5.3 Example of implied infant mortality rate check

Table 5.2 shows implied infant mortality rates calculated from the target population numbers in table 5.1. Because Table 5.1 gives a number of surviving infants for each vaccine except BCG, one implied rate may be calculated for each year and each non-BCG target population number. In this example four implied infant mortality rates may be calculated for 2000-2008 and six for 2009. the last column of the table shows *World Population Prospects* IMR estimates.

Table 5.2

Implied infant mortality rates: Country B

Year	IIMR DTP1	IIMR DTP3	IIMR MCV	IIMR Pol3	IIMR Hib3	IIMER HebB3	WPP IMR
2000	0	0	-319	0	u	u	62
2001	0	0	0	0	u	u	59
2002	123	123	123	123	u	u	57
2003	123	123	123	123	u	u	54
2004	123	123	123	123	u	u	49
2005	120	120	120	120	u	u	47
2006	109	109	109	109	u	u	45
2007	81	81	81	81	u	u	42
2008	81	81	81	81	u	u	42
2009	80	80	80	80	80	80	40

Note Calculated from target population numbers in table 5.1. The implied infant mortality rate (IIMR) in each column but the last is calculated from the target population for BCG (births) and the target population for the vaccine in the column heading (surviving children). The last column shows the WPP IMR estimate. See text for explanation. "u" = undefined value.

Examination of table 5.2 shows the following.

1. With one exception, the implied IMRs for 2000 and 2001 are zero
2. The exception is MCV for 2001, which is negative—a logical impossibility

3. For 2002-2009, the implied IMRs in each row are equal
4. Differences between the implied IMRs and the WPP estimated IMRs for 2002-2009 are very large, 50% or more

In view of the last observation in particular, it is appropriate to investigate how the target population numbers in Table 5.1 were estimated and attempt to improve the procedure by the methods discussed in Part II, Improving Accuracy.

Chapter 6. Checking *Sum of Subnational* Equals *National*

When births, surviving infants, or other counts are given for a complete set of subnational areas, the sum of the numbers over all subnational areas should equal the number for the country. This observation provides a useful check for the quality of national and subnational data.

More generally, checks of this kind may be carried out whenever counts are available for (a) some well-defined geographic area and (b) any set of mutually exclusive and collectively exhaustive subdivisions of this area.

The calculations are simple arithmetic. The challenges in practice tend to be knowing that subdivisions are mutually exclusive and collectively exhaustive and knowing how to interpret non-zero differences.

6.1 Example of *national* equals *sum of subnational* check

Table 6.1 shows EPI estimates of births in the provinces of a hypothetical country for three years, 2010, 2011 and 2012, and census projection estimates of total births for the same three years. The three bottom rows compare the sum of the EPI estimates over the 20 provinces with the national census projection estimate.

For each of the three years, the sum of the EPI provincial estimates is just under one third larger than the census projection estimate, as shown by the "Ratio" row in the lower part of the table. This indicates a large error in the census projection estimate, the EPI estimates, or (somewhat smaller) errors in both.

If there is no reason to believe that the accuracy of the estimates compared is different, their average may be taken as the best estimate of the true value. Error and relative error may then be calculated (section 2.3).

Table 6.1 shows adjusted estimates for 2012 only. The adjusted estimate of total births is the average of the original census and EPI estimates. The adjusted EPI estimates for each province are obtained by multiplying the given estimates by a factor that brings their total to this average—the average divided by the sum of the unadjusted EPI estimates. Adjusted estimates for other years may be calculated in the same way.

Table 6.1
Comparison and adjustment of EPI estimates of births for provinces and census projection estimate of total births

	Year			
	2010	2011	2012	Adj 2012*
EPI Estimate: Province 1	7,931	8,415	8,709	7,679
EPI Estimate: Province 2	11,717	12,819	13,267	11,698
EPI Estimate: Province 3	7,278	7,893	8,114	7,154
EPI Estimate: Province 4	35,751	40,061	41,425	36,524
EPI Estimate: Province 5	30,482	28,670	29,387	25,911
EPI Estimate: Province 6	14,116	14,089	14,484	12,771
EPI Estimate: Province 7	23,908	26,058	26,865	23,687
EPI Estimate: Province 8	16,117	15,946	16,456	14,509
EPI Estimate: Province 9	30,397	29,029	29,755	26,235
EPI Estimate: Province 10	36,438	37,506	38,557	33,996
EPI Estimate: Province 11	20,306	22,845	23,508	20,727
EPI Estimate: Province 12	54,636	46,373	47,392	41,786
EPI Estimate: Province 13	33,943	31,088	31,746	27,990
EPI Estimate: Province 14	14,030	13,487	13,770	12,141
EPI Estimate: Province 15	3,912	4,465	4,648	4,098
EPI Estimate: Province 16	27,760	27,338	28,022	24,707
EPI Estimate: Province 17	14,638	14,662	15,117	13,329
EPI Estimate: Province 18	21,680	20,345	20,772	18,315
EPI Estimate: Province 19	11,166	11,037	11,269	9,936
EPI Estimate: Province 20	4,480	5,137	5,338	4,707
EPI total	420,686	417,263	428,601	377,898
Census projection total	315,373	321,181	327,195	377,898
Ratio, EPI/projection	1.33	1.30	1.31	-
Average, EPI and projection	368,030	369,222	377,898	-
Error: EPI	52,657	48,041	50,703	-
Error: Projection	-52,657	-48,041	-50,703	-
Relative error: EPI	14.3	13.0	13.4	-
Relative error: Projection	-14.3	-13.0	-13.4	-
Ratio, Average/EPI	0.875	0.885	0.882	-

Notes Hypothetical Data. Error calculations and adjustment based on assumption that errors in EPI and census projection estimates are approximately comparable, so that best estimates of total births are averages of EPI and census projection totals. *Rightmost column shows adjusted estimates for 2012.

Chapter 7. Plotting and Analysing Time Series

Examining trends over time provides useful insights about possible errors in target population numbers. This chapter provides several examples.

7.1 Plotting coverage trends for multiple vaccines

Figure 7.1 plots coverage for Country B for 2000-2009 for five vaccines: BCG, DTP1, DTP3, MCV, and Pol3 (data from table 5.1). Data for HepB3 and Hib3 is available only for the last year of the period and therefore provides no trend information.

Showing coverage for the five vaccines together allows potentially useful comparisons between them. Because several values are above 100%, the vertical scale is extended and a reference line at 100% is provided.

Examination of figure 7.1 shows the following.

1. Administrative coverage rises rapidly overall during 2000-2006, drops sharply in 2007, and rises gradually overall during 2007-2009
2. Coverage rises from 20-40% in 2000 to over 100% in 2006. The programme evidently expanded very rapidly during these years.
3. Prior to 2005, coverage levels distinguish two groups of vaccines, DTP3-Pol3, and BCG-DTP1-MCV. The first group has a coverage of 23% in 2000 and rises slightly more rapidly. The second group has a coverage of about 35% in 2000 and rises slightly more slowly. This bifurcation disappears after 2005.
4. For 2006, all coverage levels—for BCG, DTP1, DTP3, MCV and Pol3—are over 100%. This suggests a problem with the denominator estimates.
5. Values over 100% disappear after 2006 with the minor exception of DTP1 in 2009. This suggests that a target population problem was recognized and addressed.

The very rapid rise during 2000-2006 in combination with the values uniformly over 100% at the end of this period suggest that target population estimates may have increased too slowly over this period.

The sharp decline in coverage between 2006 and 2007 might be real, but it might also reflect an upward adjustment of the target population in 2007, perhaps stimulated by the greater-than-100% coverage in the preceding year. From 2007, the upward trend in coverage continues at a slightly lower rate than before 2006.

7.2 Plotting target populations, vaccinations and coverage

Coverage for a vaccine is a quotient, the number of children in the target population who received the vaccine, the numerator, divided by the number of children in the target population, the denominator. Plotting numerator (target population),

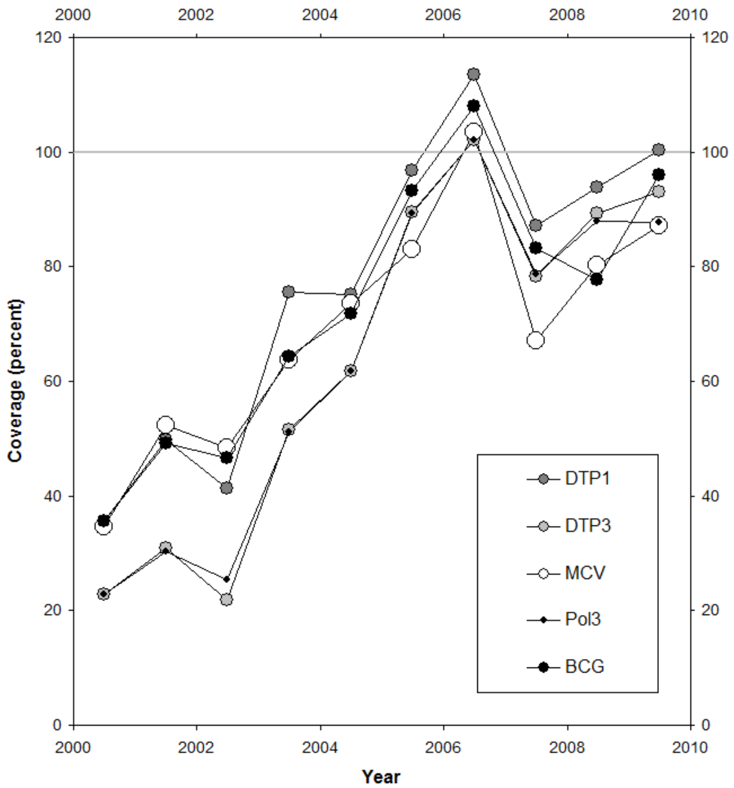

Figure 7.1
Coverage trends for five vaccines: Country B

denominator (vaccinations) and quotient (coverage) provides insight into errors in target population estimates.

Figure 7.2 plots target populations, vaccinations and coverage for BCG for Country D (table 5.1). The scale on the left vertical axis is for numbers of children vaccinated and numbers in the target population. The scale on the right vertical axis is for immunization coverage. Similar plots may be made for other vaccines.

The scale on the right has been chosen to facilitate comparison of the vaccinations and coverage series. The rationale for this is that target population numbers tend to change slowly and reasonably smoothly. When this is the case, fluctuations in coverage mainly reflect fluctuations in numbers of immunizations.

Examining figure 7.2 gives the following observations.

1. Target population numbers rise slowly from just under 400,000 in 2000 to about 450,000 in 2006, jump to about 600,000 in 2007, and rise approximately linearly and slightly more rapidly to just over 600,000 in 2011. A “blip” is seen for 2002 and a slight downturn between 2005 and 2006.

2. Numbers of children vaccinated rise rapidly from about 130,000 in 2000 to nearly 500,000 in 2006, remain near this level in 2007 and 2008, jump to nearly 600,000 in 2009.

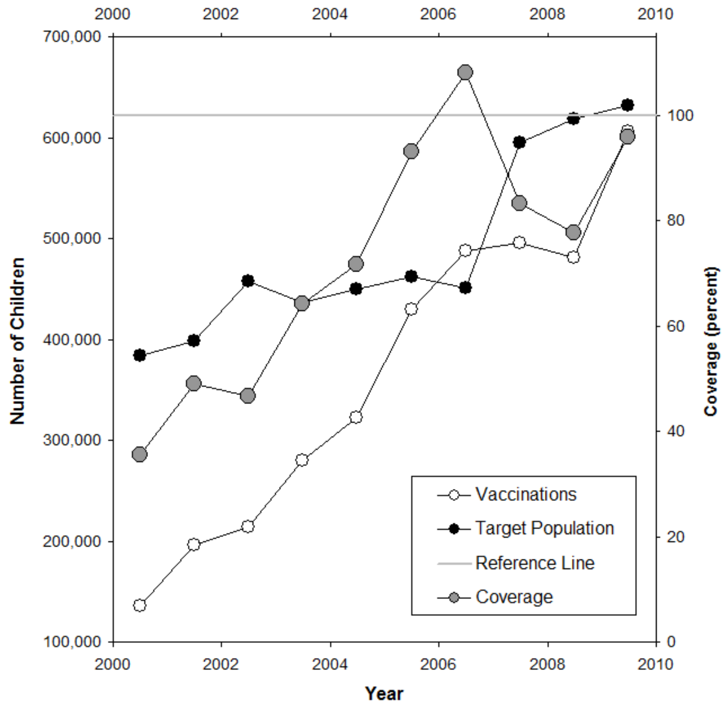

Figure 7.2
BCG target population, vaccinations and coverage:
Country B

3. Coverage rises rapidly from about 35% in 2000 to over 100% in 2006, then falls equally sharply between 2006 and 2007. A modest further decline between 2007 and 2008 is followed by a sharp jump to about 95% in 2009.

7.3 Example of trend analysis

The target population numbers in figure 7.2 may be viewed as an increase from just 400,000 in 2000 to about 600,000 in 2009, broken by a period of nearly constant numbers between 2002 and 2006.

It is possible that this progression is real, but the general tendency of target population numbers to change reasonably steadily suggests that the estimates for 2002-2006 err in not rising in line with the overall increase during the period.

If this is the case, better estimates of denominators between 2000 and 2007 might be given by linear interpolation between the values for these years, as shown by the dotted line in figure 7.3.

Recalculating coverage using using these alternative target population numbers gives the trend shown by the large grey circles in figure 7.3. The rise in coverage is still very rapid, but the level in 2006 is about 85% rather than over 100%.

The alternative target population numbers might be considered more plausible that the original numbers. The alternative coverage values are as plausible as the original

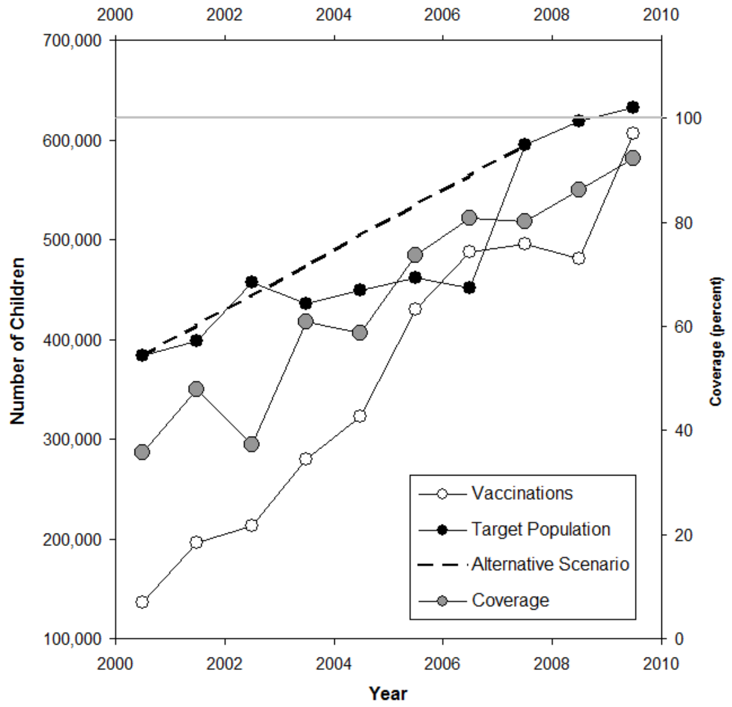

Figure 7.3
Alternative Scenario: BCG target population, vaccinations and coverage: Country B

values, with the advantage of remaining below 100%.

None of this establishes the validity of the alternative series, but it does make a case for investigating how target population numbers for 2000-2006 were estimated, an investigation that might conclude that the alternative numbers in figure 7.3 are better estimates than the original numbers.

Chapter 8. Checking Documentation

Immunization coverage data, like all data, consists of numbers whose meaning is defined by associated text matter. When data are presented in tabular form, this text matter includes the table title, column headings and rows labels, and table notes. When data are contained in a database, the text matter may be contained in one or more separate documents.

8.1 Documentation of geographic areas

Are the geographic area to which target populations and numbers of vaccinations refer clearly specified?

Target populations and vaccinations may be understood to refer to the entire country unless the contrary is specifically indicated.

Exclusion of parts of the country may occur in some years due to natural disasters or civil conflict. It may also occur because a vaccine is introduced in only part of the country prior to roll-out to the entire country. This is illustrated in table 3.1. The note to the table indicates that HepB was introduced in one province in 2003 and extended to the entire country only in 2006.

Geographic specification for the country as a whole tends not to be problematic because the national territory is defined by established international boundaries that change infrequently.

Geographic specification for subnational areas may be more difficult. Most if not all countries have an established division of the national territory into first level administrative units, with each first level administrative unit subdivided into second level units, and so on to smaller and smaller subnational geographic areas.

Establishing the geographic area corresponding to an administrative unit requires a map showing the boundaries of the unit. Up-to-date maps may not be available, particularly for smaller units.

Subnational administrative units may change over time, by sub-division of existing units into two or more new units, or by combinations of existing units to a single new unit. These changes may occur with or without changes in the boundaries of the existing units. When the changes do not involve any change in previously existing boundaries, definitions of new units formed by combination are given simply by the names of the units combined. Definitions of new units formed by subdivision require maps showing the new boundaries.

Even in the simplest case of a new unit formed by dividing an existing unit, confusion may occur. The name given to one of the new units, for example, may be the same as the name of the old, subdivided unit, with the consequence that the same name refers to different units at different points in time.

8.2 Documentation of undefined values

Is it clearly indicated which vaccines the programme provided in each calendar year?

If the EPI programme provided a particular vaccine in a particular year, the corresponding target population, number of vaccinations, and coverage rate are ***defined values*** for which numbers should be available.

If the EPI programme did not provide a vaccine in a particular year, the corresponding target population, number of vaccinations, and coverage rate are ***undefined values*** for which numbers should *not* be available.

8.3 Example of immunization coverage data: Country C

Table 8.1 presents an example of coverage data with a large number of missing values. Note however that every cell of the table not containing a number is identified either as an undefined value cells (HepB3 for 2000-2004 and Hib3 for 2000-2001) or as a missing value cell.

Table 8.1
Example of immunization coverage data: Country C

	Vaccine						
Year	BCG	DTP1	DTP3	MCV	Pol3	HepB3	Hib3
			TARGET POPULATION (000)				
2000	m	m	m	m	m	u	u
2001	m	894	894	894	894	u	u
2002	m	m	m	m	m	u	m
2003	894	894	894	894	894	u	m
2004	805	m	m	m	m	u	m
2005	m	805	805	805	805	805	m
2006	m	805	805	805	m	805	805
2007	m	894	894	894	m	894	894
2008	m	894	894	894	m	894	894
2009	m	894	894	894	894	894	894
			VACCINATIONS (000)				
2000	m	m	m	m	0	u	u
2001	715	626	m	m	0	u	u
2002	m	m	m	m	0	u	m
2003	715	715	m	m	0	u	626
2004	m	m	m	m	0	u	m
2005	805	626	179	m	0	179	715
2006	715	626	626	m	m	626	626
2007	805	715	715	m	m	715	715
2008	805	715	715	m	m	715	715
2009	894	805	805	m	0	805	805
			COVERAGE (%)				
2000	m	m	m	m	m	u	u
2001	m	73	68	82	65	u	u
2002	m	m	m	m	m	u	m
2003	m	81	75	69	74	u	m
2004	0	m	m	m	m	u	m
2005	m	107	83	91	84	27	18
2006	m	94	84	83	84	84	84
2007	m	96	84	85	84	84	84
2008	m	95	85	86	m	84	84
2009	m	96	92	88	84	87	87

Note Target population for BCG is number of births. Target population for other vaccines is number of surviving infants. HepB3 vaccination was introduced in 2005. Hib3 vaccination was introduced in 2002. “u” = undefined value “m” = missing value.

8.4 Documentation of target population and vaccinations specification

Immunization coverage is generally understood as the proportion or percentage of children in a specified target population who receive a particular vaccination. This characterization is conveniently simple, but it is not consistent with the coverage calculations made by many EPI programmes.

This section addresses subtleties of specifying target populations and groups of children vaccinated and of definitions of coverage rates. The quantitative impact of the distinctions may be negligible in some situations, but it is important in others. Definitional clarity is also important when designing special analyses, such as cohort studies, and when designing data structures for electronic immunization registries.

Numbers of children in target populations and numbers of children vaccinated may be obtained in two different and to some extent incommensurable ways, "aggregation by time period" and "aggregation by birth cohort". EPI programmes generally use data aggregated by time period. Target populations are numbers of births or children reaching age one year ("surviving infants") during some calendar year and numbers of vaccinations refer to vaccinations occurring in the same calendar year.

The mechanisms of data collection lend themselves to this mode of aggregation, and it has the advantage of providing statistics that refer to calendar years.

Coverage defined as

$$\frac{Vaccinations}{Target\ Population} \tag{8.1}$$

is not a proportion, however, and the relationship

$$Surviving\ Infants = Births - InfantDeaths \tag{8.2}$$

is in general false, though approximate equality may hold.

When events are aggregated by birth cohort, numbers of births are children born during a particular year, as for aggregation by time period, but numbers of surviving infants, infant deaths, and vaccinations refer to the cohort of children born in the given year. Children born during any year reach one year of age, if they survive, during the following year. Vaccinations of these children before they reach one year of age may occur either in the year they were born or in the following year, depending on when the child was born and the age at which the child was vaccinated.

Vaccinations of a child born on the first day of the year, for example, will occur in this year, but vaccinations of a child born on the last day of the year (BCG excepted if given at birth) will occur in the following year. Children born during the middle six months of the year will receive some vaccinations during the year in which they were born and some in the following year. The same is true of infant deaths: some occur in the year of birth, some in the following year.

Formula (8.2) is valid when the counts of surviving children and infant deaths are aggregated by cohort, provided that there is no migration of children. Even with cohort aggregation, the ratio (8.1) is not strictly a proportion because some of the children counted in the numerator will not survive to one year of age and so will not be counted in the denominator. A proportion is by definition a quotient in which the numerator counts a subgroup of the group counted by the denominator. The effect of this difference on the calculated level of coverage will usually be negligible because the number of children involved will usually be very small.

Aggregation by time period and aggregation by birth cohort give numbers of vaccinations and infant deaths that are not directly comparable, but it is possible to establish a relationship between them by considering the following four groups of vaccinations.

(1) Vaccinations in year Y to children born in year $Y - 1$

(2) Vaccinations in year Y to children born in year Y

(3) Vaccinations in year $Y + 1$ to children born in year Y

(4) Vaccinations in year $Y + 1$ to children born in year $Y + 1$

Combining groups (1) and (2) gives vaccinations in year Y, corresponding to period aggregation. Combining groups (2) and (3) gives vaccinations of the cohort of children born in year Y. Combining groups (3) and (4) gives vaccinations in year $Y + 1$.

Period and cohort aggregation give identical results for what is known in demography as a “stationary” population, for which births and mortality risks are constant over time. For a stationary population, the number of children born each year is constant and the number of events in groups (1) and (3) are identical. It follows that numbers of events given period and cohort aggregations are the same.

EPI programmes may calculate immunization coverage using formula (8.1) with vaccinations and target population obtained by period aggregation. The quotient in this case is a ratio, not a proportion. If the population is stationary, the ratio is identical to the proportion given by formula (8.1) when vaccinations and target population are obtained by cohort aggregation.

For non-stationary populations, the ratio may closely approximate the proportion. In some circumstances, however, the difference between period and cohort aggregation is crucial. Suppose for example that a new subnational administrative area is created in a particular year. There will be no births in this area during the preceding year because the area did not exist. The number of surviving infants for the given year is therefore zero and formula (8.1) cannot be used.

Coverage proportions can be calculated from data aggregated by time period. The method, known as a “product-limit life table”, is far too complicated to be used with traditional coverage data, but could be applied to data from electronic immunization registries that contain full dates (year/month/day) of birth, vaccinations, and infant death.

EPI programmes may calculate numbers of surviving infants from formula (8.2) using numbers of infant deaths aggregated by time period. In the population is stationary, the result will be precisely what is wanted, the number of children reaching age one year during the given year. For non-stationary populations, the result may closely approximate this number. In some circumstances, however, the approximation may be poor, as illustrated by the above example of a newly created subnational area.

Chapter 9. Checking Missing Values

Missing values indicate problems with EPI programme procedures for estimating numerators, denominators, and immunization coverage. This chapter presents methods for analysing missing values in the estimates. The results help identify areas in which procedural improvements are needed.

Identifying missing values requires distinguishing between defined and undefined values. If this is not possible, the frequency of missing values is indeterminate. The documentation check described in section 8.2 should therefore be applied and any inclarity rectified before applying the checks described in this chapter.

The following section describes how to calculate the frequency of missing values in a given set of coverage data and provides examples of the calculation. Section 9.3 discusses the analysis of missing values when coverage data is stored in computer databases.

9.1 Definition of missing values

Missing values are defined values (see 8.2) for which no number is available. Missing values are numbers that should be, but are not, available. Undefined values are "numbers that should not be available". A number provided for an undefined value is a defect in the data, just as a missing value is a defect.

When data are presented in tabular form, it is convenient to refer to cells of the table as ***defined value cells*** and ***undefined value cells***. Defined value cells may be subdivided into ***missing value cells*** and ***non-missing value cells***.

9.2 Frequency of missing values

What is the frequency of missing values?

The frequency of missing values may be defined as the number of missing values divided by the number of values that should not be missing, that is, defined values. Note that the denominator excludes undefined value cells.

For the Country C data in Table 8.1, determination of the frequency of missing values is complicated by the large number of missing values. There are 27 missing values for the target population panel of the table, 26 for the vaccinations panel, and 24 for the coverage panel, a total of 77 missing values. There are 65 defined value cells in each of the three panels, for a total of 195 defined value cells and a missing value frequency of 76/195 = 0.390, or 39%.

For the Country B data in Table 5.1, there are only 3 missing value cells and 12 undefined value cells in each panel. There are 70 cells total in each panel (10 rows $\times$ 7 vaccines), so the number of defined value cells is 70 - 12 = 58. The frequency of missing values is 3/68 = 0.044, or 4.4%.

Table 9.1
Selected target populations for Country B:
Standard tabular format

	Vaccine			
Year	BCG	Pol3	HepB3	Hib
2005	1,118	1,045	772	u
2006	1,138	m	1,064	u
2007	1,083	m	1,083	u
2008	1,102	m	1,102	u
2009	1,013	1,013	1,013	1,013

Note See note to table 5.1. BCG target population = births. Other target populations = surviving infants. u = undefined, m = missing.

For the Country A data in Table 3.1 of the following chapter, there are a total of 13 missing value cells and 183 defined value cells for a missing value frequency of 0.71, or 7.1%.

9.3 Missing values in database-format tables

Do computer database tables include rows for defined values that are missing?

When the immunization coverage data are stored in a computer database, the format of the tables may change. Table 9.1 shows a subset of data in the first panel of table 5.1.

Table 9.2 shows the same information in a format typical of computer databases. It contains one row for each defined value cell of table 9.1.

The ordering of rows in Table 9.2 corresponds to reading Table 9.1 from left to right and top to bottom, but the order of rows in database format tables is not considered to be significant. They are intended for processing by computer database applications, which may reorder rows on the basis of values in any of the columns.

Table 9.2 includes rows for missing value cells ("m") in table 9.1. A database format table may omit rows for undefined values ("u")without loss of information, but it should not omit rows for missing values. Missing values provide important information. If missing value rows are not shown, the frequency of missing values cannot be determined.

Table 9.2
Target population data for Country B:
Database table format

Year	Vaccine	Target Population	Geographic Area	Number
2005	BCG	Births	Country	1118
2005	Pol3	Surviving Infants	Country	1045
2005	HepB3	Surviving Infants	Provinces A/B/C	772
2006	BCG	Births	Country	1138
2006	Pol3	Surviving Infants	Country	m
2006	HepB3	Surviving Infants	Country	1064
2007	BCG	Births	Country	1083
2007	Pol3	Surviving Infants	Country	m
2007	HepB3	Surviving Infants	Country	1083
2008	BCG	Births	Country	1102
2008	Pol3	Surviving Infants	Country	m
2008	HepB3	Surviving Infants	Country	1102
2009	BCG	Births	Country	1013
2009	Pol3	Surviving Infants	Country	1013
2009	HepB3	Surviving Infants	Country	1013
2009	Hib3	Surviving Infants	Country	1013

Note Data in table 9.1 rendered in database table format. Each row of this table corresponds to a defined value cell of table 9.1. m = missing.

Chapter 10. Checking Consistency

This chapter presents several consistency checks that can be applied to immunization coverage data. Inconsistencies may be useful in identifying the source of errors in target population estimates and for assessing the quality of the EPI programme operations that produce the estimates.

10.1 Consistency of target populations

For each year, are the target populations for non-BCG vaccines identical? Or if not identical, does documentation indicate why?

Table 10.1 shows multiple non-BCG target populations numbers for every year but the last. Because no explanation for the differences is given, all non-BCG target populations in the same row should be equal. These unequal values are inconsistencies.

If the target population for vaccines other than BCG is the number of surviving infants, the target population table should show identical numbers for non-BCG denominators in every row. Or, if this is not the case, the note(s) to the table should explain or refer to an explanation for the exceptions.

The frequency of target population inconsistencies is a useful summary of the check, but the calculation is less than obvious because the check applies to pairs of table cells rather than to individual cells.

A row of identical target populations indicates no inconsistency. A row containing two distinct target population numbers indicates a single inconsistency, a row containing three distinct values indicates two inconsistencies, and so on. The number of inconsistencies in a row is one less than the number of distinct values in the row. The total number of inconsistencies may therefore be calculated as the number of distinct non-BCG target population estimates in each row, minus one, summed over all rows.

In table 10.1 there are 4 distinct values in rows 1 and 2 giving $(4 - 1) \times 2 = 6$ inconsistencies; 3 distinct values in rows 3, 5, 6, 7 and 9 giving $(3 - 1) \times 5 = 10$ inconsistencies; 2 distinct values in rows 4 and 8, giving 2 inconsistencies; and no inconsistencies in row 10; for a total of 18 inconsistencies.

The maximum possible inconsistencies in any row is one less than the number of defined value cells in the row for non-BCG vaccines. For table 8.1 this $(5-1)\times 9 = 36$ for the first 9 rows, and there are no inconsistencies in the 10th row of the table.

The frequency of inconsistencies is thus $18/36 = 50\%$.

10.2 Consistency of missing values: First check

Does every missing value cell in the coverage table correspond to a missing value cell in the target population table OR in the vaccinations table, or both?

Table 10.1
Example of Immunization Coverage Data: Country D

	Vaccine						
Year	BCG	DTP1	DTP3	MCV	Pol3	HepB3	Hib3
			TARGET POPULATION (000)				
2000	285	308	291	256	291	285	u
2001	261	296	296	186	285	291	u
2002	244	279	314	308	314	314	u
2003	256	256	238	256	238	m	u
2004	238	238	244	267	244	244	u
2005	209	221	227	244	227	227	u
2006	215	215	221	227	m	221	u
2007	203	203	203	221	m	203	u
2008	192	192	198	209	m	198	u
2009	198	198	198	198	198	198	198
			VACCINATIONS (000)				
2000	256	296	285	238	285	m	u
2001	221	285	291	174	279	m	u
2002	232	273	308	296	308	302	u
2003	244	250	232	238	232	m	u
2004	232	232	238	261	238	238	u
2005	203	221	227	238	227	227	u
2006	209	209	215	215	m	215	u
2007	203	203	203	215	m	198	u
2008	192	192	198	203	m	198	u
2009	192	192	192	192	192	192	192
			COVERAGE				
2000	90	95	98	92	98	m	u
2001	84	97	98	95	98	m	u
2002	94	98	98	96	98	96	u
2003	95	97	97	93	97	m	u
2004	97	98	97	96	98	99	u
2005	98	99	98	97	97	98	u
2006	97	98	97	95	97	98	u
2007	98	98	98	97	99	98	u
2008	100	99	99	98	m	99	u
2009	97	99	98	97	98	98	98

Note Target population for BCG is number of births. Target population for other vaccines is number of surviving infants. “u” = undefined value.

Coverage for a particular vaccine and year can be calculated only if the number of vaccinations and the target population are available. A missing value cell in the coverage table for which corresponding cells in the target population and coverage table both contain numbers is one of two possible missing value inconsistencies.

In Table 8.1, for example, coverage for BCG in 2003 is missing, but the corresponding target population (894) and vaccinations (715) are both given. This is an inconsistency because coverage can be calculated—715/894 = 80%.

Examination of the table shows that this is the only such inconsistency for the 24 missing value cells. The frequency of these inconsistencies is thus $1/24 = 0.042$, or 4.2%.

10.3 Consistency of missing values: Second check

Does every cell in the coverage table that contains a number correspond to cells in the target population and coverage tables that contain numbers?

The second missing value consistency check is the logical inverse of the first. A number in a cell in the coverage table implies the presence of numbers in the corresponding cells of the target population and coverage table. The second missing value inconsistency occurs when this is not the case.

In table 8.1, for example, a coverage of 68% is shown for DTP3 for 2001, but the corresponding number of vaccinations is shown as missing. Examining the 38 coverage values in the table shows 13 additional instances of this inconsistency: BCG for 2004; DTP3 for 2003; Hib3 for 2005; MCV for 2001, 2003, and 2005-2009; and Pol3 for 2005-2007; a total of 14 inconsistencies.

The frequency of these inconsistencies is their number divided by the number of non-missing value cells in the coverage panel of the table, 38 in this case. The frequency of these inconsistencies is thus $14/38 = 0.368$, of 37%.

Part II

Improving Accuracy

Checklist for Improving Accuracy

Begin by specifying the consecutive calendar years for which target population estimates are needed. It is advisable to generate estimates for a minimum of five calendar years.

STEP 1 Obtain estimates of ***national numbers of births*** for the required calendar years.

Step 1a Are national annual numbers of births available from civil registration? If yes, use these numbers for all years for which they are available and estimate national births for subsequent years as described in Chapter 14, Estimating Future Births. If no, go to *Step 1b*.

Step 1b Are estimates of national annual numbers of births available from population projections based on a census taken within the past 15 years (see example in Table 11.1)? If yes, use these estimates. If no, go to *Step 1c*.

Step 1c Is an estimate of national births and/or the national Crude Birth Rate (CBR) available from a census taken within the past 15 years? If yes, use the published estimate of births or births calculated as $Total\ Population \times Crude\ Birth\ Rate$ for the census year and estimate births for future years as in Step 1a. If no, go to *Step 1d*.

Step 1d Are survey estimates of CBR around the time of the census available? If yes, estimate the census year CBR as described in Section 13.3, Estimating crude birth rates (CBRs) from surveys and estimate births as in *Step 1c*. If no, go to *Step 1e*.

Step 1e Is it feasible to estimate national births for the census year by the method of Section 14.5, Growth rates by abbreviated projection (this method requires expertise that may not be available in the EPI Programme)? If yes, use this method. If no, go to *Step 1f*.

Step 1f Use *World Population Prospects* estimates of national annual births (see Annotated List of Resources).

STEP 2 Obtain estimates of ***subnational numbers of births*** for the required calendar years.

Step 2a Are subnational annual numbers of births available from civil registration? If yes, use these estimates through the most recent available year. Estimate national numbers of births for future years by the method of Chapter 14, Estimating Future Births. Estimate subnational births for future years by the method of Chapter 15, Estimating Births for Subnational Areas. If no, go to *Step 2b*.

Step 2b Are subnational annual numbers of births available from population projections based on a census taken within the past 15 years (as for the Malawi example shown in Table 11.1)? If yes, use these estimates. If no, go to *Step 2c*.

Step 2c Estimate subnational annual numbers of births as described in Chapter 15, Estimating Births for Subnational Areas.

STEP 3 Estimate annual subnational numbers of surviving infants from annual subnational numbers of births by the method of Chapter 16, Estimating Surviving Infants. Calculate national annual national surviving infants by summing subnational numbers.

Chapter 11. Introduction to Improving Accuracy

EPI programmes will preferably obtain the target population estimates they need from the country's National Statistical System (NSS), with no need for primary data collection or estimation by the programme. Few EPI programmes have the expertise and resources needed to produce accurate target population estimates independently.

The actions an EPI programme may take to obtain more accurate target population estimates depend on the existing situation. What primary data sources are being used? What methods are being used to produce estimates from these data? Are there other sources and/or methods that may produce more accurate estimates?

Most national statistical systems include a "National Statistical Office" (NSO) or "Statistical Institute" designated by legislation as responsible for overseeing and coordinating the production of official statistics. The name of this organization varies from country to country. "NSO" will be used here as a generic designation.

The first task of an EPI programme aiming to improve the accuracy of target population estimates is to learn what resources the national statistical system can provide for estimation of target population numbers. Chapter 12 suggests a step-by-step approach to working with the national statistical system to accomplish this task.

This chapter provides brief overviews of the three principle sources for target population estimates: a civil registration and vital statistics system, population projection estimates, and a population census taken within the past 15 years.

11.1 Target population estimates from civil registration data

The most accurate target population estimates will generally be provided by a well-developed civil registration and vital statistics (CRVS) system. These systems collect information on births and deaths throughout the country as the events occur. This makes it possible to obtain the most accurate, current, and geographically detailed data possible.

The registration method is inherently superior to using retrospective questions in population censuses or surveys because it does not require respondents to remember information about events that may have occurred a year or more ago.

It is however necessary that the CRVS system register all or nearly all births and infant deaths. Infant deaths are required for the calculation of infant mortality rates, which are in turn required for the calculation of surviving infants.

CRVS system data for a calendar year tends to be published in the following calendar year, but delays in processing and publication may result in later publication. If the EPI programme requires estimates for later years, national level estimates may be estimated as described in Chapter 14, Estimating Future Births. Estimates may also be produced by extrapolation of annual numbers births and infant mortality rates from the CRVS.

If CVRS numbers are unpublished or published in insufficient detail, the EPI programme may work with CRVS authorities toward having suitably detailed data made available in the future.

Use of civil registration data for subnational areas requires attention to the difference between place of occurrence of a birth or infant death and place of usual residence of the mother of the child. The administrative area in which a birth occurs may not be the same as the area in which the mother lives. A mother living in an administrative area with no hospital, for example, may have travelled to give birth in a hospital in another area. Differences between place of occurrence and place of usual residence tend to be larger for smaller administrative areas.

Births classified by place of occurrence may be the most appropriate target populations for BCG. Births classified by place of usual residence of mother may be more appropriate target populations for vaccinations given later in infancy.

11.2 Target population estimates from population projections

If estimates of births and infant deaths cannot be obtained from a civil registration and vital statistics system, the most accurate estimates will come from population projection estimates based on a national population census taken within the past 15 years.

It is however necessary that the population projections provide estimates of annual numbers of births and infant mortality rates, as in the examples of Table 11.1 and Table 16.1. Births should *never* be estimated by applying a fixed "conversion factor" to projection estimates of total population, as this will usually give severely erroneous estimates (section 13.2).

Population projections may be produced by national statistical authorities following a population census. Given the time required to process the census returns and produce the projections, however, population projection results may not be available for several years following the census.

Population projection estimates are an expedient to be used pending development of a civil registration and vital statistics system. They cannot be expected to be as accurate as estimates from a well-developed CRVS system. The accuracy of population projection estimates may be high for the country as a whole for the years immediately following the census, but accuracy tends to decline for future years and to be lower for smaller subnational areas.

Population projection methodology makes it possible to produce many of the target population estimates EPI programmes need, but published reports reports may not include these estimates. Estimates presented in population projection reports vary widely along three main dimensions.

1. *What estimates are presented?* Publications may present estimates of population by sex and age groups only, or estimates of a wide array of demographic indicators, including annual births and infant mortality rates.

Table 11.1
Projection Estimates of Annual Births:
Malawi, Districts, and Cities, 2008–2016

	Year						
District/City	2008	2009	2010	2011	2012	2013	2014
Malawi	609,487	619,587	629,928	642,301	655,231	668,801	682,962
Chitipa	9,011	9,180	9,353	9,521	9,686	9,848	10,003
Karonga	12,509	12,739	12,970	13,193	13,425	13,670	13,928
Nkhata Bay	9,232	9,481	9,739	10,002	10,269	10,539	10,809
Rumphi	7,710	7,857	8,005	8,153	8,298	8,441	8,579
Mzimba	31,112	31,795	32,503	33,234	33,992	34,777	35,585
Likoma	375	373	372	372	371	370	369
Mzuzu City	7,000	7,464	7,929	8,392	8,854	9,316	9,775
Kasungu	31,981	32,771	33,591	34,445	35,346	36,300	37,304
Nkhotakota	16,720	14,845	12,922	12,783	12,728	12,741	12,826
Ntchisi	12,112	12,346	12,584	12,826	13,072	13,323	13,579
Dowa	27,701	28,742	29,789	30,833	31,870	32,893	33,900
Salima	16,741	17,023	17,324	17,643	17,983	18,352	18,747
Lilongwe	60,929	61,537	62,202	62,910	63,693	64,553	65,483
Mchinji	23,452	24,008	24,582	25,172	25,783	26,413	27,060
Dedza	29,586	29,869	30,163	30,467	30,793	31,146	31,526
Ntcheu	22,238	22,529	22,848	23,177	23,509	23,841	24,171
Lilongwe City	30,165	31,757	33,321	34,839	36,361	37,906	39,473
Mangochi	36,933	37,756	38,615	39,531	40,518	41,589	42,744
Machinga	22,513	22,890	23,282	23,706	24,176	24,704	25,289
Blantyre	14,343	14,578	14,822	15,075	15,334	15,597	15,861
Zomba Rural	26,275	26,405	26,556	26,733	26,934	27,163	27,420
Chiradzulu	12,328	12,357	12,391	12,427	12,468	12,514	12,564
Mwanza	4,600	4,632	4,667	4,700	4,733	4,763	4,791
Thyolo	23,468	23,824	24,191	24,555	24,941	25,352	25,786
Mulanje	23,326	23,327	23,336	23,360	23,410	23,496	23,616
Phalombe	14,635	14,750	14,865	15,000	15,168	15,383	15,646
Chikwawa	20,080	20,387	20,708	21,047	21,417	21,826	22,271
Nsanje	11,845	11,985	12,142	12,316	12,509	12,721	12,950
Balaka	14,648	14,976	15,328	15,698	16,084	16,485	16,898
Neno	5,082	5,422	5,759	6,089	6,413	6,729	7,035
Zomba City	4,020	4,300	4,582	4,858	5,128	5,389	5,641
Blantyre City	26,817	27,682	28,487	29,244	29,965	30,661	31,333

Source *Population Projections for Malawi*, Annex A2, pages 56-87.

2. *For what times/time periods are estimates presented?* Publications may present estimates only at five year intervals and for five year age groups, or for calendar years and at one year intervals.

3. *For what geographic areas are estimates presented?* Publications may present estimates for the whole country only, or for first, second, and lower level administrative areas.

Table 11.1 illustrates the detail population projection methodology is capable of providing. The report from which these estimates are taken does not provide corresponding estimates of surviving infants, but estimates of surviving infants may be calculated by the method of Chapter 16, Estimating Surviving Infants.

If population projection estimates are available but do not include annual births and infant mortality rates, the EPI programme may work with the statistical authority that produces the projections to see if these estimates can be made available in the future.

11.3 Target population estimates from population census data

Producing population projection estimates is a complex process best carried out by the most expert persons available in the national statistical system. If population projections based on the most recent population census are not yet available, however, or if they are available but do not provide the estimates the EPI programme requires, and the most recent population census was taken within the past 15 years, the EPI programme may calculate estimates from data provided this census. The estimation requires the following steps.

1. Estimate the number of births for the country as a whole for the census year from available census data by one of the methods of Chapter 13, Estimating National Census Year Births
2. Estimate national numbers of births by one of the methods of Chapter 14, Estimating Future Births.
3. Estimate numbers of births in subnational areas by the method of Chapter 15, Estimating Births for Subnational Areas.
4. Estimate numbers of surviving infants by the method of Chapter 16, Estimating Surviving Infants.
5. Calculate the annual number of surviving infants for the country as a whole as the sum of the numbers for subnational areas.

EPI programmes lacking the resources and/or expertise to produce target population estimates in this way may seek assistance from the NSO or another source of statistical expertise, such as a university or research institute.

11.4 Target population estimates from *World Population Prospects*

If target population estimates cannot be obtained from the sources indicated in the previous three sections, they may be obtained from *World Population Prospects* (Annotated List of Resources).

Chapter 12. Working with the National Statistical System

The national statistical system of a country consists of organizations that collect, process and disseminate official statistics on behalf of national government, and the legal and statistical infrastructure that facilitates this work.

Official statistics may be produced and disseminated by many different government and quasi-government entities. Most national statistical systems include a "national statistical office" or "statistical institute" designated as responsible for overseeing and coordinating the production of official statistics. As mentioned in chapter 11, the name of this organization varies from country to country. "NSO" is used in this manual as a generic designation.

The literature of official statistics emphasizes that these statistics are produced for the benefit of their users. It also emphasizes that consultations with users are a fundamental part of the work of official statistical organizations. EPI programmes should be aware of the *Fundamental Principles of Official Statistics* promulgated by the United Nations Statistical Commission (section 12.1).

An EPI programme aiming to improve the accuracy of its target population estimates may begin by communicating its statistical needs to the NSO and meeting with NSO staff to ensure that these needs are understood. This work is likely to be most effective if the programme begins by documenting its needs in a form suitable for presentation to the NSO (section 12.2).

The next step is to identify national statistical resources that may meet these needs (Section 12.3).

An action plan may then be formulated for work to obtain the most accurate target population estimates the national statistical system can provide (Section 12.4).

12.1 The *Fundamental Principles of Official Statistics*

The United Nations Statistical Commission is the leading organization of the global statistical system. Established in 1947, it brings together heads of national statistical offices from member states throughout the world. It is concerned with setting international statistical standards, the development of concepts and methods, and implementation of these concepts and methods in national statistical systems and international organizations. The Statistical Commission oversees the work of the United Nations Statistics Division (UNSD).

The *Fundamental Principles of Official Statistics*, first promulgated by the Statistical Commission in 1994, is intended to guide and inform the production and dissemination of official statistics. Two principles are particularly relevant to EPI programme managers and staff.

Principle 1 states in part that "... statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information".

Principle 3 states that "To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics".

12.2 Defining target population estimates required

Documentation of EPI programme statistical needs will answer at least the following questions.

Question 1. What numbers are required? The usual approach to calculation of coverage rates requires numbers of births and surviving infants. The specification should indicate precisely what kinds of denominators the programme needs for its calculation of immunization coverage.

Question 2. For what time periods are these numbers required? The usual answer to this question will be calendar years.

Question 3. For what geographic areas are these numbers required? The question is simply stated, but the answer may be complex. The Standard Geographical Classification (SGC) published by Statistics Canada, for example, runs to nearly 600 pages of details on the system of classification of geographic units, lists of place names, and standard codes.

Question 4. When are the estimates for each time period needed? Processing primary data may be time-consuming. Results may not be available for months or years after data collection ends. Recognizing this, the needs statement should state when estimates for particular years, including future years, are needed.

12.3 Identifying national statistical resources

The next step is to deetermine the extent to which the national statistical system can meet the needs identified. Identifying and reviewing pertinent statistical publications is useful preparation for meetings with statistical authorities. It is possible that some of the target population numbers needed will be found in these publications.

Initial meetings may best be arranged at the highest level, as for example by the Minister of Health approaching the head of the national statistical office. These initial high level meetings may be followed by lower level working meetings between national statistical professionals and EPI programme staff.

The first question to be addressed will be whether there is a civil registration system that can provide accurate numbers of births for some or all parts of the country. This information may require meetings with civil registration authorities rather than with the national statistical office.

If estimates cannot be provided by civil registration, the next questions will be whether there are population projection estimates providing annual numbers of births and infant mortality rates for the country as a whole, whether the projections provide estimates for subnational areas, and if so for what subnational areas.

12.4 Action plan and follow up

The aim of these consultations is to arrive at an action plan, agreed on by the EPI programme (or perhaps the Ministry of Health) and the national statistical organizations involved, aimed at generating the most accurate target population estimates possible from existing national statistical resources.

Chapter 13. Estimating National Census Year Births

This chapter presents methods for estimating total births during the year a national population census was taken from data available in census publications, perhaps supplemented with data from other sources. Use of these methods is appropriate when the following conditions are met.

1. Target population estimates cannot be obtained from a civil registration and vital statistics (CRVS) system
2. Target population estimates cannot be obtained from population projection estimates based on a national population census taken during the last 10-15 years
3. Published data from a national population census taken during the last 10-15 years is available

13.1 Estimating births from population census data

National population censuses may collect retrospective information on births. The number of births a woman had during the 12 months prior to the census is a common example.

Reporting of births in response to these questions may however be substantially incomplete. Methods exist for estimating and correcting for the omission, but use of these methods requires expertise in what are known in demography as "indirect methods". Census reports may provide estimates of births based these methods.

EPI programmes should not use unadjusted numbers of births from population census reports. To illustrate how important this may be, the *General Population Census of Cambodia 2008 National Report on Final Census Results* gives an unadjusted reported Crude Birth Rate (CBR) of 13.0 per thousand and an adjusted CBR of 26.9 per thousand. The *reported* number of births is less than half the *adjusted* number.

EPI programmes should not use reported numbers of infants from census age distributions. These numbers also may be substantially under reported. Methods exist for adjusting reported age distributions, but use of these methods requires demographic expertise.

13.2 Do not use fixed "conversion factors" to estimate births

The accuracy of numbers of annual births calculated as total population times a CBR estimate depends on the accuracy of the CBR estimates. Using a fixed "conversion factor" rather than a best estimate of the CBR will usually result in severely erroneous estimates of births.

Demographic conditions in the world today are far too variable, between countries, within countries, and over time, for any fixed conversion factor to provide acceptably accurate estimates. Thus, fixed conversion factors should not be used despite their recommendation by both *Immunization in Practice* (Geneva: World Health Organization 2004) and *Immunization Essentials* (Washington DC: US Agency for International Development 2003).

13.3 Estimating crude birth rates (CBRs) from surveys

Population census reports will always include the total population enumerated by the census, but they may not include estimates of births or CBR. In this situation it may be possible to estimate the CBR from survey estimates

Because such surveys are rarely taken in the same year as a population census, it will usually be necessary to obtain CBR estimates from two surveys and interpolate between or extrapolate beyond them to estimate a CBR for the census year.

> Consider for example the 2008 population census of Cambodia. The *General Population Census of Cambodia 2008 National Report on Final Census Results* shows an enumerated population of 13,395,682 persons (Chapter 2, page 19).
>
> The *Cambodia Demographic and Health Survey 2005* shows a CBR of 25.6 births per thousand population (table 5.1, page 61) referring to the three years preceding the survey.
>
> Survey fieldwork was carried out from 9 September 2005 to 7 March 2006 (page 7). The period to which the CBR refers may be taken as the three year period ending at the midpoint of fieldwork period, which is approximately the point in time 2005.9. The CBR of 25.6 thus refers to the period 2002.9 to 2005.9 and may be identified with the midpoint of this period, 2004.4.
>
> The *Cambodia Demographic and Health Survey 2010* shows a CBR of 23.9 (table 5.1, page 57), also referring to the three years preceding the survey. Data collection was conducted from 23 July 2010 to 20 January 2011 (page 6). Following the procedure of the previous paragraph, the CBR of 23.9 refers to the three year period ending 2010.8 and is identified with the midpoint of this period, 2009.3.
>
> Interpolating between the CBRs from the 2005 and 2010 surveys gives a CBR of 24.8 per thousand population for the year of the census, 2008. The interpolation procedure is described below (*Annex* Interpolation/extrapolation of crude birth rates). Multiplying the census population in thousands by this CBR gives an estimate of 332,221 births.
>
> The interpolation from the *Demographic and Health Surveys* is unnecessary in this case because the census report gives an estimated CBR of 23.9 per thousand.

In addition to illustrating the method, therefore, this example suggests the accuracy that may be expected of crude birth rates estimated in this way. The difference between the estimate published in the census report, 23.9, the estimate interpolated from the surveys, 24.8, is 0.9, or a relative error of 4%.

13.4 Alternative method of estimation from population census data

An alternative to the method of section 13.1 above is given in section 14.5.

13.5 *Annex* Interpolation/extrapolation of crude birth rates

This section describes a procedure for estimating the CBR for any year from crude birth rates from two sample surveys.

For the purpose of interpolation, the CBR for any time period may be identified with the time that is the mid-point of the period. The 2005 *Demographic and Health Survey* of Cambodia, for example, provides a CBR of 25.6 births per thousand population for 2004.4, that is, the point in time four tenths of a year into calendar year 2004, and the 2010 survey provides an estimate of 23.9 per thousand as of 2009.3.

"Interpolation" and "extrapolation" are methods for using estimates for two points in time to obtain estimates for an in-between point in time (interpolation) or for a later point in time (extrapolation). They involve the following steps.

1. Calculate the annual rate of change of the CBR indicated by the two surveys. The change in the above example is $23.9 - 25.6 = -1.10$. The sign is negative, indicating decline. This change occurs between 2004.4 and 2009.3, an interval of $2009.3 - 2004.4 = 4.9$ years, so the annual rate of change is $-1.10/4.9 = -0.22$ CBR points per year.

2. Identify the time for which an interpolated/extrapolated CBR is wanted. In this case the CBR for the 2008, the yeaer of the most recent Cambodian population census is wanted. As before, the time interval is identified with the point in time that is its midpoint, so the CBR to be estimated is for time 2008.5.

3. Calculate the interval between the time to which the first given CBR value refers and the time for which the estimate is to be calculated. In this example, $2008.5 - 2004.9 = 3.6$ years.

4. Calculate how much the CBR will increase over this period, given the calculated rate of change. In this example, $3.6 \times -0.22 = -0.79$ in this case.

5. Add this increase to the first given CBR. In this example, $25.6 + -0.79 = 25.6 - 0.79 = 24.8$ per thousand.

Chapter 14. Estimating Future Births

This chapter presents methods for estimating numbers of future annual births for the country as a whole from the number of births in a "base" year and supplementary information.

If births come from civil registration (section 11.1), the **base year** will be the most recent year for which data are available. If births come from national population census data (section 11.3), the **base year** will be the year of the census.

The method may be used to estimate numbers of births for any number of years into the future, but the accuracy of the estimates tends to decline for each successive year. The method should not be used to estimate births more than ten years into the future and is preferably used to estimate births not more than five years in the future.

14.1 Estimating future births by growth rate extrapolation

Numbers of births for years following the base year are estimated as follows.

(1) Estimate the annual growth rate of births

(2) Apply this growth rate to the number of births in the base year to estimate births in the first year after the base year

(3) Apply the growth rate to this number of births to estimate the number of births in the following year

(4) Repeat this process to estimate numbers of births in future years

Suppose that the base year is 2010, that there are 32,345 births in this year, and that the growth rate of births is 0.0201 per annum. The factor by which numbers of births are multiplied is one plus the growth rate of births.

Numbers of births in 2011, 2012, . . . are estimated by

$$Births\ in\ 2011 = 32,345 \times 1.0201 = 32,878$$

$$Births\ in\ 2010 = 32,878 \times 1.0201 = 33,419$$

$$Births\ in\ 2011 = 33,419 \times 1.0201 = 33,969$$

and so on for as many years as required.

These calculations are very simple. The challenge in practice is obtaining a suitable growth rate. The following sections discuss several possibilities.

14.2 Growth rates from civil registration

If national base year births come from civil registration, it is likely that numbers for previous years are available. Numbers of births for previous years may be used to estimate the growth rate of births. The simplest approach is to calculate the growth rate of births for the last two years for which numbers of births are available.

If births are available for a longer series of years it is advisable to calculate growth rates of births for each pair of successive years and examine the variability of these rates. The growth rate of births used to calculate future numbers of births may be taken as some summary measure of past growth rates of births, such as the median.

14.3 Growth rates from projected numbers of 0-4 year old children

If population projections provided numbers of 0-4 year old children at five year intervals, growth rates of births may be estimated by growth rates of children 0-4 years of age. Most population projections will provide this information, so this method may usually be used when population projections are available but do not provide estimates of annual births.

Calculation of growth rates is more complicated when the interval between the quantities for which the rate is calculated is more than one year. The calculation will be illustrated by example.

Given a base year of 2010, 139,624 children 0-4 as of mid-2010, and 154,414 children age 0-4 as of mid-2015, the growth rate is calculated as

$$\frac{\textit{Natural logarithm of } 154,414/139,624}{5} = \frac{ln(1.10593)}{5} = 0.0201$$

This growth rate of 0-4 year old children may be used as an estimate of the growth rate of births.

Scientific hand calculators will usually provide a button for computing natural logarithms ("*ln*" is a standard abbreviation). Computer spreadsheet programs provide a function for this purpose. In Excel, for example, putting the formula `=LN(A1)` in cell A2 will show the natural logarithm of the number in cell A1. Elaborating the formula to `=LN(A1)/5` will show the growth rate in cell A2.

14.4 Growth rates from *World Population Prospects*

If no projections are available, one option is to calculate the growth rate of births from the estimates of annual numbers of births provided by *World Population Prospects* (Annotated List of Resources).

The *World Population Prospects* estimates of annual births may also be used directly. Using them to calculate the growth rate of births allows the possibility of using an alternative estimate of base year births.

14.5 Growth rates by abbreviated projection

In the absence of national population projection estimates of annual births, this may be the preferred method for estimating annual births if the necessary demographic expertise is available. It requires more input information and more complicated calculations, though it is far less complicated than even the simplest full population projection calculations.

In addition to providing an estimate of the growth rate of births, this method provides an alternative approach to estimating the national number of census year births (Chapter 13, Estimating National Census Year Births).

The following input information is required.

A. The age distribution of females as of the most recent population census, by five year age groups to age 55 (age groups 0-4, 5-9, . . . , 45-49, 50-54); this will be available from census publications.

B. Estimated age specific birth rates for the 10 year period beginning 5 years before the census and ending 5 years after the census; these may be interpolated from national demographic survey estimates or from *World Population Prospects* estimates.

C. Estimated life table ${}_5L_x$ values for ages $x = 0, 5, ..., 50$ for the 10 year period beginning 5 years before the census and ending 5 years after the census; these may be available from national statistical office publications.

The method consists of three main steps.

1. Use the above information to estimate births during the 5 year period preceding the census and births during the five year period following census.

2. Estimate births during the year of the census by summing births during the 5 years preceding and the 5 years following the census and dividing by 10.

3. Estimate the growth rate of births as the natural logarithm of the ratio of number of births during the five years following the census to the number of births during the 5 years following the census, divided by 5.

Step 1 consists of the following sub-steps.

1.1 Reverse survive the census age distribution to get numbers of females in reproductive ages (15-19, 20-24, ..., 45-49) 5 years **before** the census.

1.2 Forward survive the census age distribution to get numbers of females in reproductive ages 5 years **after** the census.

1.3 Calculate the average number of females aged 15-19, 20-24, ..., 45-49 for the 5 years prior to the census; apply the age-specific birth rates to these numbers and sum over the 7 age groups to estimate births during the 5 years prior to the census.

1.4 Calculate the average number of females aged 15-19, 20-24, ..., 45-49 for the 5 years following to the census; apply the age-specific birth rates to these numbers and sum over the 7 age groups to estimate births during the 5 years following the census.

Chapter 15. Estimating Births for Subnational Areas

Counts of persons and births for the country as a whole tend to be more accurate than counts for subnational areas. One reason is that cancellation of errors may occur when summing counts over subnational areas to obtain national totals. Another reason is that numbers of persons and births for subnational areas depend on migration. The availability and accuracy of migration data is often problematic.

The preferred approach to estimating subnational numbers of annual births is therefore usually "top down". Subnational estimates are obtained by first estimating a national total and then using available data on subnational areas to "prorate" the national total to subnational areas.

This chapter presents methods for estimating numbers of annual births in subnational areas. Use of these methods may be appropriate when

1. numbers of annual births in subnational areas cannot be obtained from a civil registration and vital statistics (CRVS) system,
2. numbers of annual births in subnational areas cannot be obtained from population projections,
3. an estimate of national annual births is available, and
4. quantities suitable for use a basis for proration are available for subnational areas.

Proration and bases for proration are explained in the following section.

When subnational estimates are made for several years after the base year, numbers of births in subnational areas may change relative to total births. Section 15.6 presents a relatively simple method for taking account of this change.

Producing estimates of subnational births may be complicated by changing definitions of administrative areas over time. This issue is discussed in Section 15.7.

15.1 Proration and bases for proration

Suppose that

1. the national number of births in a country for a particular year is known,
2. numbers of births for subnational areas are not known, but
3. a quantity related to births is known for subnational areas.

The total population of a subnational area is an example of a quantity related to the number of births: areas with larger populations tend to have larger numbers of births.

Given the national number of births and the total population of each subnational area, the proportion of births in each subnational area may be estimated as the population of the subnational area as a proportion of total population. The number of births in each subnational area may then be estimated by the national number of births times this proportion (section 15.3).

Total population, the quantity related to births in this example, may be referred to as the "basis" of the proration. Estimation of births in subnational areas by proration always requires the national number of births, but any quantity related to births may be used as a basis for proration. The closer the relation, the more accurate the estimates.

The number of persons 0-4 years of age, for example, may be used as a prorating national births to subnational areas. If the number of persons aged 0-4 years is more closely related to births than total population, the estimated numbers of subnational births will tend to be more accurate than estimates based on total population.

15.2 Accuracy of proration estimates

The accuracy of proration estimates of subnational births depends on two factors: 1) the accuracy of the national total that is prorated and 2) the closeness of the relationship between the number of births in subnational areas and the basis of the proration. The estimated proportions of births in each subnational area will be exact if the ratios of the (unknown) numbers of births in each subnational area to the basis of the proration for each subnational areas are identical.

If total population is the basis for proration, these ratios are the Crude Birth Rates (CBRs) for the subnational areas. The estimated proportions will therefore be accurate if the crude birth rate is the same for all subnational areas. Variability of crude birth rates for subnational areas will result in errors in estimated numbers of births. Higher variability will result in larger errors, lower variability in smaller errors.

One way of assessing the likely accuracy of subnational numbers of births estimated by proration is to compare estimates based on two or more different bases of proration (section 15.5).

15.3 Example 1: Proration based on total population

Given an estimate of national births for a particular year and the total populations of subnational areas for the year, births in subnational areas for the year may be estimated by proration. Table 15.1 gives an example.

The national total to be prorated, 315,373 births, is shown at the top of the table. The total population of the (hypothetical) country and its 20 subnational areas are shown in the leftmost column. The middle column shows the proportions of total population in each subnational area. The rightmost column shows the estimated number of births in each area, calculated as total births times the proportions in the preceding column.

The proration for Province 1, for example, is

$$0.02219 = 151,118/6,810,558.$$

Table 15.1
Example of Proration Based on Total Population

Given total births: 315,373			
District	Total Population	Proration Proportion	Prorated Births
Total Country	6,810,558	1.00000	[315,372]
Province 1	151,118	0.02219	6,998
Province 2	230,205	0.03380	10,660
Province 3	142,713	0.02095	6,609
Province 4	611,716	0.08982	28,326
Province 5	519,911	0.07634	24,075
Province 6	254,745	0.03740	11,796
Province 7	469,777	0.06898	21,754
Province 8	287,200	0.04217	13,299
Province 9	526,412	0.07729	24,376
Province 10	678,155	0.09957	31,403
Province 11	412,661	0.06059	19,109
Province 12	285,276	0.04189	13,210
Province 13	492,131	0.07226	22,789
Province 14	245,528	0.03605	11,370
Province 15	79,728	0.01171	3,692
Province 16	495,750	0.07279	22,956
Province 17	264,331	0.03881	12,240
Province 18	370,381	0.05438	17,151
Province 19	200,922	0.02950	9,304
Province 20	91,898	0.01349	4,255

Note Prorated births calculated by multiplying given total births by proration proportions. Hypothetical data.

The proportions must sum to 1, which provides a useful check for arithmetic errors.

The estimated number of births in each province is calculated by multiplying estimated national births by the proration proportion for the province. For Province 20, for example,

$$4,255 = 315,373 \times 0.01349.$$

Estimated births for each province are shown in the "Prorated Births" column of the table.

Summing estimated births over all provinces gives total births, which provides another check for arithmetic errors. In this case, however, the sum may not equal the given number of births exactly due to rounding errors. In this example, the sum of prorated births is 315,372. The maximum possible rounding error is one half the number of rows in the table. A difference larger than this indicates an error in the calculations.

Table 15.2
Example of Proration based on BCG Vaccinations

Given total births: 315,373			
District	BCG Vaccinations	Proration Proportion	Prorated Births
Total Country	302,654	1.00000	[315,373]
Province 1	6,706	0.02216	6,988
Province 2	10,216	0.03375	10,645
Province 3	6,248	0.02064	6,510
Province 4	31,897	0.10539	33,238
Province 5	22,628	0.07477	23,579
Province 6	11,153	0.03685	11,621
Province 7	20,686	0.06835	21,555
Province 8	12,671	0.04187	13,204
Province 9	22,911	0.07570	23,874
Province 10	29,689	0.09810	30,937
Province 11	18,101	0.05981	18,862
Province 12	9,123	0.03014	9,506
Province 13	24,444	0.08077	25,472
Province 14	10,603	0.03503	11,048
Province 15	3,579	0.01183	3,729
Province 16	21,577	0.07129	22,484
Province 17	11,640	0.03846	12,129
Province 18	15,994	0.05285	16,667
Province 19	8,677	0.02867	9,042
Province 20	4,110	0.01358	4,283

Note Prorated births calculated by multiplying given total births by proration proportions. Hypothetical data.

15.4 Example 2: Proration based on Expanded Programme on Immunization numerator data

If immunization coverage rates for a particular vaccine, BCG, for example, are believed to be high and uniform over subnational areas, the number of vaccinations may be used as a basis for proration.

Table 15.2 shows a second example of the same calculation, this time using EPI programme estimates of births for each province as the basis for proration. The given total number of births is the same.

15.5 Comparing alternative proration estimates

Table 15.3 compares the estimates of subnational births in tables 15.1 and 15.2. The differences and percent differences in the rightmost two columns give an indication of the accuracy of the estimates.

For most provinces, the differences are small, but for three provinces they are large (provinces 4, 12 and 13). For Province 4, for example, the estimate based on total population proration is 28,326 births, but the estimate based on BCG vaccinations is 33,238 births.

Table 15.3
Comparison of estimates based on alternative proration bases

District	Total Population	BCG Vaccinations	Difference	Percent Difference
Total Country	315,373	315,373	12,718	0.0
Province 1	6,998	6,988	10	0.1
Province 2	10,660	10,645	15	0.1
Province 3	6,609	6,510	99	1.5
Province 4	28,326	33,238	-4,912	-16.0
Province 5	24,075	23,579	496	2.1
Province 6	11,796	11,621	175	1.5
Province 7	21,754	21,555	199	0.9
Province 8	13,299	13,204	95	0.7
Province 9	24,376	23,874	502	2.1
Province 10	31,403	30,937	466	1.5
Province 11	19,109	18,862	247	1.3
Province 12	13,210	9,506	3,704	32.6
Province 13	22,789	25,472	-2,683	-11.1
Province 14	11,370	11,048	322	2.9
Province 15	3,692	3,729	-37	-1.0
Province 16	22,956	22,484	472	2.1
Province 17	12,240	12,129	111	0.9
Province 18	17,151	16,667	484	2.9
Province 19	9,304	9,042	262	2.9
Province 20	4,255	4,283	-28	-0.7

Note 2010 census counts and EPI field counts from Tables 15.1 and 15.2, respectively. Percent difference calculated as difference divided by average of census and EPI field count estimates.

15.6 Estimating subnational births in years following the base year

If estimates of subnational births are needed only for a few years following the base year, or if the procedure described in this section cannot be applied because the required input data are lacking, the proration proportions used for the base year may be used for future years as well.

If subnational births are required farther into the future, however, it is preferable to use an estimation procedure that takes account of changing distribution of births to subnational units. This section describes a procedure that uses the distribution of total population among the subnational units for two successive censuses to estimate future changes in the distribution of births among subnational units.

Suppose that the distribution of population among subnational units is available from two censuses, the first taken in 2000, the second taken in 2010. The change in the population of each province as a proportion of total national population may be used to estimate changes following the 2010 census.

The leftmost two columns of table 15.4 show the 2000 and 2010 census counts. The following two columns show the proportions of national population in each province at the 2000 and 2010 censuses.

The "Annual Change" column shows the 2010 proportion for this province minus the 2000 proportion for the province divided by 10. Division by 10 give the average annual change between the two censuses.

The "2011 Proportion" column at far right shows estimates of the proportion population of each province as a proportion of total national population one year following the 2010 census. It is calculated by adding the number in the "Annual Change" column to the number in the column giving the proportion for the previous year, 2010 in this case. Positive annual change values increase a province's proportion, negative values decrease the provinces proportion.

For Province 1, for example, the 2011 proportion is calculated as

$$0.02058 = 0.02219 + 0.000179,$$

and for Province 5, it is calculated as

$$0.07724 = 0.07634 - 0.000100.$$

The bracketed numbers in the total row are the sums of the numbers in the columns beneath. Proportions logically sum to one, and annual change values to 0, but rounding errors may result in small departures from these numbers, as illustrated by the "0.99998" values.

15.7 Changing subnational area definitions

In the example of the preceding section the definitions of the subnational units did not change between the censuses. If definitions do change, this must be taken account of to ensure that the numbers from the two censuses are comparable.

Census publications may provide tables showing the distribution of population for the two most recent national population censuses distributed by subnational units as defined at the most recent census.

If no such table can be found, some changes can be dealt with simply. Suppose for example that Province A as defined at the first census is sub-divided Province A1 and Province A2, with no change in the original boundary, before the following census. The population of Province-A-as-defined-at-the-first-census at the time of the second census may be obtained by summing the populations of Provinces A1 and A2 at the second census.

Similarly, if Provinces A1 and A2 as defined at the first census are combined into a single Province A before the second census, the population of Province-A-as-defined-at-the-second-census at the time of the first census may be obtained by adding the populations of Provinces A1 and A2 as defined at the first census.

When changes in definition of subnational units involve changes in existing boundaries, obtaining comparable units in the two censuses requires more detailed information, including appropriately detailed maps of the units before and after the change and the distribution of population for smaller units that can be aggregated

Table 15.4
Estimating changes in proration proportions from two censuses

	Census totals		Census proportions			
District	2000	2010	2000	2010	Annual Change	2011 Proportion
Total	5,244,571	6,810,558	[0.99998]	[0.99998]	[0.000000]	[0.99998]
Province 1	107,005	151,118	0.02040	0.02219	0.000179	0.02058
Province 2	164,198	230,205	0.03131	0.03380	0.000249	0.03156
Province 3	108,322	142,713	0.02065	0.02095	0.000030	0.02068
Province 4	442,213	611,716	0.08432	0.08982	0.000550	0.08487
Province 5	405,625	519,911	0.07734	0.07634	-0.000100	0.07724
Province 6	193,640	254,745	0.03692	0.03740	0.000048	0.03697
Province 7	347,167	469,777	0.06620	0.06898	0.000278	0.06648
Province 8	209,466	287,200	0.03994	0.04217	0.000223	0.04016
Province 9	410,708	526,412	0.07831	0.07729	-0.000102	0.07821
Province 10	514,977	678,155	0.09819	0.09957	0.000138	0.09833
Province 11	311,915	412,661	0.05947	0.06059	0.000112	0.05958
Province 12	259,366	285,276	0.04945	0.04189	-0.000756	0.04869
Province 13	405,699	492,131	0.07736	0.07226	-0.000510	0.07685
Province 14	199,201	245,528	0.03798	0.03605	-0.000193	0.03779
Province 15	53,351	79,728	0.01017	0.01171	0.000154	0.01032
Province 16	387,327	495,750	0.07385	0.07279	-0.000106	0.07374
Province 17	195,775	264,331	0.03733	0.03881	0.000148	0.03748
Province 18	301,002	370,381	0.05739	0.05438	-0.000301	0.05709
Province 19	164,495	200,922	0.03136	0.02950	-0.000186	0.03117
Province 20	63,119	91,898	0.01204	0.01349	0.000145	0.01219

Note Illustrative input based on anonymized country data.

to obtain the given units at both censuses. Such work is usually best left to the national statistical office.

Problematic boundary changes may affect only a few subnational units, however, and the calculations described in the preceding section may be carried out only for units not so affected. In this case it is advisable to include in the table all subnational units defined at both censuses, with "-" in one of the two census count columns indicating that the unit is not defined for the census in question. No estimates of births in following years will be obtained for these units, but estimates for all other units may be calculated as described in the preceding section.

Chapter 16. Estimating Surviving Infants

This chapter discusses the estimation of surviving infants, for subnational areas, for the country as a whole, and for whatever years are required.

16.1 Estimating surviving infants from births and infant mortality rates

The number of surviving infants for any area and any year may be estimated using the formula

$$Surviving\ infants = Births \times (1 - Infant\ Mortality\ Rate),$$

where the $Births$ and $Infant\ Mortality\ Rate$ are for the area and year for which surviving infants are estimated.

The accuracy of estimates of surviving infants calculated in this way depends on the accuracy of the $Births$ and the $Infant\ mortality\ rate$ estimates. Infant mortality rates are typically less than 100 infant deaths per thousand births, however, so estimates of surviving infants are relatively insensitive to errors in estimated infant mortality rates.

This insensitivity may be illustrated by example. Given true values of 1,000 births and an infant mortality rate of 50 per thousand, the estimated number of surviving infants will be $1,000 \times (1-0.05) = 950$. If the estimated infant mortality rate is 75 per thousand, the estimated number of surviving infants will be $1,000 \times (1-0.075) = 925$ surviving infants. The magnitude of the error in the estimated infant mortality rate is 50%—a very large error—, but the magnitude of the resulting error in the number of surviving infants is less than 3%—a rather small error.

16.2 Source of infant mortality rates

Estimates of infant mortality rates will preferably be available for the same subnational areas and years as the input numbers of births and the output numbers of surviving infants.

Because estimated numbers of surviving infants are relatively insensitive to the values of the infant mortality rate, however, various alternatives are possible. If infant mortality rates are available only for the base year, for example, these rates may be used for subsequent years as well. If rates are available only for regions of the country, the infant mortality rate for each subnational area may be estimated by the rate for the region to which the area belongs.

Table 16.1
Projection Estimates of Annual Infant Mortality Rates:
Malawi, Districts, and Cities, 2008–2016

District/City	2008	2009	2010	2011	2012	2013	2014
	Year						
Malawi	82.0	80.0	78.1	75.8	73.6	71.4	69.3
Chitipa	55.8	53.1	50.6	48.1	45.8	43.5	41.4
Karonga	62.1	58.7	55.4	52.5	49.8	47.1	44.6
Nkhata Bay	78.4	74.2	70.2	66.6	63.1	59.8	56.7
Rumphi	56.1	51.9	48.0	45.1	42.4	39.9	37.5
Mzimba	62.8	60.1	57.4	54.8	52.4	50.0	47.7
Likoma	54.9	52.4	50.0	47.6	45.4	43.2	41.1
Mzuzu City	24.6	23.4	22.1	21.1	20.1	19.2	18.3
Kasungu	56.1	54.7	53.3	51.2	49.3	47.4	45.5
Nkhotakota	81.9	79.0	76.3	73.2	70.3	67.5	64.7
Ntchisi	77.6	75.4	73.3	70.4	67.6	64.9	62.3
Dowa	44.8	43.2	41.6	40.1	38.7	37.3	35.9
Salima	77.8	72.8	68.2	65.4	62.8	60.3	57.8
Lilongwe	121.5	119.0	116.6	112.8	109.1	105.5	102.0
Mchinji	72.4	70.5	68.5	65.9	63.4	61.0	58.6
Dedza	105.1	102.1	99.1	95.4	91.8	88.3	85.0
Ntcheu	47.9	45.9	44.0	42.7	41.4	40.2	39.0
Lilongwe City	56.5	52.4	48.6	45.7	42.9	40.3	37.9
Mangochi	62.7	60.9	59.1	56.8	54.5	52.4	50.3
Machinga	51.1	49.5	48.0	46.0	44.2	42.4	40.7
Blantyre	61.1	59.2	57.3	54.9	52.7	50.5	48.4
Zomba Rural	94.3	90.7	87.2	83.5	80.1	76.7	73.5
Chiradzulu	107.4	104.8	102.2	98.5	95.0	91.6	88.3
Mwanza	104.1	101.8	99.5	96.2	92.8	89.7	86.6
Thyolo	126.7	123.8	120.9	116.8	112.8	108.9	105.1
Mulanje	150.8	147.6	144.4	139.7	135.1	130.7	126.3
Phalombe	106.8	104.5	102.2	98.8	95.4	92.2	89.0
Chikwawa	95.5	91.9	88.5	84.4	80.5	76.7	73.1
Nsanje	137.0	134.0	131.0	126.5	122.3	118.1	114.1
Balaka	93.2	89.1	85.1	81.3	77.6	74.1	70.8
Neno	98.2	94.7	91.3	87.3	83.4	79.7	76.1
Zomba City	78.3	74.5	71.0	67.5	64.2	61.1	58.1
Blantyre City	45.4	43.2	41.1	39.1	37.1	35.2	33.5

Note Estimates from *Population Projections for Malawi*, Annex A2, pages 56-87.

16.3 Example of estimating surviving infants for subnational areas

Chapter 11 presented estimates of annual births for 2008-2016 from published population projections based on the 2008 census of Malawi (Table 11.1). The same publication provides annual infant mortality rate estimates for districts for 2008-2030. The estimates for 2008-2016 are shown in Table 16.1.

Estimated numbers of surviving infants calculated from the numbers of births in table 11.1 and the infant mortality rates in table 16.1 are presented in table 16.2.

Table 16.2

Estimated Surviving Infants: Malawi, districts, and cities, 2008-2016

District/City	Year 2008	2009	2010	2011	2012	2013	2014
Malawi	559,664	570,761	582,080	595,566	609,549	624,127	639,257
Chitipa	8,508	8,693	8,880	9,063	9,242	9,420	9,589
Karonga	11,732	11,991	12,251	12,500	12,756	13,026	13,307
Nkhata Bay	8,508	8,778	9,055	9,336	9,621	9,909	10,196
Rumphi	7,277	7,449	7,621	7,785	7,946	8,104	8,257
Mzimba	29,158	29,884	30,637	31,413	32,211	33,038	33,888
Likoma	354	353	353	354	354	354	354
Mzuzu City	6 828	7 289	7 754	8 215	8 676	9 137	9 596
Kasungu	30 187	30 978	31 801	32 681	33 603	34 579	35 607
Nkhotakota	29 362	30 182	31 028	31 924	32 861	33 850	34 890
Ntchisi	11 172	11 415	11 662	11 923	12 188	12 458	12 733
Dowa	26 460	27 500	28 550	29 597	30 637	31 666	32 683
Salima	15 439	15 784	16 143	16 489	16 854	17 245	17 663
Lilongwe	53 526	54 214	54 949	55 814	56 744	57 743	58 804
Mchinji	21 754	22 315	22 898	23 513	24 148	24 802	25 474
Dedza	26 477	26 819	27 174	27 560	27 966	28 396	28 846
Ntcheu	21 173	21 495	21 843	22 187	22 536	22 883	23 228
Lilongwe City	28 461	30 093	31 702	33 247	34 801	36 378	37 977
Mangochi	34 617	35 457	36 333	37 286	38 310	39 410	40 594
Machinga	21 363	21 757	22 164	22 616	23 107	23 657	24 260
Blantyre	13 467	13 715	13 973	14 247	14 526	14 809	15 093
Zomba Rural	23 797	24 010	24 240	24 501	24 777	25 080	25 405
Chiradzulu	11 004	11 062	11 125	11 203	11 284	11 368	11 455
Mwanza	4 121	4 160	4 203	4 248	4 294	4 336	4 376
Thyolo	20 495	20 875	21 266	21 687	22 128	22 591	23 076
Mulanje	19 808	19 884	19 966	20 097	20 247	20 425	20 633
Phalombe	13 072	13 209	13 346	13 518	13 721	13 965	14 254
Chikwawa	18 162	18 513	18 875	19 271	19 693	20 152	20 643
Nsanje	10 222	10 379	10 551	10 758	10 979	11 219	11 472
Balaka	13 283	13 642	14 024	14 422	14 836	15 263	15 702
Neno	4 583	4 909	5 233	5 557	5 878	6 193	6 500
Zomba City	3 705	3 980	4 257	4 530	4 799	5 060	5 313
Blantyre City	25 600	26 486	27 316	28 101	28 853	29 582	30 283

Note Calculated using the formula $Surviving\ Infants = Births \times (1 - IMR)$ with births in table 11.1 and infant mortality rates in table 16.1.

Chapter 17. Retrospectively Adjusting Target Population Estimates

There are two circumstances in which it is appropriate to retrospectively adjust target population estimates and corresponding immunization coverage estimates.

The first occurs when target population estimates are based on population projections and a new set of projection estimates becomes available following a new population census. If annual series of estimates from the earlier projections and the later projections are plotted together, a more or less pronounced discontinuity may be visible.

The second circumstance occurs when the method used to estimate target population numbers has changed over time. This may result in discontinuities of the kind observed in Figure 3.1 (reproduced here as Figure 17.2)

This chapter presents methods for retrospective adjustment of target population estimates for these two cases. Retrospective adjustment will improve the consistency of annual estimates of target populations and immunization coverage and is likely improve the accuracy of the estimates.

17.1 Adjusting outdated population projection estimates

Figure 17.1 plots estimated births from two projections, one based on a 2000 census and one based on a 2010 census. The discontinuity is clearly visible. The grey line and points show the adjusted 2000 census projection produced by the method described in this section.

The method is most simply presented by the example shown in table 17.1. The older projection gives estimates for the years 2000-2010. The new projection gives estimates for 2010-2019. Both projections give estimates for the years 2010–2012, but the old projection estimates for years beyond 2010 do not enter into the calculation.

The adjusted 2000 census projection estimates in the rightmost column of the table are calculated by multiplying the original 2000 census estimates by the factor shown in the “Factor” column of the table. The factor for the year 2000 is set at “1” to leave the 2000 census projection estimate for the year 2000 unchanged. The factor for the year 2010 is set to the ratio of the 2010 census projection estimate to the 2000 projection estimate, so that the adjusted 2000 projection estimate for 2010 is the same as the 2010 projection estimate. Factors for in-between years are linearly interpolated.

17.2 Adjusting target population estimates for changes in method of estimation

Assessment of the quality of target population estimates (Part I, Assessing Accuracy) may suggest that the method of estimating target populations used by the EPI program has changed over time. Figure 3.1, for example, reproduced below as

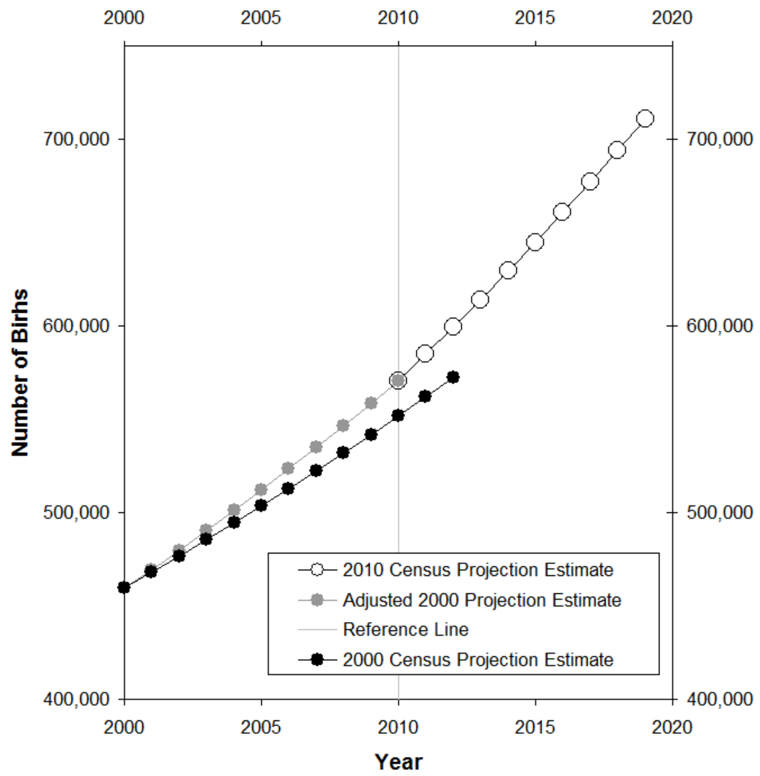

Figure 17.1
Estimated births: Old projection estimates, new projection estimates, and adjusted old projection estimates

Figure 17.2, suggests that one method was used for the years 2002-2008 and a different method for the years 2000-2001 and 2009.

When an assessment of target population estimates leads to this conclusion, it may be appropriate to retrospectively adjust the target population estimates for the change in definition so as to make immunization coverage estimates consistent over time.

The preferred approach to retrospective adjustment will usually be to recalculate denominator estimates using a single method, the method used having been selected on the basis of the assessment exercise.

If recalculation on the basis of a single definition is not possible, a procedure similar to that of Table 17.1 may be employed to adjust the level of the higher estimates (2002–2008) down to the level of the lower estimates (2000–2001 and 2009), or visa-versa.

Both approaches to adjustment require a decision on which method gives more accurate results.

Table 17.1
Adjust projection estimates of births to remove discontinuities between old and new projections

Year	Old Projection Estimate	New Projection Estimate	Factor	Adjusted Old Estimate
2000	459,403	-	1.0000	459,403
2001	467,900	-	1.0034	469,496
2002	476,554	-	1.0068	479,805
2003	485,368	-	1.0102	490,335
2004	494,344	-	1.0136	501,090
2005	503,487	-	1.0171	512,075
2006	512,799	-	1.0205	523,295
2007	522,283	-	1.0239	534,755
2008	531,943	-	1.0273	546,460
2009	541,781	-	1.0307	558,415
2010	551,801	570,625	1.0341	570,625
2011	562,007	584,755	-	-
2012	572,401	599,234	-	-
2013	-	614,071	-	-
2014	-	629,276	-	-
2015	-	644,858	-	-
2016	-	660,825	-	-
2017	-	677,188	-	-
2018	-	693,956	-	-
2019	-	711,139	-	-

Note Adjusted old projection estimates for 2000-2010 in last column calculated by multiplying original estimate in first column by the factor in the “Factor” column. Factors calculated by linear interpolation between “1” for 2000 and the ratio of the new to the old projection estimate for 2010, 570,625/551,801 = 1.0341. Data are hypothetical.

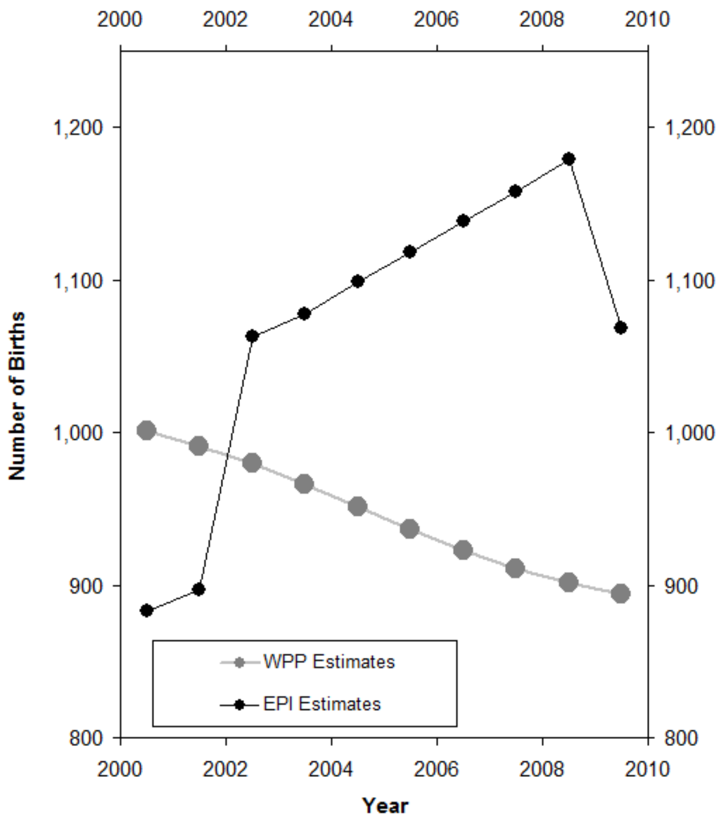

Figure 17.2
EPI and WPP estimates of births compared: Country A
(Reproduced from Figure 3.1)

Annex A. Field Enumeration

Field enumeration is the traditional method of conducting a national population census, generally regarded as the most demanding and expensive of all national data collection operations.

Population census methods have been developed and codified for over half a century, going back to the 1958 publication of *Principles and Recommendations for National Population Censuses* by the United Nations Statistics Division. The current standard reference is *Principles and Recommendations for Population and Housing Censuses*, Revision 2, published in 2008 (for more information see Annotated List of Resources).

This annex describes field enumeration methods. Its primary purpose is to indicate how difficult it is to enumerate a national population accurately and why, therefore, field enumeration at the national level should be attempted by EPI Programmes only as a last resort when the methods described in Part II, Improving Accuracy, cannot be used because (a) useable civil registration and vital statistics data is not available and (b) no population census has been taken in the past 15 years. Fortunately, these conditions apply to only a few countries.

A.1 Definitions and units of enumeration

A population census aims to obtain information on all persons in a specified geographic area, usually the entire national territory of a country, at a specified time. It is customary to distinguish between *de facto* and *de jure* enumeration.

- A *de facto* enumeration identifies persons with the place they were at the census reference time.
- A *de jure* census identifies persons with their place of usual residence at the census reference time.

Persons are enumerated in living quarters. Living quarters are either housing units or collective living quarters. Both are physical structures whose location may be indicated on a map. Collective living quarters include hospitals, orphanages, and refugee camps. Provisions may be made for enumeration of homeless persons.

Persons in housing units are identified as a members of households. A household is a group of persons, living in the same housing unit, that makes common provision for the necessities of everyday life, or a single person ("single person household") who makes his or her own provision for these necessities. A housing unit may be unoccupied, occupied by a single household, or occupied by two or more households.

A.2 Requirements for accurate enumeration of a single enumeration area (EA)

1. *The field worker must know the boundaries of the area to be enumerated.* This requires a large scale map of the enumeration area (EA) showing its boundaries

in a manner that allows the field worker to identify the boundaries in the field. The EA map facilitates exclusion of living quarters outside the EA and inclusion of all living quarters in the EA.

2. *The field worker must visit every living quarters in the EA (and no living quarters outside the EA).* This requires that the EA map show the location of every living quarters in the EA and that the field worker be provided with a list of living quarters in the EA that provides a way of identifying the location of each living quarters on the EA map.

3. *At each living quarters, the field worker must identify the persons who should be enumerated and obtain and record prescribed information for each. Information on the living quarters and occupying household(s) must also be recorded.* This requires that the field worker (a) understand which persons should be enumerated and what information should be collected for each of these persons and (b) be provided with a suitably designed data collection instrument.

4. *The field worker must receive all enumeration materials and training in their use before field work begins.* Materials usually include a manual containing instructions on how approach households, identify and interact with respondents, and record information on forms. If paper forms are used, a sufficient quantity must be provided.

5. *The organization responsible for the enumeration must receive, register and securely store all materials collected by the field worker at a suitable central location when the enumeration is complete.* This requires supporting staff and logistical arrangements and facilities.

A.3 Type of information collected

What information is collected will depend first on what statistical outputs are desired. Some additional information will be included to facilitate field worker's interaction with respondents, such as the names of household members and their relation to the head of the household. Other information may be included to facilitate quality assurance, most obviously the names of the enumerator and the supervisor.

Decisions on what information to collect will be influenced by the anticipated accuracy of information provided by respondents. Age misreporting, for example, is common in many countries, perhaps because accurate knowledge of age is not culturally significant. Responses to questions such as whether any children were born to a woman during the past 12 months may understate the number of events because the respondent does not know the exact date of the event.

A.4 Mode of collection

Information collected by field enumeration has traditionally been recorded on paper forms. In recent decades some enumerations have used mobile devices. It may be anticipated that this will become more common in the future.

A.5 Quality assurance

Quality assurance consists essentially of measures to ensure, to the extent possible, that what is supposed to happen during the enumeration does in fact happen. The first requisite for quality assurance is accordingly documentation of what is supposed to happen. Documentation must specify the work to be done in detail while anticipating the various contingencies likely to be encountered.

Quality assurance serves several purposes.

1. To improve the quality of work by field workers. Field workers tend to do better work if they know that quality control procedures are in effect.
2. To improve supervisor's identification and correction of gaps in field worker's knowledge and understanding of the work to be done.
3. To improve supervisor's identification of field worker malfeasance and disciplinary action by supervisors when appropriate.
4. To produce records of the implementation of quality control measures that provide, after the fact, a basis for assessing the quality of the enumeration, the quality of quality assurance activities, and the quality of the data collected.

Documentation of quality assurance procedures will include actions to monitor the quality of the work as it proceeds, actions to be taken when problems are identified, and information to be recorded on the results of quality assurance activities.

One component of quality assurance may be re-enumeration of a sample of living quarters in the enumeration area, sometimes called a *Post-enumeration Survey* or PES, preferably by a dedicated quality assurance field worker. The revisit will determine, for each sample living quarters, whether the field worker visited the living quarters; whether the persons and households that should have been enumerated were enumerated; and whether information recorded by the enumerator is consistent with the information recorded by the quality control field worker.

The quality assurance field worker must be able to locate sampled living quarters in the field. The enumeration area maps showing the location of living quarters and the list of living quarters noted in section A.2, Requirements for accurate enumeration of a single enumeration area (EA) are required for this purpose as well as for the original enumeration. The list of living quarters will also be used to draw the sample.

The information collected for each living quarters must be sufficiently detailed to discriminate one living quarters from another, to ensure that a re-visit will not yield the same information as the original visit purely by chance. Collecting name, sex, age, and relation to head of household for every person in the living quarters will generally suffice. Information on the number of persons in the living quarters only would not provide sufficient discrimination.

A.6 Field enumeration for a national population

Accurate enumeration of a single EA requires everything discussed in preceding sections. A national enumeration requires replicating this work for thousands, tens of thousands, or hundreds of thousands of EAs. The multiplication alone presents a formidable challenge—producing 1,000 EA maps, training 1,000 enumerators, transporting enumeration materials throughout the national territory.

The challenge is magnified by the the necessity of coordination across EAs. The EA boundaries indicated on every EA map must be consistent with the boundaries indicated on the maps of every adjacent EA. The census mapping work that precedes a national population census may involve large numbers of specialized professionals working over several years. Training large numbers of field workers and distributing and collecting enumeration materials throughout the national territory requires a high level of managerial and logistical capability.

This is why national population censuses are generally considered to be the most complex, difficult and expensive of all primary data collection operations carried out by a national statistical system.

A.7 Field enumeration for target population estimates

EPI offices located in local areas that present favourable conditions for enumeration and with sufficient staff resources may be able to carry out field enumerations that will provide reasonably accurate target population estimates.

Most national EPI programmes will be ill-equipped to replicating such local area enumerations for complete coverage of the entire national territory. The single greatest difficulty may be preparation of enumerationa area maps.

Field enumeration is in any case an inherently inferior instrument for the purpose. Retrospective questions on the number of births a woman had during the past 12 months have been used in hundreds of national population censuses with generally modest success. The resulting counts of births frequently require substantial adjustment for omission.

The problem is not that respondents do not remember that a birth occurred—obviously they do, though some cultures make it unlikely that questions about deceased children will be answered accurately. The problem is rather than respondents may not have accurate knowledge of the date of the event, making them uncertain whether it occurred within the last 12 months or more than 12 months ago.

Glossary

Accuracy

The closeness of an estimate to the unknown true value it represents.

Source ESS Handbook for Quality Reports, 2009 edition (Chapter 3, "Accuracy").

Birth/Live Birth

The complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of pregnancy, which, after such separation, breathes or shows any other evidence of life, such as beating of the heart, pulsation of the umbilical cord or definite movement of voluntary muscles.

Source United Nations Principles and Recommendations for a Vital Statistics System, (Chapter I, Section A, Paragraph 2).

Civil Registration

The continuous gathering of information on all relevant vital events occurring within the boundaries of a country or a well-defined area within a country.

Source United Nations Principles and Recommendations for a Vital Statistics System, (Chapter I, Section B, Paragraph 27).

Crude birth rate

Defined for a population observed during a given time period as the number of births to the population during the period divided by the number of person years lived by the population during the period. Person years lived during a period equals the average size of the population during the period times the length of the period.

Section 1.5, page 7, of *Source Demography: Measuring and modelling Population Processes*, by Samuel H. Preston, Patrick Heuveline, and Michel Guillot (Oxford: Blackwell, 2001.

Sometimes defined less precisely as the number of births divided by the mid-period population.

Data/Dataset

Systematic information on the individuals in some statistical aggregate. "Systematic" means that the same information is provided for every individual, with exceptions for missing and undefined values. "Statistical aggregate" refers to a collection of persons, births, infant deaths, or other real world entities defined by explicitly stated rules for inclusion.

Data consists concretely of (a) a collection of *records*, one for each entity in the aggregate, each record providing information on the entity it represents, and (b) *documentation* describing the statistical aggregate and the content of the records.

The records and the documentation may be referred to collctively as a *dataset*.

Source "Data Assessment" entry in Volumme I of the *Encyclopedia of Population*, edited by Paul Demeny and Geoffrey McNicoll. New York: Macmillan Reference USA, 2003. See also Box 1, page 9, of the *Handbook on the Collection of Fertility and Mortality Data*. New York: United Nations Department of Economic and Social Affairs, Statistics Division, Studies in Methods, Series F, No. 92, 2004.

Data Quality

Includes at least the following characteristics of a dataset.

1. The faithfulness of the correspondence between the records in the dataset and the real world entities in the statistical aggregate.
2. The incidence of missing values for characteristics included on records in the dataset, taking account of undefined values.
3. The accuracy of the information on characteristics provided by the records in the dataset.
4. Any other information required to understand the information provided by the dataset.

Source There is a substantial literature on definitions of "data quality" for official statistics. Useful background is provided in "Dimensions of Statistical Quality: A discussion note about the quality initiatives of some international organisations" by Willem de Vries, United Nations Statistics Division, August 2002 , accessed May 2015). See also Section 3.1 "Quality Definition" of the Eurostat *ESS Handbook for Quality Reports*, 2014, and Section 1.2, "Aspects of Data Quality", of the Eurostat *Handbook on Data Quality Assessment Methods and Tools*, 2007, both accessed May 2015. The characterisation above focuses on three core dimensions of quality.

Error

The difference between an estimate and the unknown true value it represents.

Source Section 2.3, Estimates, accuracy, and error.

Estimate (noun)

A number intended to represent the unknown true value of some quantity. See *accuracy* and *error*

Source Section 2.3, Estimates, accuracy, and error.

Extrapolation

A process for estimating values in a series that lie below or above a set of given values. Note also *interpolation*

Source Section 14.1, Estimating future births by growth rate extrapolation.

Field enumeration

The traditional method of carrying out a population census, used by the great majority of the world's countries. Involves sending census *enumerators* to visit dwelling units, households, and collective quarters through a country to collect information from *respondents*. See also Annex A, Field Enumeration.

Source United Nations Principles and Recommendations for Population and Housing Censuses (Chapter II, Section E, Subsection 1, "The traditional approach")

Growth rate of births, annual

The annual rate at which births in a population are growing. May be calculated approximately as one minus the ratio of the number of births in a given year to the number in the preceding year.

Source Section 4.1, Definition of target population growth rate.

Immunization coverage

The proportion or percentage of persons in a target population who are immunized with a particular vaccine.

Source Section 8.4, Documentation of target population and vaccinations specification.

Implied infant mortality rate

The infant mortality rate implied by (a) a number of births and (b) the number of these births who survive to their first birthday.

Source Section 5.1, Definition of implied infant mortality rate.

Infant

A child who has not reached their first birthday.

Infant Mortality Rate (IMR)

Generally, the number of infant deaths per thousand births. Various more specific definitions may be used, including

- the ratio of infant deaths during a year to births during this year,

- the proportion of children born during a given year who die before their first birthday times 1000, and
- in a life table, the probability of dying before age one year.

Interpolation

A process for estimating values in a series that lie in between a set of given values. Note also *extrapolation*.

Source Section 13.5, *Annex* Interpolation/extrapolation of crude birth rates.

Missing values

Quantities that are defined but not available. Compare *undefined values*.

Source Section 9.1, Definition of missing values.

National Statistical Office (NSO)

Generic designation for an organization legally designated as responsible for overall supervision of the national statistical system of a country.

Source Chapter 12, Working with the National Statistical System.

National Statistical System

The statistical organisations and units within a country that jointly collect, process and disseminate official statistics on behalf of national government.

Source OECD Glossary of Statistical Terms. See also Chapter 12, Working with the National Statistical System.

Place of occurrence

For vital events, the geographic area in which the event took place.

Source United Nations Principles and Recommendations for a Vital Statistics System, (Part I, Chapter I, Section A, Paragraph 1, and Chapter III, Section D, Subsections 1-6).

Place of registration

For vital events, the geographic area in which the event was registered.

Source United Nations Principles and Recommendations for a Vital Statistics System, (Part I, Chapter I, Section A, Paragraph 1, and Chapter III, Section D, Subsections 1-6).

Place of usual residence

For vital events, the geographic area in which a person usually resided at the time of the event.

Source United Nations Principles and Recommendations for a Vital Statistics System, (Part I, Chapter I, Section A, Paragraph 1, and Chapter III, Section D, Subsections 1-6).

Population census

The total process of collecting, compiling, evaluating, analysing and publishing or otherwise disseminating demographic, economic and social data pertaining, at a specified time, to all persons in a country or in a well- delimited part of a country.

The essential features of a population and housing census are individual enumeration, universality within a defined territory, simultaneity and defined periodicity.

Source United Nations Principles and Recommendations for Population and Housing Censuses (Chapter 2, Section A, Subsection 1, and Section B).

Population projection estimates

Estimates of future numbers of population, births and deaths produced by extrapolating past population trends.

Source Section 11.2, Target population estimates from population projections.

Proration

A process of estimating the distribution of persons in a population or events occurring during a given time period based on (a) the total number of persons or events and (b) the distribution of some quantity that serves as a proxy for the unknown numbers of persons or events.

Source Chapter 15, Estimating Births for Subnational Areas.

Relative error

The error of an estimate divided by the true value of the quantity estimated, usually expressed as a percent.

Source Section 2.3, Estimates, accuracy, and error.

Significant figures

The number of digits in an estimate, counted from left to right beginning with the first non-zero digit, that may be considered to represent the unknown true value represented by the estimate, taking account of the likely error of the estimate.

Source Section 5.2, Example of Immunization Coverage Data: Country B.

Surviving infants

The number of children in a population celebrating their first birthday during a year or other time period.

Source Section 2.2, Births and surviving infants.

Target population

A group of persons targeted for immunization.

Source Section 2.2, Births and surviving infants.

Undefined values

Quantities that are undefined because the real world entities to which they refer do not exist or do not have particular characteristics.

Source Section 8.2, Documentation of undefined values.

Vital events

Events having to do with life and death (live birth, death, foetal death) or family (marriage, divorce, judicial separation, adoption, legitimation, and recognition).

Source United Nations Principles and Recommendations for a Vital Statistics System, (Chapter I, Section A, Paragraph 2).

Vital statistics

Statistics on the occurrence and characteristics of vital events. Occurrence refers to counts of events (a) during particular years, months, or other time periods that (b) occurred or were registered in, or occurred to persons usually resident in, a particular geographic area. Characteristics refers to particular characteristics of events, such as the age at which a person died.

Source United Nations Principles and Recommendations for a Vital Statistics System, (Part I, Chapter I, Section A, Paragraph 1, and Chapter III, Section D, Subsections 1-6).

Annotated List of Resources

1. Websites

World Population Prospects
esa.un.org/unpd/wpp/

Estimates of population, births, deaths, and international migration for all countries of the world, for five year periods from 1950-1955 through 2095-2100, midyear to midyear, updated every two years by the United Nations Population Division. Most current at this writing is the 2015 revision.

Estimates of annual births and infant mortality rates at the national level are provided in a downloadable spreadsheet file as shown below.

Page	http://esa.un.org/unpd/wpp/DVD/
Link Text	Annual Demographic Indicators
File Name	WPP2015_INT_F01_ANNUAL_DEMOGRAPHIC_INDICATORS.XLS
Link	Annual Demographic Indicators

Each row of the ESTIMATES worksheet of the file provides estimates for country or region (Column C) for a calendar year (Column F) between 1950 and 2009. Births are given in Column R, Infant Mortality Rates in Column O. Each row of the MEDIUM worksheet of the file gives the same information for calendar years 2010-2100.

childmortality.org
childmortality.org

A database containing the latest child mortality estimates based on the research of the UN Inter-agency Group for Child Mortality Estimation. National data only at this writing.

Annual estimates of infant mortality rates for 1950 through (at this writing) 2015 are provided in a downloadable spreadsheet file as shown below.

Page	childmortality.org
File Name	RatesDeaths_AllIndicators.xlsx
Link	childmortality.org/files_v20/download/RatesDeaths_AllIndicators.xlsx

The Demographic and Health Surveys (DHS) Program
www.dhsprogram.com

The MEASURE DHS project (Monitoring and Evaluation to Assess and Use Results – Demographic and Health Surveys) provides technical assistance developing countries to advance global understanding of health and population trends. Final reports for 250 DHS surveys for 90 countries are available on the project website.

Home page	www.dhsprogram.com/
Data page	www.dhsprogram.com/Data/
Country List	www.dhsprogram.com/Where-We-Work/Country-List.cfm

The "Country List" page provides links to publications and datasets. Estimates of crude birth rates and infant mortality rates will generally be found in at least one of the publications listed for each country.

UNICEFs Multiple Indicator Cluster Survey (MICS)
mics.unicef.org

UNICEF assists countries in collecting and analysing data in order to fill data gaps for monitoring the situation of children and women through its international household survey initiative the Multiple Indicator Cluster Surveys (MICS). All available MICS results and datasets can be accessed on mics.unicef.org. The results from the most recent MICS5 surveys, carried out in 2012-2015, are becoming progressively available.

UNdata
data.un.org

An extensive collection of data provided by the United Nations Statistics Division. The databases listed under the "Population", "UNSD Demographic Statistics", include Live Births by month of birth. A spreadsheet file containing data for a single country may be downloaded by selecting the country name in the "Select filters:" box at upper left and then clicking "Apply Filters". The years for which data is provided varies from country to country. This database does not include infant mortality rate estimates at this writing.

2. Civil Registration and Vital Statistics Methods

Civil registration is the preferred method for obtaining accurate denominator estimates for immunization coverage. The registration method is the only method capable of giving consistently accurate denominators (population registers necessarily subsume such a system for registering births and deaths).

The United Nations Statistics Division (UNSD) has compiled and disseminated methodological guidelines for civil registration and vital statistics systems (CRVS) for over half a century. The first CRVS publication, *Principles for a Vital Statistics System*, was issued in 1953.

Principles for a Vital Statistics System, 1953. This early publication is more succinct that subsequent revisions and covers the most essential features of civil registration and vital statistics systems.

Handbook of Vital Statistics Methods This early publication on vital statistics methods is succinct and covers most essentials.

Principles and Recommendations for a Vital Statistics System, Revision 3, 2014. The most recent revision of "Principles and Recommendations" for vital statistics systems.

3. Population Censuses and Field Enumeration

When annual numbers of births and deaths are not available from a civil registration and vital statistics (CRVS) system, estimates of births and surviving infants may be provided population projection estimates based on the most recent population census. Projection estimates are commonly produced by the national statistical system following each population census.

Population censuses collect information on the national population at a particular point in time. They generally do not provide directly usable information on numbers of births. Because enumerated numbers of infants may be substantially incomplete as a result of age misreporting and/or age-selective omission, these should generally not be used to estimate denominators for immunization coverage.

Population projection estimates based on census results may however provide estimates of births and infant mortality rates. In the absence of a well-developed civil registration and vital statistics system, projection estimates will generally provide the most accurate estimates of denominators for immunization coverage. They cannot however be expected to provide estimates as accurate as may be provided by a civil registration and vital statistics system.

Field enumeration is the traditional method for carrying out censuses of population and housing. The United Nations Statistics Division (UNSD) has compiled and disseminated methodological guidelines for population and housing censuses for over half a century.

The first census methods publication, Population Census Methods, was issued in 1949. The most recent publication, *Principles and recommendations for Population and Housing Censuses*, Revision 2 (2008) is the eighth in the series. UNSD issues numerous other technical publications in support of census activities.

Several of the most pertinent publications for immunization coverage statistics are shown below. All are available in Arabic, Chinese, English, French, Russian, and Spanish at Handbooks, Guidelines and Training Manuals.

Principles and Recommendations for Population and Housing Censuses Revision 2, 2008. Most pertinent to EPI programmes is the material on planning for an executing field enumeration in Part I, "Operational Aspects of Population and Housing Censuses". See also the *Handbook on Census Management* listed below.

Handbook on Geospatial Infrastructure in Support of Census Activities, 2009. Census maps covering the national territory to be enumerated are essential for accurate coverage by a field enumeration. Most pertinent material for EPI programmes are Chapters 1-C of Part I, "Managerial considerations for heads of national statistical offices and other decision makers".

Handbook on Census Management for Population and Housing Censuses, 2001. Field enumerations covering the entire national area require coordinated activity by large numbers of persons though the national territory. This handbook addresses management issues. Most pertinent for EPI Programmes are Chapters B-D of Part I, "Overall Census Management".

4. Geographical Classification

Subdivision of the national territory into non-overlapping subnational areas is as fundamental to data collection for subnational areas as it is to immunization operations. Subnational areas are necessarily defined in part by maps showing there boundaries, so geographical classification is closely related to census and other mapping activities. Documentation of geographical classifications is particularly important when different parts of the national statistical system use different classifications.

The following *Standard Geographical Classification* illustrates what may be involved in documenting a geographical classification.

Standard Geographical Classification (SGC), Volume I, 2011. Statistics Canada.
