

# ***Chapter 8***

## ***Statistical Methods***

---

## Table of Contents

---

<b>8.1 Introduction.....</b>	<b>8-1</b>
<b>8.2 Measuring Birth Defect Occurrence .....</b>	<b>8-2</b>
8.2.1 Multiple Birth Defects in the Same Child .....	8-2
8.2.2 Counts, Ratios, Proportions, and Rates .....	8-2
<b>8.3 Calculating Prevalence at Birth.....</b>	<b>8-4</b>
8.3.1 Basic Calculation for Prevalence at Birth .....	8-4
8.3.2 Numerator and Denominator Issues .....	8-5
<b>8.4 General Types of Prevalence.....</b>	<b>8-6</b>
8.4.1 Crude Prevalence .....	8-6
8.4.2 Specific Prevalence .....	8-6
8.4.3 Adjusted or Standardized Prevalence.....	8-8
<b>8.5 Descriptive Epidemiology.....</b>	<b>8-12</b>
<b>8.6 Confidence Intervals.....</b>	<b>8-15</b>
8.6.1 About Confidence Intervals .....	8-15
8.6.2 Comparing Prevalence Values Using Confidence Intervals.....	8-19
<b>8.7 Issues to Consider If Data Reveal Unusual Patterns .....</b>	<b>8-22</b>
8.7.1 Changes in Medical Diagnoses and Technologies .....	8-22
8.7.2 Changes in Reporting and Case Ascertainment .....	8-22
8.7.3 Changes in the Population at Risk.....	8-23
8.7.4 Random Variation .....	8-23
<b>8.8 References.....</b>	<b>8-24</b>

---

## 8.1 Introduction

---

Statistics are useful to surveillance programs for:

- Summarizing and comparing surveillance data
- Assessing the potential role of chance or random variability
- Controlling for the effects of extraneous factors

The objective of this chapter is to present some common statistical concepts and tools that can be applied to surveillance data. For each tool, a definition is provided, along with background information, guidelines for use, how to calculate, and an example. This is a basic introduction only; more exhaustive treatment of these topics can be found in the reference literature.

In Section 8.2 of this chapter we discuss measures of birth defect occurrence. General issues relating to prevalence are discussed in Section 8.3, with the distinctions between crude prevalence, specific prevalence, and adjusted or standardized prevalence presented in Section 8.4. Various approaches to presenting and displaying descriptive epidemiology are described in Section 8.5, while confidence intervals and their calculation are discussed in Section 8.6. Finally, in Section 8.7 we discuss means to rule out straightforward explanations for observed changes in the prevalence of a birth defect. References cited in this chapter may be found in Section 8.8.

---

## 8.2 Measuring Birth Defect Occurrence

---

In carrying out basic epidemiologic and statistical assessment of birth defects occurrence, the analyst needs to decide what to count (issues of case definition are discussed in Chapter 3) and how to use those counts in calculations. This section presents some basic concepts, clarifies definitions with respect to analysis and reporting, and presents alternatives to the standard methods used to measure birth defects occurrence, birth prevalence.

### 8.2.1 Multiple Birth Defects in the Same Child

Analyses of birth defects surveillance data should be based on cases. An infant or fetus can have multiple birth defects and can be counted as a separate case for each defect. Thus, an infant/fetus with anencephaly and cleft lip should be counted as a case of anencephaly, and again as a case of cleft lip. When using this approach, it is important to recognize that the number of different cases cannot then be added to reach a total number of infants/fetuses.

When an infant/fetus has two or more conditions coded in the same category in an analysis, count it once only. For example, if an infant has atrial septal defect and ventricular septal defect. Count the infant once in tabulations for atrial septal defect, and once in tabulations for ventricular septal defect. Additionally, count the infant only once in tabulations for cardiac defects.

### 8.2.2 Counts, Ratios, Proportions, and Rates

The most common measures of birth defect occurrence are counts, ratios, proportions, and rates.

**Counts.** *Counts* present the simple enumeration of cases. Such information can be useful for health planning purposes, where it is important to measure the burden of birth defects on existing health care resources, to assess the need for additional resources, and for cluster investigations. However, simple counts of cases are not of value as a measure of disease risk, for which rates are necessary.

**Ratios.** A *ratio* is composed of one number (the numerator) divided by another (the denominator). Ratios can be useful for comparing the number of cases in one population group with the number in another. Proportions and rates (discussed below), and prevalence (discussed in Section 8.3) are special types of ratios.

FORMULA:  $A / B$  or numerator / denominator.

EXAMPLE: The sex ratio of cleft palate cases would be represented as notated below.

$$\frac{\text{number of male cases with cleft palate}}{\text{number of female cases with cleft palate}}$$

**Proportions.** In a *proportion*, the cases in the numerator must be included in the denominator. A *percentage* is a proportion multiplied by 100. Proportions are useful for describing basic characteristics of surveillance program data. This can help with quality control.

FORMULA for a proportion:  $A / (A+B)$

FORMULA for a percentage:  $A \times 100 / (A+B)$

EXAMPLES

The *proportion* of abstracted records with errors would be:

$$\frac{\text{the number of records with errors}}{\text{the total number of records}}$$

The *percentage* of abstracted records with errors would be:

$$\frac{\text{the number of records with errors}}{\text{the total number of records abstracted}} \times 100$$

**Rates.** In epidemiology, *rates* express the frequency with which an event occurs (e.g., the number of new cases of disease) in a defined population in a specified period of time (Last, 1995).

FORMULA for incidence rate:

$$\frac{\text{the number of new cases of a disease during a period of time}}{\text{population at risk}}$$

As will be discussed further in Section 8.3, although some investigators and studies report ‘incidence rates’ when talking about birth defects occurrence, there is general consensus that the information to determine incidence is not available (Sever, 2004). Therefore ‘prevalence’ or ‘prevalence at birth’ is the more appropriate terminology.

## 8.3 Calculating Prevalence at Birth

**Prevalence** expresses the number of existing cases of disease at a point in time divided by the total population. Prevalence is useful since it allows comparison between populations of different sizes. Prevalence may be measured at any time (e.g., X cases of spina bifida of any age on June 1 2003 divided by the entire population). However, for measuring occurrence of birth defects, it is most common to use prevalence at birth or birth prevalence. That is true even though many of the cases included may not have been live births.

Ideally, incidence rates would be used instead of prevalence to measure birth defect occurrence.

**Incidence rates** measure the occurrence of new events that occur in a population, so the formula for incidence of a birth defect would be:

$$\frac{\text{the number of new cases of birth defect A in an area and time period}}{\text{the number of conceptions at risk of developing defect A in that area and time period}} \times \text{multiplier}$$

Since the number of conceptions is unknown, as is the number of cases “lost” through spontaneous abortions, technically speaking we cannot determine incidence. Because of this, as noted above, most epidemiologists working in the area of birth defects use the term ‘prevalence’ to refer to birth defect occurrence. For a more complete discussion of this issue see Sever (2004).

### 8.3.1 Basic Calculation for Prevalence at Birth

Surveillance programs should measure birth defect occurrence using the following formula for birth prevalence. Note that both the numerator (number of cases) and the denominator (number of live births) always come from the same area and time period, that is, the same population. Usually the denominator is the number of live births to residents in the same geopolitical area from which the cases came during the same time period.

FORMULA for birth prevalence (expressed as cases of defect A per 10,000 live births):

$$\frac{\text{the number of cases with birth defect A in an area and time period}}{\text{the number of live births in that area and time period}} \times 10,000$$

EXAMPLE from Missouri:

$$\frac{193 \text{ cases with Tetralogy of Fallot statewide delivered in 1989-1995}}{532,592 \text{ live births}} \times 10,000$$

= 3.62 cases per 10,000 live births

### 8.3.2 Numerator and Denominator Issues

**Counting cases – the numerator.** For information on how to count cases for the numerator, see Section 8.2.1 on measuring birth defect occurrence (cases versus infants/fetuses).

**Pregnancy outcomes included.** Most often in analyses of birth defects surveillance data, the cases in the numerator are derived from all pregnancy outcomes collected by the program. These *may* include those listed below (see Chapter 3 on Case Definition).

- Live births
- Spontaneous fetal deaths greater than or equal to 20 weeks gestational age (GA)
- Spontaneous fetal deaths less than 20 weeks GA
- Induced terminations greater than or equal to 20 weeks GA
- Induced terminations less than 20 weeks GA
- Fetal deaths, unknown if spontaneous or induced, and/or of unknown gestational age

Sometimes analyses are restricted to certain pregnancy outcomes in comparing data from two surveillance programs that collect different pregnancy outcomes. For example, in the annual reports of EUROCAT (2002) and the International Clearinghouse for Birth Defects Monitoring Systems (2002) data are presented by live births plus stillbirths (late fetal deaths) and induced abortions (terminations of pregnancy), such that it is possible to make comparisons between prevalence based on comparable methods of case ascertainment (Sever, 2004).

The reports generated by the program should document which pregnancy outcomes are included in the numerator.

For the denominator, surveillance programs should use the total number of live births in the same area and time period from which the cases were ascertained. Although including induced and spontaneous fetal deaths would more closely approximate incidence rates calculated in other areas of epidemiology, it is impractical since these other pregnancy outcomes are often inaccurately counted compared to live births. In addition, these counts are small in comparison to the number of live births and are unlikely to affect prevalence to a large degree.

All the cases in the numerator (e.g., spontaneous and induced fetal deaths) may not come from the denominator (live births). For this and other technical reasons, birth prevalence is really a ratio and not a rate, although it is commonly referred to as a ‘rate’.

**Multiplier.** The multiplier 10,000 is used for convenience, so that prevalence of most defects will have at least one digit to the left of the decimal place. Prevalence is expressed as ‘X cases per 10,000 live births’. The most common multipliers for birth defects are 10,000 and 1,000.

EXAMPLE. 6.3 cases per 10,000 live births is easier for the reader to understand than 0.00063 cases (per birth).

## 8.4 General Types of Prevalence

As with rates and other measures of morbidity and mortality, there are three general types of prevalence. Prevalence can be categorized based on whether it:

- Applies to the whole population – *crude prevalence*
- Applies to subgroups within the population – *specific prevalence*
- Applies to the whole population, but adjusts for differing distribution of subgroups within the population – *standardized or adjusted prevalence*.

Below we discuss each of these types of prevalence in turn.

### 8.4.1 Crude Prevalence

**Definition** Prevalence calculated for the entire population without regard to possible subgroups within the population.

**When to Use** When a single, easily calculated number summarizing the occurrence of disease in a population is desired.

**How to Use** For birth defects, the basic calculation above is applied to the entire population. The area is usually the area covered by your birth defects surveillance program.

$$\frac{\text{The number of cases with birth defect A in an area and time period}}{\text{the number of live births in that area and time period}} \times 10,000$$

EXAMPLE

**Birth Prevalence of Down Syndrome, Texas, 1996/97 Deliveries**

Cases	Live Births	Prevalence*
355	300,431	11.82

\* cases per 10,000 live births      Source: Ethen and Case, 2000.

### 8.4.2 Specific Prevalence

**Definition** Prevalence calculated for subgroups or *strata* within the population, such as age groups, sex groups, or racial/ethnic groups. These then would be referred to as age-specific prevalence, sex-specific prevalence, or race/ethnicity-specific prevalence. The term ‘stratified’ is also used to refer to the prevalence among such subgroups.

**When to Use** Specific prevalence is used in looking at disease occurrence in subgroups of a population. It is also used when sufficient data are available to define and categorize the population of interest. In using specific prevalence, it is important

to consider how missing values (e.g., unknown maternal ages) would affect the interpretation of the prevalence data presented.

### How to Use

For birth defects, apply the basic prevalence calculation above to each group of interest within the population. It is necessary to have the numerator and denominator from the same group of interest. The most common groups of interest for routine birth defects reports are based on:

- Maternal age at delivery
- Maternal racial/ethnic group
- Infant sex

However, specific prevalence can be calculated for any group for which numerator and denominator data are available.

Each grouping of a variable of interest is also called a 'stratum'. For example, common strata for maternal age at delivery are:

- Less than 20 years old
- 20 – 24 years old
- 25 – 29 years old
- 30 – 34 years old
- Greater than or equal to 35 years old

Calculating maternal age-specific prevalence would then yield five values.

It is helpful to define the groups or strata in the same way vital statistics are routinely reported for the population of the area. For example, live births in Texas are commonly reported for four maternal racial/ethnic groups:

- White (non-Hispanic)
- Black
- Hispanic
- Other

Thus, those categories are used for reporting race/ethnicity-specific prevalence values.

The FORMULA for calculation is:

$$\frac{\text{the number of cases with birth defect A in group X in an area and time period}}{\text{the number of live births in group X in that area and time period}} \times 10,000$$

EXAMPLE

Prevalence of Down Syndrome by Maternal Age in Years, Texas 1996/97 Deliveries

Maternal Age (years)	# Cases	# Live Births	Prevalence*
< 20	39	48401	8.06
20 – 24	53	83398	6.36
25 – 29	45	81442	5.53
30 – 34	84	57562	14.59
35 +	134	29574	45.31

\* cases per 10,000 live births Source: Ethen and Case, 2000.

### 8.4.3 Adjusted or Standardized Prevalence

**Definition**

Prevalence calculated for the entire population (the target population) that adjusts for (eliminates the influence of) possible differences in makeup between it and some standard or reference population. It is a summary measure that is a weighted average of the stratum-specific prevalence values.

There are two types of adjusted or standardized prevalence: direct and indirect. **Direct adjustment** uses specific prevalence derived directly from the target population (hence the name) and combines them using the age distribution of the standard population. The prevalence is generated that the target population would have experienced had it had the same age structure as the standard population. **Indirect adjustment** uses age-specific prevalence figures derived from the standard population but applies those to the age distribution of the target population. This technique produces the number of cases the target population would have experienced had it had the same age-specific prevalence as the standard population. The final result is usually expressed as a ratio of the cases observed in the target population divided by the number of cases expected based on this calculation.

**When to Use**

Adjustment is used to develop a single number summarizing the occurrence of birth defects within a population compared with some other population, removing the effect of differences between populations in the distribution of the factor adjusted for. An example would be to examine the occurrence of Down syndrome in a community near a hazardous waste site where the community has a larger proportion of older mothers than a comparison community or the state as a whole. The most common characteristics adjusted for in birth defects analyses are maternal age and maternal racial/ethnic group.

Use direct adjustment:

- When information is available on both the number of cases and the number of live births in each group/stratum/level of the factor being adjusted for (e.g., in each maternal age group); or
- To compare two or more target populations with each other (e.g., prevalence for anencephaly in 20 counties [20 target populations], standardized for maternal race-ethnicity group).

Note that to compare two or more target populations, they must be standardized using the same standard population.

Use indirect adjustment when:

- Information is not available on the number of cases in each stratum of the factor being adjusted for;
- The comparison is between a target population and a standard and not with another target population (e.g., is the prevalence of anencephaly significantly different in County X compared to the whole state?);
- Statistical precision is very important (since stratum-specific prevalence used for direct standardization can sometimes vary widely if based on few cases); or
- The results are to be presented as an observed-to-expected ratio (although a prevalence can be calculated).

## How to Use

The following instructions are based on adjusting for maternal age groups. Each age group is called a ‘stratum’. The same process would be used when adjusting for other characteristics, for example, race or ethnicity.

*Direct adjustment.* For direct adjustment follow the steps below.

1. Decide on age-group categories (strata) that can be applied to both the target and standard populations.
2. Calculate age-specific prevalence for each stratum of the *target* population. Do not use the multiplier (10,000 or 1,000) for this calculation. However, the multiplier may be used for presenting the age-specific prevalence values.
3. Multiply each prevalence by the number of live births in the same stratum of the *standard* population. This gives the number of cases expected in each stratum of the standard population, had it experienced the same age-specific prevalence as the target.
4. Add up the number of expected cases across all strata.
5. Divide the total number of expected cases by the total number of live births in the *standard* population and multiply by your multiplier (e.g., 10,000).

This is the ‘directly standardized birth prevalence’.

*Indirect adjustment.* For indirect adjustment follow the steps below.

1. Decide on age-group categories (strata) that can be applied to both the target and the standard populations.
2. Calculate age-specific prevalence for each stratum of the *standard* population. Do not use the multiplier (10,000 or 1,000) for this calculation. However, the multiplier may be used for presenting the age-specific prevalence.
3. Multiply each prevalence by the number of live births in the same stratum of the *target* population. This gives the number of cases expected in each stratum of the target population, had it experienced the same age-specific

prevalence as the standard.

4. Add up the number of expected cases across all strata of the *target* population.
5. Divide the total number of observed cases in the *target* population by the calculated total number of expected cases.

This is the ‘standardized birth prevalence ratio’, sometimes called the ‘standardized observed-to-expected ratio’. When applied to mortality, the result is called the ‘standardized mortality ratio’ or SMR.

EXAMPLES. The Texas Birth Defects Monitoring Division dealt with a cluster of Down syndrome in a three-county area in Texas among deliveries in 1992–1994. Down syndrome is influenced strongly by maternal age. Thus it was necessary to adjust for maternal age to see whether the excess was still apparent when possible differences in maternal ages between these three counties and the state of Texas were removed. The three counties (1992–1994 deliveries) make up the *target population*, and the entire Texas Birth Defects Registry area (1996–1997 deliveries) the *standard population*. The years 1996–1997 were used because those were the first years with data published for most of the state.

*Direct adjustment.* The crude Down syndrome prevalence for the three counties during 1992–1994 was 31.97 cases per 10,000 live births. **Steps 1-4** are presented in the following table.

Maternal Age (years)	Target			Standard	
	# Cases	# Live Births	Prevalence*	# Live Births	Expected # Cases
	(A)	(B)	(A) / (B) = (C)	(D)	(C) x (D)
< 20	3	1032	0.00290698	48401	140.701
20 – 24	3	1666	0.00180072	83398	150.176
25 – 29	3	1498	0.00200267	81442	163.101
30 – 34	4	1028	0.00389105	57562	223.977
35 +	5	407	0.01228501	29574	363.317
Total	18	5631	0.00319659	300377	1041.272

\* Expressed as cases per live birth. To express it the usual way, multiply by 10,000.

### Step 5

$$\frac{\text{total number of expected cases in standard population}}{\text{total number of live births in standard population}} \times 10,000$$

$$= \frac{1041.272}{300377} \times 10,000$$

= the standardized prevalence

*Indirect adjustment. Steps 1-4* are presented in the following table. There are known to be 18 cases of Down syndrome in the target population, but the ages of their mothers may not be known.

Maternal Age (years)	TARGET	STANDARD			TARGET
	# Live Births	# Cases	# Live Births	Prevalence *	Expected # Cases
	(E)	(F)	(G)	(F) / (G) = (H)	(E) x (H)
< 20	1032	39	48401	0.00080577	0.83155307
20 – 24	1666	53	83398	0.00063551	1.05875441
25 – 29	1498	45	81442	0.00055254	0.82770561
30 – 34	1028	84	57562	0.00145930	1.50015635
35 +	407	134	29574	0.00453101	1.84411983
Total	5631	355	300377	0.00118185	6.06228927

\* Expressed as cases per live birth. To express it the usual way, multiply by 10,000.

### Step 5

$$\frac{\text{total number of observed cases in target population}}{\text{total number of expected cases in target population}} \times$$

$$= \frac{18}{6.0623} \times 10,000$$

$$= 2.97$$

= standardized birth prevalence ratio or standardized observed-to-expected ratio

---

## 8.5 Descriptive Epidemiology

---

**Background** Surveillance data allow the description of the occurrence of birth defects in terms of the basic epidemiologic parameters of time, place, and person. In doing so, comparisons can be made among these different parameters (e.g., comparing the prevalence of anencephaly among different maternal ages).

Note that descriptive epidemiology is really just the presentation of specific prevalence where the strata are time periods (e.g., year of delivery), areas (e.g., counties), or personal characteristics (e.g., maternal age groups or maternal racial/ethnic groups). See Section 8.4.2 above on calculation of specific prevalence.

**When to Use** To describe patterns of birth defect occurrence.

**How to Use** Choose the most appropriate measure of birth defect occurrence. This will usually be prevalence, but in some circumstances it may be counts. For example, in planning for services, the actual number of children in a population born with a defect, such as cleft lip, may be more important than the prevalence. Analyze the chosen measure according to the basic epidemiologic parameters of time, place, and person (Teutsch and Churchill, 2000; Seiffert, 1994). Tables, graphs, and maps are very useful in presenting data in an understandable form; suggested approaches can be found in Teutsch and Churchill (2000).

*Time.* Surveillance programs should clearly state the relevant time period of study from which cases are drawn. This is usually based on ‘date of delivery’. Other options include ‘estimated date of conception’, ‘estimated date of delivery’, ‘date of diagnosis’, or ‘date of incorporation into the surveillance database’. Common analyses of birth defect occurrence by time include:

- Prevalence by year: to look for long-term trends
- Prevalence by season or month within the year: to look for seasonal patterns
- Counts of cases by date of delivery or estimated date of conception: to look at birth defect clusters

*Place.* Reports should specify the geographic area of coverage for the prevalence presented in the report. Common analyses of birth defect occurrence by place include:

- Prevalence by state, county, or region
- Prevalence by zip code or census tract
- Spot maps of cases showing residence at delivery or conception

*Person.* Common analyses of birth defect occurrence by person include:

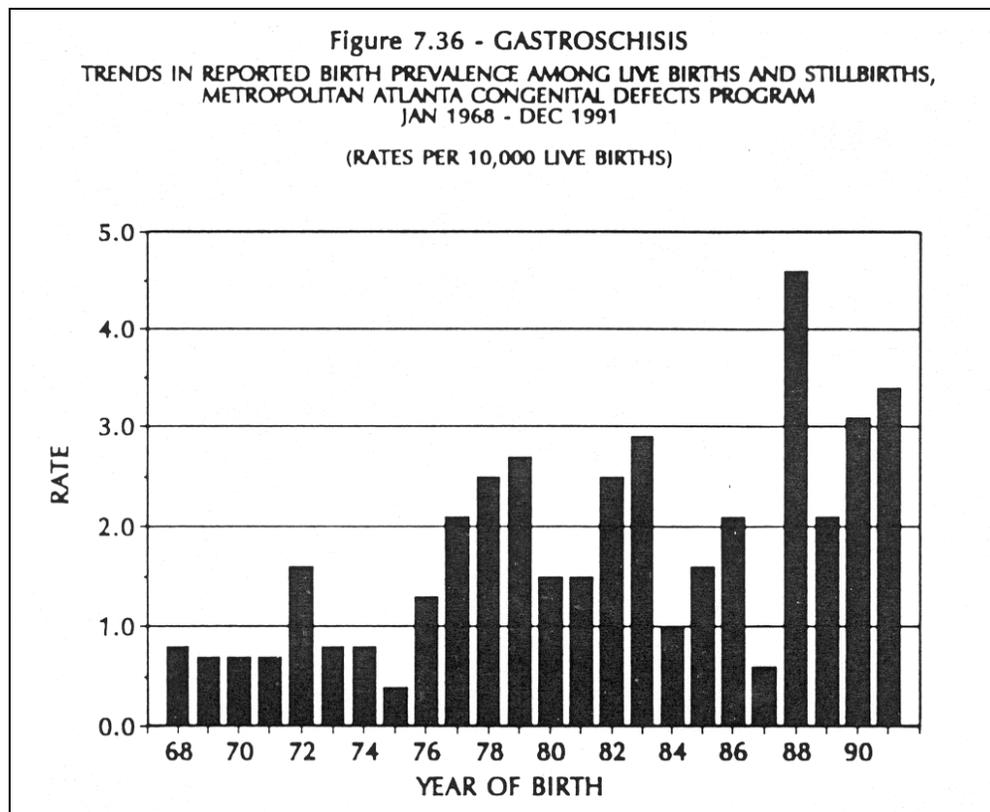
- Prevalence by maternal age (age of the mother at the time of delivery)

- Prevalence by maternal race/ethnicity (the most common definition is as stated by the mother)
- Prevalence by infant/fetus sex.

EXAMPLES

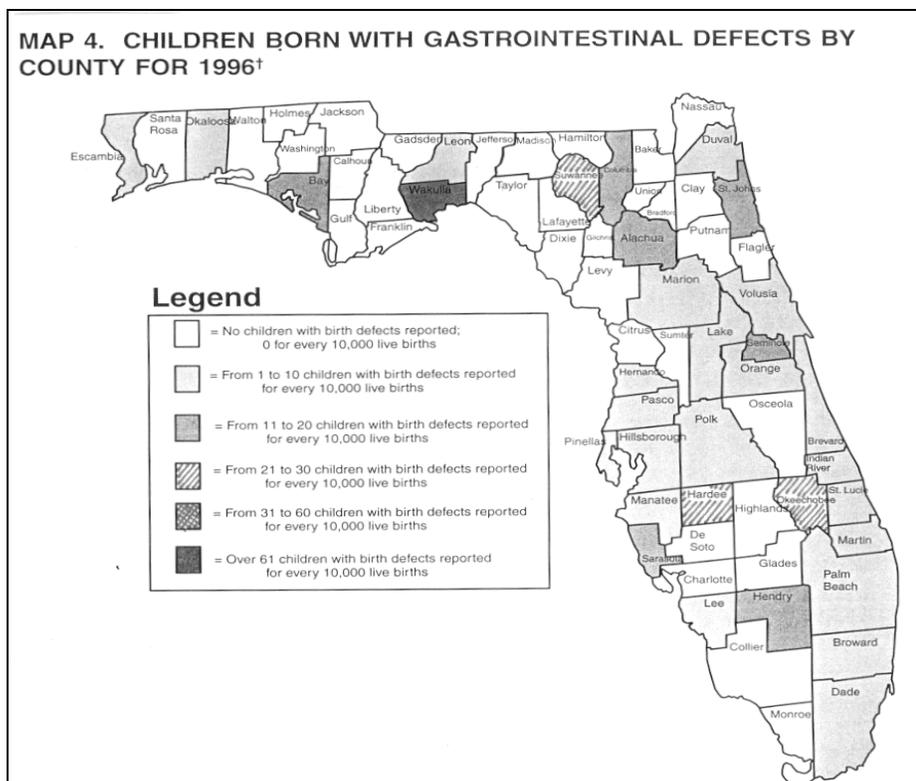
*Time* trend in the prevalence of gastroschisis over several years. The numerator (number of cases) and denominator (number of live births) for each year is determined. The prevalence for each year is calculated as:

$$\frac{\text{the number of cases with gastroschisis in Metro Atlanta in year X}}{\text{the number of live births in Metro Atlanta in year X}} \times 10,000$$



Source: Centers for Disease Control and Prevention, 1993, as cited in James et al., 1993.

Occurrence of birth defects by *place*, specifically, county prevalence of gastrointestinal defects. Numerators and denominators were collected for each county in Florida for deliveries in 1996. The results are presented in a choropleth map (a method of mapping to display quantitative information where the areas [e.g., counties] are colored or shaded according to the value of some variable [e.g., gastroschisis prevalence]).

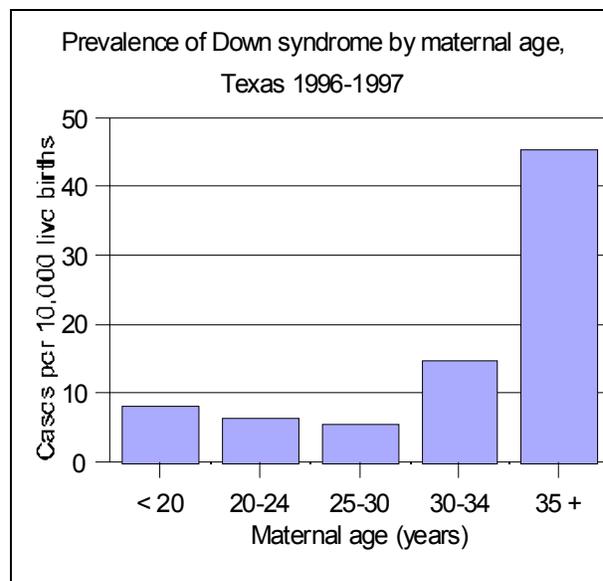


Occurrence of birth defects by *person*. Numerators and denominators were determined for five maternal age groups. Stratum-specific prevalence is calculated and presented in a table and vertical bar graph.

**Prevalence of Down syndrome by maternal age, Texas 1996-1997**

Maternal Age (years)	Cases	Live Births	Prevalence*
< 20	39	48,401	8.06
20-24	53	83,398	6.36
25-30	45	81,442	5.53
30-34	84	57,562	14.59
35 +	134	29,574	45.31

\*cases per 10,000 live births  
Source: Ethen and Case, 2000.



Source: Ethen and Case, 2000.

---

## 8.6 Confidence Intervals

---

In this section, we first discuss confidence intervals generally (Section 8.6.1). We then discuss the use of confidence intervals in comparing prevalence values (Section 8.6.2).

### 8.6.1 About Confidence Intervals

#### Definition

An interval around a statistic that contains the true underlying value of the statistic a certain amount of the time. For example, a 95% confidence interval for the prevalence of spina bifida will contain the true underlying value of the spina bifida prevalence 95% of the time.

The interval is bounded by an upper confidence limit and a lower confidence limit.

#### Background

The birth prevalence for a particular defect is estimated by the number of cases with the defect of interest ascertained from the population, divided by the total number of live births, and multiplied by some factor such as 10,000. This number is the best estimate of the true birth prevalence, which can never be known with certainty. To provide an idea about the precision of the estimated prevalence, a range of values is often calculated that is highly likely to contain the true prevalence. This range of likely values is called a *confidence interval*.

A confidence interval is calculated in such a way that, if the same procedure were to be repeated a large number of times, the proportion of intervals that contain the true prevalence would equal the confidence level. So, for example, if we choose a 95% confidence value, then 95% of all those confidence intervals will contain the true, but unknown, prevalence.

The 95% value is the conventionally used confidence interval. However, sometimes people choose other values, such as 90% confidence intervals (which are narrower than 95% confidence intervals) and 99% confidence intervals (which are wider).

Estimated confidence intervals for any given level (e.g., whether 90%, 95%, or 99%) will be narrower when their prevalence values are based on more cases.

Confidence intervals only measure random error, for example, when the occurrence of a birth defect fluctuates up and down from year to year by chance. They do not address systematic error or bias. Let's say, for example, one wanted to compare two surveillance programs. If program A does not ascertain cases from prenatal diagnosis clinics and thus consistently misses cases, while program B does ascertain cases from such clinics, a confidence interval around the prevalence of birth defects from program A should not be interpreted as accounting for the missing cases. The confidence interval merely addresses random fluctuation in the cases from program A.

**When to Use**

Confidence intervals may be calculated for any summary statistic, e.g., for proportions. However, we will only discuss confidence intervals for birth prevalence since that will be the most common statistic presented by birth defects surveillance programs. To learn about calculating confidence intervals for proportions or other types of rates, please consult one of the statistics books listed in the reference section (e.g., Dawson and Trapp, 2001; Fleiss et al., 2003; Snedecor and Cochran, 1989).

Confidence intervals can be used whenever calculating birth defect prevalence, although their use is controversial. This issue is discussed extensively in a recent commentary (Costa and Kirby, 2003) and a theory and methods paper (Correa-Villasenor et al., 2003) on the use of confidence intervals and on errors and undercounting in birth defects surveillance data. The interested reader is referred to these publications.

**Why to Use**

From a theoretical viewpoint, prevalence (e.g., X cases per 10,000 among deliveries in 1999) can also be considered to be just one sample in time, and confidence intervals give an idea of the range of values within which the true value is likely to be found. From a practical viewpoint, confidence intervals are particularly useful when dealing with small numbers of cases or where the birth defect prevalence for one group will be compared with that of other groups. This is because confidence intervals can help minimize reader concern about prevalence values that appear high or different when this is most likely due to random fluctuation. Some states have found this to be particularly helpful, for example, when looking at prevalence for counties, areas smaller than counties, or racial/ethnic groups. While the best way to compare prevalence values between different areas is always to use a statistical test, it is not practical for a surveillance system to anticipate all the comparisons readers will want to make. Confidence intervals thus provide a quick way for readers to get a rough idea of the impact of chance on the data.

**Why Not to Use**

Some surveillance programs ascertain all cases of birth defects, so that the prevalence reported is not just a sample but is considered to reflect the underlying true prevalence. Therefore, the use of confidence intervals is considered by many to be irrelevant. Calculating confidence intervals also increases statistical work for program staff. Finally, some data users, for example community groups or the media, may find confidence intervals confusing.

**How to Calculate**

Upper and lower 95% confidence limits are shown in the table below; they are the end points of the corresponding confidence intervals. Note that calculation of confidence intervals for prevalence is merely the calculation of confidence intervals for the number of cases – the denominator portion of prevalence does not change.

*For a prevalence based on a small number of cases.* For small numbers of cases (arbitrarily defined here as fewer than 30), use the Poisson distribution since birth defects are considered to be rare events.

The easiest way to use the Poisson distribution is to refer to a table that provides the upper and lower 95% confidence limits for an observed number of *cases* (reproduced below for up to 29 cases). Then follow Steps 1 through 3.

**Table of 95% Confidence Limits for the Number of Cases, for 1-29 Cases, Based on the Poisson Distribution**

Number of Cases	95% Confidence Limits	
	Lower	Upper
0	0.0000	3.6889
1	0.0253	5.5716
2	0.2422	7.2247
3	0.6187	8.7673
4	1.0899	10.2416
5	1.6235	11.6683
6	2.2019	13.0595
7	2.8144	14.4227
8	3.4538	15.7632
9	4.1154	17.0848
10	4.7954	18.3904
11	5.4912	19.6820
12	6.2006	20.9616
13	6.9220	22.2304
14	7.6539	23.4896
15	8.3954	24.7402
16	9.1454	25.9830
17	9.9031	27.2186
18	10.6679	28.4478
19	11.4392	29.6709
20	12.2165	30.8884
21	12.9993	32.1007
22	13.7873	33.3083
23	14.5800	34.5113
24	15.3773	35.7101
25	16.1787	36.9049
26	16.9841	38.0960
27	17.7932	39.2836
28	18.6058	40.4678
29	19.4218	41.6488

Source: Diem and Lentner, 1970.

**Step1.** Calculate prevalence.

$$\frac{\text{the number of cases with birth defect A in an area and time period}}{\text{the number of live births in that area and time period}} \times 10,000$$

**Step 2.** Look up the lower 95% confidence limit for the number of cases with birth defect A. Using this new number in the numerator, calculate the lower 95% confidence limit for prevalence:

$$\text{Lower 95\% CL for prevalence} = \frac{\text{lower 95\% CL for cases} \times 10,000}{\text{number of live births}}$$

**Step 3.** Look up the upper 95% confidence limit for the number of cases with birth defect A. Using this new number in the numerator calculate the upper 95% confidence limit for prevalence.

$$\text{Upper 95\% CL for prevalence} = \frac{\text{upper 95\% CL for cases} \times 10,000}{\text{number of live births}}$$

EXAMPLES using data from the California Birth Defect Monitoring Program 1983-1986.

Birth Defect	Cases		Number of Live Births <sup>3</sup>	Prevalence <sup>6</sup>	
	Number <sup>1</sup>	95% CI <sup>2</sup>		Value <sup>4</sup>	95% CI <sup>5</sup>
Anophthalmia	18	10.67 - 28.45	452,287	0.40	0.24 - 0.63
Glaucoma	27	17.79 - 39.29	452,287	0.60	0.39 - 0.87

Notes

Source: Croen et al., 1990.

- (1) Number of cases ascertained from surveillance
- (2) 95% confidence interval for that number of cases
- (3) Number of live births derived from vital records
- (4) Prevalence = [(1) / (3)] X 10,000
- (5) 95% confidence interval for the prevalence
- (6) Prevalence expressed as cases per 10,000 live births

*For a prevalence based on a large number of cases.* For a large number of cases (arbitrarily defined here as 30 cases or more), use the normal distribution. Why? As the number of cases grows larger, the Poisson distribution approximates (i.e., looks more and more like) the normal distribution. The formulae below are approximations for calculating confidence intervals using the normal distribution (Rothman and Boice, 1982, p. 29, formula 19).

Shorthand: let  $c$  = number of cases

$b$  = number of live births

1. Calculate the lower confidence limit using the following:

$$\text{Lower 95\% CL for prevalence} = c \times \left( 1 - \frac{1}{9c} - \frac{1.96}{3} \sqrt{\frac{1}{c}} \right)^3 / b \times 10000$$

2. Calculate the upper confidence limit using the following:

$$\text{Upper 95\% CL for prevalence} = (c + 1) \times \left( 1 - \frac{1}{9(c + 1)} + \frac{1.96}{3} \sqrt{\frac{1}{(c + 1)}} \right)^3 / b \times 10000$$

- 3. To determine the 90% confidence limits, replace 1.96 with 1.645. To determine 99% limits, replace 1.96 with 2.575.
- 4. To obtain confidence limits for the number of cases instead of the prevalence, apply the formulae but do not divide by births (*b*) or multiply by 10,000.

EXAMPLES using data from the California Birth Defect Monitoring Program, 1983-1986.

Birth Defect	Cases		Number of Live Births <sup>3</sup>	Prevalence <sup>6</sup>	
	Number <sup>1</sup>	95% CI <sup>2</sup>		Value <sup>4</sup>	95% CI <sup>5</sup>
Aortic stenosis	73	57.22 - 91.79	452,287	1.61	1.27 - 2.03
Cleft palate	320	285.89 - 357.05	452,287	7.08	6.32 - 7.89

Notes

Source: Croen et al., 1990.

- (1) Number of cases ascertained from surveillance
- (2) 95% confidence interval for that number of cases
- (3) Number of live births derived from vital records
- (4) Prevalence = [(1) / (3)] X 10,000
- (5) 95% confidence interval for the prevalence
- (6) Prevalence expressed as cases per 10,000 live births

*Software support.* For a few prevalence values, confidence limits (and hence the resulting intervals) can be calculated by hand or using statistical software such as PEPI. One program in PEPI (POISSON) gives the table values for any number of cases; for a large number of cases it gives a normal approximation. PEPI software and documentation are available at:

<http://sagebrushpress.com//pepibook.html>

For many prevalence values, it is useful to write programs, for example in SAS, to calculate the confidence limits along with the prevalence.

### 8.6.2 Comparing Prevalence Values Using Confidence Intervals

The best way to compare prevalence values (e.g., for different maternal ages) is to do a statistical test; one type of statistical test is described in the second section below (“When both prevalence values are based on 30 or more events”). However, in the first section below we present a quick method recommended by the National Center for Health Statistics (NCHS) that works better when one of the prevalence values is based on fewer than 30 cases. Note that the NCHS-recommended method is conservative (i.e., there will be fewer statistically significant differences than would be found by actually performing a statistical test).

**When one of the prevalence values is based on fewer than 30 cases.** First compute the 95% confidence intervals for both prevalence values. Check to see if those intervals overlap. If they *do* overlap, the difference is not statistically significant at the 95% level. If they *do not* overlap, the difference is indeed ‘statistically significant’ or unlikely to be explained by chance alone. (Method recommended by NCHS in Ventura et al. [2000].)

EXAMPLE

The prevalence of holoprosencephaly among African-American women in Texas (2.39 cases per 10,000 live births) is over three times higher than among White women in Texas (0.78 cases per 10,000 live births). Is the difference statistically significant? First compute the 95% confidence intervals.

**Holoprosencephaly Among Two Race/Ethnic Groups in Texas, 1996/97**

Group	Cases	Live Births	Prevalence*	95% CI for Prevalence
African American women	7	29,254	2.39	0.96 - 4.93
White women	8	102,193	0.78	0.34 - 1.54

\* cases per 10,000 live births

Source: Ethen and Case, 2000.

These two confidence intervals overlap. Thus, based on this approach, the difference between prevalence of holoprosencephaly in African-American women compared to White women is not statistically significant.

**When both prevalence values are based on 30 or more events.** This approach is based on calculating the confidence interval for the difference between the two prevalence values. If this interval includes 0.00, then the difference in the values is not considered to be statistically significant. Since this approach uses information from both prevalence values at the same time, it is more statistically powerful than the NCHS-recommended method. That is, if a difference truly exists, this approach will identify that more often than will the NCHS-recommended method. This approach uses the standard error for the difference between the two prevalence values (Rothman, 1986, p. 170, formulae 11-15).

Statistical software like the RATES2 program within the PEPI package can also be used to calculate this confidence interval.

Shorthand: let  $RD$  = higher prevalence - lower prevalence

$c_1$  = number of cases used to calculate the first prevalence

$c_2$  = number of cases used to calculate the second prevalence

$b_1$  = number of live births used to calculate the first prevalence

$b_2$  = number of live births used to calculate the second prevalence

1. Calculate the lower confidence limit using the following:

$$\text{Lower 95\% CL for prevalence difference} = RD - \left( 1.96 \sqrt{\frac{c_1}{b_1^2} + \frac{c_2}{b_2^2}} \times 10000 \right)$$

2. Calculate the upper confidence limit using the following:

$$\text{Upper 95\% CL for prevalence difference} = RD + \left( 1.96 \sqrt{\frac{c_1}{b_1^2} + \frac{c_2}{b_2^2}} \times 10000 \right)$$

3. To obtain 90% confidence limits, replace 1.96 with 1.645. To obtain 99% limits, replace 1.96 with 2.575.
4. If the confidence interval *does not include 0.00*, then the difference would occur by chance less than 5 times out of 100 (for 95% confidence intervals); i.e., the two prevalence values are significantly different.

EXAMPLE

Is there a statistically significant difference between African-American and Hispanic women in the prevalence of births with atrial septal defects?

**Atrial Septal Defect Among Two Race/Ethnic Groups in Texas, 1996/97**

Group	Cases	Live Births	Prevalence *
African-American women	133	29,254	45.46
Hispanic women	835	160,094	52.16

\* cases per 10,000 live births

Source: Ethen and Case, 2000.

1. The lower 95% confidence limit for the prevalence difference = -1.80.
2. The upper 95% confidence limit for the prevalence difference = 15.20.
3. The 95% confidence interval for the prevalence difference thus = -1.80 to 15.20.
4. The interval includes 0.00. Therefore, the difference between the two prevalence values is not statistically significant at the 95% confidence level; i.e., there is no statistically significant difference.

---

## 8.7 Issues to Consider If Data Reveal Unusual Patterns

---

It is very important to rule out relatively straightforward explanations for a change in the birth prevalence of a birth defect. Among the more common reasons are:

- Changes in medical diagnoses and technologies
- Changes in reporting that lead to changes in case ascertainment
- Changes in the population at risk (focus on age, period, and cohort effects)
- Random variation

Please refer to Kallen (1998, pp. 83-87) for a more extended discussion of the issues identified above.

The analytical capabilities of the surveillance system should support evaluation of the likelihood of these factors being responsible for observed changes in prevalence. Although the remainder of this section addresses “changes” in birth defect prevalence over time, it can also be applied to “differences” in birth defect prevalence between areas.

### 8.7.1 Changes in Medical Diagnoses and Technologies

To detect changes in medical diagnoses, it is important to compare isolated and multiple birth defects cases. Minor changes in the way a condition is diagnosed or reported can affect the coding and classification of specific birth defects. For example, the prevalence of neural tube defects among live-born infants may have declined during the 1980s from levels reported in the 1970s, due to the development and widespread availability of prenatal diagnostic tests, such as maternal serum alpha-fetoprotein screening and ultrasonography. The severity of spina bifida cases may be less today than in the 1970s due to the selective therapeutic termination of the more severe cases, which are more likely to be identified prenatally.

The birth prevalence of some disorders may increase due to new technologies. For example, fragile X syndrome, a chromosomal breakage disorder, is diagnosed much more often today than 10 years ago and was unknown 20 years ago.

### 8.7.2 Changes in Reporting and Case Ascertainment

Re-verification that the frequency is an un-duplicated count is also appropriate. Many suspected “clusters” reported by the media or concerned citizens involve multiple counting of the same cases. With a birth defects surveillance program that ascertains cases from multiple sources, it is important to ensure that each case is counted only once, even if reports are received from several health care providers or delivery settings. The same is true of changes in reporting. As hospitals shift to computerized diagnostic indices, reporting artifacts could decrease the numbers of cases of specific birth defects while increasing others. This is because a limited number of ICD-9-CM codes are retained in the index, and conditions that appear to be minor in the eyes of the medical records clerk may be omitted. If surveillance staff rely exclusively on the diagnostic indices to identify charts to abstract, some conditions may be missed. See also Chapter 5 on Classification and Coding and Chapter 6 on Case Ascertainment Methods.

### **8.7.3 Changes in the Population at Risk**

Population characteristics can be controlled for by using stratum-specific prevalence, age-adjustment, and similar methods. However, as most statistical surveillance methods are based on the frequency of events rather than on proportions, rates, or prevalence, separate analyses will need to be conducted to rule out changes in the population at risk. Analysts who routinely examine birth defects surveillance data will have access to detailed, current population estimates and should examine the demographic and reproductive health characteristics of all women giving birth to identify changes in the population at risk.

### **8.7.4 Random Variation**

It is also possible, and in fact very likely, that an observed difference in the frequency of a specific birth defect is due to random variability. With relatively low birth prevalence, cases of a particular condition will be quite rare, and the coincidence of two or more cases in space and/or in time may be just that: a coincidence. Confidence intervals are one way to address random variation.

---

## 8.8 References

---

- Correa-Villaseñor A, Satten GA, Rolka H, Langlois P, Devine O. Random error and undercounting in birth defects surveillance data: implications for inference. *Birth Defects Research, Part A*. 2003;67:610-616.
- Costa R, Kirby RS. Confidence intervals with birth defects surveillance data: a tempest in a teapot or honestly significant differences of opinion? *Birth Defects Research, Part A*. 2003;67:609.
- Croen L, Schulman J, Roeper P. *Birth Defects in California January 1, 1983 -- December 31, 1986. A Report of the California Birth Defects Monitoring Program*. Emeryville, CA: California Birth Defects Monitoring Program; 1990.
- Dawson B, Trapp RG. *Basic and Clinical Biostatistics*. 3<sup>rd</sup> ed. New York, NY: Lange Medical Books – McGraw-Hill; 2001.
- Diem K, Lentner C, eds. *Scientific Tables*. 7<sup>th</sup> ed. Basle, Switzerland: Ciba-Geigy Ltd.; 1970.
- Ethen M, Case A, eds.. *Report of Birth Defects among 1996 and 1997 Deliveries*. Austin, TX: Texas Birth Defects Monitoring Division, Texas Department of Health; September 2000.
- EUROCAT Working Group. *Report 8: Surveillance of Congenital Anomalies in Europe, 1980 – 1999*. Newtownabbey, Northern Ireland: EUROCAT Central Registry, University of Ulster; 2002.
- Fleiss JL, Levin BA, Paik MC. *Statistical Methods for Rates and Proportions*. 3<sup>rd</sup> ed. Hoboken, NJ: J Wiley; 2003.
- Florida Department of Health. *Annual Report on Florida Birth Defects Registry, 1996 Annual Report on Birth Defects in Florida*. Tallahassee, FL: Florida Department of Health; 1999.
- International Clearinghouse for Birth Defects Monitoring Systems. *Annual Report 2002, with Data for 2000*. Rome, Italy: International Centre for Birth Defects; 2002.
- James LM, McClearn AB, Waters GD. Metropolitan Atlanta Congenital Defects Program surveillance data, 1988 – 1991. *Teratology*. 1993;48(6):695-709.
- Kallen B. *Epidemiology of Human Reproduction*. Boca Raton, FL: CRC Press; 1988.
- Last JM, ed. *A Dictionary of Epidemiology*. 3<sup>rd</sup> ed. New York, NY: Oxford University Press; 1995.
- Rothman KJ. *Modern Epidemiology*. Boston, MA: Little, Brown and Company; 1986.
- Rothman KJ, Boice JD Jr. *Epidemiologic Analysis with a Programmable Calculator*. Boston, MA: Epidemiology Resources Inc.; 1982.
- Seiffert JE, ed. *Standards for Cancer Registries Vol III. Standards for Completeness, Quality, Analysis, and Management of Data*. Springfield, IL: American Association of Central Cancer Registries (AACCR); 1994.

Sever LE. Birth defects surveillance systems and neural tube defects. In: Wyszynski DF, ed. *Neural Tube Defects: From Origin to Treatment*. Oxford, England: Oxford University Press; 2004 (in press).

Snedecor GW, Cochran WG. *Statistical Methods*. 8<sup>th</sup> ed. Ames, IA: Iowa State University Press; 1989.

Teutsch SM, Churchill RE, eds. *Principles and Practice of Public Health Surveillance*. New York, NY: Oxford University Press; 2000.

Ventura SJ, Martin JA, Curtin SC, Mathews TJ, Park MM. Births: final data for 1999. *National Vital Statistics Reports*. 2000;48(3):96.