Chapter 2

Methodology

This chapter describes the methodology used in the two studies reported in this thesis. Graphical and statistical analyses were carried out using the R program (R Development Core Team 2007). They include methods for descriptive analysis when measuring prevalence, and odds ratios for measuring associations. Pearson's chi-squared test was used to assess the overall association between binary outcomes and categorical determinants, and logistic regression was used to model the detailed associations between these outcomes and determinants after adjusting for other determinants.

2.1 Data source and data management

Data used were taken from the database of 26,158 women who delivered singleton babies in Pattani hospital during the period from 1 October 1996 to 30 September 2005. Demographic information including complications, delivery type, birth weight, 1-minute and 5-minute Apgar scores, and records of neonatal morbidity were collected by birth attendants.

Relevant data for our studies were transferred to a MySQL database and extensively checked and cleaned to eliminate errors arising from faulty and incomplete data entry where possible. The data were then stored in text files suitable for analysis using the R statistical system (Venables and Smith 2002). This software has a suite of many appropriate functions for graphing and analysing statistical data.

The data were analyzed using neonatal morbidity (complication based-risk factors for neonatal morbidity) as an outcome for the first paper, and using caesarean delivery as an outcome for the second paper. Due to the fact that caesarean section was recorded as a complication and at most one complication was recorded for each delivery, the numbers of cases were 19,268 for the first study and 25,829 for the second study as shown in Figure 2.1.

Of 26,158 total deliveries during the period covered by the study, there were 329 twins or triplets, 3,250 transfers or referrals, and 3,311 cases with previous caesarean deliveries. These cases were excluded from the first study due to the absence of other delivery "complications" recorded in the database.

For the second study the excluded cases were only the twins or triplets: no higher order multiple births occurred. Since transferred or referred cases, previous pregnancies, and previous caesarean section deliveries are intervening variables, they could bias the associations between caesarean delivery outcomes and other determinants. Therefore, the data were stratified by new versus transferred or referred cases and analyzed separately into six groups. There are Group 1:1st New (first pregnancy mothers, 7,232 new cases), Group 2:1st Ref (first pregnancy mothers referred or transferred to the delivery room, 1,407 cases), Group 3:2+NoCNew (subsequent pregnancy and no previous caesarean, 12,036 cases), Group 4:2+NoCRef (subsequent pregnancy and no previous cesarean and referred or transferred, 1,843 cases), Group 5:2+PreCNew (subsequent pregnancy and previous cesarean delivery, 2,817 cases), and Group 6:2+PreCRef (subsequent pregnancy and previous cesarean delivery and referred or transferred, 494 cases).

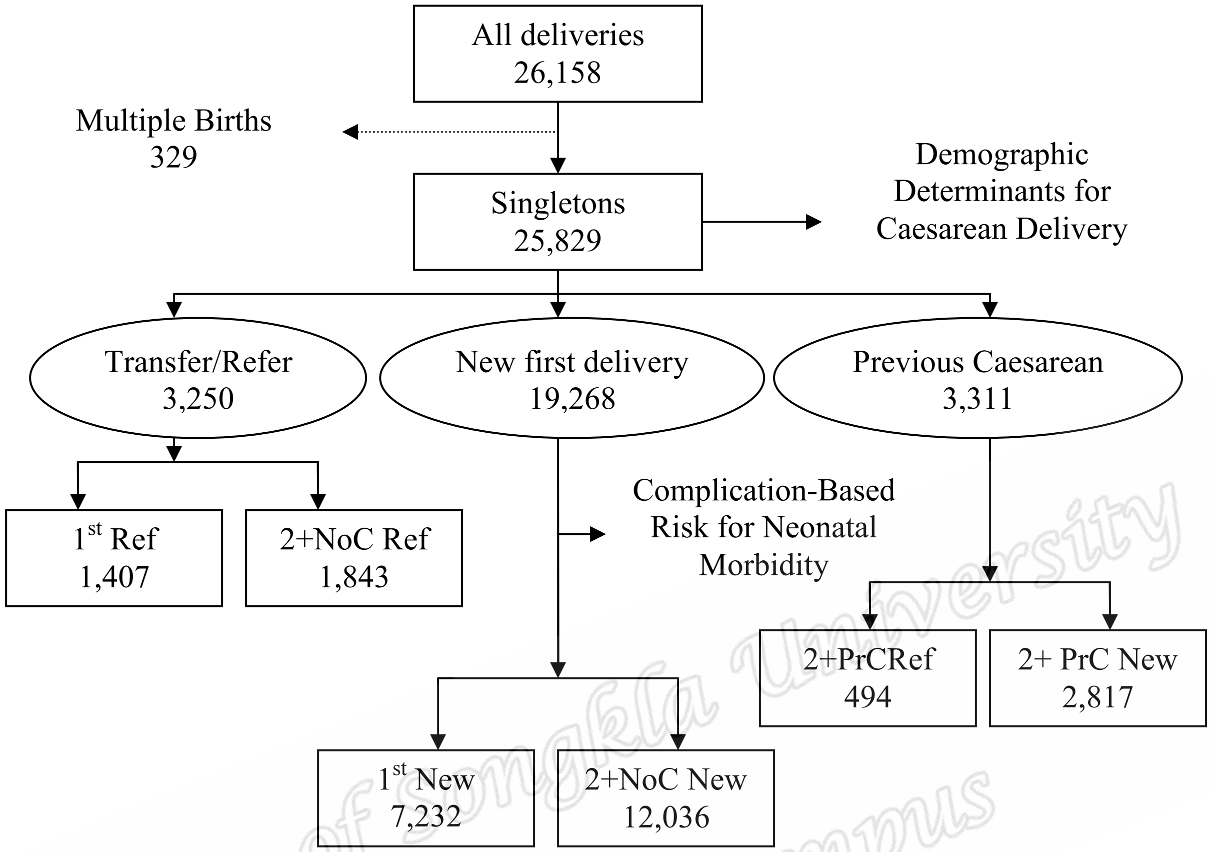

Figure 2.1: Numbers of subjects in various groups

2.2 Study variables

For the first study, a severity score for complications was assessed based on the 62 specified complications of the mothers recorded in the database.

Severity Score for Complications

A significant outcome for the study of complication-based risk factors was not appropriate because neonatal mortality was rare. While manual guidelines are used for detecting possible complications based on the mother's obstetric history, there is no objective measure of morbidity based on all the delivery complications that routinely occur. A score for predicting the morbidity risk was used based on the assessments of 11 obstetricians who worked in three different hospitals: Pattani Hospital

(4 obstetricians), Chiang Rai Hospital (4 obstetricians) and Trang Hospital (3 obstetricians). They agreed to provide their assessments when approached and gave a score from 0 to 9 for each of the 60 complications specifying the condition of the infant after birth. A complication with average score of 7 or more was defined as a substantial neonatal morbidity risk, as shown in Figure 2.2. From this information we evaluated the score criteria as that of 7 or less for risk morbidity in this study which was applied by Dr. Virginia Apgar in 1953 (Apgar 1953). The method was also used in previous studies as (Suwannachat 2004, Jerneck et al 2001 and Carla 2006), the Apgar score below 7 in newborns lead to high mortality and morbidity.

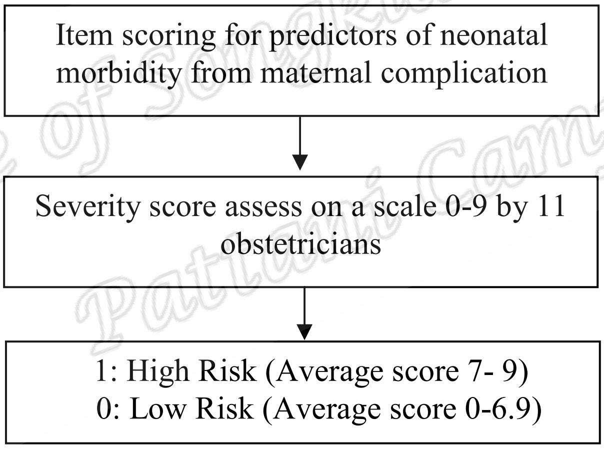

Figure 2.2: Path diagram for severity score based on complication

The obstetricians were asked to assess the risk to the baby on a scale from 0 (no risk) to 9 (most serious risk), for each of the 60 delivery complications listed in Table 2.1. The complications coded 61 and 62 are “None” and “not stated”, respectively. The 60 complications on the list are the ones routinely recorded at the hospital, in the same order and filled in routinely by the same delivery personnel in the time span of data collected. Some of the items on the list are associated with very low risk and arguably could be omitted from consideration, and others are not defined well. However, the same list as that used in the hospital was used, in the same order, to ensure the obstetricians' opinions were not pre-empted or biased by any factors other than their own personal judgments. In Table 2.1, A, B, C, …, K are codes for the 11 obstetricians, with their scores, ranges, medians and averages for each complication.

CID	*Complication*	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*range*	*Average*	*median*
1	dead fetus in utero	9	9	9	9	9	9	9	9	9	9	9	0	9.0	9
2	uterine rupture	9	9	9	9	9	9	9	9	9	8	9	1	8.9	9
3	prolapsed cord	8	9	9	9	9	9	9	9	9	9	8	1	8.8	9
4	anancephalus	9	9	9	9	9	9	8	9	9	8	8	1	8.7	9
5	eclampsia	9	8	8	8	9	9	9	9	9	8	8	1	8.5	9
6	abruption placenta	8	9	9	9	9	9	9	9	8	6	7	3	8.4	9
7	fetal distress	8	9	8	9	9	9	9	8	8	8	7	2	8.4	8
8	shoulder dystosia	8	9	9	9	9	9	8	8	9	4	9	5	8.3	9
9	thick mecomium stain	6	9	8	8	9	9	8	8	8	8	8	3	8.1	8
10	fetal anomaly	9	8	8	9	9	9	6	8	8	7	7	3	8.0	8
11	hydrocephalus	7	9	8	8	9	9	5	8	8	8	8	4	7.9	8
12	amniotic ambolism	7	8	8		9	9	9	9	9	8	6	3	8.2	9
13	pregnancy with malaria	6	5	7	7	9	9	7	8	8	8	8	4	7.5	8
14	severe PIH	8	6	7	8	9	8	8	7	7	6	7	3	7.4	7
15	pre-term labour	5	7	6	7	9	9	8	8	8	7	7	4	7.4	7
16	oligahydramios	6	8	7	7	7	7	8	7	8	8	8	2	7.4	7
17	face presentation	6	6	8	8	9	8	8	7	7	7	7	3	7.4	7
18	overt diabetes	6	6	7	7	9	9	8	8	7	6	7	3	7.3	7
19	chronic HT with accrevate	5	8	6		9	8	8	7	8	4	8	5	7.1	8
20	prolong stage 2	3	8	8	7	9	7	8	6	7	7	7	6	7.0	7
21	membrane leak > 24 hours	6	7	5	7	8	8	7	8	8	4	8	4	6.9	7
22	pulmonary edema	6	7	6	7	9	9	8	8	6	5	5	4	6.9	7
23	post-term labour	5	7	7	6	9	5	7	6	8	7	7	4	6.7	7
24	low fetal movement	2	7	5	7	9	7	6	6	8	5	8	7	6.4	7
25	chronic hypertension	4	7	5	6	9	8	7	7	6	5	6	5	6.4	7
26	premature rupture of membrane	5	6	6	6	8	5	7	6	7	7	6	3	6.3	6
27	heart disease	5	6	5	5	7	8	7	8	6	6	6	3	6.3	6
28	VDRL positive	2	7	6	6	8	7	5	4	8	7	8	6	6.2	7
29	CPD	5	8	6	6	8	8	7	7	5	4	4	4	6.2	6
30	gestational diabetes	5	4	6	5	7	7	7	8	7	6	6	4	6.2	6
31	pregnancy + thalasemia	6	5	7	7	6	5	6	7	6	6	2	6.1	6	
32	birth before admission	6	7	4	6	7	5	7	3	7	7	7	4	6.0	7
33	ante-partum hemorrhage	6	7	5	6	7	5	8	3	6	8	5	5	6.0	6
34	breech presentation	6	4	7	6	9	4	7	4	6	6	6	5	5.9	6
35	herpes	5	6	5	5	7	4	6	3	8	8	8	5	5.9	6
36	trauma	3	6	6	2	7	8	6	8	7	7	5	6	5.9	6
37	polyhydramios	6	6	7	6	8	5	7	4	5	5	5	4	5.8	6
38	epilepsy	6	6	5	2	8	8	4	7	6	5	7	6	5.8	6
39	pyelonephritis	4	4	5	6	8	6	6	6	6	5	6	4	5.6	6
40	placenta previa	4	6	7	3	8	2	9	3	6	8	5	7	5.5	6
41	mild PIH	7	4	3	5	7	6	5	6	6	5	6	4	5.5	6
42	fever	6	5	5	2	8	7	2	6	6	6	6	6	5.4	6
43	transverse lie	5	1	9	7	7	5	8	4	4	3	5	8	5.3	5
44	gut obstruction	2	3	4	4	8	8		7	6	5	5	6	5.2	5
45	prolong stage 1	2	5	6	6	5	2	7	2	7	7	7	5	5.1	6
46	appendicitis	5	3	4	2	8	7	6	6	6	5	4	6	5.1	5
47	condyloma	4	4	5	1	6	5	5	5	7	7	7	6	5.1	5
48	pregnancy with UTI	4	4	5	3	7	5	5	4	7	5	6	4	5.0	5
49	pregnancy with hyperthyroid	3	2	5	4	7	6	5	6	6	4	5	5	4.8	5
50	gestational hypertention	4	2	3	6	8	3	4	4	5	7	5	6	4.6	4
51	myoma uteri	3	1	6	2	7	5	2	3	6	6	7	6	4.4	5
52	anemia	4	1	4	2	7	6	4	3	6	5	6	6	4.4	4
53	asthma	4	2	5	3	7	6	4	5	4	4	3	5	4.3	4
54	birth with midwife at home	4	5	6	4	2	2	4	3	6	4	4	4	4.0	4
55	acute diarrhea	3	2	5	2	6	3	3	3	4	3	3	4	3.4	3
56	gastroenteritis	2	2	3	2	7	5	2	3	4	3	2	5	3.2	3
57	occiput presentation	0	0	5	1	1	1	2	3	5	4	4	5	2.4	2
58	gestational edema	1	0	2	1	0	0	0	3	2	5	2	5	1.5	1
59	post-partum hemorrhage	0	0	1	0	5	3	1	3	0	1	1	5	1.4	1
60	retain placenta	0	0	1	0	3	0	0	0	0	0	0	3	0.4	0
	Average	5	5.5	6	5.4	7.5	6.4	6.2	5.9	6.6	5.9	6.1	2.5		

Table 2.1: Data recorded from 60 complications based on opinions of 11 obstetricians

For the second study, caesarean delivery was taken as the binary outcome. It is the type of delivery where surgical delivery of a baby is performed through an incision in the abdomen and the uterus.

Common determinants for the two studies are religion, education, occupation, mother's age, budget year, and residence. Residence was classified into 15 categories for the first study but nine categories for the second study due to insufficient cases for statistical analysis. Pattani city was sectioned into six groups comprising: 1. Anuru; 2. Bana; 3. Sabarang; 4. City South; 5.City Centre and 6. City East. The other groups/areas were: 7. Narathiwat Province/Yala Province/Not stated; 8. Nong Chik district; 9. Khokpho/Maelan district; 10. Pattani East comprising (Panare district/Saiburee district/Maikan district); 11. Yaring district; 12. Rusamilae sub-district in (City); 13. Yarang district; 14. Pattani South comprising (Mayo district/Kapo dstrict/Thoungyandang district) and 15. Songkhla Province.

For residence in the second paper we classified into 9 groups. They were 1: Muang (City) comprising the 12 sub-districts of Muang City in Pattani Province, 2: Nong Chik district, 3: Khokpho and MaeLan districts, 4: Pattani East (comprising Panare district, Saiburee district and MaiKan district, 5: Yarang district, 6: Pattani South comprising Mayo district, Kapo dstrict and ThungYanDang district, 7: Songkhla Province, 8: Yaring district and 9: Narathiwat Province,Yala Province and not stated.

The number of previous pregnancies was considered as a determinant for the first study, but was taken as a stratification variable for the second study. Figure 2.3 shows a path diagram with seven demographic determinants for neonatal morbidity and six demographic determinants for caesarean delivery. The numbers in brackets are the numbers of categories. New or Transferred/Referred is an intervening variable in the path between demographic determinants and caesarean delivery outcomes.

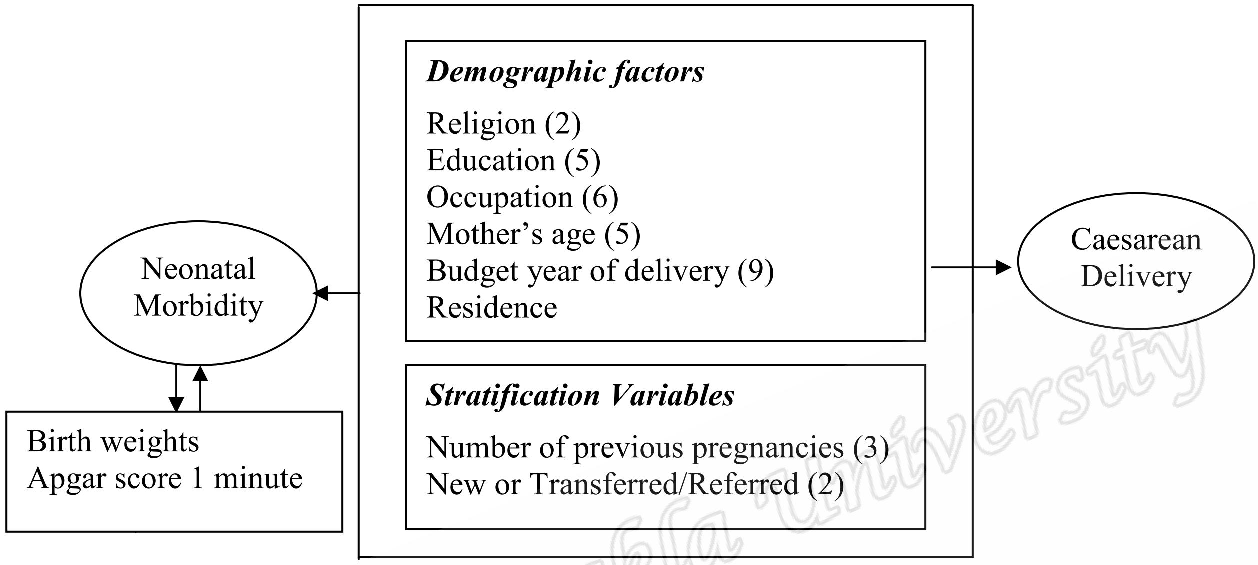

Figure 2.3: Path diagram of study variables

2.3 Statistical methods

The statistical methods comprise methods for univariate analysis and statistical modeling.

Univariate Analysis

Pearson's chi-squared test and 95% confidence intervals for odds ratios are conventionally used to assess the association between the outcome and demographics determinants. For the odds ratio, the null value is conventionally taken to be one, corresponding to equal risks of an outcome in two comparison groups. This corresponds to a null value of zero for the difference between two population proportions under the null hypothesis. The Pearson's chi-squared test gives the p-value for testing no relationship between the determinant and the outcome.

The homogeneity test is used to tell if the association could be the same in different strata, small *p*-values providing evidence to the contrary(McNeil 1998a).

A 2×2 table

To illustrate the methods, a *2×2* contingency table is constructed as follows. Let *x* be the binary determinant and *y* the binary outcome coded as zero or one, and *a*, *b*, *c*, and *d* the cell counts (McNeil 1998a, 1998b).

		y = 1	*y* = 0
x	1	*a*	*b*
	0	*c*	*d*

$n = a + b + c + d$

The odds ratio is

$$\text{OR} = \frac{ad}{bc} \tag{2.1}$$

Its asymptotic standard error is given by

$$SE(\ln OR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \tag{2.2}$$

A 95% confidence interval is thus

$$95\%\ \text{CI} = \text{OR} \times \exp\left(\pm 1.96\ SE\ [\ln \text{OR}]\right) \tag{2.3}$$

Pearson's chi-squared statistic is defined as

$$\chi^2 = \frac{(ab - bc)^2 n}{(a+b)(c+d)(a+c)(b+d)} \tag{2.4}$$

$r \times 2$ tables

In this study, some of risk factors are multi-categorical, having more than two category levels. We use non-stratified $r \times 2$ tables to compare them. For example, x is mother education level and y is neonatal morbidity (1: high risk, 0: low risk).

		y = 1	y = 0
x	1	a_{11}	a_{12}
	2	a_{21}	a_{22}
	…	..	..
	r	a_{r1}	a_{r2}

Thus the estimate of the odds ratio (OR) is

$$\mathrm{OR}_{ij} = \frac{a_{ij} d_{ij}}{b_{ij} c_{ij}} , \tag{2.5}$$

where $b_{ij} = \sum_{j=1}^{2} a_{ij} - a_{ij}$, $c_{ij} = \sum_{j=1}^{r} a_{ij} - a_{ij}$, $d_{ij} = n - a_{ij} - b_{ij} - c_{ij}$, $n = \sum_{i=1}^{r} \sum_{j=1}^{2} a_{ij}$

The standard error of the natural logarithm of the odds ratio is given by the same formula as for the 2×2 table. In general, the association is composed of $r \times c$ odds ratios, but only (r-1) ×(c-1) of them are independent.

The standard error is given by

$$SE(\ln OR_{ij}) = \sqrt{\frac{1}{a_{ij}} + \frac{1}{b_{ij}} + \frac{1}{c_{ij}} + \frac{1}{d_{ij}}} \tag{2.6}$$

A 95 % confidence interval is thus

$$95\ \%\ \mathrm{CI} = \mathrm{OR} \times \exp\left(\pm 1.96\ SE\ [\ln \mathrm{OR}]\right) \tag{2.7}$$

Pearson's chi-squared statistic for independence (i.e., no association) in an $r \times c$ table is defined as

$$\chi^2_{(r-1)(c-1)} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(a_{ij} - \hat{a}_{ij}\right)^2}{\hat{a}_{ij}} \quad (2.8)$$

where $b_{ij} = \sum_{j=1}^{c} a_{ij} - a_{ij}$, $c_{ij} = \sum_{j=1}^{r} a_{ij} - a_{ij}$, $d_{ij} = n - a_{ij} - b_{ij} - c_{ij}$, $n = \sum_{i=1}^{r}\sum_{j=1}^{c} a_{ij}$.

When the null hypothesis of the independence is true, this has a chi-squared distribution with $(r-1)\times(c-1)$ degrees of freedom For a binary outcome, $c = 2$ (McNeil 1998a).

Logistic Regression

Logistic regression analysis is used for adjusting the association between several demographic determinants and neonatal morbidity outcome for the first study, and caesarean delivery outcome for the second study. Logistic regression is a method of analysis that gives a particularly simple representation for the logarithm of the odds ratio describing the association of outcome with factors, and when fitted to data involving binary outcome and multiple determinants, it automatically provides estimates of odds ratio and confidence intervals for specific combinations of the risk factors (McNeil 1996).

For the first study, logistic regression (Hosmer and Lemeshow 2000, Kleinbaum and Klein 2002) was then used to estimate the proportions of neonatal morbidity outcomes in cells defined by combinations of the seven demographic factors, using the additive model with the form

$$\ln\left(\frac{P_{ijklmno}}{1-P_{ijklmno}}\right) = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_m + \xi_n + \kappa_o \,. \qquad (2.9)$$

In this model μ is a constant and the terms α_i, β_j, γ_k, δ_l, ε_m, ξ_n and κ_o refer to religion, age group, residence, education, occupation, budget year, and number of previous pregnancies respectively. To avoid over-specification of the parameters, each set of coefficients was constrained to have a mean equal to 0. To calculate the proportion of caesarean deliveries for each factor after adjusting for the effects of the other factors, equation (2.9) was used with the terms associated with the other factors replaced by a constant, chosen to make the sum of the expected number of cases based on the model equal to the observed number, using a Newton-Raphson iterative procedure with Marquardt damping.

For the second study of caesarean deliveries the model in equation (2.9) is reduced to six factors of determinants and it takes the form (Hosmer and Lemeshow 2000, Kleinbaum and Klein 2002)

$$\ln\left(\frac{P_{ijklmn}}{1-P_{ijklmn}}\right) = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_m + \xi_n \,. \qquad (2.10)$$

It is also informative to plot confidence intervals for the adjusted proportions.