sciendo

DOI: 10.2478/bile-2024-0009

Biometrical Letters
Vol 61 (2024), No. 2, 137-146

Inconsistency between conditional and marginal analyses

Chang Feng[1], Honghong Liu[2], Hongyue Wang[2], and Changyong Feng[*,2,3]

[1]School of Environment and Health
Jianghan University, Wuhan 430056, Hubei, China
[2]Departments of Biostatistics and Computational Biology
[3]Department of Anesthesiology and Perioperative Medicine
University of Rochester, Rochester, NY 14642, USA
[*]e-mail: changyong_feng@urmc.rochester.edu

Summary

Conditional and marginal analyses are widely used in clinical studies. However, the results of these two methods may occasionally contradict each other. For instance, marginal analysis may show that the treatment group outperforms the control group, while conditional analysis may suggest the opposite. We examine the causes of this inconsistency and provide general sufficient conditions for ensuring consistency.

Key words: Conditional expectation; Simpson's paradox; Conditional analysis; Marginal analysis.

1. Introduction

Two elementary schools, School 1 and School 2, in a town each have 100 fourth-grade students who took the same math exam. Suppose we are told that the average scores of both girls and boys in School 1 are higher than those in School 2. Should we then expect School 1 to have a higher overall average score? Intuitively, the answer might be yes, but this is not necessarily the case.

Consider the hypothetical numerical examples in Table 1. In all scenarios, both girls and boys in School 1 have higher average scores. In Scenario A, the overall average score in School 1 is indeed higher than in School 2. However, in Scenario B, School 2 has a higher overall average score. This demonstrates an inconsistency between the *conditional analysis* (considering girls and boys separately) and the *marginal analysis* (considering all students together) in Scenario B. In Scenario C, both schools have the same proportion of girls, and we can prove that in this case, the inconsistency will not occur.

Table 1. Average scores of a math exam in two schools

		Girls		Boys		
Scenario	School	n	Average	n	Average	Overall average
A	1	52	91.0	48	86.0	88.6
	2	49	90.5	51	85.5	88.0
B	1	45	91.0	55	86.0	88.2
	2	60	90.5	40	85.5	88.5
C	1	52	91.0	48	86.0	88.6
	1	52	90.5	48	85.5	88.1

As discussed above, if the two schools have the same proportions of girls (Scenario C in Table 1), the results from conditional and marginal analyses are consistent. However, this condition is too restrictive and rarely occurs in practice. Scenario A in Table 1 shows that marginal and conditional analyses can still be consistent even when the proportions differ. Therefore, having the same proportions of girls is a sufficient but not necessary condition for consistency. In this paper, we derive more general sufficient conditions for consistency.

The paper is organized as follows. Section 2 defines the notations used in the discussion. We derive general sufficient conditions for consistency in section 3, followed by conclusions in section 4. This work is a theoretical extension of Wang et al. (2022).

2. Some notations

Suppose we have measurements of an outcome from subjects in two groups (Group 1 and Group 2). In Group i, the measurement Y_i is integrable. Subjects in Group i are divided into two subgroups based on the value of a binary variable X_i, with $P\{X_i = 1\} = p_i \in (0, 1)$, $i = 1, 2$. In the math score example in Table 1, let Y_i denote the score of a randomly selected student in School i, and let $X_i = 0$ (or 1) indicate whether the selected student is a girl (or boy). Subpopulation j consists of subjects with $X_1 = j$ in Group 1 and $X_2 = j$ in Group 2, where $j = 0$ or 1. Let

$$a_i = \mathrm{E}[Y_i \mid X_i = 0], \quad b_i = \mathrm{E}[Y_i \mid X_i = 1],$$

the mean values (averages) of measurements of subjects in two subgroups within group i.

We define the following quantities:

(a) *Within-group differences* (the difference of mean values of two subgroups within each group):

$$\delta_i = a_i - b_i, \quad i = 1, 2.$$

(b) *Subpopulation between-group differences* (conditional differences):

$$\Delta_0 = a_1 - a_2,$$
$$\Delta_1 = b_1 - b_2.$$

(c) *Overall difference* (marginal difference between two groups):

$$\Delta = \mathrm{E}[Y_1] - \mathrm{E}[Y_2].$$

It is easy to see that

$$\Delta_0 - \Delta_1 = \delta_1 - \delta_2. \tag{1}$$

These differences are summarized in Table 2.

Table 2. Some differences

	Subgroup		
Group	0	1	Within-group difference
1	a_1	b_1	$\delta_1 = a_1 - b_1$
2	a_2	b_2	$\delta_2 = a_2 - b_2$
Between-group differences	$\Delta_0 = a_1 - a_2$	$\Delta_1 = b_1 - b_2$	

The *conditional analysis* compares

$$\mathrm{E}[Y_1 \mid X_1 = 0] \quad \text{vs.} \quad \mathrm{E}[Y_2 \mid X_2 = 0], \tag{2}$$
$$\mathrm{E}[Y_1 \mid X_1 = 1] \quad \text{vs.} \quad \mathrm{E}[Y_2 \mid X_2 = 1]. \tag{3}$$

There are two sub-analyses in the conditional analysis.

Although the X_i are 0–1 valued, (2) and (3) do not imply that $\mathrm{E}[Y_1 \mid X_1]$ and $\mathrm{E}[Y_2 \mid X_2]$ are compared in conditional analysis.

The *marginal analysis* compares

$$\mathrm{E}[Y_1] \quad \text{vs.} \quad \mathrm{E}[Y_2]. \tag{4}$$

From the results in Table 1 we know that even if

$$\mathrm{E}[Y_1 \mid X_1 = 0] > \mathrm{E}[Y_2 \mid X_2 = 0], \tag{5}$$
$$\mathrm{E}[Y_1 \mid X_1 = 1] > \mathrm{E}[Y_2 \mid X_2 = 1], \tag{6}$$

generally we do not have

$$\mathrm{E}[Y_1] > \mathrm{E}[Y_2]. \tag{7}$$

In fact, (5) and (6) do not mean that

$$\mathrm{E}[Y_1 \mid X_1] > \mathrm{E}[Y_2 \mid X_2]. \tag{8}$$

Otherwise, from (8) we should have (7) as $\mathrm{E}[Y_i] = \mathrm{E}\{\mathrm{E}[Y_i \mid X_i]\}$.

Since we cannot obtain (7) from (5) and (6) in general, there is inconsistency between conditional and marginal analyses. This phenomenon is a special case of *Simpson's Paradox* in statistics (Simpson, 1951; Pearl and Mackenzie, 2008) and is very prevalent in data analysis. In this paper we give both intuitive and theoretical interpretations of the paradox, and explore all inconsistencies between conditional and marginal analyses.

From the definition of conditional expectation (Dudley, 2022) we have

$$\mathrm{E}[Y_i \mid X_i] = (1 - X_i)\mathrm{E}[Y_i \mid X_i = 0] + X_i\mathrm{E}[Y_i \mid X_i = 1], \quad i = 1, 2.$$

Hence the marginal difference is

$$\begin{aligned}
\Delta &= [(1 - p_1)a_1 + p_1 b_1] - [(1 - p_2)a_2 + p_2 b_2] \\
&= (a_1 - a_2) - p_1(a_1 - b_1) + p_2(a_2 - b_2) \\
&= \delta_1(p_2 - p_1) + (1 - p_2)\Delta_0 + \ p_2\Delta_1 \qquad (9) \\
&= \delta_2(p_2 - p_1) + (1 - p_1)\Delta_0 + \ p_1\Delta_1. \qquad (10)
\end{aligned}$$

From (9) and (10) we can see that if

$$\delta_1\delta_2(p_2 - p_1) = 0, \tag{11}$$

then the marginal difference Δ is a linear combination of between-group differences. If Δ falls between Δ_0 and Δ_1, the result of marginal analysis is consistent with the result from conditional analyses. Therefore, we consider the following four cases:

(i) **Case I**: $p_1 = p_2\,(= p)$. Then

$$\Delta = \Delta_0(1 - p) + \Delta_1 p.$$

The marginal difference Δ is uniquely determined by Δ_0, Δ_1, and p, and is always between Δ_0 and Δ_1. Although $p_1 = p_2$ may be the most anticipated case, it rarely holds in practice. It is a sufficient but not necessary condition for consistency. See Case 1 in Table 4.

(ii) **Case II**: $\delta_1 = 0$, i.e., there is no difference between two subgroups within group 1. Then $\Delta_0 - \Delta_1 = -\delta_2$ and

$$\Delta = \Delta_0(1 - p_2) + \Delta_1 p_2.$$

The marginal difference is uniquely determined by Δ_0, Δ_1, and p_2, and is always between Δ_0 and Δ_1. Inconsistency between marginal and conditional analyses does not occur in this case.

(iii) **Case III**: $\delta_2 = 0$, i.e., there is no difference between the two subgroups within group 2. Then $\Delta_0 - \Delta_1 = \delta_1$ and

$$\Delta = \Delta_0(1 - p_1) + \Delta_1 p_1.$$

The marginal difference is uniquely determined by Δ_0, Δ_1, and p_1, and is always between Δ_0 and Δ_1. Similarly, the inconsistency between marginal and conditional analyses does not occur.

(iv) **Case IV**: $\delta_1\delta_2(p_1 - p_2) \neq 0$. Then $\Delta_0 - \Delta_1 \neq \delta_1$, and $\Delta_0 - \Delta_1 \neq -\delta_2$. Now Δ is not simply determined by Δ_0 and Δ_1.

Case IV is the most common scenario encountered in practice. Determining whether inconsistency occurs in such cases is not straightforward. This will be discussed in detail in the following sections.

3. Sufficient conditions for consistency

The inconsistency between conditional and marginal analyses is with respect to the signs of Δ_0, Δ_1, and Δ. A typical instance of inconsistency is $\Delta_0 > 0$ and $\Delta_1 > 0$ but $\Delta < 0$. Since each of them may be > 0, $= 0$, or < 0, Table 3 lists all 27 possible combinations of their signs. We use $x = +\,(-)$ to denote that $x > 0\,(< 0)$.

Table 3. All possible combinations of signs of Δ_0, Δ_1, and Δ

Case	Δ_0	Δ_1	Δ	Case	Δ_0	Δ_1	Δ
1	+	+	+	15	−	−	−
2	+	+	0	16	−	−	0
3	+	+	−	17	−	−	+
4	+	0	+	18	−	0	−
5	+	0	0	19	−	0	0
6	+	0	−	20	−	0	+
7	+	−	+	21	−	+	−
8	+	−	0	22	−	+	0
9	+	−	−	23	−	+	+
10	0	+	+	24	0	−	−
11	0	+	0	25	0	−	0
12	0	+	−	26	0	−	+
13	0	0	+	27	0	0	−
14	0	0	0				

Since we are only interested in the sign of Δ given the signs of Δ_0 and Δ_1, there are many symmetric pairs in Table 3. For example, Cases 1 and 15 are the same if we exchange the labels of the two groups. There are also some symmetric pairs if we exchange the labels of subgroups. For example, Cases 4 and 10 are symmetric. Therefore, we only need to consider Cases 1–9, 13 and 14 in Table 3. Given $\Delta_0 = \Delta_1 = 0$, $\Delta = 0$ if and only if (11) holds. We will not consider Case 14 in the following discussion.

If $\Delta_0 \cdot \Delta_1 < 0$, the inconsistency does not occur irrespective of the value of Δ. Without loss of generality, to study the inconsistency between Δ, Δ_0, and Δ_1, we assume $\Delta_0 \geqslant \Delta_1 \geqslant 0$. The inconsistency occurs if and only if one of the following statements is true:

$$\Delta_0 > 0 \quad \text{and} \quad \Delta_1 \geqslant 0 \quad \text{but} \quad \Delta \leqslant 0,$$
$$\Delta_0 = 0 \quad \text{and} \quad \Delta_1 = 0 \quad \text{but} \quad \Delta > 0.$$

These correspond to Cases 2, 3, 5, 6, and 13 in Table 3.

In Cases I–III of section 2, the marginal difference is of the form

$$\Delta = (1-r)\Delta_0 + r\Delta_1, \quad 0 < r < 1. \tag{12}$$

The inconsistency does not occur if Δ_0, Δ_0 and Δ satisfy (12). This means that a necessary condition for inconsistency is

$$\delta_1\delta_2(p_1 - p_2) \neq 0. \tag{13}$$

Note that given Δ_0 and Δ_1, Δ is a continuous function of p_0, p_1, and δ_1; see (9). Assume $\Delta_1 > \Delta > \Delta_0$ and $p_1 = p_2$. If we make a very small change to p_1, the inequalities still hold. For this reason, $p_1 \neq p_2$ is a necessary but not sufficient condition for inconsistency.

We first give the following result.

Theorem 1. *Assume that (13) holds. Then all cases (except 14) in Table 1 are possible.*

In Table 4 we use numerical examples to show that all cases 1–9 and 13 are possible under (13). Among them, Cases 1, 4, 7, 8, and 9 do not show inconsistency between conditional and marginal analyses. Case 1 shows that even if they are consistent, Δ is not necessarily between Δ_0 and Δ_1.

Theorem 1 indicates that (13) is a necessary but not sufficient condition for inconsistency. Since the inconsistency never occurs if Δ is a convex combination of Δ_0 and Δ_1, we prove the following result.

Theorem 2. *Given Δ_0 and Δ_1, for any $p_1, p_2 \in (0,1)$, there always exists a $p \in (0,1)$ such that $\Delta = (1-p)\Delta_0 + p\Delta_1$ if and only if $p_1 = p_2$ or $\delta_1\delta_2 \leqslant 0$.*

Proof. From (9) and (10), if there exists a $p \in (0,1)$ such that $\Delta = (1-p)\Delta_0 + p\Delta_1$, then

$$(\Delta_1 - \Delta_0)(p - p_1) = \delta_2(p_2 - p_1),$$
$$(\Delta_1 - \Delta_0)(p - p_2) = \delta_1(p_2 - p_1).$$

Table 4. Numerical examples of Δ_0, Δ_1, and Δ under Condition IV

Case	p_1	p_2	δ_1	δ_2	Δ_0	Δ_1	Δ
1	0.3	0.4	1.8	1.8	0.9	0.9	1.1
2	0.3	0.4	-9.0	-9.0	0.9	0.9	0.0
3	0.3	0.4	-18	-18	0.9	0.9	-0.9
4	0.3	0.4	2.7	1.8	0.9	0	0.8
5	0.3	0.4	-5.0	-5.9	0.9	0	0
6	0.3	0.4	-18	-18.9	0.9	0	-1.3
7	0.3	0.4	4	2.5	0.9	-0.6	0.7
8	0.52	0.3	1	-1.2	1	-1.2	0
9	0.3	0.4	-18	-19.8	0.9	-0.9	-1.6
13	0.4	0.5	18	18	0	0	1.8

(i) If $\Delta_0 = \Delta_1$, then $\Delta = \Delta_0 = \Delta_1$. We have

$$\delta_1\delta_2(p_2 - p_1) = 0.$$

(ii) If $\Delta_0 \neq \Delta_1$, i.e., $\delta_1 - \delta_2 \neq 0$, then from (9) we have

$$p = p_1 - \frac{(a_2 - b_2)(p_2 - p_1)}{\Delta_0 - \Delta_1} = \frac{\delta_1}{\delta_1 - \delta_2} p_1 - \frac{\delta_2}{\delta_1 - \delta_2} p_2.$$

Let

$$\lambda = \frac{\delta_1}{\delta_1 - \delta_2}.$$

Then

$$p = \lambda p_1 + (1 - \lambda) p_2.$$

Note that $p \in (0, 1)$ for all $p_1, p_2 \in (0, 1)$ if and only if $\lambda(1 - \lambda) \geqslant 0$, i.e.,

$$\delta_1\delta_2 \leqslant 0.$$

It is clear that the converse is true. □

We summarize our main result in the following theorem.

Theorem 3. *Inconsistency between conditional and marginal analyses does not occur if one of the following conditions holds:*
(i) $\Delta_0 \cdot \Delta_1 < 0$*;*
(ii) $p_1 = p_2$*;*
(iii) $\delta_1\delta_2 \leqslant 0$.

Remark 1. Since (11) implies $p_1 = p_2$ or $\delta_1\delta_2 \leqslant 0$, and the converse is not true, Theorem 3 gives more general sufficient conditions for consistency than (11), i.e.,

$$\Delta_0 \cdot \Delta_1 < 0 \text{ or } p_1 = p_2 \text{ or } \delta_1\delta_2 \leqslant 0 \Longrightarrow \text{Consistency},$$

or, equivalently,

$$\text{Inconsistency} \Longrightarrow \Delta_0 \cdot \Delta_1 \geqslant 0 \text{ and } p_1 \neq p_2 \text{ and } \delta_1\delta_2 > 0. \tag{14}$$

That Δ can be written in the form of (12) is only a sufficient but not necessary condition for consistency. In this case, Δ is between Δ_0 and Δ_1. However, consistency does not mean that Δ should be between Δ_0 and Δ_1. For example, in Case 9 in Table 4, $\Delta_0 = 0.9$ and $\Delta_1 = -0.9$. However, $\Delta = -1.6$. In Case 1, $\Delta_0 = \Delta_1 = 0.9$, $\Delta = 1.1$. Since all cases in Table 4 are possible when $p_1 \neq p_2$ and $\delta_1\delta_2 \neq 0$, it is impossible to give a general necessary condition for consistency.

As with the scenarios in Table 1, the case that both Δ_0 and Δ_1 are greater than 0 but Δ is less than 0 is likely the most common inconsistency observed in practice. This may be due to the significant difference between p_1 and p_2. However, If p_1 and p_2 are sufficiently close, as indicated by (9) and (10), such inconsistency can be avoided.

4. Conclusion and discussion

Simpson's paradox is well known within the statistical community. In this paper, we systematically review all possible cases in which inconsistencies between marginal and conditional analyses may occur, and derive a tight necessary condition for inconsistency (14). This condition is applicable to any type of outcome, provided that it is integrable. Unfortunately, it is not possible to establish a sufficient condition for inconsistency.

Both marginal and conditional analyses are valid statistical methods. As shown in equation (4), marginal analysis utilizes the entire dataset, while equations (2) and (3) demonstrate that conditional analysis operates on subsets of the data. Given that different data subsets often yield different conclusions, it is unsurprising that inconsistencies can arise. Therefore, results from marginal and conditional analyses should always be interpreted independently, regardless of whether they are consistent.

A common cause of inconsistency is a significant difference between p_1 and p_2. Such differences are frequently observed in observational studies. However, in randomized clinical trials, p_1 and p_2 can often be made sufficiently similar, rendering inconsistencies relatively uncommon.

References

Dudley R. M. (2002): *Real Analysis and Probability.* New York: Cambridge University Press.

Pearl J., Mackenzie D. (2018): *The Book of Why: The New Science of Cause and Effect.* New York, NY: Basic Books.

Simpson E.H. (1951): The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society*, Series B. 13: 238–241.

Wang H., Wang B., Tu X. M., Feng C. (2022): Inconsistency between overall and subgroup analyses. *General Psychiatry.* 35(3): e100732.