

1974 WFS 妊娠歷 資料의 一致性 水準

| | |
|--------|---------------|
| 目 次 | I. 序 |
| | II. 資料 및 分析方法 |
| | III. 一致性水準 |
| | IV. 結 |

I. 序

本 論文은 1978年에 研究 發表한 바 있는 “出産力資料의 一致性 檢討”에¹⁾ 대한 심층분 석이다.

인구분야의 出生과 死亡資料중 이제까지는 주로 出産力에 集中하여 研究되어 왔다.

그런데 出産力에 관한 자료는 調查對象 婦人들의 기억에 의존하여 수집된 資料 (retrospective data)가 대부분이었다. 그러한 資料는 調查當時 調查員의 能力이나 應答婦 人의 心理狀態등에 따라 같은 質問인데도 그 應答內容이 달라질 수 있고, 특히 妊娠이나 그 結果에 관한 정보수집에 있어서는 더욱 그러하다.

우리나라에서 家族計劃 및 出産力에 관해 규모가 큰 現地調查를 실시해온지도 15年이 가 까워올만큼 역사가 길어졌다. 그러므로 調查資料의 質的 (data quality)인 문제에도 研究的 관심이 경주될 필요가 있다.

本 稿에서는 「1974년 세계출산력조사」의 本調查 (main survey)와 1977년 追究調查 (follow-up survey)등 두 資料를 이용하여 出生 및 妊娠結果에 대한 一致水準이 추정되었다.

II. 資料 및 分析方法

1) 資 料

1977년 追究調查는 1974 韓國出産力調查(KNFS)²⁾의 特別分析을 위해 실시되었다. 이 調査는 농촌지역만을 대상으로 31個 調查區에서 465명의 婦人이 調査되었고 이들의 총임신 수는 2,322件이었다. 「1974 WFS」는 全體 327個 調查區중 농촌지역은 125個 調查區였는데

*韓國人口保健研究院 研究員.

1) 金弘淑, “出産力資料의 一致性 檢討”, 家族計劃論集, 第 7 號 (1980. 11), pp. 24-33.

2) The 1974 Korean National Fertility Survey.

1977년 追究調査와 동일한 調査區의 同一한 婦人에게서 2,292件의 妊娠이 報告되었다.

또한 두 調査 모두 妊娠과 出産에 관한 妊娠歷 (pregnancy history)이 수록되어 있다. 따라서 出産力 資料에 대한 두 調査間의 一致性水準 (consistency level)을 알아보기 위해서는 위의 두 調査資料가 가장 적합한 것이라고 생각된다.

2) 分析方法

本稿를 위해서는 두단계의 分析方法이 필요하다. 하나는 두 調査間의 妊娠資料를 비교하는 (matching) 과정에 대한 것이고, 다른 하나는 그것을 가지고 두 調査資料에 대한 一致性水準을 決定하는 방법을 말한다.

前者는 이미 발표된 논문³⁾에서 상세히 다루어졌으므로 여기서는 두 資料를 비교한 기준만을 간단히 언급코자 한다.

두 조사자료의 一致性을 비교함에 있어서 正常出生妊娠 (fertile pregnancy)⁴⁾과 非正常出生妊娠 (infertile pregnancy)⁵⁾에 대한 一致水準을 약간 다르게 적용하였다. 두 조사자료에서 임신결과별로 “一致된 妊娠”을 찾아내는데 정상출생임신의 경우는 妊娠年·月을 기준으로 하였으나, 그 임신시기가 약간의 차이를 보이더라도 동일한 임신이라고 판단된⁶⁾ 임신이면 “一致된 妊娠”에 포함시켰다. 그런데 비정상출생임신의 경우 妊娠年·月을 맞추어본 결과 단 한건도 같은 임신이 없었다. 이는 비정상출생임신에 대한 기억오차 (memorial error)가 얼마나 큰가를 보여주는 단편적인 자료이다. 따라서 이를 감안하여 비정상출생 임신의 경우는 1974년 이전에 응답자가 갖고 있는 비정상출생임신의 件數를 기준으로 두 조사자료를 비교했다.

두 調査資料의 一致性水準을 決定하는 데는 “Chandra Sekar and Deming Method”를 적용하였다. 이 方法을 이용하여 두 調査 모두에서 누락되어버린 妊娠數가 추정되었다.

本稿에서 一致性水準을 위해 使用하는 用語는 두 종류가 있다. 하나는 두 調査間의 “一致性水準” (consistency level)으로서 두 調査의 총임신이 一致되는 비율 (matching rate)을 의미한 것이고, 다른 하나는 각각의 調査에서 資料가 報告된 水準 (reporting rate)을 의미한다. 前者는 “一致水準” (matching level)으로 後者는 “報告水準” (reporting level)으로 불려지게 될 것이다.

3) 金弘淑, “出産力資料의 一致性檢討,” 前掲書, 1980.

4) 임신결과가 정상출생으로 끝난 임신.

5) 임신결과가 사산, 자연유산, 인공유산 등으로 끝난 임신.

6) 판단에 사용한 변수는 임신시작 년·월·일, 조사당시 출생아의 사망여부, 임신순위, 출생순위변수 등을 고려하였다.

Ⅲ. 一致性水準

1. 두 調査資料에서 누락된 妊娠數推定

두개의 現地調査資料 (survey data)를 가지고 妊娠을 一致시키는 데는 두 단계가 필요 한데 첫째는, 基準年度 以前에 發生한 妊娠數를 두 資料를 갖고 비교하며, 둘째는 두 調査에서 報告된 同一한 妊娠結果를 놓고 “order by order”로 두 資料를 비교하여 一致性水準을 결정한다.

그런데 두 調査資料를 놓고 一致性水準을 결정하는데에는 두 調査 모두에서 누락되어 버린 妊娠數를 推定하여 基準年度 (1974) 以前의 全體 妊娠數를 확정해야하는 문제가 따른다. 문제는 양쪽에서 다 빠진 數를 어떻게 추정 (estimate) 하느냐가 되는데 이 숫자는 表 I에서 “x”로 표시했고, 총 임신수는 “?”로 표시하였다. 이 숫자들을 추정하기 위하여 간단한 전제 (assumption)를 이용하는 “Chandra Sekar and Deming Method”⁸⁾를 적용하였다.

Table 1. Frequency Distribution of Matching Pregnancy-outcome; 1974 and 1977 Data
同一한 妊娠結果에 대한 頻度數 分布 : 1974년과 1977年 資料

| | | 1977 Data | | | | Subtotal | Unreported pregnancy | Total |
|-----------|----------------------|-----------|----|----|-----|----------|----------------------|-------|
| | | LB | SB | SA | IA | | | |
| 1974 Data | LB | 2,029 | 5 | | 1 | 2,035 | 47 | 2,082 |
| | SB | 3 | 6 | 1 | | 10 | 12 | 22 |
| | SA | | | 30 | 2 | 32 | 28 | 60 |
| | IA | | | 1 | 92 | 93 | 35 | 128 |
| | Sub-total | 2,032 | 11 | 32 | 95 | 2,170 | 122 | 2,292 |
| | Unreported Pregnancy | 43 | 4 | 27 | 78 | 152 | “x” | |
| | Total | 2,075 | 15 | 59 | 173 | 2,322 | | “?” |

※ LB: Live Birth
SB: Still Birth
SA: Spontaneous Abortion
IA: Induced Abortion

Source: 金弘淑, “出産力 資料의 一致性檢討”, 前掲書, 1980.

- 7) Norman B. Ryder, “Consistency of Reporting Fertility Planning Status”, *Studies in Family Planning*, Vol. 10, No. 4 (April 1979), p. 116.
- 8) Chandra Sekar and Deming, “On a Method of Estimating Birth and Death Rates and the Extent of Registration”, *Journal of the American Statistical Association*, Vol. 44, No. 245 (1949), pp. 101 - 115.

Chandra는 다음과 같이 표가 작성되었을 때 II에 enumerate될 확률이 I에 enumerate되는 경우와 안되는 경우 모두를 독립적(independent)이라고 가정한다.

| | | | | |
|---|---|-------|-------|---------------|
| | | II | | |
| | | + | - | |
| I | + | a | b | a + b |
| | - | c | x | c + x |
| | | a + c | b + x | a + b + c + x |

따라서

$$\frac{a}{a+b} = \frac{c}{c+x}$$

가 성립해야 하고, 또

$$\frac{a}{a+c} = \frac{b}{b+x}$$

등으로 표시할 수 있다. 따라서 $\frac{a}{c} = \frac{b}{x}$ 가 성립해야 하므로

$$x = \frac{bc}{a}$$

로 양쪽 調査에서 누락된 數를 구할 수 있게 된다. 그러면

$$\frac{a+b}{a+b+c+x}$$

가 74년도 調査의 완전한 총 임신수가 되고

$$\frac{a+c}{a+b+c+x}$$

는 77년도 추구조사의 완전한 총임신수가 된다.

이를 이용하여 “x”와 “?”를 구하면 表 2가 된다.

Table 2. Completers Distribution of Table 1

표 1의 완성된 分布

| | | 1977 Data | | Total |
|-----------|-------|-----------|-----|-------|
| | | Yes | No | |
| 1974 Data | Yes | 2,150 | 122 | 2,292 |
| | No | 152 | 9 | 161 |
| | Total | 2,322 | 131 | 2,453 |

※ “Yes” means enumerated.

“No” means not enumerated.

2. 妊娠結果別 一致性水準

1) 全體妊娠의 一致性水準

앞節의 表 2에 의해 두 調査間의 전반적인 一致性水準을 결정할 수 있다. 즉 表 1에서 妊娠結果別로 一致된 숫자는 정상출생 2,029件, 사산 6件, 자연유산 30件과 인공유산 92件으로 총 2,157件이 一致되었고, 表 2에서 추정한 총 임신건수는 2,453件이었다. 따라서 두 調査資料의 一致水準(matching level)은 87.9퍼센트임을 알 수 있다(表 3).

Table 3. Matching Level of Two Survey Data; 1974 and 1977 Data
두 調査資料의 一致水準

| | % | N |
|-----------------------|-------|-------|
| Matched Pregnancies | 87.9 | 2,157 |
| Unmatched Pregnancies | 12.1 | 296 |
| Total | 100.0 | 2,453 |

이들 두 調査 각각에 대해 “報告된 水準(reporting level)”을 보면 表 4와 같다. 基準年度인 1974年 調査에서는 총임신 2,453件중 93.4퍼센트가 보고되었고 1977年 추구조사에서는 94.7퍼센트가 調査되었다. 전체 임신건수로만 보면 1977자료가 1974자료보다 많은 임신수가 조사되었다. 그런데 이 두 자료를 임신결과별로 보면 一致性水準이 달라짐을 알 수 있다.

Table 4. Comparison of the Reporting Pregnancies; 1974 and 1977 Data
두 調査 各各의 報告水準 比較

| | 1974 Data | 1977 Data |
|-----------------------|---------------|---------------|
| Reporting Pregnancy | 93.4 (2,292) | 94.7 (2,322) |
| Unreporting Pregnancy | 6.6 (161) | 5.3 (131) |
| Total | 100.0 (2,453) | 100.0 (2,453) |

※ ()=Number of pregnancies.

2) 妊娠結果別 一致性水準

表 5는 妊娠結果別 一致水準(matching level)을 결정하기 전에 우선 각각의 調査에서 一致된 임신(matching pregnancies)수가 報告된 임신(reporting pregnancies)수에 대해 갖는 비율을 임신결과별로 나누어본 것이다.

1974년 資料에서 보고된 임신중 一致된 정상출생임신은 97.5퍼센트, 1977년 추구조사에서는 97.8퍼센트로 거의 같은 수준을 보였고, 인공유산의 경우 1974년 자료에서는 71.9퍼센트, 1977년 자료에서는 53.2퍼센트의 일치수준을 보였다. 이는 인공유산의 경우 조사자

Table 5. Percent Distribution of Matching Pregnancy, by Pregnancy-outcome; 1974 and 1977 Data

妊娠結果別 一致된 妊娠의 百分率 分布

| | 1974 Data | | 1977 Data | |
|----------------------|--------------|-----------------|--------------|-----------------|
| | % (N) | N. of Sub-total | % (N) | N. of Sub-total |
| Live Birth | 97.5 (2,029) | 2,082 | 97.8 (2,029) | 2,075 |
| Still Birth | 27.3 (6) | 22 | 40.0 (6) | 15 |
| Spontaneous Abortion | 50.0 (30) | 60 | 50.8 (30) | 59 |
| Induced Abortion | 71.9 (92) | 128 | 53.2 (92) | 173 |
| Total | 94.1 (2,157) | 2,292 | 92.9 (2,157) | 2,322 |

료의 신빙성에 대해 심각한 회의를 느끼게 해주는 것이 아닐 수 없다. 따라서 우리가 인공 유산자료를 사용함에 있어 이 점을 깊이 인식한 상태에서 연구, 분석되어야 할 것이다. 사산과 자연유산의 경우 그 절대수가 적었기 때문에 일치성수준을 논하기에 미흡하다 하겠다.

表 6 은 表 5 를 참고로 하여 妊娠結果別로 “Chandra Sekar and Deming Method” 를 이용하여 決定된 一致水準 (matching level) 이다. 주의할 점은 表 6 에 나타난 각 임신결과별 임신수를 더하면 전체임신에 대한 총수가 아니라는 점이다. 왜냐하면 정상출생의 경우 두 調査 모두에서 누락된 임신수는 1 件이었다. 그런데 이 1 件이 정상출생으로 양쪽에서 누락은 되었지만 자연유산의 임신에 “unmatching” 임신으로 들어가 있을 수도 있기 때문이다.

Table 6. Matching Level of Two Survey Data, by Pregnancy-outcome

妊娠結果別 두 調査資料의 一致水準

| | Percent of | | N | | |
|----------------------|------------|---------------|------------|---------|-------|
| | Matched | Preg. Matched | Unreported | Missed* | Total |
| Live Birth | 91.1 | 2,029 | 99 | 1 | 2,129 |
| Still Birth | 10.9 | 6 | 25 | 24 | 55 |
| Spontaneous Abortion | 25.4 | 30 | 59 | 29 | 118 |
| Induced Abortion | 38.2 | 92 | 117 | 32 | 241 |
| Total | 87.9 | 2,157 | 287 | 9 | 2,453 |

※ “missed” means number of missing pregnancies in two survey all.

각 임신결과별로 가장 一致率 (matching rate)이 높은 것은 정상출생으로 91.1퍼센트였다. 가장 낮은 一致率은 사산으로 10.9퍼센트밖에 안되었다. 자연유산은 25.4퍼센트, 인공유산은 38.2퍼센트의 一致水準을 보여 주었다.

따라서 정상출생을 제외한 비정상출생으로 끝난 임신들의 경우 調査報告에 대한 신뢰도가 낮음을 충분히 감안하여 資料活用을 해야할 것이다.

Ⅲ. 結

본 분석결과 「1974年 WFS」 농촌資料와 1977年 追究調査의 一致性水準은 87.9퍼센트임이 밝혀졌다. 즉 두 조사자료에서 같은 임신으로 一致된 것은 87.1퍼센트에 해당된다는 것이다. 임신결과별로 보면 정상출생은 91.1퍼센트, 인공유산은 38.2퍼센트의 一致水準(matching level)을 보여 주었다.

각 調査別로 보면 전체임신에 대해 1974年 本調査는 93.4퍼센트, 1977年 追究調査에서는 94.7퍼센트의 報告率(reporting rate)을 보이고 있었다. 각 임신결과별로는 정상출생이 1974年 資料에서는 97.5퍼센트, 1977年 資料에서는 97.8퍼센트, 인공유산의 경우 1974年 資料에서는 71.9퍼센트였고 1977年 調査에서는 53.2퍼센트만이 一致된 임신을 갖고 있었다. 따라서 정상출생과는 달리 인공유산이나 기타 비정상출생임신에 관한 조사자료를 분석함에 있어 이와같은 사실은 연구자가 사전에 고려해야만 할 것이다.

또한 본 논문에서 밝힌 一致性水準은 “Chandra Sekar and Deming Method”에 의한 것인데, 앞서 언급한 바와 같이 두 조사 모두에서 누락된 임신수는 양쪽“system” 및 각 항목들이 서로 “독립적(independent)”이라는 가정하에 추정된다. 그런데 실제로는 두 조사 및 각 항목에 보고되는 확률이 독립적이지 아니므로 추정된 누락임신수 및 총임신수는 일반적으로 低推定(under estimate)이 된다. 그러므로 두 조사자료에서 실제로 일치되는 수준은 본 분석에서 밝힌 것보다 더 낮은 상태가 된다.

참 고 문 헌

- 1) 金弘淑, “出産力資料의 一致性檢討”, *家族計劃論集*, 第7號, 1980. 11.
- 2) Ryder, Norman B., “Consistency of Reporting Fertility Planning Status”, *Studies in Family Planning*, Vol. 10, No. 4, April 1979.
- 3) Sekar, Chandra and Deming, “On a Method of Estimating Birth and Death Rates and the Extent of Registration”, *Journal of the American Statistical Association*, Vol. 44, No. 245, 1949.

(Abstract)

Consistency Level of Pregnancy History of the 1974 WFS Data in Korea

Hong-Suk Kim*

I. Introduction

The purpose of this paper is to determine the level of discrepancies of pregnancy information between two survey data, given by same respondents in Korea.

II. Data and Analytic Method

The data were collected in the Korean National Fertility Survey in 1974 (the 1974 WFS) and the follow-up survey which was carried out in the sample of thirty-one census enumeration districts in rural area.

A total of 465 respondents reported 2,322 pregnancies before 1974 in the 1977 follow-up survey. And the 1974 survey data reported 2,292 pregnancies from the same respondents as in the 1977 survey.

In preparation for the comparison of pregnancy histories, a procedure for matching pregnancy informations in two surveys was required.

The first step was to compare the number of fertile pregnancies (pregnancies ended in live birth) reported in 1977 as occurring prior to the first interview with the number reported in 1974. The next step in matching was to compare the dates reported in the two surveys, for the same fertile pregnancies, order by order.

The Chandra Sekar and Dewing Method was used to estimate both surveys in 1974 and in 1977. That method has an assumption that the chance of an event being missed

* Researcher, Korea Institute for Population and Health

in one survey may be independent of its chance of being missed on the other survey.

III. Findings

According to matching pregnancy histories of two surveys, the matching rate is 87.1 per cent in the total pregnancies. By pregnancy-outcome the matching rate of live birth is 91.1 per cent and that of induced abortion is 38.2 per cent.

The reporting rate is 93.4 per cent in the 1974 survey and is 94.7 per cent in the 1977 survey, respectively.

It must be considered that the matching level of two survey data in fact was lower than the estimate level because the probability of an event being missed in both systems is not independent.