

# Chandrasekar-Deming 技法에 의한

## 正確한 人口動態水準 測定

1. 緒 論  
2. Chandrasekar-Deming 技法의 理論  
3. C-D 技法의 推定值에 미치는 諸誤差의 影響  
4. C-D 技法 適用에 의한 우리나라 人口動態率 推定  
5. 結 論

### 1. 緒 論

人口動態란 人口의 時間的 變化에 起因한 事象이며 그것은 具體的으로 出生·死亡·移動 및 婚姻 등 人口의 代置過程 내지 動態의 樣相을 의미한다. 이러한 人口의 動的 現象은 過去의 社會的 變動의 結果로서 將來 人口構造에 影響을 주며 다시 現在의 人口構造는 將來의 人口動態 혹은 社會發展의 條件이 되기도 한다.

따라서 한 나라의 社會政策의 目的을 遂行하기 위하여는 正確한 人口의 動態的 把握이 要求되며 그렇기 때문에 대부분의 나라는 人口動態 申告를 義務化하고, 申告(registration)를 통하여 人口動態統計를 얻으려 하고 있다. 恒久的이고 계속성 및 最近性的 精確한 人口動態統計는 申告方法에 의하여 生産되기 때문이다.

우리나라는 人口動態 申告水準 自體가 낮고 申告資料의 精確性도 낮아서 申告方法이 人口動態 統計의 生産에 전혀 기여하지 못하고 있다.

經濟企劃院이 推定한 1966年의 期待發生數에 대한 當年 申告率은 出生에서 34%, 死亡에서 2%였다.

梁<sup>1)</sup> 등의 調査結果에 의하면 1965年 金泉市 및 그 주변 人口를 대상으로 한 調査에서 1965年 出生의 當年申告率은 32%, 死亡申告率은 27%였다.

전국 25個 地域을 대상으로 한 本 研究事業 地域의 1969年 9月~1971年 8月 2個年間的 出生申告率은 도시지역에서 39%, 농촌지역에서 39%였고 死亡申告率은 도시지역에서 46%, 農村地域에서 37%였다.

한편 經濟企劃院<sup>2)</sup>이 발표한 1969年 總出生申告數 중 當年度 申告比率은 29%에 불과했고 申告된 대부분이 지연신고였다. 臺灣<sup>3)</sup>의 1970~71年 總出生申告中 당년도 申告比率은 92%였다.

이와같이 우리나라 人口動態申告水準은 그간에 별 變化없이 낮았고, 낮은 主原因은 지연신고에 있음이 지적되었다. 이에 따라 申告方法에 의하여 利用可能한 人口動態 統計는 현재까지 生産되지 못하였다.

우리나라는 1945年 8·15解放 以後 申告制度의 改善과 人口動態 統計의 계속적 수집 등을 위한

1) 梁 在模 外 5人, 韓國家族計劃事業의 支援方案과 正確한 人口動態統計의 把握을 위한 研究, 1969, p. 65~66.  
2) 經濟企劃院, 1970年 總 인구 및 주택조사보고, 제2권, 出生率. p. 252  
3) Ministry of Interior, Republic of China, 1971, Taiwan Demographic Fact Book, Nov. 1972, p. 29

여러 노력을 부단히 경주하여 왔다.

인구동태통계의 주관 部處는 美軍政 官房, 保健厚生部, 公報處, 內務部 등을 거쳐 1961年 7月 에 經濟企劃院으로 변화하여 오늘에 이르렀다.

經濟企劃院은 1962년에 統計法의 制定과 人口動態調查規則을 시행하여 法的 體制를 정비하였고 人口動態申告率 向上을 위해 1961년부터 人口動態申告 강조기간을 설정하여 신고계몽 및 이행에 힘썼다. 한편 1963년부터 週期的 標本調査(年 4回實施)와 몇가지 시험 또는 特別調査로서 계속적으로 최근의 人口動態水準의 파악에 노력하였다.

이 외에 여러 研究機關에서 人口動態統計 改善을 위한 研究와 改善案이 제시되었다. 人口問題 研究所<sup>4)</sup>, 연세대학교<sup>5)</sup>, 崔<sup>6)</sup>, 趙<sup>7)</sup> 등의 改善案 및 연구가 있다.

그러나 이러한 노력에도 불구하고 아직 이용가능한 人口動態統計가 申告制度를 통하여 生産되지 못하고 있다.

Marks<sup>8)</sup>는 현재까지의 우리나라 人口動態統計의 不振理由를 多角的으로 分析, 評價하여 계속적인 개선된 人口動態 申告制度 발전을 위하여 다음 事項으로 구성된 制度를 건의하였다.

1) 지난 기간의 出生 및 死亡을 파악하기 위한 週期的 面接調査가 실시될 계속적 標本家口의 設定

2) 家口調査의 完全성을 추정하고 누락된 人口動態事件의 調整을 위한 보고(面接調査)된 出生 및 死亡과 신고된 出生 및 死亡의 Matching

3) 10~15個 邑, 面의 標本에서 出生 및 死亡 申告의 向上을 위하여 活動할 特別造成員(經濟企劃院 또는 保健社會部 소속)의 配置.

이 制度는 「파키스탄」의 PGE project 를 통하여 80~85%의 完全率을 성취한데 근거를 둔 것으로서 주의 깊은 훈련과 감독으로 韓國에서도 동일한 좋은 成果가 얻어질 것으로 기대된 것이다. 이 制度는 현실적으로 導入할 수 있는 合理性과 實踐性을 겸비한 것으로 고려된다.

1971년~73년의 18개월간 家族計劃研究院에 의하여 실시된 「保健組織網을 통한 人口動態改善을 위한 研究」는 Marks의 建議事項을 참고로 하여 人口動態統計의 改善을 위해 活用해 본 일은 없으나 그 潜在力이 큰 保健要員을 활용한 최초의 연구였다. 全國에 郡, 市, 區, 行政單位로 보건소가 있으며 邑, 面, 行政單位로 保健支所가 조직되어 있고 각 보건지소에 結該, 母子保健, 家族計劃 要員등 3인이 배치되어 있다.

人口動態 統計의 主要 利用者는 保健關係 機關이며 이는 末端機關에서 住民과의 밀접한 접촉 및 活動으로 보건요원은 人口動態, 統計의 生産에 크게 기여할 잠재력을 갖고 있는 것이다.

이 研究結果<sup>9)</sup>에 의하면 18개월 간의 申告改善을 위한 집중적 노력에도 불구하고 人口動態 申告率은 이용가능한 統計生産이 가능한 수준으로 향상되지 못 하였다. 早期申告의 履行에 사회·문화적 요인이 완강히 저항하고 있음이 지적되었다. 그러나 報告된 出生 및 死亡資料는 期待發 生數에 접근하며, 이 報告資料와 다른 獨立된 자료수집(家口面接調査) 方法에 의한 資料를 C-D

4) 人口問題研究所, 人口動態統計改善方案, 1967.

5) 梁在模 外 5人, 韓國家族計劃事業의 支援方案과 正確한 人口動態統計의 把握을 위한 研究, 延世大學校, 1969.

6) 崔仁鉉, 韓國人口動態統計의 現況과 問題點, 서울大學校 文理科大學 附設 人口研究所 1969.

7) Cho, Lee-Jay, The Level of fertility and mortality in Korean rural community, Korean Journal of Sociology, Vol. 2, 1966, pp. 143~146.

8) Eli S. Marks, Vital Statistics and Population Growth Measurement, Recommendations, 1970, Population and Housing Census and Vital Statistics (Mimeo), 1970, p. 103~109.

9) 宋建鏞·林鍾權, 全國保健網을 통한 人口動態申告 改善을 위한 研究—綜合報告一, 家族計劃研究院, 1974年 9月.

技法을 적용하여 얻은 人口動態率은 실제치에 접근되는 것으로 보고되었다.

世界 人口의 45~66%(中共을 포함하는 경우 66%)는 出生, 死亡, 結婚등의 申告資料가 不完全한 地域에 거주하는 것으로 추정되고 있다. 이들 不完全한 人口動態 申告制度를 갖고 있는 대부분은 開發途上國이다. 이들 國家는 이 統計의 改善을 위한 努力을 여러 측면에서 기울이고 있다.

1968년 西「파키스탄」에서 개최된 CENTO(Central Treaty Organization) 5개국인 가진 Demographic Statistics 심포지움 결과에 의하면 「터키」<sup>10)</sup>, 「이란」<sup>11)</sup>, 東「파키스탄」<sup>12), 13)</sup> 등 국가에서 人口動態申告의 問題點과 이의 改善策의 강구에 대단한 關心을 갖고 努力을 경주하고 있음을 알 수 있고, Kenya<sup>14)</sup>에서도 人口動態統計의 발전을 위한 努力을 하고 있다.

Seltzer<sup>15)</sup>는 아시아 5개국에서 1945년~67년에 실시된 人口動態統計에 관한 研究를 比較·檢討한 바 있다. 이들 研究는 二元化資料蒐集方法 (dual record system)과 Chandra Sekar-Deming formula<sup>16)</sup>를 적용하여 최근의 정확한 人口動態 諸率을 추정하는데 目的을 둔 것이었다.

人口動態事件은 單回調査(Single round retrospective Surveys)에서 정확히 보고되기 어렵다. 이 調査에서 응답자는 자행적으로 12個月 이전의 일정 기간에 발생된 모든 出生과 死亡에 대하여 기억을 살려 응답해야 한다. 그러므로 이 方法은 漏落誤差와 時期誤差 (errors of dating) 양자를 범하게 된다<sup>17)</sup>. 開發途上國에서 總誤差率(Gross errors)이 出生에서 20%, 死亡에서 40%<sup>18)</sup>에 달함이 보고 되고 있다.

이 方法의 誤差源(Sources of errors)의 약관을 극복하기 위한 시도로 多回追求調査(Multiround follow-up Survey)<sup>19)</sup>가 있다.

人口動態率의 결정을 위해 많은 국가에서 이 方法이 사용되었다. 印度는 1958~59年<sup>20)</sup>, 「브라질」은<sup>21)</sup> 1961년, 「모로코」<sup>22)</sup>는 1961~62年, 「인도네시아」<sup>23)</sup>는 1961~62年, 「나이지리아」<sup>24)</sup>는 1965~66年, 「튀니지아」<sup>25)</sup>는 1968년에 각각 이 方法을 사용한 바 있다.

- 10) Sabahaddin Alpat, Current Practices and Problems in Demographic Statistics, Turkey, CENTO, Proceedings of a Simposium Karachi, West Pakistan, Nov., 1968, p. 57~64.
- 11) Manouchehr Rais-Zadeh, Vital Registration System, Problems and Practices, Iran, Ibid., p. 65~68.
- 12) M. A. Chaudhury, An Overview of the Sources and Method of Collection of Vital Statistics in East Pakistan, Ibid., p. 69~76.
- 13) Mahbud-ud-Din Ahmad, Problems and Prospectives of Vital Registration in East Pakistan, Ibid., p. 77~82.
- 14) D. L. Huxtable, Vital Statistics Aid in Developing Countries, Amer. J. Pub. Health, 57:504~8, 1967.
- 15) William Setzer, Some Results from Asian Population Studies, Population Studies, Nov., 1969, p. 395~406.
- 16) C. Chandra Sekar & W. E. Deming, On a Method of Estimating Birth and Death Rates and the Extent of Registration, Laboratories for Population Statistics Reprint Series No. 1, May 1971.
- 17) Christopher Scott, Technical Problems of Multi round Demographic Surveys, Reprint Series No. 11, Laboratories for Population Statistics, Sept. 1973, p. 5.
- 18) Government of India, Preliminary Estimate of Birth and Death Rates and of the Rate of Growth of the Population, National Sample Survey, No. 48 (New Delhi, 1961)
- 19) G. Sabagh and C. Scott, A Comparison of Different Survey Techniques for Obtaining Vital Data in a Developing Country, Demography, Vol, IV (1967), p. 759.
- 20) Government of India, op. cit.,
- 21) Guanabara Demographic Pilot Survey (UN Publication Sales No. 64, XIII.3)
- 22) G. Sabagh and C. Scott, op. cit.
- 23) Vaino Kannisto, Population Increase in Indonesia (Djakarta, Central Bureau of Statistics, Statistical Research and Development Center, 1963) Mimeo.
- 24) Nigeria, Federal Office of Statistics, Rural Demographic Sample Survey. 1965~66(Lagos, 1968)
- 25) J. Vallin, C. Paulet and C. Tarifa, mimeo.

이 방법은 高價의 經費가 投入되나 누락 및 時間 양자의 오차를 감소시킬 수 있는 반면, 모든 문제를 해결하는 것은 아니다. 특히 첫번 調査후 出生한 嬰兒가 다음 調査 以前에 死亡되거나 調査 間 標本地域을 떠난 사람과 관련된 事件의 發見에 어려움이 많다.

그러므로 「아프리카」와 아시아 여러 국가는 多回追求 調査와 몇가지 共通의 特性을 갖고 있는 Dual Record System 을 이용한 人口動態 諸率 推定 事業을 C-D 技法에 의해서 실시하고 있다.

가장 잘 알려져 있는 事業은 「파키스탄」의 the Population Growth Estimation (PGE) Project 이며, 이외에 the Indian Sample Registration System, the Thailand Survey of Population Change, the Turkish Demographic Survey, the Liberian Population Growth Survey 등이 있다.

「파키스탄」의 Population Growth Estimation Project (1962~65年)는 出生率, 死亡率, 人口增加率을 추정하고 자료수집의 여러가지 方法의 실험, 기존 申告制度의 評價, 改善을 위한 建議등을 위하여 실시되었다. 印度의 Sample Registration System (1964~)은 洲(State) 및 全國(National levels)의 年間 出生 및 死亡率의 신속 精確한 추정을 위해 실시되고 있다.

The Turkish Demographic Survey (1965~)는 지역 및 전국적인 신뢰도 높은 최근의 出生 및 死亡統計를 수집하고 人口移動年齡, 性別分布 및 婚姻, 文盲, 教育狀態 등에 관한 年間 資料를 제시하기 위해 실시되고 있다. 이외에 이 Survey의 목적에 申告制度의 補完과 여러 保健事業의 계획, 측정, 評價를 위한 기초자료의 제공등이 포함된다.

The Liberian Population Growth Survey (1968~)는 「리베리아」政府에서 精確한 최근의 出生 및 死亡 추정치의 제공과 出生 Patterns, 人口移動, 年齡, 性別 分布, 婚姻, 文盲, 就學등 자료를 제시하기 위해 실시되고 있다. 이외에 이 事業의 目的에 現在 및 未來의 病院, 學校, 道路 職業, 電力, 水力등을 평가할 기초자료의 제공이 포함된다<sup>26)</sup>.

Lauriat<sup>27)</sup>는 Dual Record System 은 one system 만의 자료보다 훨씬 신뢰도 높은 자료를 제공할 수 있고 이 방법은 人口動態申告가 부진하거나 存在하지 않는 국가에서 성공적으로 적용될 수 있음을 지적하였다.

Linder<sup>28)</sup>는 粗出生率 및 死亡率 측정에 있어 Dual Record System 의 效果에 대하여 이 方法의 結果는 개발도상국에서 申告資料에 계산된 諸率보다 우월하고 精確하며 개발도상 또는 개발된 국가에서 실시된 面接 調査에서 계산된 諸率보다 더욱 精確하다고 보고 하였다.

그는 Dual Record System 의 능력<sup>29)</sup>은 粗出生率과 死亡率의 단순한 측정을 넘어 어느 이전의 System 보다 강력한 統計的 道具가 된다고 하였고 先進國에서 이 方法이 고려되어야 한다고 강조하고 있다.

이와 같이 여러 國家는 자기 국가적 要求와 條件에 따라 精確한 人口動態統計와 기타 부수된 자료의 수집에 Dual Record System 을 利用하고 있음을 알 수 있다.

특히 우리나라는 現實적으로 短期間에 人口動態 申告水準이 현저히 증가되어, 申告를 통하여 精確한 人口동태 통계가 생산되는 것을 기대할 수 없다.

26) J. R. Abernathy and A. S. Lunde, Subject matter Coverage in the Dual-Report Systems of India, Pakistan, Turkey, and Liberia, Laboratories for Population Statistics Science Series No. 3, Feb. 1973, p.7

27) P. Lauriat, Field Experience in Estimating Population Growth, Demography, Vol. 4, No. 1, pp. 228~243.

28) Forrest E. Linder, Vital Event Numeration System as a New Tool for Measuring Population Change, Bulletin of the International Statistical Institute, Vol. 44, Book 1, 1971, p. 377~393.

29) F. E. Linder, A Proposed New Vital Event Numeration Unitary System for Developed Countries, Milbank Memorial Quarterly, Oct. 1970, pp. 77~85.

合理性和 實踐性を 겸비한 代案은 保健組織網을 活用해서 人口動態統計의 긴박한 需要를 충족하는 恒久的 制度를 확립하는 것이다. 이 代案은 몇 개 開發途上國家에서 성공하고 있는 Dual Record System 과 C-D 技法의 適用이다.

本稿는 이용가능한 人口動態統計의 生産可能性은 이러한 側面에서 검토하기 위하여 Chandrasekar-Deming 技法의 理論的 背景 및 이 技法 適用上의 問題를 제시하고, 이 技法이 우리나라 人口動態改善에 주는 效果 등을 分析 제시한 것이다.

## 2. Chandrasekar-Deming 技法의 理論<sup>30)</sup>

### 1) Chandrasekar-Deming formula

C-D 技法의 適用은 Dual record system 을 前提로 한다. Dual record system 은 同一 期間, 同一 地域에 대한 두개의 독립된 資料蒐集 方法으로 구성된다. 흔히 한 方法은 계속적인 관찰하에서 人口動態 事件(出生과 死亡)을 記錄하며, 다른 하나는 定期的 家口面接調査(reference period 는 3개월, 6개월, 또는 1年)를 통하여 人口動態事件을 조사한다.

C-D 技法은 이들 두 方法에 의하여 수집된 자료를 다음과 같은 4개의 事件集團으로 구분하여 두 자료수집방법에 의하여 누락된 事件을 추정하여 總發生數를 추정하고, 동시에 정확한 人口動態 水準을 추정하는 方法이다. 即

<첫째> 두개의 자료수집 System에서 공통으로 報告된 事件集團("C" events)

<둘째> System 1(예컨대 申告資料)에서만 報告된 事件集團(events)

<셋째> System 2에서만 報告된 事件集團("N<sub>2</sub>" events)

<넷째> 兩 System에서 모두 報告되지 않은 事件集團("Y" events)

등 네가지 경우로 區分된다.

이와같이 兩 System에서 蒐集된 資料의 對照過程을 거쳐 事件을 네개의 部類로 區分한 다음 Chandra Sekar-Deming formula 에서 "Y" events 는  $\hat{Y} = \frac{N_1 N_2^{31)}}{C}$  로써 推定되고 總事件發生數

30) C. Chandrasekaran and W. Edwards Deming, On a method of estimating birth and death rates and the extent of registration, Journal of the American Statistical Association, (1949).

31)  $\hat{Y} = \frac{N_1 N_2}{C}$  는 다음과 같은 근거에 의해서 推定된다.

System 1 (R 表)	System 2(I 表)		計(Σ)
	報告된 것	報告 안된 것	
報告된 것	C	N <sub>1</sub>	C+N <sub>1</sub>
報告 안된 것	N <sub>2</sub>	Y	N <sub>2</sub> +Y
計(Σ)	C+N <sub>2</sub>	N <sub>1</sub> +Y	N(=C+N <sub>1</sub> +N <sub>2</sub> +Y)

即 위 表에서 다음의 關係를 얻는다.

1)  $\frac{C+N_2}{N} = I$  表에서 報告되는 確率

$\frac{N_1+Y}{N} = \text{非 } I$  表에서 " "

$\frac{C+N_1}{N} = R$  表에서 " "

$\frac{N_2+Y}{N} = \text{非 } R$  表에서 " "

2) 따라서 다음과 같은 等式으로 부터  $\hat{Y} = \frac{N_1 N_2}{C}$  가 導出된다

( $\hat{N}$ )는 다음의 代數學的 方法에 의하여 추정된다. 即

$$\hat{N} = C + N_1 + N_2 + \hat{Y} = C + N_1 + N_2 + \frac{N_1 N_2}{C} = \frac{1}{C} (C + \hat{N}_1) (C + N_2) \quad (1)$$

但  $C$  = System 1 (R 表)과 System (I 表)에서 共通으로 報告된 事件數

$N_1$  = System 1 (R 表)에서만 報告된 事件數

$N_2$  = System 2 (I 表) " "

$\hat{Y}$  = System 1 (R 表)과 System 2 (I 表)에서 모두 漏落된 事件數의 推定值

그리고  $\hat{N}$ 의 分散 (Variance)<sup>32)</sup>과 標準誤差는 다음과 같이 表示된다. 即 分散은

$$\sigma_{\hat{N}}^2 = \frac{Nq_1q_2}{p_1p_2} \quad (2)$$

이고 따라서 標準誤差는 다음과 같다.

$$\sigma_{\hat{N}} = \sqrt{\frac{Nq_1q_2}{p_1p_2}} \quad (3)$$

但  $\sigma_{\hat{N}}^2 = \hat{N}$ 의 分散

$\sigma_{\hat{N}} = N$ 의 標準誤差

$N$  = 特定期間의 總發生事件數

$\hat{N}$  =  $N$ 의 推定值

$p_1$  = 發生事件이 R 表에서 報告될 確率

$p_2$  = 發生事件이 I 表에서 報告될 確率

$q_1 = 1 - p_1$

$q_2 = 1 - p_2$

위의 (1)式에서  $\hat{N}$ 은  $N$ 가 크고 R 와 I에서 事件들이 漏落될 確率 間에 서로 統計적으로 獨立의이라는 假定下에서 不偏推定值이며 이것은 또한 最尤推定值 (Maximum Likelihood Estimate)가 된다.

$\hat{N}$ 의 分散 (Variance)을 表示하고 있는 (2)式과  $\hat{N}$ 의 標準誤差인 (3)式에서는  $p_1$ 과  $p_2$ 가 높을 수록 그리고  $q_1$ 과  $q_2$ 가 낮을수록 人口動態水準의 推定值  $N$ 의 精度는 높아지는 關係이다.

$\hat{N}$ 의 相對精度 (Relative Precision)를 Chandra Sekar-Deming 은 다음과 같이 나타내고 있다

$$(CV)\hat{N} = \sqrt{\frac{p_1q_2}{Np_1p_2}} \quad (4)$$

但  $(CV)\hat{N} = \hat{N}$ 의 相對精度

$N$  = 總發生事件數

31)  $C = N \left( \frac{C+N_1}{N} \times \frac{C+N_2}{N} \right) = R$  表, I 表에서 同時에 일어나는 度數.

$N_1 = N \left( \frac{C+N_1}{N} \times \frac{N_1+Y}{N} \right) = R$  表, 非 I 表에서 " " "

$N_2 = N \left( \frac{C+N_2}{N} \times \frac{N_2+Y}{N} \right) = I$  表, 非 R 表에서 同時에 일어나는 度數.

$Y = N \left( \frac{N_1+Y}{N} \times \frac{N_2+Y}{N} \right) =$  非 I 表, 非 R 表에서 " " "

32)  $\hat{N}$ 의 分散 (The Variance of  $\hat{N}$ )은 다음과 같이 하여 얻어진다. 即  $\hat{N}$ 의 分散은 事件  $x$ 의 函數  $f(x)$ 의 分散이  $V_{f(x)} = \left( \frac{\partial f}{\partial x} \right)_E^2 V(x)$ 의 形式을 취하는 것으로 부터 求해진다.

여기서  $C+N_1$ 과  $C+N_2$  및  $N$ 가 一定하다면  $E(C)$ 와  $V(C)$ 는  $E(C) = Np_1p_2$ ,  $V(C) = Np_1q_1p_2q_2$ 와 같이 되며, 同一 條件下에서  $V(\hat{N})$ 은  $V(\hat{N}) = (C+N_1)^2(C+N_2)^2V\left(\frac{1}{C}\right)$ 이 된다.

그러므로  $V(\hat{N}) = \frac{Nq_1q_2}{p_1p_2}$ 와 같이 된다.

$p_1$  = 發生事件이 R 表에서 報告될 確率

$p_2$  = 發生事件이 I 表에서 報告될 確率

$q_1 = 1 - p_1, q_2 = 1 - p_2$

위 (4)式은 一稱 變動係數(Coefficient of Variation)라고 하는 것으로  $N$ 가 크면 클수록 그리고  $p_1$ 과  $p_2$ 가 높을수록 精度가 좋아지는 關係를 보는 內容이다.

한편 System 1(예컨대 申告水準)의 尺度  $p_1$ 의 推定值  $\hat{p}_1$ 과 그 變動係數는 다음과 같다.

$$\hat{p}_1 = \frac{C}{C + N_2} \quad (5)$$

$$(CV) \hat{p}_1 = \sqrt{\frac{q_1}{(C + N_2)p_1} \cdot \frac{N - C - N_2}{N - 1}} \quad (6)$$

但,  $\hat{p}_1 = p_1$ 의 推定值

$C = (\text{II}-1)$ 式 參照

$N = (\text{II}-3)$ 式 參照

$p_1 = (\text{II}-3)$ 式 參照

$N_2 = (\text{II}-1)$ 式 參照

$(CV) \hat{p}_1 = \hat{p}_1$ 의 相對精度

위 (6)式에서  $C + N_2$ 가 增加할수록 變動係數는 減少되고  $C + N_2 = N$ 일 때 申告水準은 誤差없이 正確하게 推定된다는 關係이다.

## 2) C-D 技法 適用上的 諸問題

前節에서 記述한 C-D formula의 單純하고 公式化된 推定理論을 實際에 있어서 그대로 적용하기에는 약간의 問題가 있다. 그 問題點을 要約하면 다음과 같다.

### (1) 申告의 不正確性

R 表上에서  $C$  혹은  $N_1$ 으로 分類되기 어려운 誤字등의 애매한 事件이 있을 수가 있는데 이렇게 分類되기 어려운 事件들을 올바른 R表上的 發生事件으로  $C$ 나 혹은  $N_1$ 에 分類되어 計算處理되는 경우 申告水準은 過大推定(over-estimate)되고 그 反對의 경우 申告水準은 過少推定된다

### (2) 調査의 不完全性

I 表上에서 일어나는 問題點은 調査의 漏落, 調査의 重複, 多數의 無應答者 등을 들 수 있다. 調査漏落이 일어나는 主된 경우는 旅行者, 水上生活者, 浮浪者 등 移動中에 있는 人口 또는 家口, 調査直前의 出生兒, 作名前의 嬰兒, 私生兒등의 嬰兒層(Infants)등이다.

또한 集團的 公共施設(公共施設의 公舍 같은 것)에서 發生되는 事件이 어느 特定된 家口에서 定住되고 있지 않다는 內容에서 調査는 不正確하게 될 수 있다.

R表에서는 대체로 이러한 事件을 포함하고 있는 것이 보통이나 調査員에 의한 家口調査結果(表 I)에서는 이들이 漏落되기 쉽다.

### (3) 相關性的 問題

兩 System(R表와 I表)에서 漏落된 事件 間에 일어나는 相關關係의 問題이다. 즉, 조사원이 조사한 家口表(I表)가 R表上에서 누락된 사건을 발견할 條件附確率이 I表가 R表에서 報告된 事件을 發見할 條件附確率과 다를 때 兩 System 間에는 相關性이 存在함으로써 推定值는 偏倚

(bias)되는 것이다.

이러한 相關性은 調査過程에서 혹은 Matching 過程에서 發生하기 쉽다. 相關性은 母集團을 여러 개의 同質的인 group 으로 區分하여 推定함으로써 어느 정도 弱화시킬 수 있다. 즉 몇개의 同質的인 group 으로 母集團을 區分하고 각 group 에 대한 推定值를 얻은 다음 다시 發生事件의 總數는 이들 group 別 合計를 總計함으로써 얻는 것이다.

여기서 母集團을 몇 개의 同質的인 group 으로 區分하여 推定한 事件總數와 全母集團을 對象으로 하는 總體的 推定值와의 差를 求해 보면 다음과 같다. 즉,  $N$ 를  $k$ 個의 同質的인 group ( $N_i, i=1, 2, \dots, k$ )으로 區分하면  $N_i q_1^i q_2^i$ 는  $i$ 번째 group 에서의 兩 System 에 의해서 누락된 사건수가 되고 따라서 그 총 누락수는  $\sum N_i q_1^i q_2^i$ 가 된다. 그리고 全母集團 事件을 同時的 總體的으로 취급하는 경우 兩 System 에서 누락되는 모든 사건수는 (註 31)의 근거에 의거하여  $\frac{[\sum N_i p_1^i q_2^i][\sum N_i q_1^i p_2^i]}{\sum N_i p_1^i p_2^i}$ 와 같이 된다. 그리하여 前者와 後者의

差는

$$\frac{[\sum N_i p_1^i q_2^i][\sum N_i q_1^i p_2^i]}{\sum N_i p_1^i p_2^i} - \sum N_i q_1^i q_2^i = -\frac{N^2 S_1 S_2 r}{\sum N_i p_1^i p_2^i} \quad (7)$$

$$\text{但, } N = \sum_1^k N_i$$

$p_1^i$  = R 表에서의 사건발견 확률 ( $i$  번째 group 에서)

$p_2^i$  = I 表에서의 사건발견 확률 ( " )

$$S_1^2 = \frac{\sum N_i (p_1^i - \bar{p}_1)^2}{\sum N_i}$$

$$S_2^2 = \frac{\sum N_i (p_2^i - \bar{p}_2)^2}{\sum N_i}$$

와 같이 表示된다.

그리고  $p_1$  과  $p_2$  의 相關關係의 尺度는 다음의 (8)式에서의 相關係數( $r$ )로 表示된다. 즉,

$$r = \frac{S_{12}}{S_1 S_2} = \frac{\sum N_i (p_1^i - \bar{p}_1) (p_2^i - \bar{p}_2)}{S_1 S_2 \sum N_i} \quad (8)$$

$$\text{但, } \bar{p}_1 = \frac{\sum N_i p_1^i}{\sum N_i}$$

$$\bar{p}_2 = \frac{\sum N_i p_2^i}{\sum N_i}$$

여기서  $r > 0$ 이면 同時的 總體的 推定值는 過少推定(underestimate)되고 System 1(申告水準)은 過度推定(overestimate)된다. 반대로  $r < 0$ 일 때에는 그 反對의 현상이 일어난다.

이와같이  $r=0$ 가 아니고 相關關係가 兩 System 間에 依存할 때에는 母集團을 同質的인 group 으로 적절하게 區分할 필요가 있는 것이다. 적절한 區分이란 애매한 표현이나 그렇다고 5년간격 年齡階層의 區分이 10년간격 年齡階層보다 더 좋다 혹은 더 나쁘다 혹은 영아사망의 경우는 個別的으로 처리되어야 한다는 등이 확일적으로 어느 경우에나 적용되는 것은 아니다. 그것은 실제의 자료를 시험함으로써 결정이 가능하게 되는 문제이다.

### 3) C-D 技法과 Singur Health Centre(India)

C-D 技法은 1947년에 印度의 Singur Health Centre 地域에서 1945년과 1946 年の 人口動態 및 申告水準을 推定하기 위하여 適用되었다.



Singur Health Centre 는 4개의 隣接된 Union Board(Singur, Balarambati, Bora, Begumpur)로 構成되었다. 4개의 Union Board 中에서 Singur 는 이 地域의 中心地域이며 Culcutta 로 부터 約 21마일 떨어져 있다. Centre 가 管轄하는 地域의 總面積은 約 33平方마일이고 68개의 마을과 人口 64,000名 12,000家口 그리고 8,300개의 住宅으로 構成되어 있었다.

1944년에 이 地域은 公衆保健方法論 (Public Health Methology)의 實驗을 위하여 印度의 衛生 및 公衆保健院 (All India Institute of Hygiene and Public Health)의 特定調査地域으로 造成되었다.

Centre 地域에서 出生 및 死亡申告는 里長의 定期的인 報告로 부터 이루어졌으며 申告表 作成을 위한 情報은 마을 住民들의 自發의 申告와 母子厚生省의 出生 및 嬰兒死亡表로 부터 補完되었다.

1945年과 1946年の 人口動態水準에 관한 調査는 1947年 2月 17日부터 11週간 계속되었다. 調査員들은 常住人口와 訪問客들의 年齡, 性, 男便의 身分 등을 調査하기 위하여 全家口를 訪問하였다.

이와 같이 하여 획득한 調査結果表는 申告表(R 表)와 對照(matching)시켰으며 對照된 結果表는 表1과 같다. 前述한 바와 같이 여기에서도 誤字나 不完全한 調査, 人口의 移轉, 集團의 公共施設에서의 出生등이 약간의 문제가 되었다.

따라서 表1에서는 非常住의 集團의 公共施設에서의 出生(Institutional Birth) 및 死亡은 除

表 1. R表와 I表의 比較表(Singur Health Centre)

Event	Year	Total number of event	R 表				No. non-verifiable, illegible, incomplete etc,	Number incorrect	I 表 N <sub>2</sub>
			Total	C	N <sub>1</sub>				
Birth (excluding non-resident institutional)	1945	1,748	1,504	794	710	156	88	741	
	1946	2,659	2,242	1,506	736	228	189	1,009	
Death (excluding non-resident institutional)	1945	1,356	1,083	350	733	190	83	372	
	1946	1,052	866	439	427	117	69	421	

外되어 個別的으로 考慮되었는데 非常住의 集團의 公共施設에서의 出生은 1945년에 約 8%였고 1946년에는 約 3%에 不過하였다.

出生 및 死亡의 推定方法은 가장 좋은 不偏推定値를 얻기 위하여 全體를 몇개의 同質的인 group 으로 區分하여 推定하는 方式도 採用되었다. 即 그 推定方法을 보면 다음과 같다.

- <첫째> 全 Centre에 대한 總體的 推定(出生 및 死亡)
- <둘째> 各 Union Board에 대한 個別的인 推定(出生 및 死亡)
- <셋째> 全 Centre에 대한 性別 推定(死亡)
- <넷째> 全 Centre에 대한 性別 年齡階層別 推定(死亡)

위의 네가지 推定方法에 의해서 推定된 1945年の 總死亡數는 各各 2,234, 2,238, 2,245 그리고 2,418이고 이들에 對한 標準誤差(Standard error)는 約 70이었다. 1946年の 그것은 各各 1,696, 1,684, 1,698, 1,265이며 그 標準誤差는 約 40이었다.

全 Centre 의 總體的 推定方法에 의한 出生數는 1945년에 2,908名이고 1946년에는 3,744名이

었다. Union Board 를 個別的으로 推定하여 合計한 出生數는 1945年과 1946年에 各各 2,915名과 3,775名이었다.

出生率 및 死亡率을 計算하기 위한 分母는 家口調査에 의한 人口를 使用하였으며 申告水準은

表 2. Singur Health Centre 의 出生 및 死亡率(1945, 1946)

	1 9 4 5		1 9 4 6	
	率	標準誤差	率	標準誤差
出生率(1,000人當)	46.1	0.8	59.8	1.0
死亡率(1,000人當)	37.7	1.2	27.5	0.7
男子 死亡率	36.4	1.6	27.3	1.0
女子 死亡率	39.2	2.1	27.8	1.0

表 3. 出生 및 死亡申告率(1945, 1946) (單位: %)

Union Board	出生申告		死亡申告	
	1945	1946	1945	1946
Singur	60.4~67.9	70.9~77.1	38.1~46.9	42.0~49.1
Balarambati	51.5~55.8	53.3~57.8	45.8~55.9	50.8~58.0
Bora	53.1~61.3	56.0~66.0	54.9~66.5	52.6~63.4
Begumpur	47.4~50.3	61.3~64.7	42.6~46.4	44.9~48.1

註 (1) The range is due to non-verified entries on the R-lists.

(2) The figures are subject to a standard error of about 3 per cent.

總事件推定値와 R 表上에서 發見된 錯誤없는 事件總數로 計算되었다.

表2에서 1946年の 出生率이 매우 높게 나타나고 있는데 그 理由는 經濟狀態의 改善과 調査의 不完全性 2次大戰後의 軍除隊등으로 說明되고 있다.

### 3. C-D 技法의 推定値에 미치는 諸誤差의 影響<sup>33)</sup>

C-D 技法에 의하여 人口動態水準을 推定하려 할 때 그 推定値는 資料蒐集過程에서 부터 事件數의 推定에 이르기까지 여러가지 標本誤差 내지 非標本誤差의 要因에 의해서 影響을 받게 된다.

물론, 量的 分析의 實際的 目的은 選擇의 責任이 있는 사람에게 選擇의 結果를 어느 程度의 確率을 갖고 提示하는 것이며 따라서 어떤 推定値에 어느 程度의 誤差가 存在하느냐 하는 것을 提示하는 것은 計量的 手法의 普遍的 役割이라 하겠다.

그런데 Chandraseker-Deming 는 처음부터 이 技法에 의해서 發生事件의 不偏推定値를 算出 하려는 基本的인 必要條件을 提示한 바있다. 即,

<첫째> 兩 System에서의 蒐集資料가 모두 造作的 報告에 依한 것이 아닐 것. 또는 모든 造作된 報告는 事件推定 以前에 除去시킬 것.

<둘째> 對照節次(Matching procedure)는 完全無缺할 것 즉 가짜의 matches 나 가짜의

33) William Seltzer & Arjun Adlakha, On the Effect of Errors in the Application of the Chandra Sekar-Deming Technique op. cit., No. 14, Apr. 1974.

non-matches가 없을 것.

<셋째> 兩 System의 資料蒐集節次는 統計的으로 獨立的인 것.

그는 또 아무리 發生事件數가 正確하게 推定되었다 하더라도 出生率 死亡率 및 自然增加率의 推定値는 이들을 계산하는데 分母로써 사용되는 基礎人口(Base Population)의 計算上의 誤差에 의해서 일어나는 偏倚 (Bias)를 指摘하였다.

그러나 實際로 이러한 基本條件이 제 아무리 嚴守된다 하더라도 誤差는 發生하는 것이며 따라서 推定値의 誤差에 關한 普遍性은 C-D 技法에 依한 推定値에도 例外일 수는 없는 것이다.

그리하여 William Seltzer 와 Arjun Adlakha 는 特히 上述한 바 네가지 基本條件이 지켜지지 않는 경우의 發生可能한 誤差에 關하여 論及하여 誤差測定의 統計的 方法을 提示하였다.

결국 C-D 技法의 基本的인 必要條件이란 실제에 있어서 基本的인 推定値의 誤差要因이 되는 것이다. 이러한 誤差들을 個別的인 要因別로 볼 때 各各의 誤差中에는 負의 값을 取하는 것도 있고 正의 값을 취하는 것도 있다. 그리하여 總誤差(total bias)는 個別的인 誤差보다 작게 나타나는 것이 일반적이다. 그러므로 여기서는 諸誤差들의 個別的인 屬性和 아울러 이들 誤差들 사이의 相互作用에 關한 內容을 보고자 한다.

### 1) 造作的 報告에 의한 相對誤差

事件의 造作的 報告는 地域的 혹은 時間的인 不可避性에 의해서 만들어지기도 하며 死産을 出生으로 記入하는 등의 事件發生에 關한 概念의 曖昧性에 의해서 이루어지기도 한다.

이러한 造作的 報告에 의한 相對誤差(relative bias)의 크기는 두 System間的 統計的 獨立性和 對照誤差(matching error)가 일어나지 않는다는 假定下에서 다음과 같이 計算된다.

$$B_s = \frac{u}{V_1 V_2} - 1 \quad (9)$$

但,  $u$  = 두 System에서 共通으로 發見된 事件中 妥當한 報告率

$V_1$  = System 1에서의 妥當한 報告率

$V_2$  = System 2에서의 妥當한 報告率

위 (9)式에서 만약 對照된 事件報告가 眞실로 모두 妥當한 것이라면 위의 (9)式은 다음과 같이 表示될 수 있다.

$$B_s = \frac{1}{V_1 V_2} - 1 \quad (10)$$

但, 記號 說明은 (9)式과 동일함.

그리고 두 System에서 共通으로 發見된 事件中에서의 妥當한 事件報告率이 어느 한 System에서의 妥當한 그것보다 크다면 造作된 報告에 의한 C-D 推定値의 偏倚規模(size of bias)의 限界는 다음과 같다.

$$\left(\frac{1}{V_2} - 1\right) \leq B_s \leq \left(\frac{1}{V_1} - 1\right) \quad (11)$$

但, 여기서의  $V_2$ 는 두 System에서의 妥當한 比率이 낮은 System의 妥當한 事件報告率임.

(11)式의 內容은 圖1에서 볼 수 있는 關係이다.

### 2) 對照誤差(Matching Errors)

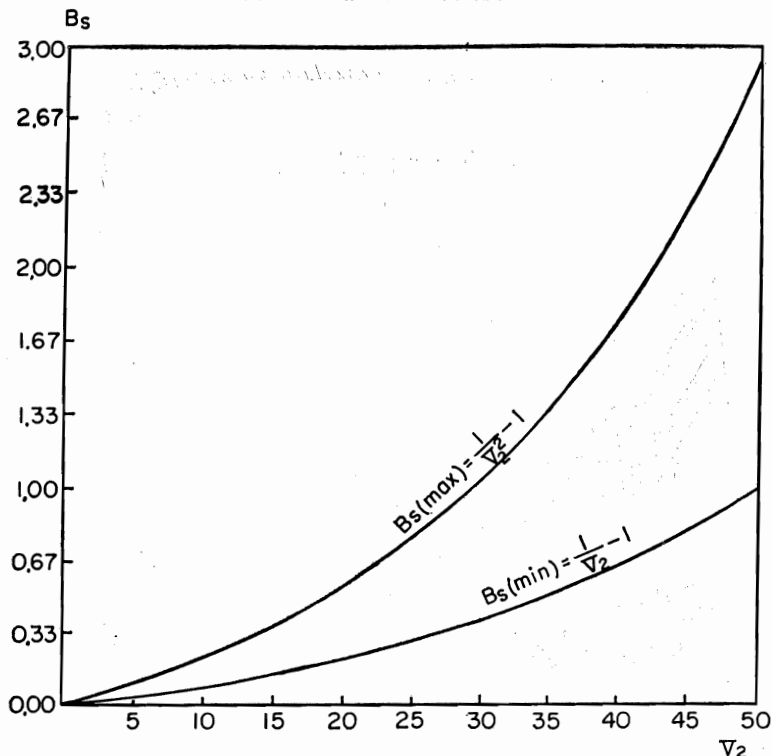
對照誤差는 兩 System의 發生事件을 對照하는 過程에서 일어나는 誤差이다. 理論的으로 兩 System에서 共通으로 對照된 事件 "C"는 하나이다. 그러나 實際로 이들의 對照節次는 複雜한

뿐만 아니라 對照過程이 매우 지루한 것이므로 誤差는 발생하게 마련이다.

根本적으로 Matching Error의 問題는 事件屬性에 관한 不完全하고 不正確한 情報로부터 發生된다. 事件에 관한 不完全하고 그릇된 識別은 Matching Error의 有意的 水準을 크게 한다.

Matching Error에는 가짜의 matches와 가짜의 非 matches가 있다. 前者는 다시 兩 System 혹은 하나의 System에서의 가짜의 matches中 진짜의 matches가 있는지 없는지에 따라서 두 개의 group으로 區分된다.

圖 1 C-D 推定値의 相對誤差



가짜의 非 matches는 "C"를 過少推定(underestimate)되게 하고 N를 過大推定되게 한다.

한편 가짜의 matches에 진짜의 非 matches가 兩 System에 다 있는가 혹은 하나의 System에만 있는가 혹은 兩 System에 모두 없는가에 따라 "C"의 값은 偏倚(bias)되기도 한다.

이와 같이 複雜한 問題들을 單純化시켜 "C"가 가짜의 非 matches보다 가짜의 matches를 더 많이 내포하고 있는지 또는 그 反對의 경우인지의 均衡關係에 의해서 "C"에서의 純誤差를 포착하는 方法을 보면 다음과 같다. 即, 兩 System에서 造作된 報告가 일어나지 않고 또 兩 System이 統計적으로 獨立的이라는 假定下에서 C-D 推定値의 純 對照誤差(Net Match Error)는

$$B_m = -\frac{m}{1+m} \tag{12}$$

但,  $B_m$  = C-D 推定値의 相對誤差

$m$  = 對照事件의 純相對誤差

$$\cdot \left( = \frac{C}{C_{true}} - 1 \right)$$

와 같이 表示된다.

여기서 만약  $m$ 가 正值(+ )이면 가짜의 非 matches보다 가짜의 matches가 더 크고  $\hat{N}$ 는 下 偏向(down bias)되고  $m$ 가 負值(-)를 取하면 그 反對의 경우가 되고  $\hat{N}$ 은 上 偏向(upward bias)된다.

다음의 表 4는 純對照誤差의 서로 다른 값에 대한  $\hat{N}$ 의 相對誤差를 보는 比較表이다.

表 4 純對照誤差에 對應하는  $\hat{N}$ 의 相對誤差

$m$	-25	-20	-15	-10	-5	+5	+10	+15	+20	+25
$B_m$	33.3	25.0	17.6	11.1	5.3	-4.8	-9.1	-13.0	-16.7	-20.0

### 3) 獨立性 缺如에 의한 誤差

獨立性的 缺如에 關한 問題는 두 System 間의 相互作用(例컨대 두 System에서의 個別的인 情報交換 등) 또는 事件自體의 特性(하나의 System에서만 일어나기 쉬운 事件등)에 依해서 일어난다.

이렇게 獨立性이 缺如되는 경우 推定值의 相對誤差는 對照誤差와 造作的 報告가 없다는 假定下에서 다음과 같이 表示된다.

$$B_c = q_1 \left( \frac{p_{22}}{p_{21}} - 1 \right) \quad (13)$$

但,  $p_{21} = R$  表에서 發見되고 또 I 表에서 발견될 條件附確率

$p_{21} = R$  表에서 漏落되고 I 表에서 發見될 條件附確率

위의 式에서  $B_c$ 는  $q_1$ 과  $q_{21}$  그리고  $p_{22}$ 에 의해서 決定되는 것이며 따라서 만약 條件附確率과  $p_{21}$ 과  $p_{22}$ 가 같다면  $B_c$ 는 零(Zero)이 되고 推定值는 不偏推定值가 되며  $p_{21} > p_{22}$ 이면  $N$ 은 過少推定되고  $p_{21} < p_{22}$ 이면 過大推定이 될 것이다.

한편 兩 System 間의 獨立性이 缺如되어 있을 때 System 1(申告水準)의 推定值의 相對誤差는 다음과 같이 表示된다.

$$B_p = \frac{(p_{21} - p_{22}) p_1 q_1}{(p_1 p_{21} + q_1 p_{22})} \quad (14)$$

但,  $p_{21} = (13)$  式에서와 동일

$p_{21} = (13)$  式에서와 동일

$p_1 = (11)$  式과 동일

$q_1 = 1 - p_1$

위의 式에서 만약  $p_{21} > p_{22}$ 이면 申告水準은 過多推定이 될 것이고  $p_{21} < p_{22}$ 이면 過少推定이 될 것이다.

이와 같이 獨立性 缺如에 의해서 일어나는 偏倚(bias)는 無相關(Zero correlation)의 假定下에서 이루어지는 推定值의 標準誤差보다 더 重要性이 強調되기도 한다.

### 4) Base population에 의한 誤差

粗出生率(粗死亡率)은 出生數(死亡數)를 人口로 나누어준 것과 같다. 그러므로 이 比率은 事件(出生, 死亡)과 人口의 두 要素를 포함한다. 따라서 어느 한 쪽만의 誤差라도 粗率(出生率 혹은 死亡率)에 영향을 준다.

그러나 人口計算과 事件(出生, 死亡)의 두 경우를 고려할 때 推定事件發生率은 매우 근사하게 眞값(true value)에 접근할 수도 있는데 그것은 人口와 事件이 대체로 同等한 相對誤差를 가지고 같은 方向으로 偏向될 때 可能한 것이다. 即 事件發生數와 人口數에 關한 誤差의 方向을 안다고 假定하면 推定人口動態率의 偏倚(bias)는 表5의 分割表와 같다

또한 出生과 死亡數 그리고 人口에 依해서 決定되는 自然增加率은 이들 構成要素들 중에 어느 하나라도 誤差가 發生하면 그 推定值에 偏倚가 일어난다. 即, 自然增加率은 粗出生率과 粗死亡

表 5. 人口動態率의 偏倚의 方向 分割表

人口推定値	事件推定値出生死亡		
	過少推定	適度推定	過多推定
過少推定	變動 <sup>1)</sup>	多推定過	過多推定
適度推定	過少推定	適度推定	過多推定
過多推定	過少推定	過少推定	變動 <sup>2)</sup>

- 1) 人口動態率은 人口推定値의 相對誤差가 人口動態事件推定値의 相對誤差보다 크거나 같거나 작음에 따라 각각, 過大, 適度, 過少推定値가 된다.
- 2) 人口動態率은 人口推定値의 相對誤差가 人口動態事件推定値의 相對誤差보다 작거나 같거나 큼에 따라 각각 過多, 適度, 過少推定되게 된다.

率의 差로서 定義된다. 따라서 粗出生率과 粗死亡率의 誤差의 方向을 알게 되면 自然增加率의

表 6. 自然增加率의 偏倚의 方向 分割表

粗出生率	粗死亡率		
	過少推定	適度推定	過多推定
過少推定	變動 <sup>1)</sup>	過少推定	過少推定
適度推定	過多推定	適度推定	過少推定
過多推定	過多推定	過多推定	變動 <sup>2)</sup>

註 (1) 自然增加率의 推定値는  $\frac{d}{b}$ 가  $\frac{CBR}{CDR}$ 보다 크거나 같거나 작거나 함에 따라 각각, 過多推定, 適度推定, 或은 過少推定된다. 여기서  $CBR$ =粗出生率  $CDR$ =粗死亡率,  $b=CBR$ 의 相對誤差  $d=CDR$ 의 相對誤差

(2) 自然增加率의 推定値는,  $\frac{d}{b}$ 가  $\frac{CBR}{CDR}$ 보다 작거나 같거나 크거나 함에 따라 각각 過多推定 適度推定 或은 過少推定된다.

偏倚의 方向을 알 수 있다(表 6 參照).

### 5) 諸誤差要因間의 相互作用

造作된 報告와 matching error 그리고 統計的 獨立性의 缺如가 存在할 때 C-D의 推定値<sup>34)</sup>는 다음과 같다.

34) 諸誤差要因間의 相互作用이 存在할때 C-D 推定値는 다음과 같이 하여 얻어진다.

$\hat{R}_1$ =System 1에서의 報告

$\hat{R}_2$ =System 2 " "

$\hat{C}$ =Common Report

$N = \frac{\hat{R}_1 \hat{R}_2}{\hat{C}}$  (N의 偏倚된 推定値)

$R_1$ =System 1에서의 妥當한 報告

$R_2$ =System 2에서의 妥當한 報告

$C$ =妥當한 matches

$M_1$ =妥當한 非 matches

$M_2$ =虛僞 matches

$L$ =虛僞 非 matches

$m_1 = \frac{M_2}{C}$

$$\hat{N} = \frac{N(1+S_1)(1+S_2)(q_1 p_{22} + p_1 p_{21})}{p_{21}(1+m_1) - q_{21} m_2} \quad (15)$$

但,  $N$  = 總事件發生數

$\hat{N}$  =  $N$ 의 推定值

$$S_1 = \frac{\hat{R}_1 - R_1}{R_1}$$

34) 계 속 :  $m_2 = \frac{L}{M_1}$

$\hat{R}_1 - R_1$  = 造作된 報告

$$S_1 = \frac{\hat{R}_1 - R_1}{R_1}$$

$$S_2 = \frac{\hat{R}_2 - R_2}{R_2}$$

$q_1$  = System 1에서 發生事件이 漏落될 確率

$p_{22}$  = System 1에서 漏落된 事件을 System 2가 포착할 條件附確率

$p_{21}$  = System 1에서 포착한 事件을 System 2가 또 포착할 條件附確率

$$N = \frac{R_1 R_2}{C} - N q_1 \left( \frac{p_{22}}{p_{21}} - 1 \right)$$

$$\hat{C} = C + m_1 C - m_2 M_1$$

$$\hat{R}_1 = R_1 (1 + S_1)$$

$$\hat{R}_2 = R_2 (1 + S_2)$$

$$\hat{N} = \frac{R_1 (1 + S_1) (1 + S_2) R_2}{(C + m_1 C - m_2 M_1)}$$

$$= \frac{(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \cdot \frac{R_1 R_2}{C}$$

$$= \frac{(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \cdot \frac{R_1 R_2}{C} - \frac{(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \cdot N q_1 \left( \frac{p_{22}}{p_{21}} - 1 \right)$$

$$+ \frac{(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \cdot N q_1 \left( \frac{p_{22}}{p_{21}} - 1 \right)$$

$$= \frac{(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \cdot \left[ \frac{R_1 \hat{R}_2}{C} - N q_1 \left( \frac{p_{22}}{p_{21}} - 1 \right) \right] + \frac{(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \cdot N q_1 \left( \frac{p_{22}}{p_{21}} - 1 \right)$$

$$= \left[ \frac{(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \right] \cdot N + \left[ \frac{(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \right] \cdot N \cdot q_1 \left( \frac{p_{22}}{p_{21}} - 1 \right)$$

$$= \frac{N(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \left[ 1 + q_1 \left( \frac{p_{22}}{p_{21}} - 1 \right) \right]$$

$$= \frac{N(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \left[ \frac{p_{21} + q_1 p_{22} - q_1 p_{21}}{p_{21}} \right]$$

$$= \frac{N(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \left[ \frac{p_{21} (1 - q_1) + q_1 p_{22}}{p_{21}} \right]$$

$$= \frac{N(1 + S_1) (1 + S_2)}{\left(1 + m_1 - \frac{m_2 M_1}{C}\right)} \left[ \frac{p_{21} p_1 + q_1 p_{22}}{p_{21}} \right]$$

$$= \frac{N(1 + S_1) (1 + S_2)}{\left(1 + m_1 - m_2 \cdot \frac{N p_1 q_{21}}{N p_1 p_{21}}\right)} \left[ \frac{p_{21} p_1 + q_1 p_{22}}{p_{21}} \right]$$

$$= \frac{N(1 + S_1) (1 + S_2)}{\left(1 + m_1 - m_2 \cdot \frac{q_{21}}{p_{21}}\right)} \left[ \frac{p_{21} p_1 + q_1 p_{22}}{p_{21}} \right]$$

$$= \left[ \frac{N(1 + S_1) (1 + S_2)}{\left(1 + m_1\right) p_{21} - m_2 q_{21}} \right] \left[ p_{21} p_1 + q_1 p_{22} \right]$$

$$S_2 = \frac{\hat{R}_2 - R_2}{R_2}$$

$m_1$  = 妥當한 matches 와 虛偽 matches 의 比率

$m_2$  = 妥當한 非 matches 와 虛偽非 matches 의 比率

$p_{21}$  = (13) 式과 동일

$p_{22}$  = (13) 式과 동일

$q_{21} = 1 - p_{21}$

$q_1 = 1 - p_1$

이 경우 推定值의 絕對誤差와 相對誤差는 다음과 같이 定義된다.

$$\text{絕對誤差}(B_s) = \frac{N(1+S_1)(1+S_2)(q_1 p_{22} + p_1 p_{21})}{p_{21}(1+m_1) - q_{21} m_2} - N \quad (16)$$

$$\text{相對誤差}(B_r) = \frac{(1+S_1)(1+S_2)(q_1 p_{22} + p_1 p_{21})}{p_{21}(1+m_1) - q_{21} m_2} - 1 \quad (17)$$

여기서 만약  $(1+S_1)(1+S_2) > \frac{p_{21}(1+m_1) - q_{21} m_2}{q_1 p_{22} + p_1 p_{21}}$  이면

推定值 ( $\hat{N}$ ) 는 上偏向되고 만약  $(1+S_1)(1+S_2) < \frac{p_{21}(1+m_1) - q_{21} m_2}{q_1 p_{22} + p_1 p_{21}}$  이면

$\hat{N}$  은 下偏向된다.

다시 推定值의 相對誤差는 위에서 言及한 바 1) 獨立性的 缺如 2) matching error 3) 造作된 報告등의 個別的 誤差를 볼 수 있으며 同時에 이들 1), 2), 3) 들 사이의 各組合別 相互作用에 의한 誤差를 볼 수 있는데 그 內容은 綜合的으로 다음과 같다.

1) 統計的 獨立性的 缺如에 의한 誤差 (a)

$$= q_1 \left( \frac{p_{22}}{p_{21}} - 1 \right)$$

2) 對照誤差(matching error) (b)

$$= \frac{m_2 q_{21} - m_1 p_{21}}{p_{21}(1+m_1) - m_2 q_{21}}$$

3) 造作된 報告에 의한 誤差 (c)

$$= S_1 + S_2 + S_1 S_2$$

4) 諸誤差間의 相互作用

가)  $a \times b$  = 獨立性 缺如와 matching error 의 相互作用에 의한 誤差

나)  $b \times c$  = matching error 와 造作的 報告의 相互作用에 의한 誤差

다)  $a \times c$  = 獨立性 缺如와 造作的 報告의 相互作用에 의한 誤差

라)  $a \times b \times c$  =  $a, b, c$  相互作用에 의한 誤差

따라서 이들을 모두 合計하면, 즉 가)+나)+다)+라) =  $(a+1)(b+1)(c+1) - 1$  와 같이 된다.

## 4. C-D 技法 適用에 의한 우리나라 人口動態率 推定

恒久的이고 繼續的인 最近性的 精確한 人口動態統計는 우리나라에서 申告를 통해 短期間에 얻기 어렵다. 이의 代案으로 申告資料에서 人口動態 統計가 생산되기 까지 이러한 目的을 충분히 달성할 수 있는 Dual Record System 에 의한 人口動態統計 生産이란 側面을 고려할 수 있다.

우리나라에서 가능한 Dual Record System 은 다음과 같은 System 1과 System 2로 이루어지는 6개의 combination 을 생각할 수 있다.



먼저 System 2(家口面接調査)에서 reference period 를 1年 또는 1年 이내로 하는가 하는 문제는 우리나라에서 수차 실시된 出生力調査에서 reference period 가 1년이 넘는 경우에도 응답자의 기억상실로 인한 누락보고는 크지 않았던 것으로 보아, 年 1회 실시하는 것이 경제적이며, 실천가능한 것으로 생각된다. 특히 出生과악에 있어 妊娠歷調査에 기초를 둘 때 누락수준은 훨씬 낮아진다.

System 1	System 2 (家口面接調査)	
	年 1 回	年 1 回以上
① 申告資料		
② 人口센서스		
③ 家口觀察		

다음 System 1에는 가능한 3개 資料源이 있다.

첫째, 申告資料는 加外的 노력없이 邑·面 또는 市·區廳에서 얻을 수 있는 利點이 있는 반면 다음과 같은 결정적 결함을 갖고 있다.

① System 2 자료와 對照하기에 申告水準이 너무 낮다.

② 申告의 대부분이 本籍地에서 이루어지기 때문에 현주지의 申告數와 실제 발생수 간에 현저한 차이가 있다<sup>35)</sup>. 忠北報恩에서 調査된 全出生 중 他地域에 本籍을 둔 出生은 16%였고, 報恩郡 2개面에 1973년 6~12월에 신고된 出生의 36%, 死亡의 30%가 他地域 居住者로서 本籍地인 報恩郡에 신고되었다. 그러므로 어떤 特定地域이 申告資料는 申告의 本籍地主義가 강하게 이행되고 있는한 System 2 자료와 對照되기 어렵다.

③ 申告資料 중 특히 發生記錄이 실제와 다른 比率이 크다.<sup>36)</sup> 報恩에서 조사된 1973년 이후 신고된 자료에 기록된 發生年月日이 실제 發生日과 일치하는 比率은 出生에서 50%, 死亡에서 48%였다.

둘째, 人口센서스는 5년 또는 10년 간격으로 이루어지고, 美國의 경험에 비추어 센서스間 발생된 出生과 死亡 과악에 많은 누락이 있음이 지적되고 있다.

셋째, 家口觀察 資料의 이용 가능성은 높다.

현재 經濟企劃院 統計局 主管下에 전국 代表的 標本에 대하여 실시하고 있는 CDS(continuin, demographic survey)는 別個의 家口調査 資料와 對照하여 精確한 動態水準 추정이 가능한 有力한 資料이다. 그러나 이 survey 를 하나만을 위해서 莫大한 資金이 投入되고 있어 經濟的 側面에서 어려운 점이 있다.

다른 하나의 家口觀察 資料는 保健要員의 活動에 의해서 이를 담당지역의 出生과 死亡을 發見하는 것이다.

全國保健網을 통한 人口動態 申告改善 研究<sup>37)</sup>에서 保健所 家族計劃要員이 1개 面の 2/3 지역(가구수 平均 844 家口, 최대 1,333가구, 최소 383가구)을 담당하여 出生·死亡·人口移動 事件의 發見과 이들의 申告권장을 위한 追加的 活動을 전개하였다. 특히 推定發生數에 대한 發生報告率은 높았다.

表 7은 1969년 9월~1971년과 1973년 1년간의 人口動態申告造成員(家族計劃要員)이 담당하고 있는 실험지역에서의 出生報告<sup>38)</sup>와 별개의 독립적 家口調査(이들 담당지역 가구의 1/10 標本家口)에 의한 結果를 C-D 技法으로 누락된 出生을 추정 한 것이다.

35) 宋建鏞·林鍾權, 報恩地域 人口動態改善研究— 1次中間報告—, 家族計劃研究院, 1974年 3月, p. 8.

36) 宋建鏞·林鍾權, 上揭書 p. 9.

表 7 C-D 技法 適用에 의한 누락 出生數 推定 : 1969~72年

年度 및 地域	百 分 率 分 布				
	計(推定出生數)	Both Systems	System 1*	System 2**	Missed by Both Systems
1969. 9~71. 8					
市 部	100% ( 72)	90. 3%	8. 3%	1. 4%	0
郡 部	100 (384)	83. 1	12. 8	3. 6	0. 5
1972年					
市 部	100 ( 55)	83. 6	10. 9	5. 5	0
郡 部	100 (215)	64. 2	22. 8	9. 8	3. 3

\*面接調査員의 家口調査 結果임.

\*\*造成員의 報告 結果임.

市部地域에서 System 2(造成員 報告)에 의한 出生報告率은 全推定 出生數의 92~89%, 郡部地域에서 87~74%였다.

이와 같이 C-D 技法 適用에 의하여 調査地域의 粗出生率은 1969~71년 기간에 市部에서 人口 1,000人當 29.8, 郡部에서 31.7이였고, 1972년에는 市部 郡部 양지역에서 28.1로 각각 추정되었다.

이들 推定値가 갖는 標準誤差( $\sigma_{\hat{N}}$ )는 1969~71년 기간 出生의 경우 市部에서 0.25, 郡部에서 1.55였고, 1972년 出生의 경우 市部에서 0.60, 郡部에서 3.36이었다. 이와 같이 標準誤差가 극히 낮은 것은 matching rate가 큰데 기인된다.

이와같은 研究經驗에 비추어 保健要員의 계속적 관찰에 의한 出生 및 死亡의 報告資料와 別개의 獨立된 reference period 1年の 家口調査로서 이루어지는 Dual Record System은 우리나라의 精確한 人口動態統計를 생산하는 유력한 方法으로 고려된다.

이 System 실시에 있어 다음 사항이 고려되어야 할 것이다.

첫째, 각 邑·面에 배치된 3명(가족계획·결핵·모자보건 요원)의 要員은 多目的의 要員化하여 邑面 地域을 3分割한 각각의 담당지연에서 出生·死亡 事件의 發生을 잘 알고 있는 里 및 부락단위로 어머니회장, 里長, 자원지도자 등으로 協助網을 조직하여 이들의 적극적 협조를 얻어야 한다.

둘째, 이들 保健要員이 발견한 出生 및 死亡 事件은 이들 固有業務의 効率化에 이용되도록 추가적인 人口動態 事件의 發見과 保健活動을 體系化해야 한다.

셋째, 全國을 대표할 수 있는 地域을 선정해서 이러한 活動을 전개하되, 郡部地域은 전지역에서 이 活動이 전개되고 市部地域에서 單은 事業의 性格上 標本地域 선정이 필요하다.

넷째, 別個의 獨立된 家口調査는 全國을 대표할 수 있는 標本에 대하여 年 1回 定期的으로 실시한다.

이들 事項이 고려된 우리나라의 Dual Record System은 精度 높은 人口動態統計의 生産에 적은 費用으로 크게 기여할 수 있으며, 이 System下에서 人口動態 申告水準의 改善도 크게 영향을 받아 申告資料를 통한 人口動態統計 生産이 가능한 期間을 短縮할 수 있다.

37) 宋建鏞·車正根, 全國保健網을 통한 人口動態申告 改善을 위한 研究(第1輯), 家族計劃研究院, 1972年 5月, p. 42.

38) 宋建鏞·林權鍾, 전게서(최종보고), 家族計劃研究院, 1974年 9月, (內部資料)

## 5. 結 論

정확한 人口動態統計의 要求는 國家 및 社會的으로 높고, 이 統計가 모든 統計의 가장 基礎가 된다는 데서 긴급히 利用 가능해야 한다. 그러나 人口動態 申告가 不振하여 申告制度를 통하여 이 統計가 생산되지 못하고 있고, 가까운 장래에도 신고수준의 급격한 개선이 기대되지 않는다.

이러한 情況下에서 恒久的으로 계속적인 最近性的 정확한 人口動態統計의 生産은 Dual Record System의 실시로서 가능하다는 것이 지적되었다. 이 System은 印度를 위시한 몇 개 開發途上 國에서 좋은 成果를 올리고 있다.

Dual Record System과 C-D 技法의 理論的 背景과 우리나라에서의 適用 가능성이 검토되었고 그 결과로 다음과 같은 우리나라가 갖고 있는 利點을 충분히 活用한 자료수집 方法으로 구성된 Dual Record System이 申告를 통한 人口動態統計 生産이 가능해질 때 까지의 期間에 人口動態 統計 生産에 크게 기여할 수 있다는 결론에 도달하였다.

첫째, 多目的要員化한 邑·面 保健要員이 이들 基本業務의 하나로 이들이 담당할 지역의 出生 및 死亡을 發見報告한다. 都市地域에서는 標本地域을 선정하여 역시 保健要員이 出生과 死亡을 發見報告한다.

둘째, 全國을 代表할 수 있는 標本에 대한 家口調査를 年 1回 실시한다.

셋째, 이들 두 方法에 의하여 수집된 자료는 C-D 技法에 의하여 정확한 人口動態水準을 추정한다.

이러한 형태의 Dual Record System은 全國的 조직망을 갖고 주민과의 밀접한 접촉을 통해 전개되는 保健活動의 潜在力이 우리나라 人口動態統計의 획기적 개선에 크게 기여할 것으로 기대된다.

〈ABSTRACT〉

## An Application of Chāndrasekar-Deming Techniques for Estimating Current Fertility Level in Korea

by **Kun Yong Song**  
**Young Hoon Lee**

This paper aims to propose an useful scheme to produce current vital data at a continuing basis in Korea.

A promoter system was introduced using public health field workers as special health statistics promoters for 18 months in Korea. The system was built on the existing vital registration scheme rather than replacing it.

1. Although the promoter system did considerably increase the level of registration, it was not complete enough to yield useful data.

2. On the other hand, the data notified by the promoters were of high quality and completeness. They may be used for estimating current vital rates through dual record system.

The vital events recorded by the promoters were matched with data obtained by interview in the study area. The matching indicated that; the proportions of births covered by both methods were 90 percent of the total estimated births in urban areas and 83 percent in rural areas: the estimated proportions missed by both methods of the total estimated births were negligible in urban areas and 0.3% in rural areas. Annual crudebirth rates per 1,000 population for 1969-71 were estimated as 29.8 in urban areas and 31.7 in rural areas, with standard errors of 0.25 and 1.55 respectively.

3. It appears that a dual record system, which will be consisted of health statistics promoter's report and data obtained from interviewing, will produce currently accurate vital data at a continuing basis in Korea.