

資料處理過程에 對한 統計學的 檢討

—一部 醫學雜誌에 揭載된 論文例를 中心으로—

서울大學校 醫科大學 豫防醫學教室

安 允 玉

서울大學校 保健大學院

高 應 麟

—Abstract—

Statistical Resumption for Health Articles in One Series of Medical Journals

Yoon Ok Ahn, M. D.

Dept. of Preventive Medicine, College of Medicine, Seoul National University

Ung Ring Ko, M. D., Ph. D.

School of Public Health, Seoul National University

One hundred and thirty one health articles (as "original articles") reported in medical journals published in Korea during the period from Jan. to Dec. 1971. were selected as the study materials for the purpose of resumption and criticism.

Of a total of 131 articles, thirty-nine were not required the statistical application for the data processing and drawing their conclusions and ninety-two, 70.2%, were needed to apply. The 92 articles were divided into three groups by the types of study: 54 for experimental study, 30 for clinical case analytic study and 8 for field study. Each report was resumed and criticised under the statistical aspects in terms of data processing and the validity of the conclusions.

The results of resumption of the 92 articles needed statistical application are summarized as follow;

1. Of the 54 articles of experimental studies resumed, thirty-one articles, 57.4%, were considered acceptable. Among the 38 articles classified as clinical case analytic and field studies, only 28.9 per cent were considered acceptable. In total 45.7 per cent or less than half were considered acceptable.
2. Types of errors encountered are as shown in Table 3.
3. The most frequently encountered error was that the conclusions were drawn without having made any statistical test. This type of error was found in 14 out of 35 errors presented in experimental studies, 11 of 35 in clinical case analytic studies and 5 of 9 in field studies and 30 out of 97, 38.2%, in total.

緒 論

個個의 物理的, 生物學的 現象이나 多變의 生體反應으로부터 그 一般性을 찾고 그것의 機轉과 原理를 糾明하는 것은 科學의 重要한 課程이요 內容이라 하겠다. 이런 分野의 研究結果가 發表되는 科學論文의 結論은 그것을 說明하고 뒷받침 할 만한 充分한 根據의 提示가 隨伴되어야 할 것은 當然한 事實이다. 研究者들에 依해 내리진 結論을 뒷받침 할 수 있는 여러가지 根據의 提示에 있어서는 特別히 醫學, 齒醫學, 藥(理)學 및 看護學 分野에서는 肉眼的 혹은 顯微鏡의 所見이나, 生化學的 反應에 依한 所見, 더 나아가서 多角的인 解析에 依한 理論展開, 그리고 統計的 處理方法 等等이 適用되고 있다. 이러한 여러가지 方法中 現在 科學論文에서 가장 많이 使用 혹은 適用되고 있는 것 中의 하나가 統計的 方法이라 하겠다.

모든 生物學的 現象이나 生體反應을 數量的으로 表示될 수는 없어 數理的인 解析(mathematical proof)으로 說明할 수 있는 것은 매우 制限되어 있으나³⁾ 統計(Statistics)는 觀察可能한 現象을 限定된 觀察結果(Specific observation)로써 一般性(Generalities)을 推論할 수 있는 方法이 될 수 있다²⁾. 따라서 統計的 分析 方法은 모든 科學研究分野에 있어 結論의 妥當性을 證明하는데 가장 基本的이고 普遍的인 方法의 하나이다.

科學, 特別히 醫學分野 研究에 있어서는 統計的 方法 適用 및 資料處理方法에 關한 著書는 現在 國內外로 많이 나와 있으나²⁻⁷⁾ 實際 그 適用이 얼마나 正確히, 適切하게 이루어지고 있는가에 對한 報告가 國內에서는 아직 發表된 바 없었고 外國의 경우에서도 매우 드물어 1966年 Schor & Karten 이 처음으로 報告한 바 있었다¹⁾.

이에 本 著者는 國內 科學論文을 對象으로 그 資料處理過程에서의 統計的 方法의 適用 및 여기에서 얻어진 結論들에 對하여 主로 統計學的 見地에서 檢討하였는 바 약간의 結果를 얻었으며 이 結果가 앞으로의 科學論文, 特別히 統計的 方法 適用이 必要한 論文에 보다 正確하고 適切한 統計的 方法 適用이 이루어지는데 多少의 보탬이 되고져 이에 報告하는 바이다.

對象 및 方法

1) 對象

1971年 中國內에서 發刊된 一部 醫學綜合雜誌 24卷에 「原著」로 掲載된 醫學, 齒醫學, 藥(理)學, 看護學分野의 論文 131編을 그 對象으로 하였다. 本 著者가 擇한 醫學綜合雜誌가 國內에서 發刊되는 모든 醫學雜誌를 대

表할 수는 없지만 여러分野의 學者들에게 比較的 자주 실리고, 읽히고 있는 까닭에 月刊 醫學綜合雜誌를 對象으로 하였다.

2) 方法

먼저 131編의 論文을 그 內容에 있어 理論展開形式, Review Reports, 病理, 組織學, 微生物學 等等의 分野에서와 같이 研究計劃이나 結論誘導에 있어 統計的 方法 適用이 不必要했던 論文과 그 以外의 統計的 方法 適用이 必要했던 論文으로 區分하였다. (Table 1 參照)

Table 1. Classifications of articles by the type of study (articles published in 1971)

Articles not required statistical application	39
Articles required statistical application	92
experimental studies	54
clinical case analytic studies	30
field studies	8
Total	131

또한 統計的 方法 適用이 必要로 했던 92編의 論文을 그 研究內容에 따라

1. 實驗的 研究(Experimental type study)
2. 病院患者를 對象으로 한 症例分析 (Clinical case analytic study)
3. 人口集團을 對象으로 한 Field study 로 分類하였으며 (Table 1 參照) 이들 各各에 對하여 主로 1. 研究計劃에 合當한 資料處理過程과 2. 여기에서 誘導된 結論들의 妥當性에 對하여 統計學的 見地에서 檢討하였다.

結 果

1) 92編中 42編(45.7%)의 論文에 있어서는 매우 適切하고 合當한 統計的 方法이 適用되어 그 結論들의 妥當性을 認定할 수 있었다. (Table 2. 參照)

2) 나머지 50編에 있어서는 研究計劃에 따르는 資料處理에 있어서나 그 結論誘導에서도 統計的 方法 適用이 不合理하였거나 또는 不充分하였다. (Table 2 參照)

3) 한편 50編의 再檢討되어야 할 論文에서 發見된 不合理點이나 未備點(Errors) 들을 大略 分類해 보면 아래와 같았다. (Table 3. 參照)

1. 統計的 檢定을 行한 後에 내버려져야 할 結論이 다만 나타난 數值로써 比較結論지은 경우.

例로서 A와 B를 各各 따로 處置한 두 實驗動物群의

Table 2. Result of resumption by types of study

Types	Number resumed	Acceptable		Revised	
		No.	%	No.	%
Experimental	54	31	57.4	23	42.6
Clinical case analytic and Field	38	11	28.9	27	71.1
Total	92	42	45.7	50	54.3

肝에서 Galactose dehydrogenase 의 activity 의 差를 論할 때 단지 各各의 實驗動物群에서 얻어진 平均實驗成績值로서만 行한 경우이다. 이런 경우에는 各各의 實驗成績의 平均值와 標準偏差(標準誤差도 同)를 求하여 t-test를 하거나 比較하려는 實驗群이 많은 경우는 分散分析法을 利用하여 그 差異의 有意性을 表示해 주어야 할 것이다.

2. 研究內容上 그 計劃에 있어 對照群의 設定이 必要한데도 對照群의 成績없이 研究가 이루어진 경우.

예로써 어느 疾患患者의 性比를 求해보니 男女比가 6:4로 算出되었을 때 이는 곧 그 疾患이 男性에게 많다고 言及된 경우이다. 이때에는 그 病院에 來院하는 患者의, 혹은 그 地域住民의 性比를 表示하여 이를 對照群으로 比較되어 言及되어야 할 것이다. 즉 病院에 오는 患者나 그 地域住民의 性構成의 特性에 따라 조사된 患者의 性比가 달라질 수 있기 때문이다.

Table 3. Types of insufficient statistical analysis frequently encountered in the articles

1. No statistical test
2. Omission of control group
3. Too small sample size or/and the size not mentioned
4. No evidences for justification of conclusions
5. Not appropriate design of study
6. Improper conclusion drawn as the result of the application of statistical tests
7. Use of measures of sensitivity without measures of specificity
8. Statement of significance of differences without description of test applied or level of significance used
9. All others

3. 標本의 크기 (Sample size)가 不合理하게 작거나 標本의 크기를 明示하지 않은 경우.

現在 行하여지고 있는 모든 研究의 大部分은 標本에 의하여 이루어지고 있음이 事實인 바 全數調査나 全數研究가 아닌 경우에는 研究에 使用될 標本이 그 研究內容의 全體를 代表할 수 있는 適合한 크기를 지녀야 하고 또한 그 크기에 對하여 明示하여야 할 것이다.

4. 結論을 言及함에 있어 이를 뒷받침 할 만한 根據가 모호한 경우로써 나타내 주고 있는 數値가 平均值인지, 아니면 特殊한 數値인지, 얼마만한 크기의 標本에 얻어진 結果인지, 統計의 檢定이나 處理가 되었는지 等に 對한 아무런 言及이 없는 結論이었다.

5. 研究計劃이 不充分한 경우.

만약 A 라는 新藥이 既存의 B藥보다 그 治療效果가 좋다고 할 때 A藥投與患者群에서 얻은 治療成績을 過去 다른 研究者에 의해 發表된 B藥의 治療成績과 比較한 경우로써 이때는 年齡, 性別 및 기타 治療에 影響을 미칠 수 있는 諸條件이 비슷한 두 患者集團에 各各 A, B藥을 投與하여 얻은 治療成績을 統計的方法으로 比較檢討되어야 할 것이다.

6. 適切한 統計的 檢定을 하고서도 다른 結論을 기술한 경우.

두 數値의 差를 論함에 있어 t-test 結果 $P > 0.01$ 이라고 해 놓고 그 差가 有意하였다고 하거나 $P > 0.05$ 라고 表示하고 그 差의 解析에 있어서는 몇 % 增加 혹은 減少하였다고 說明한 경우이다.

7. 治療效果나 診斷正確도에 對한 研究에 있어 sensitivity 만으로 判斷한 경우.

특히 診斷方法의 價値에 있어서 case 를 case 라고 밝혀내는 率(sensitivity)이 重要하지만 그 診斷過程에 있어서 多少의 危險이나 經濟的 負擔이 있을 때에는 non-case 를 non-case 로 밝혀내는 率(specificity)도 같이 알아내어 이 두 率로 그 價値를 論해야 할 것이다.

8. 統計學的 分析(檢定)方法과 判定基準(有意水準)에 對한 明示가 없이 統計的 結果만 表記해 놓은 경우.

예로 무슨 項目을 比較하였든지, 有意水準얼마로 무슨 test 의 統計的 檢定을 行하였는지를 말하지 않고 단지 "Significant" 라고만 表記했던 경우로써 적어도 $P < 0.05$ (t-test), $P < 0.05$ (χ^2 -test) 等等의 表示로써 讀者로 하여금 有意水準과 test 方法을 알 수 있도록 明記되어야 할 것이다.

9. 그 외 諸般 統計學的 見地에서 發見된 未備點이나 잘못 된 點들이다.

考察 및 結論

1966年 Stanley Schor & Irving Karten 이 美醫學協會誌에 發表한 바에 依하면 1964年에 美國內 醫學雜誌에 發表된 醫學論文中 295編을 對象으로 Analytic type study 와 case description type study 로 分類하여 結論들에 對하여 統計學的 側面에서 檢討한바 Table 4.에서 보는 바와 같이 analytic type studies 에서는 149編中 41編(27.6%)에서만이 統計學的으로 妥當하였고, case studies 에서는 78.1%에서 妥當하여 全體的으로 52.5%에서 妥當하였다고 하였다.

本 著者의 結果는 Table 2 에서와 같이 全體的으로 總 92編中 45.7%인 42編에서 妥當한 結果로 나타났다. Schor & Karten 의 檢討方法과 本 著者의 方法이 一致하지 않을 뿐아니라 論文內容區分에 있어서도 같지 않기 때문에 엄밀히 두 結果를 比較할 수는 없지만 上記의 條件이 같다는 假定下에 그 結果를 比較할 때 그 差異는 有意하지 않아 ($P[x^2] > 0.05$) 비슷한 結果로 나타났다.

Table 4. Results of Evaluation by Type of Study
(by Schor & Karten, JAMA Vol. 195, No. 13, 1966)

Types	Number Read	Acceptable		Revised		Rejected	
		No.	%	No.	%	No.	%
Analytic	149	41	27.5	101	67.8	7	4.7
Case description	146	114	78.1	32	21.9	—	—
Total	295	155	52.5	133	45.1	7	2.4

한편 各 研究內容別로 볼 때 Table 2에서 보는 바와 같이 資料處理過程이나 結論이 統計學的 見地에서 妥當하다고 認定되는 論文은 實驗의 研究內容에서 57.4%, 症例分析이나 field study 內容에서 28.9%로 實驗의 研究內容에 있어서 그 成績이 더 좋았다. ($x^2=6.18$, $P < 0.05$) 다시 말하면 症例分析 研究論文에 있어서는 그 論文이 雜誌에 掲載되기 前에 그 內容을 統計學的 側面에서의 再檢討가 필요하다고 思料되는 것이다.

實驗의 研究에서나 症例分析 및 field study 의 再檢討되어야 할 50編의 論文에서 가장 많이 나타난 未備點은 研究成績에 對한 統計的分析(檢定)의 缺如로 實驗의 研究에서는 全 errors 의 40%, 症例分析 및 field study 의 경우에서는 36%로 全體的으로 볼 때는 總 79個의 未

備點이나 잘못된 點中에서 30個로 38.2%이다. 實際 이러한 error를 지녔던 論文들의 研究成績에 對하여 統計的分析(檢定)을 行하여 본바 各 研究者들의 結論과 相異한 結果로 나타난 경우는 매우 적었다. 따라서 이들의 結論을 妥當한 것으로 判斷한다면 全體的으로 65% 以上の 論文에서 妥當하다는 結果가 된다.

Table 5. Number and Type of Errors by Type of Study

Error No.	Type of study			All
	Experimental	Case analytic	Field	
1.	14	11	5	30
2.	—	5	1	6
3.	7	2	—	9
4.	2	2	—	4
5.	1	3	—	4
6.	4	—	—	4
7.	—	3	—	3
8.	2	—	—	2
9.	5	9	3	17
Total	35	35	9	79

또한 實驗의 研究에서 두번째로 많이 發見된 error는 標本의 크기가 不合理하게 작거나 標本의 크기를 明示하지 않았던 點이었다. 그러나 症例分析 研究에서는 對照群設定 缺如이었다. 對照群의 設定없이 이루어진 症例分析 研究는 實際로 매우 많은 例에서 볼 수 있었으나 여기에 밝힌 數値는 그 研究內容上 꼭 對照群이 必要로 했던 경우이다.

以上 1971年 國內에서 發刊된 一部 醫學綜合雜誌에 「原著」로 發表된 論文 131編中 그 論文作成에 있어 統計的方法 適用이 必要로 했던 92編을 對象으로 資料處理過程과 여기에서 誘導된 結論에 對하여 統計學的 見地에서 그 妥當성을 檢討한 바 몇가지의 結果를 얻을 수 있었다.

1. 實驗의 研究 論文들 中에서는 57.4%, 症例分析 및 field study 論文에서는 28.9%에서 妥當하고 適合한 統計的方法이 適用되었었고 全體的으로는 45.7%이었다.

2. 雜誌에 掲載되기 前 再檢討되어야 할 未備點中 38%에서 研究成績에 對한 統計的分析(檢定)의 缺如이며 實際로 統計的檢定을 行한 結果 各 研究者들의

結論과 一致하지 않았던 例는 매우 적었다.

3. 論文 完成에 앞서 간단한 統計學的檢討를 함으로써 大部分의 未備點들이 補完될 수 있다고 思料된다.

(이 論文이 完成되기까지 여러 資料 및 物心兩面으로 協助하여 주신 保健大學院의 金正根副教授, 李瑛煥선생님, 姜吉遠선생님 特히 서울醫科大學 豫防醫學教室의 金仁達教授任, 權彝赫教授任 그리고 尹德老教授任께 深甚한 감사를 드립니다.)

參 考 文 獻

1. Stanley Schor, Irving Karter: *Statistical Eva-*

luation of Medical Journal Manuscripts. J. A. M. A. Vol. 195 No. 13, pp. 1123-1127, 1966.

2. Dixon, Massey: *Introduction to Statistical Analysis.* 3rd Ed. McGraw Hill, 1969.

3. Norma T. J. Bailey: *The Mathematical Approach to Biology and Medicine.* 1st Ed. Wiley, 1970.

4. Bradford Hill: *Principles of Medical Statistics,* 8th Ed. Oxford Univ. Press, 1967.

5. Bancroft: *Introduction to Biostatistics.* 1st Ed. Hoeber-Harper, 1957.

6. Goldstein: *Biostatistics; An Introductory Text.* 8th Printing, 1971.

7. 鄭英鎮: 近代 統計學의 理論과 實際. 寶晉齋, 1963.

