原 著

内リンパ嚢手術の再評価

—Thomsen 論文の問題点—

矢沢代四郎

Reevaluation of the Endolymphatic Sac Surgery for Meniere's disease

—Problems with the 1981 Article by Thomsen et al.—

Yoshiro Yazawa

Department of Otolaryngology, Shiga University of Medical Science

Since the publication of the article entitled “Placebo effect in surgery for Meniere's disease. A double-blind, placebo-controlled study on endolymphatic sac shunt surgery” by Thomsen et al. in 1981, endolymphatic sac surgery for Meniere's disease has tended to receive a lower evaluation. Investigators who read only its abstract were likely to understand that this study was well planned and performed under a double-blind, placebo-controlled study and that the conclusion would be reliable. The results showed that only a minor difference ($p < 0.05$) was observed between endolymphatic sac surgery (active groups) and mastoidectomy (placebo group), and, in contrast, the greatest difference ($p < 0.01$) was found in both groups when preoperative and postoperative scores were compared. Thomsen et al. concluded from their results that the great difference ($p < 0.01$) between preoperative and postoperative scores in each surgery was more important than the minor postoperative difference ($p < 0.05$) between the two surgeries. In this study, however, they made a mistake in which p-value derived from paired-test was compared with the p-valued derived from the non-paired-test. They neglected the postoperative difference ($p < 0.05$) between the two surgeries. Therefore, their conclusion would be wrong. Our statistical reevaluation demonstrated that endolymphatic sac surgery (active group) was more effective than regular mastoidectomy (placebo group) in some aspects such as dizziness & vertigo, total scores, tinnitus, 250 Hz hearing threshold and AAOO evaluation. We must also note that Thomsen et al. were strongly influenced by Tork who considered that any treatment has 60–80% effectiveness in Meniere's disease. Thus, Thomsen et al. initiated this study expecting that endolymphatic sac surgery would have non-specific placebo effect. Finally we have to add that their research using human beings was a violation of the Helsinki Agreement on Medical Ethics in 1964.

Key words: Meniere's disease, endolymphatic sac surgery, reevaluation

はじめに

1970年代から1980年代初頭に盛んに行われていたメニエール病に対する内リンパ嚢手術であったが、1981年に Thomsen らが発表した論文¹⁾²⁾で「内リンパ嚢手術はプラセボ効果に過ぎない」と批判したことを契機にして内リンパ嚢手術はやや下火になって行った経緯がある。Thomsen らは、メニエール病患者30名を2群に分けて二重盲検法で A (active) 群15名には内リンパ嚢手術を行い、P (placebo) 群15名には乳突削開術を行い、1年後の効果判定の結果、両群間には大差が無かったと結論を出している。二重盲検法で2群間の手術効果を科学的に比較検定試験を行っていることからその説得力は大きく、この論文を契機として「内リンパ嚢手術はプラセボ効果に過ぎない」という批判が広まった。しかし、Thomsen らは当初から「メニエール病ではどのような治療を行っても60-80%の有効率がある。従って内リンパ嚢手術も非特異的なプラセボ効果に違いない」とする前提のもとにこの研究を行っていることがその序文や考察から読みとることができる。

Thomsen らの論文では、患者さんには内リンパ嚢手術と乳突削開術のどちらの手術を受けるかについて伏せて行われたことに対する倫理的な批判や効果判定の統計処理方法に対する批判が論文発表当初から出されていたが^{3)~5)}、2000年に Welling ら⁶⁾の内リンパ嚢手術を再評価する論文が発表されてから Thomsen らの論文の問題点が再度指摘されるようになってきている。そこで1981年の Thomsen らの論文¹⁾²⁾に掲載されている図表を参考にして統計的検討を行い、Thomsen 論文の統計的検討の問題点を明確にする。

対象と方法

1981年に発表された Thomsen らの2論文¹⁾²⁾を検討の対象論文とした。

Thomsen らは、メニエール病患者30名を15名ずつの2群に分けて A 群15名には内リンパ嚢手術を行い、P 群15名には乳突削開術を行い、術後1年後の効果を統計的に判定している。手術を受ける場合に、患者さんにはメニエール病の2種類の別々の手術治療の有効性を検定する試験に参加してもらうことを報告しているが、内リンパ嚢手術を受けるのか乳突削開術を受けるのかは患者さんにはわからない状況で行っている。さらに研究

者にも個々の患者さんがどちらの手術を受けたかわからない状態にして（二重盲検法）、手術した施設とは別の施設で術後の経過観察は行った。

自覚的難聴、耳鳴、めまい、吐気、耳閉塞感の評価項目については患者さん自身が毎日0~3の4段階のスコアで記録し、月平均のスコア (mean monthly score) を算出して比較している。一方、患者さんの全般的症状については研究者が0~10のスコアで評価し、聴力検査としては250, 500, 1000 Hz について評価している。これらのスコアや数値について手術 A 群、P 群それぞれの術前・術後および A 群・P 群の2群間で統計的に比較検討している。特に患者さん自身のスコアについては、術前の月平均スコア (mean monthly score) と術後1年間の月平均スコアをもって統計的に比較検討している。統計法としては Mann-Whitney U 検定、Friedmann 検定、Kolmogorov-Smirnov two-sample test を用いている。

さらに患者さん自身および研究者の手術に対する術後評価の印象を「良、まずまず、不良」で評価している。また AAOO (1972) の評価基準で手術効果を χ^2 検定で評価している。

今回、これらのスコアは論文中の図に表示されているので、図から術前・術後のスコアを判読し、これらのスコアをもとにして術前・術後の比較にはノンパラメトリック検定の Wilcoxon 検定 (対応あり) を、A 群・P 群の2群比較にはノンパラメトリック検定の Mann-Whitney U 検定 (対応なし) を用いた。一部、聴力域値の検定には対応ある t 検定と対応のない t 検定も用いた。AAOO (1972) の評価基準に基づく検定には χ^2 検定を用いた。そして、今回我々が再検定した結果と Thomsen らが発表した結果を比較検討した。

結 果

(1) 全スコア (嘔気・嘔吐、めまい、耳閉塞感、耳鳴、自覚的難聴の5項目の総スコア) :

全スコアの術前・術後の変化を示すグラフは図1左 (文献1の Fig. 2) に示されている。図1左グラフから個々のスコアを判読して表にすると図1右表になるが、この値を基にして A 群、P 群それぞれの術前・術後の p 値を Wilcoxon 検定で求めると、A 群では $p=0.0008$ であり、P

図1 全スコア (嘔気・嘔吐、めまい、耳閉塞感、耳鳴、自覚的難聴の5項目の総スコア)

左グラフは Thomsen 論文¹⁾ の Fig. 2 より改変、右表は左グラフから判読したスコア、右下は再検討結果を表す。

群では $p=0.4419$ であった。また、A 群と P 群の有意差を求めると Mann-Whitney U 検定では $p=0.0294$ であった。Thomsen らの論文では「A 群、P 群ともに術後は術前に比較して有意差 $p<0.01$ をもって改善し、A 群と P 群の2群間には有意差がなかった」と述べているが、P 群には術前・術後の有意差が認められず、A 群と P 群間には明らかな有意差をもって A 群がすぐれていることが示され、Thomsen らの総スコア評価結果とまったく異なっていた。

(2) 嘔気・嘔吐：

嘔気・嘔吐の術前・術後のスコア結果は図2左 (文献1の Fig. 3) のグラフに示されている。図2左グラフから個々のスコアを判読して表にすると図2右表になるが、この値を基にして A 群、P 群それぞれの術前・術後の p 値を Wilcoxon 検定で求めると、A 群では $p=0.0033$ であり、P 群では $p=0.0554$ であった。また、A 群と P 群の有意差を求めると Mann-Whitney U 検定では $p=0.2146$ であった。Thomsen らの論文では「A 群、P 群ともに術後は術前に比較して有意差 $p<0.01$ をもって改善し、A 群と P 群の2群間には有意差がなかった」と述べているが、A 群

には術前・術後の有意差が認められる ($p=0.0033$) もの、P 群では有意差を認めていない ($p=0.0554$)。再計算の結果でも A 群と P 群の2群間には有意差がなかった。

(3) めまい：

めまいの術前・術後の結果は、図3左グラフ (文献1の Fig. 4) に示されている。図3左グラフから術前後の個々のスコアを判読して表にすると図3右表になるが、この値を基にして A 群、P 群それぞれの術前・術後の p 値を Wilcoxon 検定で求めると、A 群では $p=0.0010$ であり、P 群では $p=0.0231$ であった。また、A 群と P 群の有意差を求めると Mann-Whitney U 検定では $p=0.0446$ であった。Thomsen らの論文では「A 群、P 群ともに術後は術前に比較して有意差 $p<0.01$ をもって改善し、A 群と P 群の2群間にはわずかな有意差 ($p<0.05$) しかなかった」と述べているが、P 群の術前・術後の有意差は $p=0.0231$ であり、Thomsen らの論文中の記述 ($p<0.01$) と異なる。再検討でも両群間に有意差 ($p<0.05$) を認めた。

(4) 耳鳴：

耳鳴の術前・術後の結果は、図4左グラフ (文

図 2 嘔気・嘔吐
左グラフは Thomsen 論文¹⁾ の Fig. 3 より改変, 右表は左グラフから判読したスコア, 右下は再検討結果を表す。

図 3 めまい
左グラフは Thomsen 論文¹⁾ の Fig. 4 より改変, 右表は左グラフから判読したスコア, 右下は再検討結果を表す。

献 1 の Fig. 5) に示されている。図 4 左グラフから術前後の個々のスコアを判読して表にすると図 4 右表になるが、この値を基にして A 群、P 群それぞれの術前・術後の p 値を Wilcoxon 検

定で求めると、A 群では $p=0.0076$ であり、P 群では $p=0.8504$ であった。また、A 群と P 群の有意差を求めると、Mann-Whitney U 検定では $p=0.0309$ であった。Thomsen らの論文で

図4 耳鳴

左グラフは Thomsen 論文¹⁾ の Fig. 5 より改変, 右表は左グラフから判読したスコア, 右下は再検討結果を表す。

は「A 群は術後は術前に比較してスコアが減少しているが, P 群のスコアは同じような減少を認めていない。A 群と P 群の 2 群間には術後有意な差は認めなかった」と述べているが, A 群の術前・術後では有意に改善を示し, 2 群間についても A 群は P 群よりも有意差 ($p<0.05$) をもって改善している。

(5) 耳閉塞感:

耳閉塞感の術前・術後の結果は, 図 5 左グラフ (文献 1 の Fig. 6) に示されている。図 5 左グラフから術前後の個々のスコアを判読して表にすると図 5 右表になるが, この値を基にして A 群, P 群それぞれの術前・術後の p 値を Wilcoxon 検定で求めると, A 群では $p=0.0262$ であり, P 群では $p=0.0283$ であった。また, A 群と P 群の有意差を求めると Mann-Whitney U 検定では $p=0.5664$ であった。Thomsen らの論文の中では「A 群, P 群ともに術後は術前に比較して有意差 $p<0.01$ をもって改善し, A 群と P 群の 2 群間に有意差がなかった」と述べているが, A 群, P 群ともに術前・術後の有意差は $p<0.05$ であり, Thomsen らの論文中の記述 ($p<0.01$) と異なる。また 2 群間には有意差は認めなかった。

(6) 自覚的難聴:

自覚的難聴の術前・術後の結果は, 図 6 左グラフ (文献 1 の Fig. 7) に示されている。図 6 左グラフから術前後の個々のスコアを判読して表にすると図 6 右表になるが, この値を基にして A 群, P 群それぞれの術前・術後の p 値を Wilcoxon 検定で求めると, A 群では $p=0.1093$ であり, P 群では $p=0.2477$ であった。また, A 群と P 群の有意差を求めると Mann-Whitney U 検定では $p=0.1211$ であった。Thomsen らの論文の中では「P 群では術後スコアの減少を認めなかったが, A 群ではわずかな減少を認めた。しかし, A 群と P 群の有意差はわずかな ($p<0.05$) なものであった」と述べているが, 再検討結果では 2 群間に有意差を認めなかった。

(7) 研究者による全般的スコア:

研究者による全般的スコアの術前・術後の結果は, 図 7 左グラフ (文献 1 の Fig. 8) に示されている。図 7 左グラフから術前後の個々のスコアを判読して表にすると図 7 右表になるが, この値を基にして A 群, P 群それぞれの術前・術後の p 値を Wilcoxon 検定で求めると, A 群では $p=0.0015$ であり, P 群では $p=0.0010$ であった。また, A 群と P 群の有意差を求めると Mann-Whitney U 検定では $p=0.9130$ であった。

図 5 耳閉塞感
左グラフは Thomsen 論文¹⁾ の Fig. 6 より改変, 右表は左グラフから判読したスコア, 右下は再検討結果を表す。

図 6 自覚的難聴
左グラフは Thomsen 論文¹⁾ の Fig. 7 より改変, 右表は左グラフから判読したスコア, 右下は再検討結果を表す。

Thomsen らの論文では「A 群, P 群ともに術後は術前に比較して顕著な有意差 $p<0.005$ をもって改善したが, A 群と P 群の 2 群間に有意差がなかった」と述べ, 再検討結果とほぼ同様な結

果であった。
(8) 聴力検査値 (250, 500, 1000 Hz の平均):
聴力検査値の術前・術後の結果は, 図 8 左グラフ (文献 1 の Fig. 9) に示されている。図 8 左

図7 研究者による全般的スコア

左グラフは Thomsen 論文¹⁾ の Fig. 8 より改変, 右表は左グラフから判読したスコア, 右下は再検討結果を表す。

グラフから術前後の個々のスコアを判読して表にすると図8右表になるが、この値を基にしてA群、P群それぞれの術前・術後のp値を対応あるt検定で求めると、A群では $p=0.5930$ であり、P群では $p=0.0694$ であり、Wilcoxon検定ではA群では $p=0.9773$ であり、P群では $p=0.0639$ であった。また、A群とP群の有意差を求めると、「対応のないt検定」では $p=0.0657$ 、Mann-Whitney U検定では $p=0.1009$ であった。Thomsenらの論文の中では「A群とP群の2群間に有意差がなかったが、A群の方がP群よりも術後に聴力が改善する症例が多かった。また周波数別に見ると250 Hzのみが術後にわずかな有意差($p<0.05$)で改善していたが、他の500, 1000, 2000, 4000 Hzでは有意差を認めなかった」と述べている。A群では術前後でほとんど聴力変化がないが、P群では術後の聴力悪化傾向を認める。しかし2群間には有意差を認めなかった。

(9) AAOO (1972) 評価:

図9には文献1のtable 2を示しているが、分類のA(めまい消失と聴力改善)、B(めまい消失と聴力不変)、C(めまい消失と聴力悪化)、D(めまい不変)の内で通常はAとBが手術有効

と判定され、C、Dは手術不成功と判定される。この判定法に従って χ^2 検定を用いて評価するとA群はP群より有意に($p=0.003$)有効であった。しかし、Thomsenらの論文の中では「Cは聴力が悪化しているがめまい消失なので手術有効と判定されるべきであり、従ってA、B、CクラスとDクラスで χ^2 検定を用いて検討するとA群とP群間には有意差がなかった」と述べ、A、B、C、Dのクラス分類の段階での見解の相違を示している。

(10) めまいに関するA群、P群それぞれの術前・術後のp値と、A群・P群2群間のp値について:

図10aにはめまいスコアに関するThomsenらの記述¹⁾から作成した統計結果の図を示し、図10bにはめまいスコアに関するデータを統計的に再検討を加えた結果を示した。めまいスコアのThomsenらの記述では「A群、P群ともそれぞれに術後は術前に比べて有意($p<0.01$)にスコアが改善している。両群間を比較するとA群がわずかな有意差($p<0.05$)でP群よりまさっている。しかし、 $p<0.01$ は $p<0.05$ よりも統計的有意差としては優れているので、A群もP群も

図8 聴力検査値

- a : Thomsen ら¹⁾ のめまいに関する A 群, P 群それぞれの術前・術後の p 値と A 群・P 群 2 群間の p 値のまとめ。
- b : 再検討しためまいに関する A 群, P 群それぞれの術前・術後の p 値と A 群・P 群 2 群間の p 値のまとめ。

AAOO* Class	Treatment			
	Active		Placebo	
	No.	%	No.	%
A	1	7	1	7
B	12	80	6	40
C	0	0	3	20
D	2	13	5	33
Total†	15	100	15	100

*AAOO indicates American Academy of Ophthalmology and Otolaryngology.

†Results not significant by χ^2 test.

----- 今回の分類(AB/CD)

----- Thomsenらの分類(ABC/D)

図9 AAOO 基準に基づく効果判定の区分方法の問題

Thomsen らは ABC と D で区分して 2 群間比較をしているが、本来は AB と CD で区分するのが通常の方法である。

ともに示した術前に対する術後の大きな (greatest) 有意差 ($p < 0.01$) に比べると両群間の有意差 ($p < 0.05$) は小さな (minor) 差に過ぎない。従って A 群も P 群もともに有効であることを示したことになり、メニエール病に対する内リンパ嚢手術の効果はプラセボ手術の効果と同じである」と結論付けている。一方、図10bに示すように A 群の術前後の有意差は $p=0.0010$ であつたのに対して P 群の術前後の有意差は $p=0.0231$ であり、Thomsen らの記述のように P 群の有意差は $p < 0.01$ には該当していない。さらにまた術前・術後を統計的比較するときには対応ある検定が行われ、2 群間の統計的比較には対応のない検定が行われたはずであり、それぞれで求められた p 値は性質の異なるものであるから対応ある検定で得られた p 値と対応のない検定で得られた p 値は相互に比較できる性質のものではない。

(11) 項目別に求められた p 値

表1には、項目別に Thomsen ら¹⁾²⁾ が求めた p 値と今回我々が再計算して求めた p 値を示した。Thomsen ら¹⁾²⁾ の p 値の表現は $p < 0.005$, $p < 0.01$, $p < 0.05$ または nos (not significant)

a. Thomsenらのめまいに関する結果 (1981)

b. 今回のめまいに関する再検討結果

図10 Thomsen らのめまいに関する結果 (a) と今回の再検討結果 (b) の比較

P 群の術前・術後の差は $p=0.0231$ で Thomsen らの $p<0.01$ とは異なる。A 群は有意に P 群より有効であることを示している。

表1 A 群, P 群それぞれの術前後ならびに A 群・P 群 2 群間の有意差の Thomsen らの検定結果と今回の再検定結果

項目	A群術前後差 (対応あり)		P群術前後差 (対応あり)		2群間差(対応なし)	
	Thomsenのp値	Wilcoxon (再検定のp値)	Thomsenのp値	Wilcoxon (再検定のp値)	Thomsenのp値	Mann-Whitney U 検定 (再検定)
1 全スコア	<0.01**	0.0008***	<0.01**	0.4419	nos	0.0294*
2 嘔気・嘔吐	<0.01**	0.0033**	<0.01**	0.0554	nos	0.2146
3 めまい	<0.01**	0.0010**	<0.01**	0.0231*	<0.05*	0.0446*
4 耳鳴		0.0076**		0.8504	nos	0.0309*
5 耳閉塞感	<0.01**	0.0262*	<0.01**	0.0283*	nos	0.5664
6 自覚的難聴		0.1093		0.2477	<0.05*	0.1211
7 研究者の全般的スコア	<0.005**	0.0015**	<0.005**	0.0010**	nos	0.9130
8 聴力検査値		0.9773		0.0639	nos	0.1009
250Hz	<0.05*		nos		<0.05*	
9 AAOO判定					ABC/D: nos	AB/CD: 0.003** (χ^2 test)

nos: not significant, ***: $p<0.001$, **: $p<0.01$, *: $p<0.05$

のいずれかであり、細かい p 値は記載されていない。今回再検定してみると Thomsen らの p 値と異なる箇所がいくつか認められた。とりわけ P 群に対する評価点はあまい傾向があり、有意差を認めない項目においてさえ $p<0.01$ として記載されている箇所が多い、また 2 群間の比較においては「めまい」と「自覚的難聴」において $p<$

0.05 の有意差を認めるのみであったと記載し、A 群と P 群間には大きな差がないことを強調しようとしているが、今回の再検定では「全スコア」「めまい」「耳鳴」「AAOO 通常分類」の 4 項目において A 群は P 群より有意差をもって優れた結果であったことを示している。

考 察

(1) Thomsen らの意図

Thomsen らの論文の序文に記述されているように「メニエール病にはどのような治療を行っても60～80%の有効率を示し、これはまさに本来の治療効果というよりはプラセボ効果である」とする Torok⁷⁾ の意見に同調する文章で始まっている。そして、メニエール病の治療においては確かにプラセボ効果が現れやすいことをも記述⁸⁾ している。従って、Thomsen ら¹⁾²⁾ の論文は、当初よりメニエール病の治療におけるプラセボ効果を二重盲検法の比較試験を用いて科学的に証明しようとする意図で始められた研究である。

内リンパ嚢手術と乳突削開術の2種類の手術法が二重盲検法で比較検討された報告は、科学的であり、たいへんに貴重な研究結果であるが、医学倫理に関するヘルシンキ宣言(1964)には違反する可能性があることから今後は二度と行われることのない比較試験である。従ってこの二重盲検法の比較試験の結果は厳密に再検討されて正確に理解され評価されることに大きな意味がある。

ところが内リンパ嚢手術群と乳突削開術群の2群間で比較すると「めまい」「自覚的難聴」「聴力検査の250 Hz」において $p < 0.05$ の有意差をもって内リンパ嚢手術が優れていることが Thomsen らの検定で判明した。しかし、この結果を隠すために「乳突削開術に見られた2例を除けばこのような結果にはなっていない」と弁解している。また「自覚的難聴」の差もわずかなものであると主張している。さらに「聴力検査の250 Hzで見られた内リンパ嚢手術群の良好な聴力成績は、わずかに5～10 dBであり、この値は検査上の誤差の範囲である」と述べている、このように Thomsen ら自身が科学的に行った統計的有意差に対しても、その結果を無視し、曲解しようとして記述された論文である。とりわけ、AAOO(1972)の評価基準から考えると「めまいがとまり聴力が不変以上である」クラス A, B が有効群であり、「めまいがとまっても聴力が悪化した群(C) やめまいが不変群(D)」は無効群であり AB と CD を区分して扱うのが本来の比較方法であるのに、本論文では、Cのように聴力が悪化してもめまいがとまれば有効に属すとして ABC と Dをもって区分する方法をあえて取って χ^2 検定

を行って有意差のないことを証明しようとしている。これらの事実を考察すると内リンパ嚢手術の有効性が予想に反して出てしまったが、「メニエール病にはどのような治療法を行っても有効性が認められて、それはプラセボ効果である」と恣意的に主張しようとした論文と言わざるをえない。

(2) 統計学的結果の解釈の問題点

Thomsen らの論文¹⁾²⁾ における最大の問題点は、種々の統計的計算結果から得られる p 値の解釈方法の間違いである。同一の統計的計算方法で得られた p 値の場合は p 値の小さい方がより統計的な意味が大きい、異なった統計的計算方法で得られた p 値を相互に比較することには意味がないどころか大きな間違いを引き起こす。Thomsen ら¹⁾²⁾ が論文で主張する p 値の比較は、おそらく対応ある検定で得られた p 値 ($p < 0.01$) と対応のない検定または Mann-Whitney U 検定で得られた p 値 ($p < 0.05$) を比較して、抄録の中で述べているのは「 $p < 0.01$ (対応ある検定の p 値) は $p < 0.05$ (対応のない検定または Mann-Whitney U 検定の p 値) よりも統計的有意差としては優れているので、A 群も P 群も共に術前・術後の大きな (greatest) 有意義 $p < 0.01$ に比べると2群間の有意差 $p < 0.05$ は小さな (minor) 差に過ぎない」とする間違った解釈が根本にあることが最大の問題である。

(3) 統計的検討方法

Thomsen ら¹⁾ の論文中使用された統計法としては Mann-Whitney U 検定, Friedmann 検定, Kolmogorv-Smirnov two-sample test が記載されているが、その統計的結果においては p 値が正確に求められてはいない。その代わりに記載されている記述が $p < 0.05$ とか $p < 0.01$ と言った1%または5%の危険率での有意差を示している。しかし、再度検定をしてみると表1に示したように多くの点において Thomsen らの結果と我々の検討結果との間には相違が認められた。特に、P 群の術前・術後の Wilcoxon 検定においては Thomsen らは5項目で $p < 0.01$ の有意差を示したと記載しているが、再検定において $p < 0.01$ の有意差を示したのは表1の(7)の研究者の全般的スコアのみであり、(3)めまいと(5)耳閉塞感において $p < 0.05$ の有意差である。その他

の(1)全スコア(2)嘔気・嘔吐においては全く有意差を認めていない。

(4) 二重盲検比較試験の結果について

二重盲検比較試験の意図するところは、対照患者を盲検法で2群に分けて、2群間に有意差が無いかどうかを科学的、統計的に純粋に判定する方法である。従ってA群やP群のそれぞれの術前・術後の変化を観察することにはそれほどの重要性が無く、最も重要な問題は二重盲検で行った上での2群間に観察されるかもしれない有意差である。Thomsenらは当初は2群間に大差が無いものと予測してこの比較検討を行っているが、彼らの検定結果においては少なくとも3項目(めまい、自覚的難聴、250 Hz)においてはA群がP群よりも有意であることが示されている。この結果はThomsenらの当初の予測を裏切る結果でもあったと思われるが、そのためか有意差が出てきたことに対して弁明の記述がいくつか見られる。我々が改めてMann-Whitney U検定を用いて2群間の検定を行うと4項目(全スコア、めまい、耳鳴、AAOO判定)について有意差が確認された。従って9項目の内で4項目においてA群はP群よりも有意に有効であったことを科学的な統計法で証明したことになり、この事実を意図的に曲げて解釈した点において1981年発表のThomsenら論文¹⁾²⁾は大きな間違いを犯していると言える。

まとめ

Thomsenらの論文の序文に記述されているように「メニエール病にはどのような治療を行っても60～80%の有効率を示し、これはまさに本来の治療効果というよりはプラセボ効果である」とするTorok⁷⁾の意見に彼らは同調し、これを科学的に証明しようとした論文である。また、二重盲検比較検討試験を人を対象として行うことは医学倫理に関するヘルシンキ宣言(1964)に違反する行為であり今後二度と行われることのない比較検討法である。その意味においてはたいへんに貴重な研究結果の論文であるが、Thomsenらの当初の「内リンパ嚢手術は乳突削開術と同様なプラセボ効果に過ぎない」とする意図と反する「いくつかの点で内リンパ嚢の乳突洞削開術より有意に有

効性がある」とする事実が判明したために、Thomsenら論文中にはいくつもの弁解や弁明がみられる。当初のように科学的に二重盲検法比較検討試験で得られた結果をそのまま尊重すれば、内リンパ嚢手術の有効性を証明するものであり、混乱を巻き起こすことも無かったはずである。

文 献

- 1) Thomsen J, Breatlau P, Tos M, et al: Placebo effect in endolymphatic sac surgery for Meniere's disease. Arch Otolaryngol 107: 271-277, 1981
- 2) Breatlau P, Thomsen J, et al: Endolymphatic sac shunt surgery for Meniere's disease. A placebo-controlled double-blind study. Meniere's disease. Pathogenesis, diagnosis and treatment. In eds Vosteen KH, Schuklencht H, et al. pp 230-236, Geroche Thieme Verlag, New York, 1981
- 3) Arenberg IK: Placebo effect for Meniere's disease sac shunt surgery disputed (Letter to the Editor) Arch Otolaryngol 107: 773, 1981
- 4) Vaisrub N: Summary statement to Placebo effect in endolymphatic sac surgery for Meniere's disease. (Letter to the Editor) Arch Otolaryngol 107: 773-774, 1981
- 5) Pillsbury HC, Arenberg IK, Ferraro J, et al: Endolymphatic sac surgery. The danish sham surgery study: an alternative analysis. Otolaryngol Clin N Am 16: 123-127, 1983
- 6) Welling DB, Nagaraja HN: Endolymphatic mastoid shunt: a reevaluation of efficacy. Otolaryngol Head Neck Surg 122: 340-345, 2000
- 7) Torok N: Old and new in Meniere's disease. Laryngoscope 87: 1870-1877, 1977
- 8) Thomsen J, Bech P, Geisler A, et al: Lithium treatment of Meniere's disease. Acta Otolaryngol (Stockh) 82: 294-329, 1976

(原稿到着：平成14年6月10日
別刷請求先：矢沢代四郎
〒520-2192 滋賀県大津市瀬田月輪町
滋賀医科大学耳鼻咽喉科学教室)