

私の考える臨床医レベルが行う医学統計

七島 篤志

[令和2年9月5日入稿, 令和3年3月3日受理]

はじめに

2019年10月にこの項をカンファレンスで語ったときに、質の高い論文を輩出する世界の研究機関ランキングに沖縄県の恩納町に設立された沖縄科学技術大学院大学OISTが世界の9位に急成長しacademiaの脚光を浴びているというWebニュースを見ました。丁度、そこで学会が開かれる直前でしたので興味を持っていくと、宮崎大学の様に山と海、自然に恵まれた場所に整然と存在していました。外国人研究者比率が85%というちょっとした留学気分が味わえそうでした。また今年2020年9月現在は、産業医科大学の被引用率が世界で急上昇し注目されています。要は都市圏の旧帝大および類似大学でなくとも大学の実績をあげられる機会は十分あると思います。

さて、そんな研究や論文作成では統計が求められます。統計の誤った使用方法がいろいろな論文に用いられているという現状があるために、2019年3月Nature誌に800人の科学者が“統計的有意差なし”もうやめませんか？という投稿がされたそうです。そのことが科学の評価に深刻な存在をもたらしているとのこと。医学でもEBMこそがすべてなりの時期が2000年代おりましたが、次第に診療・社会的に影響が出るようになり、数字情報のみを頼り実証データ（その後リアル・ワールドといわれるようになりました）を表せないという論争がこの数年起きています。診療ガイドラインは、統計処理が基礎にあるEBMから作られますが、そのガイドラインに固執すると、個々の診療や臨床的知見を得るうえで自主性が失われてしまい、新たな発見

を臨床試験以外で生み出せなくなってしまうことに危惧しています。大学の日常臨床においてガイドラインがすべてであるという主張には、私自身は納得がいかないことがしばしばあります。そのような中、同じ外科学講座で心臓血管外科学分野の中村都英教授が、今米国ではEBMよりもValue based medicine (VBM) が発展して、近く日本もそうなるはずだと予告されています（@外科カンファレンス）。Value=Outcomes/Costで世界や地域経済に合った、情報過多の影響を受けにくい医療・科学の実践です。話を統計に戻しますが、従来から対立仮説検定が前提で5%、1%の有意水準で対立仮説が棄却（AがBよりいい、悪いは誤り）されます。P値(probability)、信頼区間、他の統計尺度いろいろあるけど統計的有意差の有無ということばはやめようということのようです。でも、これでないとな学会報告もできませんよね、とりあえず。医学的に真実に近い仮説は何か？に取り組んだ時に、1)母数や群間数で結果が違ってくる可能性もある、2)群の背景の違いに影響を受ける可能性など考えておく必要があると思います。私は、論文を読むときに、母数によってはある程度真偽を疑わない結果かもしれないと判断しています。

I. 統計の準備

1. 統計ソフト：大学まで統計を座学で学んでは来ましたが、数学が苦手な人にはできませんでした。しかし、PCの発展と共に手助けしてくれる統計ソフトが出てきて、一般の私たちも自分で行えるようになりました。信頼のおける統計ソフトはやはり高額で、かつては統計学科の持つSAS[®]、SPSS[®]などを用いた計算を行脚

してお願いしに行き、数カ月して結果をもらったこともありました。その代り統計者の介入があるので非常に信頼度が高い有益なものであることは間違いありません。私は簡単な統計処理は自前でソフトを買って行ってきましたが、発表用にグラフ化しやすいソフトを選んだり、自分の頭で理解しやすいものを、解析手法に応じて複数使い分けてきました（これらはEscher, Fisher, SAS, StatiView, Excel統計, Statistica, SPSSなどですが、多くはWindows XPでしか動きません。今や現代の新しい統計ソフトにまで使い方を覚えきれない！と限界を感じています）。これからのひとはこんな不経済なことはやめましょう！

2. **統計入門 (本)**：本を間違うと一気にやる気が覚めてきます。自分は統計専門家になるための勉強をしたいのか、いやそこそこの理屈がわかって、実用的であればというのが多くの臨床医の本音でしょう。一方、統計にはまって自分は、統計は良く勉強して十分理解しているとうそぶく医療者もいますが、私見では統計専門家に鼻で笑われますよと言いたいです。医学統計に限っても専門家と付き合ってみると、まったくレベルは違いますし、考え方も臨床医レベルと違うことがわかります。統計をマスターした臨床医になることは不可能ではないでしょうか。もう時代はネットでソフトの使い方、統計手法の選択なんて数分で調べられます。私を含め多くの人が無駄本をたくさん買ったはずで、**統計学も医学同様に急速に進歩**しています、Propensity score matchingなども2000年代に報告された手法ですが、Pitfallも理解しないといけないことを統計家から指摘されています。ある学会のデータベース委員をしています、医療統計専門者（医学部卒）と統計に自信のある臨床医との間で、データ解析がうまく進行しない事例に遭遇しました。第三者的にみると互いに相手の状況をまず理解して、共同作業による重要なビッグデータ解析を行う必要性を感じました。その後うまくはまると一気に素晴らしい結果に結びつきました！思い起こすと自

分も20年前に気軽に赤池情報量基準AIC解析を基礎医学の統計教室の教授にお願いに言ったら、3時間基本統計量の指導を数回受けました。本気で指導して頂けたので四分位点の必要性など覚えることができていい思い出です（結局、当時の優しい准教授が2,3日PCを用いて解析してくれました！）。皆さんも統計専門家に習うときは謙虚に行きましょうね、第一印象が大事です。

3. **判定**：臨床医に統計の理解度が足りないとしても、学会発表や雑誌投稿で統計的な内容を厳しく指摘されることは少ない気がします。自験例では、1) 司会者からこの治療群のパワーは？と聞かれた発表者が、パワーポイントを使いましたと答え、爆笑されていたこと見たことがあります。パワーは統計的検出力の意味でした。2) 論文で対応の無い多群間比較もしくは対応のある対象の多時相間の比較において、2群間比較ではなく多重比較検定を用いるよう指示されました（図1 a）¹⁾。
4. **論文上の記載**：Continues data (parameter, variable連続変数) were expressed as mean \pm SD. Differences between groups were examined using repeated measure ANOVA and Scheffe's multiple comparison test (多重比較検定). Differences in the categorical data (カテゴリー変数や階層変数) between groups were examined using χ^2 test. The correlation (連続変数同志の変数の関係：相関) between the two variables was examined by Pearson's correlation analysis. A two-tailed (両側) p value of <0.05 was considered significant. The software for analysis was used …くらいが一般的でしょうか。雑誌ではmedian (minimum-maximumもしくは四分位点Quartile) を良い雑誌では求めてくる傾向があります。線維化スコアの1, 2, 3, 4などの階層変数で、平均値で比較するひとがいますが正しくないのではないのでしょうか。連続変数や階層変数で比較する結果が混在する表では、図1 bのように私は示しています²⁾。

a) Immediate and late changes in leukocyte count and serum IL-6 after laparoscopic cholecystectomy

	PP group	AWL group
Leukocytes (cells/mm ³)		
before	6,822 ± 2,465	5,280 ± 1,272
immediately after	10,556 ± 4,420*	8,173 ± 2,517*
POD 1	7,767 ± 2,271	7,040 ± 2,282
Interleukin-6 (pg/ml)		
before	2.48 ± 2.33	1.33 ± 0.69
immediately after	11.49 ± 9.66*	29.07 ± 18.78*
POD 1	24.33 ± 32.24*	51.95 ± 42.85*
POD 2	11.88 ± 14.4*	11.13 ± 9.93*

* p < 0.01 versus preoperative value

b) Comparison of patient demographics, surgical record, and postoperative outcomes

	LHM group (n=11)	Non-LHM group (n=16)	p-value
Age	64±13	60±9	.272
Gender (Male/female)	8/3	14/2	.323
Background liver			
Normal/chronic hepatitis, cirrhosis/fatty, alcoholic, CASH	1/0/10	0/3/13	0.166
Preoperative liver functions			
Liver Damage Grade A/B*	11/0	12/4	0.122
Indocyanine green retention rate at 15 minutes (%)	17.1±9.5	11.6±6.1	0.150
Liver uptake ratio by the GSA liver scintigraphy	0.91±0.04	0.93±0.03	0.241
Surgical records			
Thoracotomy (No/Yes)	8/3	10/6	0.692
Blood loss (ml)	1,269±1,407	2,197±1,281	0.039
Red cell transfusion (No/Yes)	4/12 (333 ml)	7/4 (1,340 ml)	0.061
Total operation time (minutes)	429±149	544±148	0.071
Transsection time (minutes)	39.7±10.5	48.1±15.1	0.270
Procedures (hemihepatectomy/extended hemihepatectomy/trisectonectomy)	9/1/1	8/7/1	0.172
Inferior vena cava clamping (yes/no)	11/0	15/1	1.0
Cutting device (none/ultrasonic coagulator/Ligasure [®])	4/2/5	15/0/1	0.453
Patient outcome			
Morbidity (No/Yes)	7/4 (36%)	2/14 (88%)	0.011
Hepatectomy related complications (No/Yes)	7/4 (36%)	10/6 (63%)	0.252
Uncontrolled ascites (No/Yes)	9/2	10/6	0.401
Bile leakage (No/Yes)	9/2	13/3	1.0
Intraabdominal abscess formation (No/Yes)	8/3	13/3	0.662
Hospital stay (days)	29.2±21.4	34.0±14.9	0.121

*Liver Damage grade guided by the General Rules for the Clinical and Pathological Study of Primary Liver Cancer in Japan.^{1,2}

GSA: galactosyl serum albumin

図1. 2～多群間の比較に用いる表。

II. よく使う統計の実際：筆者の臨床研究論文を題材に私見を述べます。

1. 検出力：統計では解析に必要な適正な症例数の設定があります。ただこれが難しいですよ、特に限られた症例数で正規分布に従ってないことが多いので。少ないと過誤エラーがおきやすくなります。検定の結果が正しいとする一方、間違ふ可能性もあります。1) 第一種過誤 (αエラー)：偽陽性。有意差はでるけど真実で差はないこと。結果として重要なので有意水準 (significance level) はよく5% (偶然誤りが起こる確立) に設定。2) 第二種過誤 (βエラー)：偽陰性。有意差はでてないけど真実で差があること。よく20%に設定されているようです。これらはサンプルサイズに影響されるので、過誤を有意水準未満におさえられを計算しなければいけません。

* 図1 a の論文でType I errorと書かれていましたが当時Macintoshのパソコンを初めて購入するとよく「過誤 (エラー) が発生しました」と画面に出していました。そこでMacの会社や雑誌に問い合わせても、なんだか辻褄が合わずにいたお馬鹿な時代を思い出します。

2. 基本統計量：1) 連続変数では、平均値 (mean), 中央値 (median), 標準偏差, 標準誤差, 母数 (n), 中央値, 最小値, 最大値, 四分位点, 変動係数, 欠損数などを, ヒストグラム, 分布図, 生起確率プロット, 箱ひげ図などで集団

の傾向を見る必要があります。標準偏差SDは良く使われますがデータのばらつきをみるもの、標準誤差SEは平均値の推定精度を見るとされデータのばらつきがない実験系で使われるようです。連続変数の2群の平均値比較をまずみんなやりますよね！でも群別のデータの分布の検定が必要で分散分析ANOVAを用います。ANOVAが順位変数でノンパラメトリック検定で比較するにはMann-Whitney U testが行われますが、後者が検出力が劣るそうです。

2) カテゴリー変数では、群A, B, C, Group I,II,III,IV...などであらわし、各群の度数分布, ヒストグラム, 欠損数, 分散, 尖度, 歪み度などをみます。2群間比較では分割表 (二元表) をよく使います。しかし図1 b のカテゴリー変数などのようにm*n表が多くなります。期待値が少ないとKruskal-WallisやU testなどが良く使われ、分割が多いほど例数が減るのでp値は大きくなります。多群間で比較する際はχ²検定では判定は困難です。

3. 多重・多群検定：独立した群が3群以上である群とある群を比較するときに第一種過誤を出にくくするため使われます (図2 a)。Dunnet, Tukey, Bonferoni, Duncan, Scheffeいろいろありますが、比較対象が決まっているとき、群数, 正規分布, 分散などで目的が違います。余裕がある人は勉強してください！私は何か有意なものがないかと探すだけなのでScheffeを使

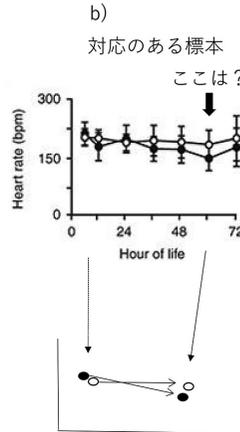
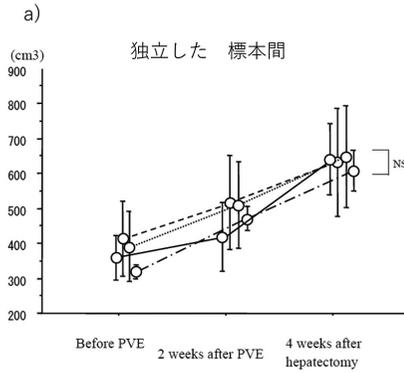


Table 3. Immediate and late changes in arterial blood gases and acid-base after laparoscopic cholecystectomy

	PP group	AWL group
pH		
before	7.46 ± 0.05	7.45 ± 0.06
immediately after	7.38 ± 0.1	7.3 ± 0.05
POD 1	7.41 ± 0.02	7.40 ± 0.04
POD 2	7.37 ± 0.02	7.41 ± 0.01
PO ₂ /FIO ₂ (mmHg)		
before	433.4 ± 138.1	535.8 ± 309.1
immediately after	592.5 ± 319.3	622.3 ± 473.9
POD 1	380.9 ± 83.9	389.6 ± 65.4
POD 2	349.2 ± 22.9	387.2 ± 38.4
SiO ₂ (%)		
before	99.2 ± 0.8	99.4 ± 0.6
immediately after	99.3 ± 0.4	98.8 ± 1.8
POD 1	95.8 ± 1.6*	95.6 ± 1.8*
POD 2	94.9 ± 0.8*	95.8 ± 1.1*
PCO ₂ (mmHg)		
before	35.5 ± 6.9	34.9 ± 5.6
immediately after	42.6 ± 6.1	41.3 ± 4.3
POD 1	41.8 ± 4.5	39.4 ± 4.0
POD 2	44.3 ± 2.2	43.2 ± 1.7
HCO ₃ (mmol/L)		
before	24.2 ± 2.0	24.2 ± 1.8
immediately after	25.0 ± 2.2	24.1 ± 1.5
POD 1	26.3 ± 2.7	24.3 ± 1.6
POD 2	25.3 ± 2.4	25.4 ± 3.4
BE (mmol/L)		
before	0.97 ± 1.67	1.08 ± 1.42
immediately after	0.00 ± 1.44	-0.22 ± 1.58
POD 1	2.18 ± 2.65	-0.11 ± 1.52
POD 2	0.16 ± 2.48	2.55 ± 0.78

* FIO₂ = 0.21 before operation and at POD 1 and POD 2, 0.4 immediately after operation
* p < 0.05 versus preoperative value

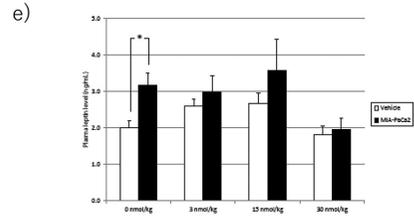
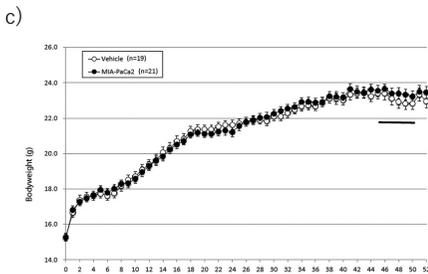


Figure 5. Relationship between plasma leptin levels and the administered CHG concentration of the control and transplanted groups.

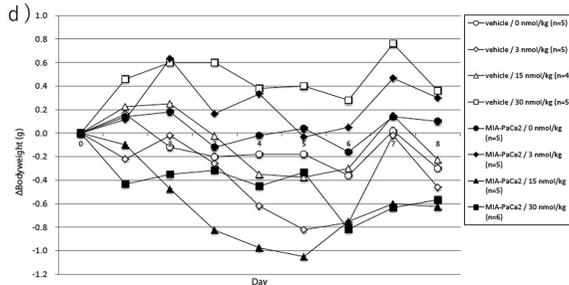


Figure 2. Changes in body weight of the control and transplanted groups receiving various concentrations of CHG for 8 days after the final injection.

図 2. 多群間の多重検定.

っています。ANOVAも反復測定として解析していました。図 2b はよくあることです。A 群と B 群の経時的なデータ比較でこの図だと 60 時間目に差がありそうにみえます。そこだけ 2 群間比較すると $p < 0.05$ だったとすると、何か 60 時間目に起こっているのだ！と喜ぶかもしれません。しかし多重検定すると差はないですし、たまたまここだけ何かの条件か値に影響を受けた、いわゆる過誤が起きる可能性があります。もし 60 時間目が何かの関心領域の時間帯なら、別のグラフにして 0 時と 60 時間でのみ差を比較するべきです。図 1 a の論文の別の表

です。これは多重検定で差が出ました¹⁾。グラフの全体の山の面積に有意差がでるかを考えるとうわりやすいと習いました。図 2 c - e も別の例の多重検定です³⁾。図 2 c ではマウスの 44 - 50 時間目の体重の変化にのみ関心があることを説明して、結果の有意差を示しました。棒グラフなどいろんな表現がありますが、最もわかりやすいものを選択してください。

4. 相関correlation：2つの連続変数の関連性をみて、相関があれば回帰直線式 $y = A * x + B$ を出します (図 3 a)。臨床では肝切除前の ICG15分停滞率の補正值を出す際に用いていま

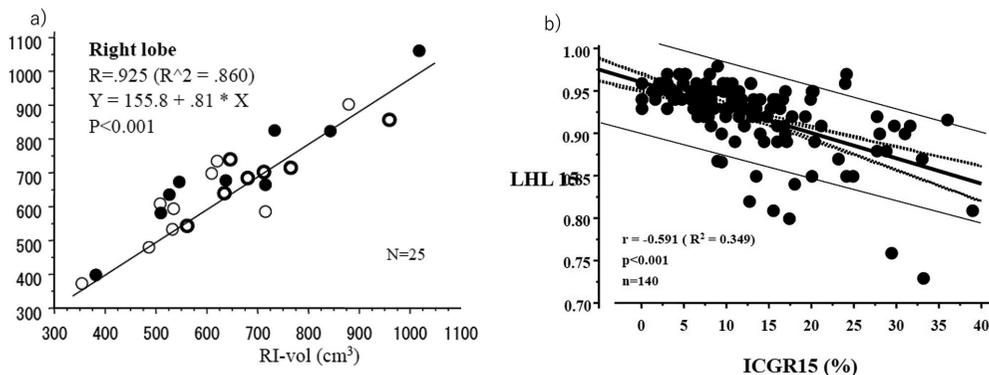


図3. 連続変数の相関分布と直線回帰.

す。逆相関の場合 r 値はマイナスになりますが、二乗した r^2 で示す場合もあります。図3 b のように99%信頼区間, 95%信頼区間をだせます。新しい肝機能指数と従来の検査値との相関についてですが95%信頼区間を外れたものを、逆に背景を解析して新しい指数の問題点を定義してacceptされました。カテゴリー変数の相関では、Kendallのランク相関など分布する濃度でみるものもあります。

* p 値 (ピーチ) は魔法の言葉ですね。前述の様に検定testの指標でprobability value, 偶然起こりうるデータの差を表し, 小さいほど偶然起こりうる確率が減り信頼性が増します。 p 値はよく5%とか1%未満とあたかも決まっているかのように思われていますが, ある統計者の意見では, 研究者の仮定によって p 値はどの値にでも設定できるそうです (10%未満, 20%未満のように)。あらかじめ有意水準を決め, 結果がそれより小さければ差がある (有意) と判断されます。一般論からいけば $p < 0.05$ で有意としていいのでしょうか。また結果の表で p 値が0.05以上の場合にns (not significant) とし, 有意な場合だけ数値を書くひとがいますが厳禁です。すべての p 値は表記しなければ読者はどの程度の傾向か判断できません。人によって $p=0.2$ や 0.3 でも何か差があるのではないかと考えます。差があるという仮説とないという仮説のどちらかを選べば片側, 両方とも p 値を足したものが両側検定ですが, 通常両側検定を用います。ちなみにメディアの世論調査の統計結果の解

釈も気を付けて理解する必要があると思います。

5. 回帰regression: 良く使われているのはある従属要因の因子を明らかにするための多変量解析で用いられていますが, Logistic解析が多いようです (若い頃, 物流のlogisticsと意味を間違えていました!)。ここで重要なのは元データで, 独立変数を2値化することです。事象を1とするか0とするか, 考えておかないと逆数が出てきます。連続変数などを要領よく2値化するにはexcelソフトのformula機能が一番だなといつも思います。連続変数のカットオフ値では安易に平均値を用いたりせず, Receiver Operating Characteristic analysis (ROC) 解析で適正な基準を設定するようにしましょう。SPSSでは簡単です。また, 因子には交互作用・交絡作用interactionがある因子は一緒に解析しないようにしましょう。例えば癌の進行度stageとT因子や腫瘍径などが単変量で差があったとします。進行度は生存率に差が出やすいように使われているので強力な因子です。そこに関わる他の因子を同時に解析しても結果が出なくなります。いくつかの因子からなるスコアや分類は解析しないほうがいいと思います。SPSSだと図4 a のように出てきます。ひとまず p 値が0.05未満とするとヒアルロン酸値HAが除かれます。Areaが最も多きものが正診率が高いと考えられます。Odds比もできますから次に有意差のある因子で減少法を用いてmodelをいくつか作成し, 残ったもので回帰式から予測式を作り

a) Area under the receiver-operator curve analysis between fibrotic markers or the Vs by ARI and histological cirrhosis

Parameters	Area	Standard deviation	Significant possibility	95% confidence interval		Cut-off value	Sensitivity (%)	Specificity (%)
				Lower limit	Upper limit			
HA	0.621	0.096	0.210	0.433	0.809	105	64	53
FIB-4 index	0.870	0.065	<0.001	0.743	0.996	3	82	80
APRI	0.955	0.032	<0.001	0.000	1.000	0.2	91	65
Vs (right liver)	0.874	0.053	<0.001	0.771	0.977	1.40	89	58
Vs (left liver)	0.831	0.068	0.001	0.698	0.964	1.63	85	66

b)

Multiple Linear Regression Stepwise Method Output Using Data of 265 Patients without Biliary Obstruction

Model	Unstandardized coefficients		Standardized coefficient		t value	Significance (P value)
	β	Standard error	β			
Model 1 (constant)	37.78	25.17			1.50	0.135
Total cholesterol	0.006	0.015	0.021		0.388	0.691
HH 15*100	0.31	0.11	0.255		20.79	0.006
LHL15*100	-0.50	0.22	-0.221		-20.28	0.023
AST	0.028	0.037	0.085		0.76	0.45
ALT	-0.029	0.028	-0.116		-10.12	0.27
Platelet count	0.073	0.052	0.076		10.42	0.16
Prothrombin activity	-0.008	0.045	-0.010		-0.178	0.86
Hyaluronic acid	0.021	0.005	0.271		40.12	<0.001
Model 2 (constant)	410.41	240.67			10.679	0.094
HH 15*100	0.276	0.104	0.230		20.648	0.009
LHL15*100	-0.501	0.213	-0.222		-20.35	0.020
Hyaluronic acid	0.020	0.005	0.263		40.135	<0.001

Dependent variable: step 6 total.

$$Y \text{ (converted ICGR15)} = 0.02 \times \text{hyaluronic acid level (ng/mL)} + 0.276 \times (\text{HH15} \times 100) - 0.501 \times (\text{LHL15} \times 100) + 41.41$$

図4. ロジスティック回帰解析から予測式へ。

ます(図4b)。後ろ向きデータはtraining set ですので、次にこれを使って前向きコホート・介入試験でvalidation set analysisすることでこの研究の目的が達成されると思います^{4,5)}。

6. 生存解析：よく治療後の予後を重要視する外科領域では生存解析は必須ですね。生存か、死亡かというイベントの有無、様々なドロップアウト中断による打ち切りがあります。また群間の治療が行われた時期の違いがある場合は、最初に行われた治療開始のスタート日と、その後の治療開始日との期間の差を算出し、四分位点用いて群間の治療開始日の差を評価します。また治療後の観察期間の影響(術後は最後に行われた治療から12-24か月のフォローが必要)などが問題とされます。通常Kaplan-Meyer法で累積生存率を計算し、生存確率による生存曲線を評価します。誰かが死亡するまでの患者をフォローしていると考えます。○年生存率、50%生存期間・中央値などで評価し、全期間を追跡して差を検定するには一般にLog-rank試験が用いられます。図5のように解析ソフトで図に違いがあります。私は統計ソフトFisher[®]が好きでずっと使っています。群別の生存の層別化discriminationはLog-rankや中間での差をみるにはWilcoxon試験を用いて比較します。生存解析で有意差がある予後因子が絞られたら、

次は多変量解析で指数回帰、対数正規回帰、正規回帰、パラメタ推定値、パラメタ分散/共分散行列、層比分析を基にCox比例ハザード回帰が良く用いられます。やはり独立変数は交絡因子の排除、2値化して、イベントも1とするかどうか検討します。図6aのように単変量と多変量の結果を並べてリスク比で示します。まだ今のようなソフトがない時代、私の先輩がexcelに β 値とSEをいれれば計算できる式を作成され(図6b)私に提供して頂きました。だいたい使用したので感謝するばかりです。図6cはいろんな予後分類でどの層別化がいいかという前述の赤池情報量基準AICによる結果です。AICが小さいほど層別化が良いとされています⁶⁾。

7. 多変量解析：多数の既存のAに対する新たな少数Bを比較する場合最近では3つほど解析方法があるように思います。1) データをまとめて昔ながらの多変量解析を行う場合、手法は簡単ですがBが少なすぎたり背景や時代が違っていると真の可能性が減ってしまいます。2) 前向きに行うランダム化比較試験は信頼性が高いが、データ収集と結果がでるまでに時間がかかります。最近では3) 傾向(Propensity)スコアマッチングが良く用いられるようになり、後ろ向きコホートデータで共変量(独立変数、背景因子)を概ね一致して揃えることができます。

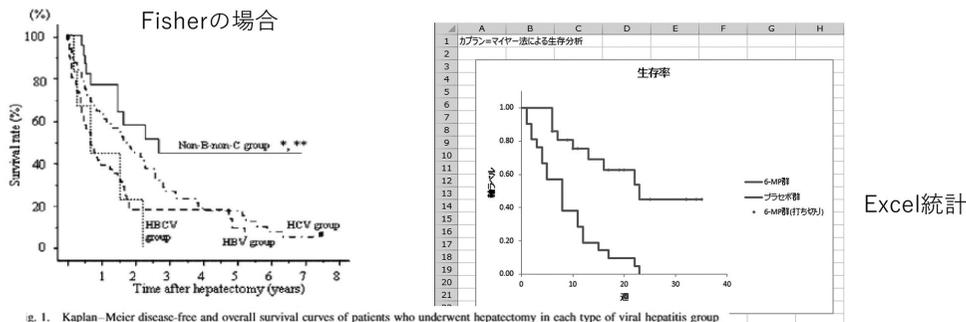


Fig. 1. Kaplan-Meier disease-free and overall survival curves of patients who underwent hepatectomy in each type of viral hepatitis group [SP: mean survival period. * $P=0.039$ versus HCV group. ** $P=0.001$ versus HBV group. *** $P=0.013$ versus HBCV group. $P=0.047$ versus BCV group. $P=0.019$ versus HBCV group. $P=0.025$ versus HBCV group. [Color figure can be viewed in the online issue, available at www.terscience.wiley.com.]

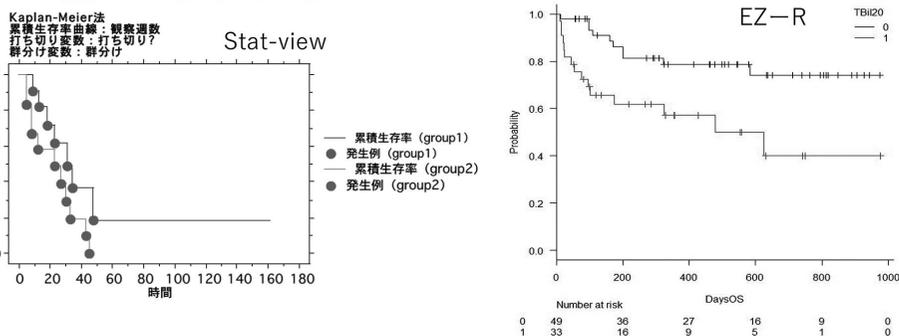


図5. 生存確率曲線のいろいろ.

a)

Variable	Tumor-free survival			
	Univariate analysis		Multivariate analysis	
	RR* (95% CI)	p value	RR (95% CI)	p value
Number of tumors (solitary/multiple)	1.85 (1.38-2.46)	<0.001	1.29 (0.89-1.87)	0.18
Size of tumor				
<2 cm				
2-5 cm	1.73 (1.30-2.30)	<0.001	1.67 (1.08-2.56)	0.020
≥5 cm	1.45 (1.07-2.43)	<0.001	1.53 (1.06-2.21)	0.025
Japan TNM classification				
1				
2	1.47 (1.14-1.89)	0.003	1.67 (1.12-2.44)	0.011
3	1.75 (1.31-2.33)	<0.001	0.88 (0.54-1.45)	0.62
4a	2.80 (1.87-4.19)	<0.001	1.26 (0.69-2.33)	0.35
Liver cirrhosis				
No				
Yes	1.21 (0.94-1.55)	0.14		
Background liver				
Non-viral				
Viral	1.70 (0.98-2.58)	0.064		

b)

Cox regression analysis							
Variable	β	標準偏差	標準化係数	Wald	Risk ratio	95%CI	p
0.13929	0.381450	0.37	0.13	1.15	0.54	0.243	0.286
0.40289	0.269410	1.11	1.23	1.51	0.75	0.311	0.385
0.83423	0.614124	1.36	1.85	2.30	0.69	0.288	0.001
0.56580	0.434127	0.81	0.57	1.50	0.56	0.305	0.207
-0.22520	0.483927	-0.47	0.22	0.79	0.80	0.207	0.378
0.10817	0.268202	0.30	0.09	1.12	0.54	0.230	0.629
0.54884	0.289221	0.79	0.62	1.78	0.44	0.179	0.001
-1.14927	0.258202	-2.14	4.59	0.22	0.11	0.81	0.001
0.93795	0.433163	2.27	5.16	2.71	1.15	0.642	0.001
		#DV/#	#DV/#				
0.008535	0.318655	0.05	0.30	1.01	0.54	1.08	0.930
0.546128	0.297742	1.93	3.96	1.73	0.96	0.309	0.001
0.400273	0.463706	0.85	3.74	1.50	0.60	0.274	0.001
0.178461	0.257194	0.50	0.25	1.20	0.59	0.241	0.629
-0.102695	0.240011	-0.30	0.09	0.90	0.46	0.176	0.001
-0.144840	0.266173	-0.47	0.22	0.67	0.47	1.58	0.412
-0.512755	0.244013	-0.80	0.63	0.60	0.17	2.12	0.001
-0.238213	0.266704	-0.32	0.85	0.71	0.35	1.46	0.642
-0.183442	0.265075	-0.52	0.20	0.62	0.41	1.68	0.642

Table 4. The Akaike Information Criteria (AIC) statistic was used to compare different models influencing overall survival of HCC patients and show the most appropriate model among staging systems

	AIC
Japan TNM stage	637.4
CLIP score	636.5
Modified CLIP score	634.8
JIS score	635.8
Modified JIS score	634.3

図6. 生存率の回帰解析.

RCTを行った同様の効果が期待され、記録された共変量を何度でも組み直し解析でき試行錯誤できるとされますが、やはり後ろ向きデータに違いはありません。期待する共変量が欠落しマッチングで偏った比較(バイアス)が生じる

ことが問題視されます。十分なBの数が必要なのでビッグデータの解析には適しているようです。外科領域で、昔と今の治療法を比較する場合は、昔の手技にはなかなか戻れませんし、施設間で手技が違うときにマッチングさせるには

傾向スコアマッチングは適しているとされています。

III. さいごに

今回は、あくまでも数学が苦手だった統計には今でも素人である一般外科医が30年かけて会得した私見でした。大きく間違ったことはないと思いますが、若い人たちの統計入門の参考にしていただくと幸いです。まとめます。1. まず臨床して研究する。2. 一つ一つ慌てず勉強する。3. 統計ソフトは自分に合ったものに絞る (いくつも買わない!)。4. 聞く? 調べる? 学びに行く? 5. 統計家とケンカしない。6. 有意差ある結果が出なくても意味がある。それでもなんとか論文にできますよ! 7. やはり、体力、不眠と気力・精神力が必要(働き改革は上、労働時間には入りません)。このところ様々なオンライン会議の手法で研究会を行っています。劇場型になって集まった感じが得られたら悪くない気がします。またVol.1で述べたようにパワーポイントのスライドショー機能で音声を入れて時間内にきちんと終了する学会発表形式も出てきました。またZoomのβ版機能として、パワーポイントを共有する際に、自分の画面を発表画面に重ねる機能が加えられ、講演している感覚が付加されました(2020年9月5日現在)。どこまでオンラインや拡張現実(MR)が進歩していくのでしょうか? 台風10号の迫る中楽しんでます。

第5回は“私の考える臨床医の研究費獲得”を掲載予定です。

著者のCOI開示：本論文発表内容に関連して特に申告なし。

文 献

- 1) Nanashima A, Yamaguchi H, Tsuji T, et al. Physiologic stress responses to laparoscopic cholecystectomy. A comparison of the gasless and pneumoperitoneal procedures. Surg Endosc 1998 ; 12 : 1381-5.
- 2) Nanashima A, Sumida Y, Tominaga T, et al. Significance of liver hanging maneuver for anatomical hepatectomy in patients with a large hepatocellular carcinoma compressing intrahepatic vasculatures. Ann Hepatobiliary Pancreat Surg 2017 ; 21 : 188-93.
- 3) Nanashima A, Kodama T, Murakami G, et al. Effects of Compounded Human Ghrelin in a Mouse Model of Pancreatic Carcinoma. Journal of Pancreas 2016 ; 17 : 53-59.
- 4) Nanashima A, Sakamoto A, Sakamoto I, et al. Usefulness of evaluating hepatic elasticity using artificial acoustic radiation force ultrasonography before hepatectomy. Hepatol Res 2014 ; 44 : 1308-19.
- 5) Nanashima A, Abo T, Arai J, et al. Functional liver reserve parameters predictive for posthepatectomy complications. J Surg Res 2013 ; 185 : 127-35.
- 6) Nanashima A, Sumida Y, Abo T, et al. Modified Japan Integrated Staging is currently the best available staging system for hepatocellular carcinoma patients who have undergone hepatectomy. J Gastroenterol 2006 ; 41 : 250-6.