

厚生労働科学研究費補助金
医療技術評価総合研究事業

国民の視点に基づく適切な診断過程の確立に関する研究

平成 17・18 年度 総合研究報告書

平成 19 年(2007)年 3 月

主任研究者 峯 徹 哉

目 次

I. 班員構成	1
II. 班会議 開催日程一覧（平成 17 年度・平 18 年度）	2
III. 統括研究報告	
国民の視点に基づく適切な診断過程の確立	3
主任研究者 峯 徹哉	
IV. 分担・共同研究報告	
1. 患者の視点から見た適切な消化器疾患診断プロセスの研究	9
事前確立調査からのガイドライン作成手順まで	
神奈川歯科大学 教授 森實 敏夫	
2. プロトンポンプ阻害薬有効例をゴールドスタンダードとした胃食道逆流症患者の解析(最終報)	29
横浜市立大学 稲森正彦	
3. 肝生検をゴールドスタンダードとした NASH 診断法の開発	33
横浜市立大学 米田 正人	
4. 慢性肝疾患における肝生検の意義	41
横浜市立大学 斉藤 聡	
5. 肝細胞癌診断における各種検査の感度・特異度について	47
東海大学 加川 建弘	
v. 研究成果に対する一覧表	53

I. 班員構成

厚生労働科学研究費補助 医療技術評価総合研究事業
「国民の視点に基づく適切な診断過程の確立に関する研究」

主任研究者	峯 徹哉	東海大学医学部消化器内科
分担研究者	森實 敏夫	神奈川歯科大学内科
	中島 淳	横浜市立大学消化器内科
	大川 伸一	神奈川県立がんセンター消化器内科
研究協力者	松橋 信行	NTT 東日本関東病院消化器内科
	玉井 拙夫	神奈川県津久井保健福祉事務所
	桐越 博之	横浜市立大学消化器内科
	斉藤 聡	横浜市立大学消化器内科
	島田 英雄	東海大学医学部消化器外科
	米田 正人	横浜市立大学消化器内科
	加川 建弘	東海大学医学部消化器内科
	柴田 実	NTT 東日本関東病院消化器内科
	松嶋 成志	東海大学医学部消化器内科

Ⅱ. 班会議 開催日程一覧

<平成17年度>

第1回班会議	平成17年9月17日	神奈川県立神奈川歯科大学	横浜研修センター
第2回班会議	平成17年10月31日	神奈川県立神奈川歯科大学	横浜研修センター
第3回班会議	平成17年12月19日	神奈川県立神奈川歯科大学	横浜研修センター
第4回報告会	平成18年2月20日	持田製薬本社	ルークホール

<平成18年度>

第1回班会議	平成18年6月26日	神奈川県立神奈川歯科大学	横浜研修センター
第2回班会議	平成18年9月25日	神奈川県立神奈川歯科大学	横浜研修センター
第3回班会議	平成18年11月27日	神奈川県立神奈川歯科大学	横浜研修センター
第4回班会議	平成19年1月29日	神奈川県立神奈川歯科大学	横浜研修センター
第5回報告会	平成19年2月19日	持田製薬本社	ルークホール

Ⅲ. 統括研究報告書

厚生労働科学研究費補助金（医療技術評価総合研究事業）

国民の視点に基づく適切な診断過程の確立に関する研究

統括研究報告書

主任研究者 峯 徹哉

東海大学医学部消化器内科

教授

研究要旨：まず、分担研究者森實敏夫が胃食道逆流症（GERD）および C 型慢性肝炎患者を対象に診断プロセスの適切性に関するアンケート調査を行なった。その結果、診断確定までの診断の進め方、時間について適切と評価する患者の割合は疾患によって異なることが明らかになった。更に、診断に関する研究のエビデンスレベルに分類法も新たに設定した。しかし世間では消化器疾患の診断プロセスの最初のステップを標準化されているものは殆んどない。現在この点を中心とした消化器疾患のガイドライン作成にも着手した。大川らは膵癌患者における診断および治療導入までの流れについて検討を行ない、膵癌患者の診断過程とステージ分類との関係を解析した。胃食道逆流症については研究協力者の稲森正彦らが行った。稲森らは胸やけという主訴の患者のうちプロトンポンプ阻害薬が効果がある群を抽出するのに有効な方法を見つけ無駄な検査を省くことが出来ないか検討した。胸やけを主訴とする患者の 6 割がプロトンポンプ阻害薬投与のみで十分であった。しかも病納期間、PPI テスト、内視鏡所見との関連も認めた。肝臓については肝生検を gold standard にした NASH 診断法の開発について研究協力者の米田らが検討を行なった。NASH(non-alcoholic steatohepatitis)とはメタボリックシンドロームとの直接的な関与もありその診断方法は現在一定していない。NASH を診断する際に高感度 CRP、アディポネクチン、4 型コラーゲン 7 S が相関を認めることが明らかとなった。更に B 型・C 型慢性肝炎患者の肝生検を gold standard として血小板数等との比較を研究協力者の齊藤らが行なった。C 型慢性肝炎患者においては肝線維化の指標としてやはり血小板数が参考になることが確認された。しかし、B 型慢性肝炎疾患においては血小板数よりむしろプロトロンビン時間やコリンエステラーゼによく相関した。次に肝細胞癌診断における各種検査の感度・特異度について研究協力者の加川らが検討を行なった。病理学的に確認された肝細胞癌に対し検討を行なったが、US、CT は PIVKA-II よりも明らかに優れていた。これらの臨床結果は評価して如何に診断過程を迅速に行なうかについて検討すると同時にその診断のガイドラインの作成に着手した。

A. 研究目的

一般国民が医療機関を受診した際に期待する医療サービスの中で、診断を確定するために行われる問診、診察、検査といった診断確定にいたる過程と、EBM に基づく適切な診断過程、さらに、現行の医療機関でおこなわれている診断過程の実態は必ずしも一致していない。実際には、EBM に基づいた診断過程が確定していない疾患が多く、実際の診断過程では多くの不必要な検査が行われているのが実態である。特に一般健康人を対象に保健所等で行われるスクリーニング検査は、その有用性が証明されていないものがほとんどであり、またその有用性を証明するためのEBM に基づいた研究もほとんど行われていない。また、診断のために行われる検査は検尿のように患者への負担および費用ともわずかな検査から、生検や血管造影のように、これらの負担が大きいだけでなく、リスクも伴う検査までさまざまである。一般国民の視点からは、検査を受けることによって、診断が100%確定するかの様に認識されているが、EBM に基づいた診断に関する研究はほとんど行われていない現状で、それは非科学的な認識である。消化器領域では、内視鏡をはじめとして血管造影など、信憑性の高い検査が多く行われるため、EBM に基づく適切な診断過程の確立は特に必要とされる。国民が医療機関を受診した際に、診断を確定する過程で行われるさまざまな検査に対して、その必要性を納得しているか、また、それぞれの検査が患者に受容できるものであるかはきわめて重要である。さらに、不必要な検査が漫然と行われることは、医療資源の浪費

でもある。診断についても不必要な検査が多い場合、現在はDPCの導入も殆んど病院で行なわれているので診断に時間がかかり、進行が進んでいる場合も多い。本研究の目的はさまざまな事前確率のさまざまな疾患の診断過程で、一般国民が納得し、高い満足度が得られ、なおかつ、EBM に基づいた適切なものを明らかにすることである。医療は従来、治療が優先され診断は必要性が不明な検査も含め、総花的に行われ、それが許容されてきた。その一因はEBM に基づく、論理的診断体系に関する研究がほとんど行われてこなかったことにある。さまざまな事前確率のさまざまな疾患の最も適切な診断過程を明らかにすることが必要とされている。本研究で成果が得られれば、さまざまな疾患において、根拠に基づいた推奨すべき診断過程が明らかになるであろう。事前確率については消化器専門医師と一般医師とのギャップがあるので年齢とか性別、履歴等を含んだ上で症状から疾患の診断を確立をもっと高めるようなプログラムを組むことを行なった。昨年度の研究から、現在消化器ガイドラインの作成も準備段階にはいつている。

B. 研究方法

現在、後向き研究であるが各々の疾患においてgold standardを決め、臨床検査の感度及び特異度について検討した。診断プロセスについてはアンケート調査を行ない、その診断の進め方を検討した。消化器疾患のガイドラインについては膨大な文献からエビデンスレベルを決定してその作成を行なっている。

C. 研究結果

まず、分担研究者森實敏夫が診断プロセスの適切性に関するアンケート調査を行なった。主にGERD(胃食道逆流症)と肝疾患についてとりあげた。その結果、両疾患によって診断確定までの診断の進め方、時間について適切と評価する割合は疾患によって異なっていた。更に消化器診断ガイドライン作成を考えており、エビデンスレベル分類法も新たに設定した。大川らは膵癌患者の診断過程とステージ分類との関係を解析したが、診断過程に時間がかかった群とかがかからなかった群ではステージ分類には差がなかった。次に研究協力者の稲森らが最近、疾患の数の増加している胃食道逆流症の患者について解析した。胸やけという主訴の患者うちプロトンポンプ阻害薬が効果がある群を抽出するのに有効な方法を見つけ無駄な検査を省くことが出来ないかを検討した。胸やけを主訴とする患者の6割がプロトンポンプ阻害薬投与のみで十分であった。しかも病悩期間、PPIテスト、内視鏡所見との関連も認めた。肝臓については肝生検をgold standardにしたNASH診断法の開発について研究協力者の米田らが検討を行なった。NASH(non-alcoholic steatohepatitis)とはメタボリックシンドロームとの直接的な関与もありその診断方法は現在一定していない。NASHを診断する際にHLD コレステロール、血清フェリチン、高感度CRP、4型コラーゲン7Sが相関を認めることが明らかになった。今回、アディポネクチンが関わっていることが明らかとなった。

更にB型・C型慢性肝炎患者の肝生検をgold standardとして血小板数等との比較を研究協力者の斉藤らが行なった。C型慢性肝炎患

者においては肝線維化の指標としてやはり血小板数が参考になることが確認された。しかし、B型慢性肝炎疾患においては血小板数よりむしろプロトロンビン時間やコリンエステラーゼによく相関した。次に肝細胞癌診断における各種検査の感度・特異度について研究協力者の加川らが検討を行なった。病理学的に確認された肝細胞癌に対し検討を行なったが、US、CTは30mm以下では80%以上の感度・特異度を有していた。MRIの感度は70%ぐらいであった。AFP、PIVKA-IIについては画像診断より診断価値が劣っていた。

D. 考察

それぞれの疾患の診断過程について検討をさらに続けることによって新たな診断ガイドラインができると思われる。胃食道逆流症については患者に対する負担を減らすようなPPIテストの有効性をアピールしてその診断方法の確立を検討したい。NASHについてはその存在の意義付けと同時に診断方法の確立と、いかに肝硬変への阻止を行なうかが重要となる。肝硬変についてはB型・C型による肝硬変の線維化の違いおよび血小板の有用性と限界をアピールしていきたい。肝細胞癌の診断について α Feto、PIVKA-II、US、CT、MRIの陽性率をもとにどのような診断体系がベストなのかを考えていきたい。膵癌についても症状毎の診断過程の解析を行なっていきたい。

E. 結論

(1)様々な症状に基づく疾患確立の値は専門医がどうかによってかなりのばらつきがあるので消化器専門医のさまざまな症状に基

づく疾患確率の代表値を明らかにしていくことが消化器診断プロセスには重要である。更にEBM等を付け加えることによって標準化できる可能性があると思われるので、この作業を続けていく。

(2)胃食道逆流症については標準GERD予測の陽性所見になりうる様々な因子を精査するような前向き検討を検討中である。

(3)NASHについてはその重要な検査をしてHDLコレステロール、血清フェリチン、高感度CRP、4型コラーゲン7Sが診断に重要な検査所見であると思われるのでこれらの測定値とNASHとの関係を明らかにすることができると思われる。

(4)B型・C型肝炎の診断についてC型については血小板と明らかな関係があることが判明した。B型についてはプロトンビン時間やコリンエステラーゼが重要でありその因果関係についてももっと掘り下げるつもりである。

(5)HCC（肝細胞癌）については α Feto,PIVKA-II,CT,超音波,MRIの利点が見られているが各々の診断確率がどうなのか、どのような診断手順を踏めばいいのかが明らかとなる。

(6)膵癌については紹介までのかかった時間と予後についても検討中である。

F. 健康危険情報

全て保険診療内で行なっており、研究結果が人体の健康に及ぼす可能性はない。

G. 研究発表

1. Mine T. Is post-endoscopic retrograde cholangiopancreatography pancreatitis the same as acute clinical pancreatitis? J

Gastroenterol. 2007 ;42(3):265-6.

2. Mine T. Is hepatic vena cava disease an endemic type of the Budd-Chiari syndrome?Hepatol Res. 2007 ;37(3):170-1.

3. Mochizuki K, Kagawa T, Numari A, Harris MJ, Itoh J, Watanabe N, Mine T, Arias IM. Two N-linked glycans are required to maintain the transport activity of the bile salt export pump (ABCB11) in MDCK II cells. Am J Physiol Gastrointest Liver Physiol. 2007 ;292(3):G818-28.

4. Mine T. Eupatilin has an antiapoptotic action on hepatocytes, in contrast to apoptotic actions on other cells.J Gastroenterol. 2006 ;41(8):818-9.

5. Mine T. Effect of polycarbophil Ca on IBS. Nippon Rinsho. 2006;64(8):1483-6.

6. Watanabe N, Takashimizu S, Shiraishi K, Kagawa T, Nishizaki Y, Mine T, Akatsuka A, Selmi C, Gershwin ME. Primary biliary cirrhosis with multinucleated hepatocellular giant cells: implications for pathogenesis of primary biliary cirrhosis. Eur J Gastroenterol Hepatol. 2006 ;18(9):1023-7.

7. Mine T. Ductal plate malformation, congenital hepatic fibrosis and Calori's disease are the same entities or not? Hepatol Res. 2006 ;35(2):77-8.

8. Sakamoto C, Sugano K, Ota S, Sakaki N, Takahashi S, Yoshida Y, Tsukui T, Osawa H, Sakurai Y, Yoshino J, Mizokami Y, Mine T, Arakawa T, Kuwayama H, Saigenji K, Yakabi K, Chiba T, Shimosegawa T, Sheehan JE, Perez-Gutthann S, Yamaguchi T,

- Kaufman DW, Sato T, Kubota K, Terano A. Case-control study on the association of upper gastrointestinal bleeding and nonsteroidal anti-inflammatory drugs in Japan. *Eur J Clin Pharmacol.* 2006 ;62(9):765-72.
9. Mine T. Influence of melatonin and acetylsalicylic acid on lipid peroxidation and antioxidant enzyme activities in gastric mucosa. *J Gastroenterol.* 2006 ;41(5):507-8.
10. Mine T. What is the role of lamellae in solubilizing of biliary cholesterol in hepatic bile? *Hepatol Res.* 2006 ;35(3):151. Epub 2006 Jun 21.
11. Mine T. The role of magnifying endoscopy in the diagnosis of early gastric carcinoma. *J Gastroenterol.* 2006;41(4):397-8.
12. Matsuhisa T, Kawai T, Masaoka T, Suzuki H, Ito M, Kawamura Y, Tokunaga K, Suzuki M, Mine T, Takahashi S, Sakaki N. Efficacy of metronidazole as second-line drug for the treatment of *Helicobacter pylori* Infection in the Japanese population: a multicenter study in the Tokyo Metropolitan Area. *Helicobacter.* 2006;11(3):152-8. Erratum in: *Helicobacter.* 2006 ;11(4):306.
13. Matsushima M, Suzuki T, Kurumada T, Watanabe S, Watanabe K, Kobayashi K, Deguchi R, Masui A, Takagi A, Shirai T, Muraoka H, Kobayashi I, Mine T. Tetracycline, metronidazole and amoxicillin-metronidazole combinations in proton pump inhibitor-based triple therapies are equally effective as alternative therapies against *Helicobacter pylori* infection. *J Gastroenterol Hepatol.* 2006 ;21(1 Pt 2):232-6.
14. Takahashi H, Kagawa T, Kobayashi K, Hirabayashi H, Yui M, Begum L, Mine T, Takagi S, Saheki T, Shinohara Y. A case of adult-onset type II citrullinemia -deterioration of clinical course after infusion of hyperosmotic and high sugar solutions. *MedSciMonit.* 2006 ;12(2):CS13-5. Epub 2006 Jan 26.
15. 峯徹哉. 機能性ディスペプシアに対するクエン酸モサプリドとテプレノンの前向き比較試験. *消化器科* 2006;43(6):508-511
16. 峯徹哉. 色素内視鏡を 30 秒で終わらせるために. *消化器内視鏡* 2006;18(12):1804-1808
17. 鈴木孝良, 松嶋成志, 峯徹哉. 便秘の診断と一般的な治療 代表的な便秘薬と対症療法法の注意点. *Medicina* 2006;43(13):2038-2040
18. 加川建弘, 峯徹哉. 【一般医、消化器科医が知っておきたい肝胆膵疾患診療のエビデンス】 肝疾患の診断と治療 肝細胞癌診断における腫瘍マーカーと画像診断. *EBM ジャーナル* 2006;8(1):22-27
19. 川口義明, 峯徹哉. 【一般医、消化器科医が知っておきたい肝胆膵疾患診療のエビデンス】 胆道疾患の診断と治療 胆嚢結石・総胆管結石併存例の診断治療方針. *EBM ジャーナル* 2006;8(1):48-53
20. 峯徹哉. 【胆道疾患治療の現況】 胆石の治療 胆嚢結石の治療選択. *医学と薬学* 2006;56(5):649-651
21. 峯徹哉. 内視鏡専門医に必要な基礎知識

内視鏡治療時に使用する抗凝固薬、抗血小板薬.消化器内視鏡 2006;18(11)1767-1770

22. 峯徹哉.【ERCP 後膵炎とその予防】

ERCP 後膵炎.胆と膵 2006;27(8)525-528

23. 峯徹哉.【過敏性腸症候群(IBS) 基礎・臨床研究の新展開】 臨床研究 治療法 薬物療法 ポリカルボフィルカルシウム.日本臨床 2006;64(8)1483-1486

24. 峯徹哉.【消化器心身症 update】 肝炎と心身医療.心療内科 2006;10(4)245-247

25. 峯徹哉, 川口義明【EST30年】 胆管癌診断における EST 細胞診,生検を含む.消化器内視鏡 2006;18(6)993-997

26. 峯徹哉.偶発症と合併症を使い分ける.消化器内視鏡 2006;18(5)693

27. 長田成彦, 津田道雄, 大久保朋一, 三澤さおり, 白井孝之, 峯徹哉, 白石武昌.過体重者に対する低カロリー食の体重等への影響と血液マーカーの変動について.肥満研究 2006;12(1)54-64

28. 峯徹哉.内視鏡と抗凝固薬,抗血小板薬使用における問題点と症例報告.治療学 2006;40(3)333-334

29. 峯徹哉.【消化器疾患ガイドライン 最新の診療指針】 胆・膵疾患 急性胆管炎.今月の治療 2006;13 臨増 S197-S199

30. 松嶋成志, 鈴木孝良, 白井孝之, 峯徹哉.【高齢者消化器疾患の治療指針】高齢者出血性消化性潰瘍の治療指針.消化器科 2006;42(1)8-12

31. 峯徹哉, 寺岡宏倫, 斉藤哲朗.【肝門部胆管癌に挑む】 肝門部胆管癌の診断における ERCP,IDUS,経口胆道鏡の役割.消化器内視鏡 2006;18(1)57-60

国民の視点に基づく適切な診断過程の確立に関する研究

「患者の視点から見た適切な消化器疾患診断プロセスの研究：事前確率調査からガイドライン
作成手順まで」

分担研究者 森實敏夫

神奈川歯科大学内科

教授

研究要旨：消化器疾患の主な症状に基づいてそれぞれ想定される疾患の主観的な疾患確率について、調査を行った。その結果、症状に基づく疾患確率の値は、医師によりおおきなばらつきがあり、消化器専門医とそれ以外で差が認められる場合があった。消化器専門医の症状に基づく疾患確率の代表値を明らかにすることは、消化器疾患診断プロセスの最初のステップを標準化するために必要であると考えられた。

消化器疾患の診断に関する文献検索をPubMedで行い、さまざまな検索式の結果を比較検討した（2006年3月24日実施）。diagnosisとdiagnostic procedureの両MeSH TermをORで組み合わせ、診断能の指標を表す語句であるsensitivity、specificity、likelihood ratio、roc、receiver operatingを組み合わせた検索式で19,183件の文献が引き出された。さらに、これにclinical trial[pt]を組み合わせると、2,000件、ランダム化比較試験とメタアナリシスに限定すると846件、メタアナリシスに限定すると67件であった。

胃食道逆流症およびC型慢性肝炎患者を対象に行った調査で、全体としての問診、診察、検査の進め方、および、診断確定までの時間について適切だったかどうかについてたずねた結果、いずれも1～5の5段階評価での中央値は前者では3.まあそう思う、後者では4.そう思うであった。患者の視点で見た場合、診断プロセスの適切さに関する評価は疾患により異なることが示された。

診断法を実施すべきかどうかの決断は、事前確率などさまざまな要因によってなされるが、実施することによって患者にとって意味のある、すなわち臨床的意義のあるアウトカムが改善することが証明されていれば、実施すべきと決断することが必然的となる。そこで診断法に関する研究の分類として、診断法実施という介入が行われた研究かどうか、次いでアウトカムが臨床的意義のあるものかどうか、さらに、ランダム化比較試験またはメタアナリシスカ、非ランダム化比較試験か、コホート研究あるいは症例対照研究かで分類する。一方で、介入が行われていない研究は診断能の解析が主題かどうかで分類する。以上から全体としてType IからVまで分類することが可能である。

新しい、エビデンスレベル分類法とそれに対応した推奨度分類を用いて、診断に関する診療ガイドラインの作成が可能と思われた。そのための、構造化抄録のフォームおよびクリニカルクエスションとそれに対する推奨のフォームを作成した。

研究協力者

柴田実 NTT 東日本関東病院消化器内科

A. 研究目的

平成 17 年度は、エビデンスに基づいた消化器疾患診断における事前確率を明らかにするための第一歩として、消化器系のいくつかの症状に関して、医師の主観的疾患確率(事前確率)を調査した。さらに、消化器疾患診療ガイドライン作成を視野に入れ、消化器疾患診断に関する文献検索法を確立するために、PubMed で消化器疾患診断に関する文献の調査を行った。

平成 18 年度は、医学的のみならず患者にとっても適切な消化器疾患診断プロセスの確立のために、消化器疾患の診断プロセスが患者の視点から見た場合に、適切と評価されているかどうかを、胃食道逆流症と C 型慢性肝炎を取りあげて調査した。また、診断に関する研究は治療法に関する研究とは異なり、ランダム化比較試験の施行が困難な場合が多く、エビデンスレベルの分類として特異的なものが必要である。そこで、新しいエビデンスレベル分類法を考案した。さらに、消化器疾患診断ガイドライン作成の方法論について、作成手順を検討した。

B. 研究方法

アンケート調査法

消化器疾患診断における事前確率についてはウェブフォームを用いたアンケート調査を施行した。単純化した例における想定疾患の疾患確率について Visual Analogue Scale (VAS)を用いて、疾患確率を%で解答を求めた。その値の分布、10パーセンタイル、90パーセンタイルおよび中央値を算出した。調査対象者は消化器専門医12名、その他の専門領域の医師5名、歯科医4名である。これら3群間の疾患確率の値の差はKruskal-Wallis法で解析した。

症状は、嚥下困難、胃の痛み、胸焼け、右季肋部痛、右下腹部痛、血便である。これらの症状と年齢を条件として50歳以上と50歳未満を組み合わせた。想定する疾患としては、食道癌、胃癌、消化性潰瘍、機能性ディスペプシア、胃食道逆流症、胆石症、虫垂炎、大腸癌である。

文献調査のための検索式

消化器疾患を表す語句として、MeSH Term には digestive system diseases と gastroenterology があり、これらの下位用語には肝疾患、胆道疾患、膵疾患などが含まれているので、この2つの語句を OR で結合した検索式を基本とした。さらに、diagnosis と diagnostic techniques and procedures が診断法に関する MeSH Term なので、これらを用いて検索式を作成した。

さらに、sensitivity、specificity、likelihood ratio、roc、receiver operating の語句を OR で結合し、診断法に関する論文の中で、少なくともタイトルあるいはアブストラクト、インデックス語句中に、これらの語句が含まれる文献が引き出されるようにした。

また、研究デザインについては、Publication Type で randomized controlled trial、controlled clinical trial、meta-analysis を指定した検索を行った。すなわち、エビデンスレベルの高いとされるメタアナリシスとランダム化比較試験に限定した検索も行った。

また、人に関する研究に限定するためには、MeSH Term である human を組み合わせ、言語は english[lang] OR japanese[lang]を組み合わせ英語と日本語に限定した。なお、検索を実施したのは 2006 年 3 月 24 日である。

診断プロセスの適切性に関する調査

診断が確定した時点で、初診からのプロセスを振り返り、患者がそれを適切と判断しているかどうかについて、アンケート調査を行った。対象は、横浜市立大学医学部消化器内科を 2006 年度に受診した、胃食道逆流症 (Gastro-Esophageal Reflux Disease ,GERD)40 例と C 型慢性肝炎 38 例である。なお、C 型慢性肝炎例は全例肝生検を受けていた

質問票

質問票では、項目 1.全体としての問診、診察、検査の進め方、項目 2.診断確定までの時間について適切だったと思うかをたずね、患者は 5 段階(1.まったくそう思わない、2.あまりそう思わない、3.まあそう思う、4.そう思う、5.強くそう思う)で評価した。また、受けた検査の中で、項目 3. しないですめばその方がよいと思う検査があったかどうか、項目 4. 受けた検査の中で、とくにつらかった検査があったかどうかについても質問し、あった場合は検査名を記入してもらった。最後に、自由に意見を書き込んでもらった。

診断に関する研究から得られる知見

診断に関する研究からは、1) 診断能、2) 適切なアウトカムに対して証明された臨床的有用性、の 2 種類の知見が得られる。そこで、介入として診断法の実施が行われた研究かどうか 2 つに分類するべきと考えた。介入が行われた研究については、臨床的な意義のあるアウトカムへの影響が解析されているかどうかさらに 2 つに分類し、介入が行われていない研究については、診断能の解析が主題かどうかで 2 つに分類することが適切と考えた。介入が行われた研究については、さらに研究デザインによって細分化することができる。

診断法が適用される被験者の事前確率による分類

被験者の想定される疾患の事前確率が低い場合は、健常者を対象として行われる健康診断や人間ドックのような場合で、スクリーニング (Screening) として分類される。たとえば、肝硬変のように肝細胞癌のリスクが高い被験者で、肝細胞癌の診断をするような場合には、事前確率が低～中であり、サーベイランス (Surveillance) として分類される。さらに、患者が何らかの症状を訴え、医療機関を受診するような場合には、事前確率が中から高であり、診断 (Diagnosis) と分類される。すなわち、疾患の平均的危険群 = 健常者、疾患の高危険群、疾患が想定される = 通常の診療の状況の 3 つに分類することができる。同じ検査法であっても、これら 3 つのレベルでは結果が得られた際の意義付けは異なり、いわゆる陽性的中率や陰性的中率が異なってくるので、臨床的な意義も異なってくる。

消化器疾患の診断法の診断能と適切なアウトカムに対して証明された臨床的有効性に関する論文を包括的に収集し、それらを分類し、まとめる。平成 17 年度の本班会議報告書で述べたごとく、PubMed 検索を行った。(digestive system diseases[mh] OR gastroenterology[mh]) AND (diagnosis[mh] OR diagnostic techniques and procedures) AND (sensitivity OR specificity OR likelihood ratio OR roc OR receiver operating) AND (randomized controlled trial[pt] OR controlled clinical trial[pt] OR meta-analysis[pt]) AND human[mh] AND (english[lang] OR japanese[lang]) および (digestive system diseases[mh] OR gastroenterology[mh]) AND (diagnosis[mh] OR diagnostic techniques and procedures) AND (sensitivity OR specificity OR likelihood ratio OR roc OR receiver operating) AND clinical trial[pt] AND human[mh] AND (english[lang] OR japanese[lang]) の 2 種類の検索式で重複を除いた 2,110 件の文献を得た。

一次選定

各班員が分担して、一次選定を行った。一次選定の基準は、1.消化器疾患患者が対象である、疾患群と対照群を調べている、2.標準的 Gold Standard で診断されている、3.Sensitivity, Specificity, Likelihood ratio, Odds ratio, ROC (receiver operating characteristics) analysis と AUC (Area under the curve)などの診断能の効果指標のデータが提供されている、である。なお、予後予測に関する論文は原則として除外した。

C. 研究結果

事前確率の調査

消化器専門医による症状別事前確率は表 1 に示すごとくであり、その値は広い範囲に分布していた。

たとえば、嚥下困難で食道癌を想定した場合、年齢が 50 歳以上の場合には、25%に対して、年齢が 50 歳未満の場合には 5%であり、年齢を考慮して、当然のことながら高齢の場合には悪性疾患の疾患確率を高く考えていることが分かる。しかし、最小値が 5%、最大値が 75%とかなりのばらつきが観察された。

また、消化器専門医とその他の専門領域の医師、歯科医師を比較すると、症状が胃の痛みで想定疾患が消化性潰瘍の場合、3 者で有意差が認められ、消化器専門医がより高い疾患確率を想定していることが分かった。その他、症状が嚥下困難で年齢が 50 歳未満で想定疾患が食道癌の場合にも、消化器専門医が他と比べより高い疾患確率を想定していた。その他にも 3 群でかなりの差が認められる症状・想定疾患があった(表 2)。

文献調査

(digestive system diseases[mh] OR gastroenterology[mh]) AND diagnosis[mh] AND (sensitivity OR specificity OR "likelihood ratio") AND human[mh] AND (english[lang] OR japanese[lang])と診断能の指標を表す語句を組み合わせると、18,684 件に減少した。また、ROC を組み合わせると、(digestive system diseases[mh] OR gastroenterology[mh]) AND diagnosis[mh] AND (roc OR "receiver operating") AND human[mh] AND (english[lang] OR japanese[lang])の検索式を用いると、1,236 件の文献が引き出された。従って、診断能の指標を表す語句をタイトル、アブストラクト、インデックス語句に含む文献はこれら 2 つの結果を合わせて、約 2 万件弱あることが分かった。

diagnosis の MeSH Term の代わりに diagnostic techniques and procedures の MeSH Term を用いると、(digestive system diseases[mh] OR gastroenterology[mh]) AND diagnostic techniques and procedures AND human[mh] AND (english[lang] OR japanese[lang])の検索式で、147,907 件の文献が引き出され、diagnosis の場合よりも、33%減少した。

これにさらに、診断能の指標を表す語句をすべて組み合わせると、(digestive system diseases[mh] OR gastroenterology[mh]) AND diagnostic techniques and procedures AND (sensitivity OR specificity OR

likelihood ratio OR roc OR receiver operating) AND human[mh] AND (english[lang] OR japanese[lang])で検索すると、12,879 件の文献が引き出された。当然のことながら、diagnosis を組み合わせた場合より、30%以上減少した。

研究デザインによる限定

ランダム化比較試験とメタアナリシスに限定し、(digestive system diseases[mh] OR gastroenterology[mh]) AND diagnostic techniques and procedures AND (sensitivity OR specificity OR likelihood ratio OR roc OR receiver operating) AND (randomized controlled trial[pt] OR controlled clinical trial[pt] OR meta-analysis[pt]) AND human[mh] AND (english[lang] OR japanese[lang])では、593 件と大幅に減少した。

さらに、メタアナリシスだけにすると、(digestive system diseases[mh] OR gastroenterology[mh]) AND diagnostic techniques and procedures AND (sensitivity OR specificity OR likelihood ratio OR roc OR receiver operating) AND meta-analysis[pt] AND human[mh] AND (english[lang] OR japanese[lang])では、43 件のみであった。

diagnosis と diagnostic procedure の両者を含む場合 diagnosis と diagnostic procedure の両者を OR で組み合わせ、診断能の指標を表す語句をすべて組み合わせると、(digestive system diseases[mh] OR gastroenterology[mh]) AND (diagnosis[mh] OR diagnostic techniques and procedures) AND (sensitivity OR specificity OR likelihood ratio OR roc OR receiver operating) AND human[mh] AND (english[lang] OR japanese[lang])で 19,183 件あった。

さらに、これに clinical trial[pt]を組み合わせると、2,000 件、ランダム化比較試験とメタアナリシスに限定すると 846 件、メタアナリシスに限定すると 67 件であった。

患者の視点から見た診断過程の適切性に関する調査

項目 1 については、C 型慢性肝炎では 4 のスコアがもっとも多く、GERD では 3 のスコアがもっとも多く、両群には差があった ($P < 0.0001$) (図 1)。

項目 2 についても、C 型慢性肝炎では 4 のスコアがもっとも多く、GERD では 3 のスコアがもっとも多く、両群には差があった ($P < 0.0001$) (図 2)。

項目 3 については、しないですめばその方がよい

と思う検査があると答えた割合はそれぞれ 13%と 15%で差はなかった(図 3)。

項目 4 については、とくにつらかった検査があると答えた割合は、それぞれ 24%と 58%で差が認められた ($P = 0.0024$)。内訳は、胃カメラがほとんどを占め、内視鏡検査をつらいと思っている患者が多いことが判明した(図 4)。

患者の意見からは、大学病院とその他の医療機関の医師の診断能力の差が存在すること、専門医と一般医で疾患の重要性に対する解釈が異なることがうかがわれた

診断に関する研究の分類

今回われわれが提唱する、診断に関する研究の分類は以下のごとくである。

Type I

臨床的意義のあるアウトカムに対する効果を解析した RCT または RCT のメタアナリシス

Type II

臨床的意義のあるアウトカムに対する効果を解析した非ランダム化比較試験

Type III

臨床的意義のあるアウトカムに対する効果を解析した症例対照研究、コホート研究またはそのメタアナリシス

Type IV

診断能の解析を目的とした臨床試験、横断研究など

Type V

RCT、コホート研究、症例対照研究などの研究に付随して診断法の診断能のデータが提供されているもの

すなわち、臨床的意義のあるアウトカムに対する効果を解析した研究を Type I から III に分類し、診断能に関する研究を Type IV と V に分類した。前者は研究デザインにより、後者は主題が診断能の解析かどうかでさらに分類するようにした(図 5)。

Type I, II, III の研究の分類基準について図 6、7、8 に示す。

さらに、Type IV, V の研究の分類基準について図 9、10 に示す。

これらの分類の基準となる項目は、Type I, II の場合は、研究デザイン、対象者、介入群、対照群、アウトカム、効果指標であり、Type III の場合は、研究デザイン、対象者、予知因子、アウトカム、効果指標であ

る。Type IV、V の場合は、研究デザイン、対象者、予知因子、アウトカム、対象者の疾患スペクトラム、効果指標である。

推奨レベルとの関係

上記の Type 分類は、番号の若い方がエビデンスレベルが高く、診断法を実施すべきかどうかという臨床決断の際に、実施すべきであるという強い根拠を与えることになる。とくに、Type I の研究は、診断法の実施という介入によりアウトカムの改善を証明していれば、必ず実施すべきといえる。しかし、同じように診断法の実施によりアウトカムの改善が証明されていても、Type II、III の研究の場合には、その根拠が Type I ほど強固とはいえない。対照群との比較において、いわゆる共変量のバランスが取れていない可能性があるため、証明されたアウトカムの改善が診断法の実施という介入によりもたらされたものか、明確に支持できない。

Type IV の場合は、診断能が高いことが証明されていても、それだけでは、その診断法を実施することによって、アウトカムがどれくらい改善するかを証明しているわけではないので、その結果だけをもって、実施すべきと強く推奨することは困難である。しかし、実際の個々の患者の事前確率から、実施して検査結果が得られた際に、疾患確率がどう変化するかは明確になるので、実施する意義があるということは、確実にいえるであろう。

そこで、エビデンスレベルの分類(図 11)と、推奨の関係については、表 1 のような分類を提案する(表 3)。

診断能と推奨の関係

Type IV、V の研究の場合、診断能の効果指標である、感度、特異度、陽性尤度比、陰性尤度比、ROC 曲線の AUC(曲線下面積)、平均的事前確率(検査前確率)における陽性的中率、陰性的中率、などの値によって推奨度を判断する。その基準については、図 12、13、14、15 に示す。

また、疾患によって、診断能がそれほど高くない診断法しかまだ存在しない場合もある。その場合には、診断能が高くないという理由で、使用を否定することはできない。そのような場合、古典的な診断法も含め複数の診断法がある場合、少なくとも 5 種類以上あるような場合には、効果指標の高い方の 5 分位にある診断法を推奨することは可能と考える。

しかし、Type IV の研究でも、最も高い推奨レベルで B とし、Type V の研究ではもっとも高い推奨レベルで C1 とする。

分類アルゴリズム

分類アルゴリズムを図 16 に示す。

一次選定採用文献数

641 件の文献が選定された。それらを疾患別に分類すると、大腸癌がもっとも多く、膵癌、肝細胞癌、虫垂炎がそれに続いた(表 4、表 5)。また、メタアナリシスだけに限定すると、総数 30 件で、文献数は、大腸癌、直腸癌、虫垂炎、転移性肝癌、その他の順であった(表 6)。

ガイドライン作成時のエビデンスレベル分類

これらの論文から、本報告書「診断に関する研究のエビデンスレベル分類法」で述べた、診断に関する研究のエビデンスレベル分類の Type V の研究は診断能に関する情報を表形式で、Type I から IV は構造化抄録を作成し、疾患ごとにまとめ、さらにクリニカルクエスト、推奨、理論的根拠、根拠と解説、文献の形式でガイドラインを作成すべきと考えられた。とくに、RCT のメタアナリシスを優先すべきと考えられた。

構造化抄録フォーマット

表 7 にそのフォーマットを示す。該当しない部分は削除して用いる。治療に関する論文に対する構造化抄録フォーマットとは、原則的に同じであるが、効果指標など細部で項目を変える必要がある。

ランダム化比較試験の場合には、Verhagen らのチェックリストを用い、妥当性のチェックも同時に行う。ランダム化比較試験でない場合には、この部分は削除して用いる。

ガイドライン執筆時のフォーマット

Type I、II、III、IV の研究の場合のそのフォーマットを表 8 に、Type V の場合のフォーマットを表 9 に示す。

D. 考察

今回の調査は、あくまで主観的な疾患確率を問うたものであるが、医師によってかなり大きくばらつくことが明らかとなった。専門領域によっても異なり、それぞれの医師の経験に左右されること、また、勤務する医療機関の専門性によっても影響を受けていることが考えられる。しかし、今後一定の症状を認める場合に、それぞれ想定される疾患の事前確率を標準化することは、正しい診断へ結びつけるために必要と考えられ、患者が来院した時点から、診断が確定する時点まで、前向きな研究が今後必要と考えられる。

C型慢性肝炎患者は、全例肝生検を受けているにもかかわらず、肝生検を苦痛あるいはできればしないで済む検査としてあげる患者はほとんどいなかった。一方で、上部内視鏡検査は必要と理解しても苦痛であると捕らえている患者が多いことが判明した。まず、

それぞれの疾患、病態によって、必要な検査は異なるので、疾患によって、診断のプロセスの適切性に関する患者の評価は異なることがいえる。

今回の研究は、医療水準が一番高いと考えられる大学病院を受診した患者を対象に行ったので、診断にいたるまでに、他の医療機関を受診した後に大学病院を受診した患者がおり、それらの患者の一部から、病院による臨床レベルの差をうかがわせる意見が得られた。

診断プロセスの適切性の評価は医療提供者あるいは医師とは異なるであろう事が推測される。今回は、医師側が診断プロセスの適切性をどのように評価しているか、個々のケースについて調査していないので、断定はできないが、患者の評価は医師の評価と必ずしも一致しないことが推測される。

ある患者を目の前にして、その診断法あるいは検査法を実施すべきかどうかの決断は、エビデンスレベルが高いほど実施すべきであり、エビデンスレベルが低ければ実施する必要はないといえる。たとえば、実施することによって、生存、治癒などの臨床的意義のあるアウトカムが改善することが証明されていれば、必ず実施すべきといえる。そのようなエビデンスを提供できる研究は、ランダム化比較試験のメタアナリシスあるいはランダム化比較試験である。

一方で、たとえば、感度が 90%、特異度が 90%であることだけ、すなわち診断能を報告している横断研究の場合、それだけでは、実施した方がいいか、実施する必要がないのかは判断が困難である。その診断法が適用されるであろう、対象者の平均的な事前確率の場合に、陽性的中率が十分高い、あるいは陰性的中率が十分低いことが証明されていれば、実施する価値があると考えられるであろう。論文によって、標準的な事前確率の患者に適用した場合にどれくらい疾患確率が変動するかを示し、その結果から常識的に判断して、診断法の臨床的意義について述べているものもある。

また、診断能や的中率すなわち検査後確率がどれくらい高ければ、臨床的意義があるかは、疾患の重篤度、予後、診断後に適用される治療のコストや負担、他の診断法との相対的な診断能などによって、個別に異なるので、一定の基準を適用することが困難な場合もある。他の診断法と診断能を比較して、相対的に高い群に属するものは使用しても良いという判断をせざるを得ない場合もありうる。今回提案した診断に関する研究のエビデンスレベル分類は、これらも包含した基準である。

結論

1. 消化器疾患の事前確率は医師によって異なり、それぞれの症状について、標準的な事前確率の値が設定される必要がある。
2. 患者の視点から見た場合、診断確定までのプロセス、時間について適切との判断は疾患により異なる。
3. 診断法に関する研究のエビデンスレベルの分類は、診断に特化した基準が必要であり、診断法の実施という介入により、アウトカムが改善することを証明するランダム化比較試験とそのメタアナリシスから診断能のみを示したもので分類される。
4. 診断に関するガイドライン作成には、上記のエビデンスレベルとそれに対応した推奨グレードを考慮すべきである。

E. 健康危険情報
なし。

F. 研究発表
なし。

G. 知的財産権の出願・登録状況(予定を含む。)

1. 特許取得 なし。
2. 実用新案登録 なし。
3. その他 なし。

表 1. 消化器専門医による症状別事前確率

症状	年齢	想定疾患	最小値	10%*	25%	中央値	75%	90%	最大値
嚥下困難	50歳以上	食道癌	5	5	10	25	57.5	73.5	75
嚥下困難	50歳未満	食道癌	1	1	2	5	10	47	50
胃の痛み	50歳以上	胃癌	1	2.2	5	10	35	61	70
胃の痛み	50歳未満	胃癌	0.1	0.22	1	4	16.25	58	70
胃の痛み	年齢・性別 は問わず	消化性潰瘍	15	16.5	20	25	30	67.5	75
胃の痛み	年齢・性別 は問わず	機能性ディスぺプシア	5	6.5	12.5	30	37.5	61	70
胸焼け	50歳以上	胃食道逆流症	15	25.5	50	65	80	80	80
胸焼け	50歳未満	胃食道逆流症	5	9.5	35	60	73.75	90.5	95
右季肋部痛	年齢・性別 は問わず	胆石症	10	10	11.25	27.5	37.5	78.5	80
右下腹部痛	年齢・性別 は問わず	虫垂炎	2	4.4	12.5	27.5	40	67.5	75
血便	50歳以上	大腸癌	5	5.9	12.5	30	47.5	70.5	75
血便	50歳未満	大腸癌	1	1.3	3.125	7.5	30	54	60

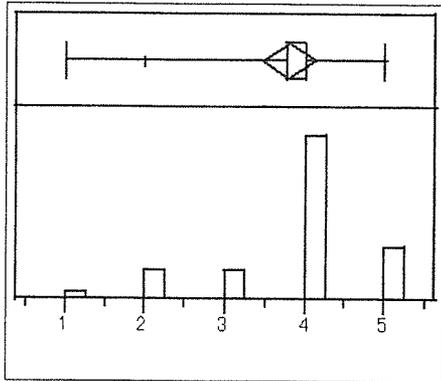
*パーセンタイル

表 2. 専門別事前確率中央値

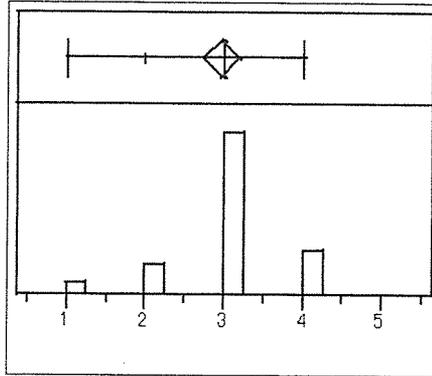
症状	年齢	想定疾患	消化器 専門医	その他医師	歯科医	P値*
嚥下困難	50歳以上	食道癌	25	5	10	0.2767
嚥下困難	50歳未満	食道癌	5	1	2	0.08
胃の痛み	50歳以上	胃癌	10	2.2	5	0.2966
胃の痛み	50歳未満	胃癌	4	0.22	1	0.2213
胃の痛み	年齢・性別 は問わず	消化性潰瘍	25	16.5	20	0.0242
胃の痛み	年齢・性別 は問わず	機能性ディスペプシア	30	6.5	12.5	0.1221
胸焼け	50歳以上	胃食道逆流症	65	25.5	50	0.0506
胸焼け	50歳未満	胃食道逆流症	60	9.5	35	0.1151
右季肋部痛	年齢・性別 は問わず	胆石症	27.5	10	11.25	0.7458
右下腹部痛	年齢・性別 は問わず	虫垂炎	27.5	4.4	12.5	0.9916
血便	50歳以上	大腸癌	30	5.9	12.5	0.7483
血便	50歳未満	大腸癌	7.5	1.3	3.125	0.8931

*Kruskal-Wallis 検定による。

CH-C



GERD



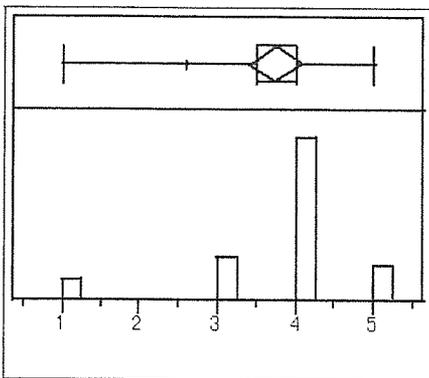
$P < 0.0001$ (Wilcoxon rank-sum test)

図1. C型慢性肝炎とGERD患者の診断プロセスの適切性に関する評価。質問「全体を通して、問診、診察、検査の進め方について、適切だった。」に対する回答。中央値は、C型慢性肝炎では4とGERD患者では3であり異なっていた。

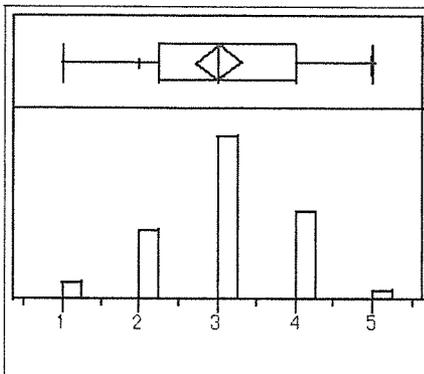
評価スコア:

1. まったくそう思わない
2. あまりそうは思わない
3. まあそう思う
4. そう思う
5. 強くそう思う

CH-C



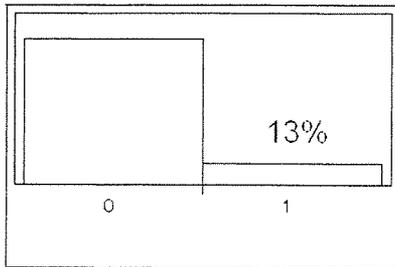
GERD



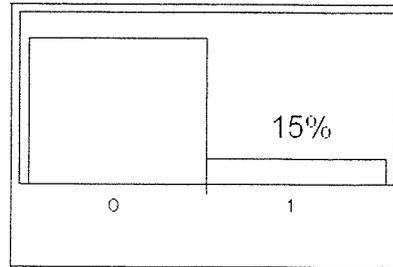
$P < 0.0001$ (Wilcoxon rank-sum test)

図2. C型慢性肝炎とGERD患者の診断プロセスの適切性に関する評価。質問「全体を通して、診断が確定するまでの時間が適切だった。」に対する回答。中央値は、C型慢性肝炎では4とGERD患者では3であり異なっていた。

CH-C



GERD



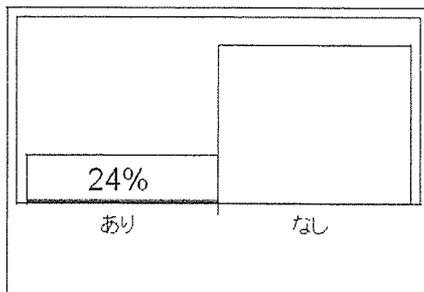
P=0.815 (Chi-square test)

- 2名肝生検
- 1名胃カメラ
- 1名複数回のX線撮影

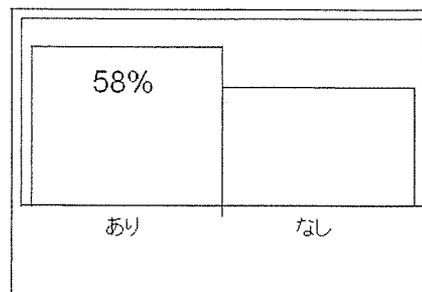
- 内訳不明

図 3. C 型慢性肝炎と GERD 患者の診断プロセスの適切性に関する評価。質問「受けた検査の中で、しないですめばその方がよいと思う検査がありましたか。」に対する回答。両群とも約 14%の患者がそのような検査があったと答えた。

CH-C



GERD



P=0.0024 (Chi-square test)

図 4. C 型慢性肝炎と GERD 患者の診断プロセスの適切性に関する評価。質問「受けた検査の中で、とくに良かった検査がありましたか。」に対する回答。C 型慢性肝炎患者 38 例中 5 例が上部消化管内視鏡、2 名が肝生検、2 名が採血をあげた。GERD 患者 40 例中 22 例が上部消化管内視鏡、1 名が pH モニタリングを上げた。