



Title	医療および診断情報の有効活用のための定義・評価方法の明確化に関する検討
Author(s)	堀, 芽久美
Citation	大阪大学, 2014, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/34323
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論文内容の要旨

〔題名〕

医療および診断情報の有効活用のための定義・評価方法の明確化に関する検討

学位申請者 堀 芽久美

近年、大規模な医療情報データベースの構築が進み、医療情報を利用した研究によって、患者、医療スタッフが有益な情報を得ることが可能となってきた。しかしながら、医療情報の利用環境が整備される一方で、医療情報からもたらされる結果に対して、多くのテーマで統一された定義・評価指標、目的に沿った評価方法が存在しない実状がある。本研究では、評価者間の一致度において一般的に知られた定義、あるいは推奨される評価方法がないことに着目し、これまでの一致度指標の問題点と新たな一致度指標の有効性の検討を行うことを目的とした。

【研究1：一致度評価指標の比較：シミュレーションによる検討】

背景：

評価者の一致度、評価の再現性については、Kappa係数が一般的に用いられてきた。しかしながら、Kappa係数は判定結果の偏りに影響を受け、過少評価されることが報告されている。そのため、このような問題に対して、Gwet's AC1やPABAKといった一致度評価指標が提案されている。本研究ではシミュレーションにより、有病率や判定結果の偏りがGwet's AC1, CohenのKappa係数, PABAK, FleissのKappa係数に与える影響を検討した。

方法：

2名の評価者間の一致度評価指標としてCohenのkappa係数, PABAK, およびGwet's AC1の比較, 3名以上の評価者間の一致度評価指標としてFleissのkappa係数とGwet's AC1の比較を行った。比較は、有病率(90%, 70%, 50%), ランダム評価割合 (0.2, 0.4, 0.6), 症例数(10, 20, 50, 100, 200), 評価者数(5名, 10名, 15名)を設定し、シミュレーションによって行った。

結果1：2名の評価者間における比較

Gwet's AC1が観察一致度と最も近い値を示し、Cohenのkappa係数は観察一致度を大きく下回った。有病率が50%のとき、症例数、観察一致度にかかわらず、3種類の指標はいずれも近い値をとった。有病率90%の場合では3種類の指標の差が大きく、特にCohenのkappa係数が他の2種類の指標と比較して小さかった。また観察一致度が高いとき、指標間での差が大きくなる傾向にあった。

3種類の指標を、10000回分の中央値として算出した場合、有病率、観察一致度の違いに関わらず症例数の変化による影響は見られなかった。

結果2：3以上の評価者間における比較

Fleissのkappa係数に対して、Gwet's AC1が観察一致度と近い値を示した。有病率、観察一致度、症例数の影響は2名の評価者間の場合と同様であった。評価者数が増えると、観察一致度が高くなり、FleissのKappa係数, Gwet's AC1のどちらの指標も上昇する傾向にあった。

まとめ：

有病率が50%に近く、判定結果の偏りが小さい場合はKappa係数, Gwet's AC1で共通の評価を与えることが可能である。有病率が50%から離れ、判定結果の偏りが大きく、さらに観察一致度が高い場合にはKappa係数の過少評価が著しく、Gwet's AC1の利用が望まれる。一致度の過少評価は診断法の有効性を過少評価することにつながる。それぞれの指標と陽性判定割合、観察一致度との関係を理解した上で利用する一致度評価指標を決定することが重要である。

【研究2：臨床データに対する一致度指標の利用可能性の検討】

背景：

WT1は多くの種類の固形腫瘍やそれらの腫瘍血管に過剰発現することが報告されている。さらに，WT1タンパクの遺伝子産物は高い免疫抗原性を示し，免疫療法の有効な標的分子として考えられる。それゆえに，腫瘍細胞におけるWT1発現の診断の一致度評価は重要である。

方法：

WT1タンパクの腫瘍組織における発現判定の一致度評価の統計学的手法を確立するため，6例の腫瘍組織の標本スライドをWT1免疫染色後、スキャンしてバーチャルスライド画像を作製した。評価者9名が独立してこれらのバーチャルスライドを観察し7つの項目：染色のバックグラウンドおよび腫瘍細胞と腫瘍血管それぞれの核、細胞質、そして細胞全体のWT1発現について判定した。これらの項目の判定それぞれについてGwet's AC1とFleissのKappa係数を計算し一致度を評価した。Gwet's AC1の頑健性を評価するために評価者の人数を減少させた場合のGwet's AC1を算出した。

結果：

染色のバックグラウンド判定の観察一致度は高かった。腫瘍細胞と腫瘍血管のいずれにおいても細胞質のWT1発現判定および細胞全体のWT1発現判定の観察一致度は0.8以上と高く，それに対して核のWT1発現判定の観察一致度は0.6未満と低かった。腫瘍細胞と腫瘍血管それぞれの細胞質のWT1発現判定，細胞全体のWT1発現判定において陽性の判定割合が高くなっていた。そのため，Gwet's AC1では観察一致度の高い（0.8以上）これら4項目のうち3項目で一致度が高い“Almost perfect”と，残りの1項目でも“Substantial”と解釈された。しかしながら，FleissのKappa係数では高い観察一致度を持つ4項目のすべてで一致度の低い“Slight”と解釈された。観察一致度の低い2項目についてはGwet's AC1とFleissのkappa係数ともに一致度が低いと解釈された。また，Gwet's AC1は評価者の人数が減るとばらつく傾向にあった。

まとめ：

腫瘍細胞におけるWT1発現の診断は判定結果が陽性に偏る傾向があった。Gwet's AC1は観察一致度とよく相関し，WT1発現判定の評価者間一致度を解析する有用な指標であるといえる。今後，遠隔診断技術の発展によって多数の評価者が増加する一致度研究の増加が予想され，一致度評価指標の解釈はますます重要になることが考えられる。

【研究1,2に基づく結論】

Kappa係数は有病率が50%に近い場合において適用可能であり，有病率が50%から離れる場合，あるいは不明な場合にはGwet's AC1の適用が望ましいことが示唆された。

陽性診断に大きく偏った腫瘍細胞へのWT1発現の診断データに対してGwet's AC1を適用した場合，Gwet's AC1は観察一致度に対応した評価を行うことが明らかとなり，臨床データに対する適用可能性が示された。しかしながら，臨床データへの適用を考えるには，評価者数の増減の問題が考えられ，評価方法の工夫が必要である。本研究で行ったように評価者数の人数をそろえるために無作為抽出を行う場合，複数回抽出を繰り返すことで得られる中央値，分布から一致度の評価を行うべきである。

論文審査の結果の要旨及び担当者

氏 名 (堀 芽久美)		
論文審査 担当者	(職)	氏 名
	主 査 教 授	大野 ゆう子
	副 査 教 授	小 泉 雅 孝
	副 査 教 授	神 本 直

論文審査の結果の要旨

医療現場では多くの診断、検査に基づいて最終的な診断が下される。しかし、この診断や検査には、最終的な診断の判定基準が曖昧である場合がある。本研究は、いまだ明確な評価方法が確立されていない評価者間の診断の一致度に焦点を当てた研究である。

一致度とは複数名の評価者が同一対象について何らかの判定を行った場合に、その判定結果がどの程度一致したかを示す指標である。これまで複数名の評価者の一致度の評価方法に定まった方法はなく、さらにこれまで利用されてきた評価指標には大きな問題があることが知られている。医療現場における診断の一致度は、その後の患者の治療方針に影響を与えるため重要であり、一致度を適切に評価することができる指標や評価方法の検討が必要である。

本研究では、シミュレーションによって、これまで利用されてきたCohenのKappa係数、FleissのKappa係数と、新たに提案されたGwet's AC1について、その有効性の比較が行われた。シミュレーションは有病率、曖昧な判定をする評価者の割合(ランダム評価割合)、症例数、評価者数を設定して行われた。このシミュレーションの結果によって、CohenのKappa係数、FleissのKappa係数は有病率が50%に近い、つまり、陽性、陰性と判定された症例の数に偏りが無い場合においてのみ適用可能であり、有病率が50%から離れた場合には、評価者の多くの判定が同じであっても、低い評価を与えることが明らかになった。それに対してGwet's AC1は、いずれの有病率の場合でも適用可能であることが示された。Gwet's AC1は観察された一致割合が大きい場合に高い評価を与えており、現実に即した指標であることが明らかになった。また、CohenのKappa係数、FleissのKappa係数は小さな診断結果の変化によって大きな影響を受けた。以上より、CohenのKappa係数、FleissのKappa係数と比較して、Gwet's AC1が高い有効性を示した。

次いで、Gwet's AC1を腫瘍細胞のWT1遺伝子(以降、WT1)発現判定結果データに適用し、臨床データへの適用可能性の検討を行った。

判定は9名の評価者が6症例について、染色のバックグラウンド、腫瘍細胞の核、細胞質、全体におけるWT1発現、腫瘍血管の核、細胞質、全体におけるWT1発現を評価している。WT1発現データへ適用することで、シミュレーションでは想定していなかった、症例によって評価者数が異なるという問題が生じた。これは、染色のバックグラウンドの評価によって、染色の状況がその他の項目の評価を行うには適切ではないと判断される症例が存在したためである。これに対する解決方法として、各症例で評価者の人数をそろえるための無作為抽出の繰り返しを提案した。抽出人数は7人、6人、5人を設定した。1回の無作為抽出では、抽出の偏りによって極端に低い評価や、高い評価を与える可能性があったが、無作為抽出を複数回行い、それら

のデータセットに対して算出されたGwet's AC1の中央値、分布を合わせて評価することで、頑健な一致度の評価を行うことが可能となった。これは高いオリジナリティとして評価できる。

WT1発現データは腫瘍細胞、腫瘍血管ともに細胞質における発現結果が陽性に大きく偏っていた。細胞質におけるWT1発現の診断は陽性で高い一致度があると考えられる。しかしながら、FleissのKappa係数は6段階で評価される一致度の強度としては2番目に低い評価であるSlightであった。一方、Gwet's AC1はもっとも高い評価であるAlmost perfectであった。Gwet's AC1がWT1発現結果の一致度についても現実 に即した評価を行い、臨床データへの適用可能性を示した。

本研究では、検査や診断方法の違いの影響ではなく、評価指標の違いによる影響のために、一致度が過小評価される場合があることを示した。これは、本来有効なはずの診断方法に対して誤った低い評価を与えることにつながる。医学分野において一致度の評価が重要である一方、評価指標自体の問題点に対する注目は小さい。指標の影響を含めた上で、適切な一致度の評価を行うことで、診断の効果が正しく認識されることは、診断を受ける患者、診断を行う医療者双方の利益へとつながると考えられる。本研究は適切な一致度の評価指標、方法を提案し、現実問題への適用上の課題にも解決策を提案しており、患者、医療者の適切な治療選択を促進させる研究であるといえる。

以上より、本研究は博士号（保健学）の学位に値するものと評価する。