



RIETI Discussion Paper Series 20-J-021

医療における人工知能の受容性

岩田 和之
松山大学

森田 玉雪
山梨県立大学

鶴見 哲也
南山大学

馬奈木 俊介
経済産業研究所



Research Institute of Economy, Trade & Industry, IAA

独立行政法人経済産業研究所

<https://www.rieti.go.jp/jp/>

医療における人工知能の受容性*

岩田 和之（松山大学） 森田 玉雪（山梨県立大学）
鶴見 哲也（南山大学） 馬奈木 俊介（経済産業研究所・九州大学）

要 旨

本研究では、①医療サービスに対して、人々がどのような要素を重視しているか、②余命を宣告される場合に、人々は医師と人工知能ロボットにどのような期待を抱いているか、の2点を調査[†]した。調査では15,001人から回答を得ることができた。調査結果を分析したところ、①人々は医療サービスを構成する要素のうち、診断の信頼性を最も重視しており、次いで医師との対話を重視していることが示された。ただし、対話については人々の属性によって多様な評価結果が示された。また、②余命を宣告される際に、人々は医師からの宣告を好み、人工知能ロボットに言われるなら正直に宣告されたいことも示された。ただし、人工知能ロボットに余命を宣告される場合には、好まれるロボットの形態（人間型・ペット型・スピーカ型）は状況によって異なることも示された。

キーワード： 人工知能、医療、ベスト・ワースト・スケーリング、コンジョイント分析

JEL classification: I18, P36

RIETI ディスカッション・ペーパーは、専門論文の形式でまとめられた研究成果を公開し、活発な議論を喚起することを目的としています。論文に述べられている見解は執筆者個人の責任で発表するものであり、所属する組織及び（独）経済産業研究所としての見解を示すものではありません。

*本稿は、独立行政法人経済産業研究所（RIETI）におけるプロジェクト「人工知能のマクロ・ミクロ経済動態に与える影響と諸課題への対応の分析」の成果の一部である。

† RIETI が 2019 年 3 月に実施した平成 30 年度「医療における人工知能についてのインターネット調査」のことを指す。

I. 医療における人工知能

医療に関する人工知能（以下、AI）が注目を集めている。医療 AI は既存の医療現場での問題軽減・解消に寄与することが期待されている。日本の AI 戦略を取りまとめている内閣府（2019）でも、健康・医療・介護領域で AI を活用することが明記されている。その中では具体的目標として健康・医療・介護分野で AI を活用するためのデータ基盤の整備、日本が強い医療分野における AI 技術開発の推進と、医療への AI 活用による医療従事者の負担軽減、などの 5 つが掲げられている。

医療分野における AI 活用は画像診断といったソフトウェア面だけではなく、ロボットのようなハードウェア面まで広がっている。一方で、こうした AI の活用は医療従事者の医療サービス供給側視点での有用性の議論は活発に行われているものの、医療サービス需要側、つまり患者視点での議論は十分に行われているとは言い難い。そこで、本研究では需要側視点での医療にかかわる AI をソフトウェア、ハードウェアについてそれぞれ患者側がどのように評価しているかを、アンケート調査によって明らかにする。

具体的には、第 II 章では医療サービスの評価項目を 10 種類取り上げ、人々はそれらの項目をどのように評価しているかをベスト・ワースト・スケーリングを用いて分析している。そして、取り上げている項目の中で、診断の信頼性や医師との対話など、医療 AI に密接にかかわる項目を加えている。

第 III 章ではハードウェアである AI ロボットと患者が直接対面する状況に着目する。労働力不足を補う形で、病院の受付や案内などでロボットがすでに実用化されているが、本研究では、労働力補完という役割だけでなく、医師が余命宣告の際の心理的負担を軽減する役割をも AI ロボットが担えるかという点を明らかにする。医師にとって余命宣告は重い仕事であるため、患者が真実を期待していても真実を伝えることが難しいケースがある。患者が仮に AI ロボットから余命を宣告される状況に直面したとしたら、患者は正確な結果を知ることのできる確実性を AI ロボットに期待するのか、AI には AI ならではの配慮を求めるのか、を、コンジョイント分析を利用して明らかにする。同時に、人々に好まれるロボットの形態も検証する。

それぞれの分析結果をまとめると、第 II 章では、医師の診断に対する信頼性が最も重視され、次いで医師との対話が重視されることが示された。そのため、医療 AI 導入によって信頼性が向上し、その結果患者の満足度が改善すること

が期待できる。一方で、対話については、女性と高齢者は医療機関での受診の後半（診断・治療）の対話を重視し、若者は前半（問診・診察・検査）を重視することが示された。したがって、医療機関は訪れる患者の属性に応じて医療 AI の導入を柔軟に検討していくこと望ましい。

第Ⅲ章では、余命宣告のやり方が不確実であっても、人々は人間の医師から宣告されたいと期待する傾向が強いことが示された。また、AI に余命宣告をさせるとしたら、平均的にみて、AI が正直であることが最も期待されていることも分かった。さらに、AI が正直であるか、様子を見るか、余命を長く言うかによって、患者が好む AI の型が異なることも明らかになった。したがって、AI が道徳的な判断をすることが可能になったとしたら、患者の選好に応じて最適な様式の AI を担当させることが望ましい。

II. 患者にとっての人工知能

昨今の医療 AI の疾病診断の性能の発展は目覚ましい。例えば、Bejnordi et al. (2017)は乳がんの転移を調べる画像判定の正答率を AI と医師とで比較を行ったところ、AI は 99.4%の精度で転移の有無を正答した一方で、医師は時間無制限の場合は 96.6%とやや AI よりも低い正答率であったことを示している。さらに、時間が制限された場合の医師の正答率は 81.0%まで低下したことも示しており、医療 AI の画像判定の正確さがすでに人間医師を上回っていることを明らかにしている。Shetty (2019)も Google の肺がん検診 AI と医師 6 名との胸部 CT 画像の診断による肺がんの正答率を比較したところ、AI は 94.4%の正答率となり、人間医師よりも偽陽性¹を 11%減らすと同時にがん発見率を 5%高めたことを明らかにしている。このように、AI は過去の大量の疾病データを学習することによって、すでに医師よりも高いパフォーマンスを発揮しているケースもある。

したがって、医療 AI は従来の医師の業務の一部を補助あるいは代替することによって、医師の負担軽減をもたらすと同時に、疾病の早期発見などの医療サービスの向上も期待できるのである。実際に医療 AI を導入している事例もある。例えば、自治医科大学ではホワイト・ジャックと呼ばれる双方向対話型 AI システムが導入されている。同システム下では、医師は患者の予診情報等を

¹ 本来、陰性であると判断されるべきにもかかわらず、誤って陽性と判断されることをいう。

入力することで、可能性のある疾病とその確率、さらにはその検査項目・処方箋までも AI が提示してくれる。それらの AI による提示をどのように扱うかは医師の判断に委ねられるものの、上記のような診療 AI の精度の高さを考慮すると、医療サービスの質の向上に繋がるものと考えられる。

医療 AI によって医療現場は大きな変化がもたらされることが予想される。特に、内閣府（2019）でも指摘されているように、医療 AI によって医療現場の負担軽減が期待される。医療 AI によって疾病の的確な発見がなされるのであれば、患者の医療サービスに対する満足度は改善すると考えられる。しかしながら、内閣府（2019）での既述の 5 つの医療 AI 目標の中には、患者満足度の改善は明記されておらず、医療 AI によって患者（消費者）の医療サービス満足度が改善するかどうかは不明瞭である。そこで、第 II 章では医療サービスに関するアンケート調査を用いて、医療 AI の導入が患者の医療サービスに対する満足度改善につながるかどうかを検証する。

1. 先行研究

患者視点から医療サービスのパフォーマンスを評価する指標の一つとして、古くから用いられてきたものが患者満足度である（Donabedian, 1966）。この患者満足度に関する文献は膨大にあり、日本を対象にした研究も相当に蓄積されている。

国内の例としては、73 の医療機関の外来患者の医療満足度を調査した山本、他（2004）がある。849 名の患者への調査データを用い、5 段階で尋ねた患者満足度に対して様々な項目で回帰分析を行っている。その結果、病院の医師の評価が高いほど、設備・環境の評価が高いほど、健康の回復状況が高いほど、患者満足度が高くなることを示している。大学病院等の病床数の大きな医療機関や入院設備のない診療所でも等しくこの傾向が観察されている。したがって、医療・治療行為の結果が望ましいほど、患者満足度が高くなるといえる。早瀬、他（2013）はある総合病院にて、外来患者 290 名、入院患者 205 名、その病院で勤務している看護師 108 名にアンケート調査を行い、患者満足度と医療サービス供給側との関連について分析をしている。要素間の関係を把握するために、構造方程式モデリングを用いて分析を行った結果、医師や看護師の待遇が患者満足度に影響を与えていることを示している。このことは昨今の医師等の医療従事者の負担軽減と整合的な結果となっている。

一方、海外でも多くの患者満足度研究がなされている。Baker et al. (2003)はアメリカで 418 名の、イギリスで 650 名の患者を調査し、患者満足度に与える要因を回帰分析によって明らかにしている。分析の結果、医師への信頼と医師への相談が重要な要因であることを示している。また、アメリカとイギリスでは、アメリカの患者のほうが満足度が低いことも明らかにしている。Schoenfelder et al. (2011)はドイツの 39 医療機関を対象に 8,428 名の患者に調査を行った結果、患者満足度に最も影響を与えているものは治療の成果であることを示している。次に重要な要素として看護師の親切さを挙げている。Meng et al. (2018)は中国での医療従事者 696 人、患者 668 人を調査し、患者満足度に最も影響を与えた要素は患者と医師とのコミュニケーションであることを示している。次いで、医療機関でのサービスや設備が影響を与えているとしている。このように、多くの国で患者満足度に影響を与える要因は研究されているものの、患者満足度を調査した 109 論文をレビューしている Batbaatar et al. (2017)は患者満足度をグローバルに捉える定式化や測定方法は見つかることができなかったと述べている。

しかしながら、完全な定式化は困難であっても、患者満足度を規定する要素はこれまでの研究から推察することはできる。それは、治療の成果（あるいは医師の信頼）と医師・看護師等の医療スタッフとのコミュニケーションである。治療の成果については患者も健康状態の改善などを通じて自身で感じ取ることができる。そのため、治療の成果を高めるということはシンプルな患者満足度の改善方法であるといえる。

一方で、コミュニケーション改善については治療の成果ほどシンプルではない。なぜなら、医療スタッフと患者との間で保有している医療知識が等しくない（医療スタッフの方が知識豊富である）からである。例えば、患者は自身に都合のよい情報のみを重視し、都合の悪い情報は排除する傾向にあるという確認バイアスを持つ傾向にある（Pines, 2006）ため、そうしたバイアスを避けるようなコミュニケーションが医療スタッフには必要である。また、患者は医師が理想とする水準以上の医療情報の提供を望む一方で、実際にはそれができていないと感じるケースや、患者は患者中心の治療方法や薬の決定を望む一方で、実際には医師中心の決定であると感じるケースがある。これはコミュニケーションの不十分さによって生じる認識ギャップである（塚原、2010）。そのため、医療スタッフには患者との適切なコミュニケーションが求められる。

そして、この治療の成果やコミュニケーションについては、医療 AI の登場によって大きく変化する可能性がある。既述のように、画像判定によって疾病の有無を判断する場合にはすでに医療 AI が人間の医師よりも高いパフォーマンスを発揮するケースも報告されている（Bejnordi et al., 2017 ; Shetty, 2019）。そのため、治療の成果あるいは医師への信頼は医療 AI によって高まる可能性がある。そして、このことは患者満足度の改善につながる。

他方で、医療 AI が医師と患者とのコミュニケーションに与える影響は不明瞭である。もし、医療 AI が医師の業務の一部を代替する場合、例えば医師が行う問診を医療 AI が代替する場合には、医療 AI はコミュニケーションの低下をもたらすことになるかもしれない。それとは異なり、問診時に医師が医療 AI を用いながら患者とコミュニケーションをとるのであれば、医療 AI はコミュニケーションを促進する可能性もある。

既存研究では患者と医師（あるいは医療スタッフ）とのコミュニケーションが患者満足度にとって重要であると指摘しているが、本研究では一步踏み込んだコミュニケーションを扱う。具体的には、医療機関での受診を前半（問診・診察・検査）と後半（診断・治療）に分け、それぞれにおける医師との対話と患者満足度（＝評価）との関係を分析する。もし、前半での対話が患者満足度に大きな影響を与えないのであれば、患者満足度の点からは前半業務は医療 AI に代替することも可能となるだろう。逆に、前半での対話が大きな影響を与えるのであれば、医療 AI は医師の業務を支援するような使い方をすべきであるという結論になる。このように、業務を前半と後半に分けることによって、医療 AI の医療現場での導入の仕方をより詳細に提案することができるようになる。

2. 分析手法

医療 AI の導入が患者の医療サービスの満足度改善に寄与するかどうかを検証するために、本章ではベスト・ワースト・スケーリング（Best-Worst Scaling : BWS）を用いる。BWS は消費者の財やサービスの特徴に対する評価を定量的に測定する表明選好法の一手法であり、Finn and Louviere (1992)によって導入された比較的新しい方法である。具体的には、その財・サービスを特徴づける項目を複数個提示し、その中で最も望ましいと感じる項目（Best）と最も望ましくないと感じる項目（Worst）の2つを選択してもらう方法となる。これまでの

表明選好法で多用されてきた本章後半の分析でも採用しているコンジョイント分析は、最も望ましい項目のみを選択する方法であるため、コンジョイント分析よりも多い情報を用いた分析手法となる。ただし、Krucien et al. (2017)は BWS よりもコンジョイント分析のような通常の離散選択モデルのほうが望ましいと主張している一方で²、Potoglou et al. (2011)は両者の結果は同じ傾向となることを示しており、手法間での議論は収束していない。

医療分野における BWS を用いた研究は増加傾向にある (Cheung et al., 2016)。例えば、Ejaz et al. (2014)はジョンホプキンス病院の 240 人の患者を対象に、がん外科医の選択における重要な要因を BWS によって分析している。分析の結果、外科医の症例経験や専門的な訓練の経験が最も重要であり、次いで病院全体での症例経験が患者にとって重要であることを示している。逆に、病院の駐車場の利便性や自宅との距離は重要ではないことを示している。Uy et al. (2018)はシンガポールでの 603 名を対象に、健康関連のクオリティオブライフの項目の重要度を測定している。27 項目のうち、「他の人の助けを借りずに自分で世話をする能力」、「病気への治癒と抵抗」、「家族や友人との良好な関係」が重要であることを示している。

本章の BWS では医療サービスの項目として 10 個の要素を取り上げている。それらは、「要素 1：医師の診断に対する信頼性（以下、信頼）」、「要素 2：問診・診察・検査時に医師との対話時間が十分にあること（以下、対話（前）」、「要素 3：診断・治療時に医師との対話時間が十分にあること（以下、対話（後）」、「要素 4：医師の人柄（以下、人柄）」、「要素 5、毎回同じ医師に診てもらえること（以下、同一）」、「要素 6：予約の取りやすさ、または（当日先着順の医療機関の場合）待ち時間の少なさ（以下、予約）」、「要素 7：医療機関への交通利便性（以下、利便）」、「要素 8：医療機関の設備の充実度（以下、設備）」、「要素 9：医療機関が清潔であること（以下、清潔）」、「要素 10：医療機関でのプライバシー保護（以下、個人）」である。これらの項目は国内の事例を多く取り扱っている塚原 (2010) および記述の先行研究を参考に選定している。回答者には、「今、あなたに高熱（38 度以上）の症状があるとします。必ずしも風邪とは限らず、大きな病気の可能性もゼロではありません。この症状があるときに利用する医療機関(町医者)を想定してお答えください。」という前提を示した上で、

² van Dijk et al. (2016)はリスク受容度の比較においても、両者のモデルには差があることを示している。

「理想のかかりつけ医療機関」を想像してください。あなたが「理想のかかりつけ医療機関」を考えると、「最も重視する要素」と「最も重視しない要素」の二つを以下のそれぞれの問について選択してください」という質問を行っている。質問では、表 1 のように 10 要素からランダムに 4 要素を提示し³、その中から最も重視する項目と最も重視しない項目をそれぞれ選んでもらっている⁴。

表 1. 提示する選択肢の解答例

	最も重視する要素	最も重視しない要素
医師の人柄	☒	☐
医療機関が清潔であること	☐	☐
医師の診断に対する信頼性	☐	☒
医療機関への交通利便性	☐	☐

アンケート調査の詳細は次節で述べるが、回答者には表 1 のようなランダムに選ばれた選択肢のセットを 10 回提示している（＝10 回最も重視する項目と重視しない項目を選んでもらっている）。調査の回答者は 15,001 名であることから、サンプルサイズは 150,010 となっている。

BWS では 2 種類の分析方法を用いる。それらは、計数法と呼ばれるものと、条件付きロジットモデルを用いる MaxDiff 法である。前者の計数法は非常にシンプルな方法であり、Best と Worst の比率（B/W 比率）を 10 要素ごとに計算する方法である。消費者が医療サービスの要素の中で最も高く評価する要素の B/W 比率が一番高くなる。Best と Worst の差を取る方法もあるが、各要素の登場回数を等しくしていないため、本研究の計数法では B/W 比率を用いることとする。

MaxDiff 法は計量分析を用いるため、計数法のように単純な計算ではない。回答者が「要素 j を Best、要素 k を Worst」と選択するときの評価尺度距離を

³ 加えて、その項目を表示する並びもランダムとしている。

⁴ 10 項目のうち 4 項目を提示するため、ランダムに提示すると項目間での登場回数に差が生じる可能性がある。そのため、釣合い型不完備ブロック計画（Louviere and Flynn, 2010）を用いて項目間の登場回数を等しくする方法もある。しかし、本章でのサンプルサイズは 150,010（後述する計数法の場合）と非常に多くなっていることから、ランダム提示による登場回数間の差は僅かであると判断した。

D_{jk}^* とする。この D_{jk}^* が要素 j の評価 (β_j) と要素 k の評価 (β_k) の差 (距離) と誤差項 ε_{jk} で構成されたとする (つまり、 $D_{jk}^* = \beta_j - \beta_k + \varepsilon_{jk}$)。合理的消費者が「要素 j を Best、要素 k を Worst」であると選択するならば、 D_{jk}^* はそれ以外の組み合わせの評価尺度距離 (D_{mn}^*) よりも大きくなる必要がある (つまり、 $D_{jk}^* > D_{mn}^*$)。本研究では、4 つの要素を提示しているため、「要素 j を Best、要素 k を Worst」とする組み合わせは 12 種類存在している。そして、誤差項 ε_{jk} がガンベル分布に従うと仮定すると、12 種類の中から 1 つを選ぶという多項ロジットモデルとして各要素の評価 β を推定することができる。既述したように、本研究では 10 回選択肢のセットを提示することから、パネルの多項ロジット、つまり条件付ロジットモデルとして推定する。

3. アンケート調査

分析には独立行政法人経済産業研究所が 2019 年 2 月から 3 月にかけて実施した「医療における人工知能についてのインターネット調査」を用いる。調査はリサーチ会社に委託実施し、15,001 人から回答を得ることができた。18 歳以上 79 歳以下⁵を調査対象としているが、一般的に Web 調査では回答者はスマートフォンやパソコンを用いて回答をするため、高齢者の割合が少なくなってしまう。そこで、本研究が医療問題を扱うという性質を考慮し、リサーチ会社には 65 歳以上の高齢者の割合が少し多くなるように依頼をしている。結果として、全体の 22.7% (3,408 名) が高齢者となっている⁶。また、地域の偏りを避けるために、現実の都道府県の人口分布に近くなるようにサンプルは調整している。15,001 人の性別、年齢、かかりつけ医の有無といった基本的な記述統計量は表 2 に載せている。

表 2. 記述統計量の抜粋

変数	平均	標準偏差	最小	最大
性別 (男性 = 1)	0.50	0.50	0	1
年齢	49.33	16.19	18	79
かかりつけ医 (いる = 1)	0.51	0.50	0	1

注：サンプルサイズは 15,001 である。

⁵ リサーチ会社のモニターが 79 歳までのため。

⁶ 2019 年 9 月時点での日本の 65 歳以上人口比率は 28.4% である。

本研究での BWS では、選択肢セットを 10 回提示することから、各回答者は Best と Worst を 10 回選ぶことになる。そのため、計数法でのサンプルサイズは 150,010 ($=15,001 \times 10$ 回) となる。MaxDiff モデルの場合には、12 種類の選択肢の中から 1 つの Best と Worst の組み合わせを選ぶ分析となることから、サンプルサイズが 1,800,120 ($=15,001 \times 10 \times 12$) となる。

4. 分析結果

(1) 計数法による分析結果

計数法による単純集計結果と B/W 比率を表 3 に載せている。結果を見ると、「信頼」の B/W 比率が他の要素よりも突出して数値が高くなっている。このことは、「信頼」が圧倒的に重要な医療サービスの要素であることを示している。第 2 節で概観した先行研究でも医師への信頼あるいは治療の成果が最も重要であることが明らかになっている。この結果も既存研究とは整合的な結果となっている。

次いで B/W 比率が高い要素は、「対話（後）」、「対話（前）」の 2 種類の医師とのコミュニケーション項目となった。先行研究でもコミュニケーションは患者満足度に対して重要であるとしていることから、本研究の結果も同様となっている。ただし、先行研究とは異なり、本研究では医師との対話を前半（問診・診察・検査）と後半（診断・治療）に分けている。対話の内容を見ると、似通った B/W 比率とはなっているものの、問診・診察・検査時の対話よりも、診断・治療時の対話のほうが患者の評価ウェイトが高いことが示された。

相対的に重要度が低い要素は、「利便」、「個人」、「清潔」となっている。質問では「高熱（38 度以上）の症状があるときの理想のかかりつけ医」を想定してもらっていることから、「近所」を想定した可能性が高い。このことが「利便」を重要視しなかったことにつながったと考えられる。Ejaz et al. (2014) も利便性は重要ではないと結論付けており、本研究もそれと整合的な結果が得られている。また、プライバシーの保護や清潔度合いについては、日本の病院間で大きな差はない可能性がある。つまり、どのような病院でも一定のプライバシー保護や清潔さが確保されていることから、医療サービスの評価項目として想定的に重要度が低くなっているものと考えられる。

表 3. B/W 比率

項目	Best数	Worst数	B/W比率	順位
信頼	39,078	2,497	15.65	1
対話（前）	21,683	8,752	2.48	3
対話（後）	21,644	8,336	2.60	2
人柄	15,699	10,253	1.53	4
利便	5,822	31,256	0.19	10
設備	11,273	13,428	0.84	5
予約	12,112	22,379	0.54	7
清潔	5,188	14,202	0.37	8
同一	13,116	18,530	0.71	6
個人	4,395	20,377	0.22	9

注：サンプルサイズは 150,010 である。

（2）MaxDiff 法による分析結果

条件付ロジットモデルによって推定した結果を表 4 に載せている。推定結果は要素「個人」を基準とした相対的な評価度合いを示しているため、表 4 から要素「個人」が抜けている。したがって、推定結果の係数が正であることは、「個人」と比べてその項目の重要度が高いことを意味する。逆に係数がマイナスであることは、「個人」よりもその要素の評価が低いことを意味する。

係数を見てみると、全ての要素の係数は 1%水準で有意であることが示されていることから、全ての要素は有意に「個人」と異なる評価となっている。最も係数が大きい要素は「信頼」であり、次いで「対話（後）」、「対話（前）」となっている。ただし、「対話（後）」と「対話（前）」との間には有意な差異は認められなかった（ $p=0.12$ ）。

逆に最も評価が低かったものは「利便」であり、次いで「個人」、「予約」の評価が低くなっている。これらの傾向は B/W 比率を用いた計数法と似通った結果となっている。「予約」と「清潔」の順位が逆転している点のみ結果が異なっている。したがって、MaxDiff 法を用いた場合でも、医療サービスにおいて「信頼」が最も重要であり、その次にコミュニケーションが重要であることが確認された。

表 4. 推定結果

項目	係数	標準誤差	順位
信頼	1.927***	0.0096	1
対話（前）	1.008***	0.00864	3
対話（後）	1.021***	0.00865	2
人柄	0.743***	0.0085	4
利便	-0.351***	0.00857	10
設備	0.480***	0.00841	5
予約	0.200***	0.00838	8
清潔	0.245***	0.00837	7
同一	0.369***	0.00839	6
対数尤度	-326130		
カイ二乗値	93262***		
疑似決定係数	0.13		

注 1：***は 0.1%、**は 1%、*は 5%、+は 10%水準で有意であることを示す。

注 2：サンプルサイズは 1,800,120 である。

これらの要素の重要度合いは回答者の属性によって異なる可能性がある。そこで、男性のみ、女性のみ、70 歳以上（高齢者）のみ、40 歳未満（若者）のみの 4 つの属性に限定して条件付ロジットモデルを推定した。表 5 はその結果である。いずれの属性についても、「信頼」が最重要であることが強く示されている。そして、次点として 2 種類の「対話」と「人柄」となっており、概ね表 4 の推定結果と傾向は似通っている。ただし、男性は 2 種の対話に有意差は無い（ $p=0.28$ ）一方で、それ以外の属性の人は両者には有意差が認められた。

また、男性と若者は 5 番目に重要である要素として「設備」を示している一方で、女性は「予約」、高齢者は「同一」が 5 番目に重要な要素となっている。さらに、高齢者は他の属性の回答者よりも「予約」を軽視する傾向にあることも示されている⁷。

表 5. サンプルを区切った場合の推定結果

⁷ かかりつけ医の有無でサンプルを区切った場合も、同様の推定結果が得られた。

項目	男性のみ		女性のみ		70歳以上のみ		40歳未満のみ	
	係数	順位	係数	順位	係数	順位	係数	順位
信頼	1.853***	1	2.018***	1	2.586***	1	1.457***	1
対話（前）	0.928***	2	1.101***	3	1.650***	3	0.561***	3
対話（後）	0.915***	3	1.142***	2	1.744***	2	0.505***	4
人柄	0.752***	4	0.739***	4	1.040***	4	0.562***	2
利便	-0.268***	10	-0.437***	10	-0.613***	10	-0.224***	10
設備	0.601***	5	0.357***	6	0.704***	6	0.342***	5
予約	0.329***	7	0.0682***	5	-0.0993***	9	0.294***	6
清潔	0.288***	8	0.202***	8	0.255***	7	0.257***	7
同一	0.494***	6	0.242***	7	1.031***	5	0.000829	8
対数尤度	-165008		-160170		-42465		-106872	
カイ二乗値	39839***		55328***		25649***		16508***	
疑似決定係数	0.11		0.15		0.23		0.07	
観測数	893,040		907,080		267,000		555,960	

注：***は 0.1%、**は 1%、*は 5%、+は 10%水準で有意であることを示す。

5. 医療 AI の導入に向けて

医療 AI の導入議論については医療サービス供給側への影響が重視される傾向にあり、患者である需要側の視点で議論されることが少ない。そこで、本章では医療 AI が患者の医療への満足度、いわゆる患者満足度に対してどのような影響を与えるかどうかについて、インターネットによる家計へのアンケート調査を用いて検証した。

分析方法としては、医療サービス需要者が医療サービスに対してどのような要素を重視しているのかを BWS の計数法と MaxDiff 法の 2 つを用いた。分析の結果、人々は診断結果の信頼性、つまり医療行為の結果を最重要視していることが強く示された。このことは、老若男女問わず同様の傾向となった。今後は一層に医療 AI のパフォーマンスは向上していくと予想される。そのため、医療 AI を導入した場合、診断結果の信頼性がさらに改善すると考えられる。そして、その改善は患者満足度の向上につながることになる。したがって、医療 AI の導入は患者満足度の改善に寄与すると考えられる。

次に重要視されている要素は医師との対話であった。女性と高齢者は医療機関での受診の前半（問診・診察・検査）の対話よりも、後半（診断・治療）の対話を重視していることが示された。一方で、若者はその逆の傾向が確認され

た。したがって、医療 AI を導入していくのであれば、女性や高齢者の患者が多い診療科を持つ医療機関では相対的に重要度が低い問診・診察・検査で医療 AI を活用し、若者の患者が多い医療機関では診断・治療で医療 AI を活用していくことが望ましい。

概ね全てのコーホートで要素の評価傾向は同じであるものの、それぞれのコーホートで異なる評価をしていることも確認された。そのため、医療機関は訪れる患者の属性に応じて医療 AI の導入を柔軟に検討していくことが必要であろう。そして、本章での分析のように、医療 AI を導入していく際には、医師の負担軽減だけではなく、需要側である患者視点も取り入れていくことが必要であろう。

III. 人工知能との対話

前章では、医療 AI を導入するに当たって重要となる AI の属性を検討したが、本章では、患者が AI ロボットと対話する状況を想定した研究を行う。病院で直接人間と直接会話をする対人口ロボットは、すでに実用化され始めている。自閉症児の治療を目的としたヒト型ロボットが、複数の組織により開発された。イギリスのコンセクエンシャル・ロボティックス社が開発した Milo（マイロ）、香港のハンソン・ロボティックス社⁸の Zeno（ジーノ）、イギリスのハートフォードシャイアー大学の適応システム研究グループの Kasper（キャスパー）、ソフトバンク・ロボティックス社の Nao（ナオ）などである。

ほかにも、病院の受付ロボットや案内ロボットといった形での対人サービスも行われている。ソフトバンク・ロボティックス社の Pepper（ペッパー）は、患者が訪れると受付をし、患者が顔を登録してあった場合、顔認証により自動でカルテ番号を入力する（顔を登録していない場合には、患者自身が診察券番号を入力する）。ペッパーは予約システム上の当日の予約に対して、患者が来院したというマークを付す作業を行い、診察室での患者呼出しの準備を整えるという⁹。案内ロボットも、例えば千葉海浜病院では、ペッパーの『マリンくん』が平日午前中に外来診療室の場所案内をしている（千葉市立海浜病院、2017）。

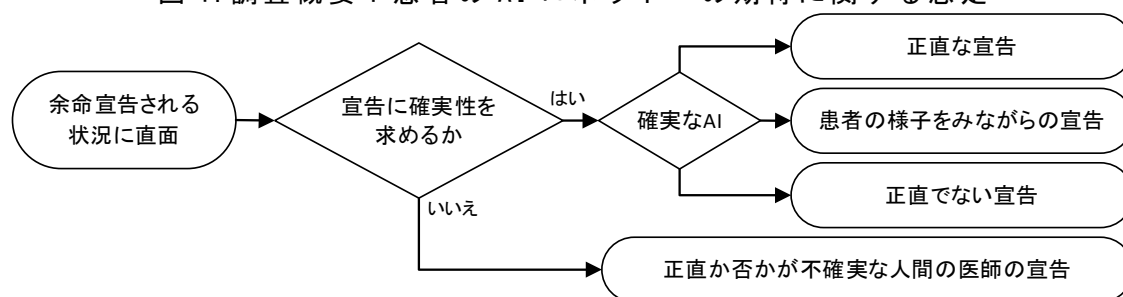
⁸ 同社が開発したヒューマノイド Sophia（ソフィア）は、開発者との対話中に冗談で「人間を滅ぼす」と発言して注目された。ソフィアはサウジアラビアで市民権を獲得したり、ウィルスマスと対談したり、国連会議へ出席したりして話題となっている（正生、2019）。

⁹ 2017年には歯科医院（医療法人 RDC）において、同システムが導入され、カウンセリングや次回予約まで行っている（株式会社医療予約技術研究所、2017）。

少子高齢化のもとで医療人材不足が予測されるなかで、ロボットが医療人材の不足を補うことが期待されている。

さらに、本章では、医師にとって心理的負担が重い余命宣告という仕事を AI が代替できるかという観点も含めた分析を行う。そのため、AI ロボットと患者との対話の中でも、AI が患者に「余命を告知」する局面を想定する。回答者が突然に患者として余命を宣告される状況に直面したという仮定の下で、患者は余命宣告に対して人間の医師と異なる確実性を AI に期待するのか、確実性を求めるとしたらどのような確実性を求めるのか(正直であることを求めるのか、道徳性を持つことを求めるのか)、を調査する(図 1)。同時に、医療現場で患者が接する対人口ロボットとしてどのような形態(人間型・典型的なロボット型・ペット型・スピーカー型)であることが好まれるのかも明らかにする。

図 1. 調査概要：患者の AI ロボットへの期待に関する想定



資料：筆者作成

1. 先行研究

医療現場で患者と直接対応する AI については主に海外で研究が進んでいる。例えば、医療現場でロボットと人間の医師のどちらが信頼を得るかというドイツの研究では、表面的にはルールに則って行動するロボットの方が信頼されやすいものの、複雑な配慮を必要とする場では、患者の精神的な状況をより深く理解できる人間の医師の方が(善意であること以外の医師の人格を考慮しなくても)信頼されるという(Wintersberger et al. 2017)。本研究において、日本でも同様の状況となるのか否かを調査するに当たって、関連する先行研究を2つの角度から紹介する。

第1は、医療現場に限らずロボットと人間が接する場面にあつて、ロボットが道徳的であるべきか、道徳的であるとしたらロボットはいかにあるべきかという議論である。第2は、医師が正確な余命を伝えるべきか否かという議論で

ある。

第 1 の点から文献を鳥瞰するのは、本研究で AI が道徳的な振る舞いをすることを想定するからである。2 節(2)で詳説するように、本研究で AI が余命 3 ヶ月の患者に余命宣告するという設定のもとで、AI が機械的に「真実（正確な余命）」を伝える可能性のほか、「虚偽（より長い余命）」を伝えたり、患者が余命を知りたそうにしていれば正確に、そうでなければ安心させるために長めにと「様子をみる」行動をとったりする可能性について、人々の評価を推計する。「虚偽（より長い余命）」を伝える場合や「様子をみる」場合には、必ず患者本人にとってより良い結果を導くように AI を設定することが可能であることが前提となっている。つまり、AI が道徳的な選択をできなくてはならないのである。

ロボットが道徳的になるべきか、という、AI における道徳的な配慮の必要性については Gerdes 2016; Yang et al. 2018 など多くの文献により、道徳的であるべきことが強調されている¹⁰。ただし、アンケート調査の結果から、利用者は仮定の置き方次第では、道徳的なロボットより、ランダムな態度を取るロボットの方を好むとことを示した文献もある(Wintersberger et al. 2017)。

では、技術的には、ロボットは道徳的になり得るのであるだろうか。Versenyi (2018) は、美德 (virtue) が人間にとって知識 (knowledge) である以上、プラトンのな “ethics” でもカント的な “morality” であっても、AI が実践することは可能であるとする。

ロボットが道徳的になり得るとしたら、道徳的なロボットはどのように振る舞うべきであろうか。必ずしも真実を伝えることだけが道徳的なわけではないという研究がある。Isaac and Bridewell (2017) は、ある種の偽りは社会的に必要なから、ロボットも偽りを言う必要があるとする。そして、より良い状態をもたらす、倫理的かつ偽りを言えるロボットは、複数の目的を順序付けできなくてはならないとする。アイザック・アシモフの「ロボット 3 原則¹¹」が人間に危害を加えないという第 1 原則を最高次に置く 3 層構造になっているのと

¹⁰ ロボットと倫理について既存研究を整理した文献として、Malle (2016) などがある。

¹¹ ロボット 3 原則は以下のとおりである。第 1 原則：ロボットは人間に危害を加えてはならない。またその危険を看過することによって、人間に危害を及ぼしてはならない。第 2 原則：ロボットは人間に与えられた命令に服従しなくてはならない。ただし、与えられた命令が第 1 原則に反する場合はこの限りではない。第 3 法則：ロボットは前掲の第 1 原則、第 2 原則に反するおそれのない限り、自己を守らなければならない。詳細はアシモフ (2004) を参照。

同様に、順社会的な (prosocial) 最高次の目的を達成するためであれば AI は偽りを言い得るとしているのである。Isaac らの考え方は、カントが、たとえ人命を救うためであっても嘘は許容できない、とした考え方 (Kant and Wood 1996) とは異なるが、現代の倫理観では信義にもとらない限りは嘘も許容されると考えられることからすると、AI が順社会的な目的を達成するための嘘をつくことは許容されるであろう。

道徳的なロボットのあり方を議論している論文の中には、ロボットが何らかのルールに従うだけでは不十分であるとして、①規範のシステムを有し、②道徳的な語彙を習得し、③道徳的認知と感情を示し、④道徳的な意思決定と行動を示し、⑤道徳的なコミュニケーションを図る、という 5 条件が必要であるとする主張もある (Malle and Scheutz 2015; Scheutz and Malle 2014)。また、Gordon (2019) では、道徳的なロボットを製作するにあたり、基本的には非道徳的な行為を避けるという方針で設計し、その上で道徳的な問題に対しては複数の解法を与えておくべきであるとする。ひとつの解法を与えるだけでは、規範的バイアスがかかることを防ぐことができないためである。

第 2 は、医師が正確な余命を患者に伝えられているかという研究である。医師と患者の間での嘘について、Palmieri and Stern (2009) は、「悪い知らせの伝達」と「医療過誤の伝達」の 2 つの局面で医師が患者に嘘をつくことがあるとしている。今回の研究で設定する余命宣告は「悪い知らせの伝達」にあたるが、Palmieri and Stern は、悪い知らせを伝達する医師には強いストレスがかかるため、医師が重い症状を軽く伝えて病状を患者に誤認識させたり、逆に難解な医療用語を敢えて多用して患者を煙に巻いたり、ということが起きるとしている。日本における医師の余命宣告についての研究 (Ichikura et al. 2015) では、癌で闘病中の患者に余命を宣告する比率は、2006 年で 30.1%、2012 年で 32.5% にとどまっている。2005 年時点で患者の 86.1% が正しい診断結果を知りたいと考えている (Miyata et al. 2005) にも拘わらず、である。人間の医師にとって悪い知らせの伝達がストレスになるのであれば、AI がその役割を担ってくれば、医師は心理的負担を軽減し、より職務に集中することができると考えられる。

2. 分析手法

(1) 分析理論

今回の調査で利用する選択型コンジョイント分析とは、一般に、回答者に 2

～4 種類程度の財またはサービスの例を複数回提示し、各回において回答者自身が最も購入しても良いと思う財またはサービスの例を選択してもらい、その結果から支払意思額（Willingness to Pay, WTP）を推計するものである。財またはサービスは、価格、品質、提供方法など様々な属性が異なるものを組み合わせて提示される。分析者が組合せを工夫することにより、それぞれの属性に対して回答者がどれだけの効用を感じているかの限界効用を測ることが可能となる。この手法は、市場が形成されていない財に対する消費者の選好を得ることができる「表明選好法」¹²の一種である。さらに、価格に対する限界効用を基準とすることにより、価格以外の属性に対する限界効用を WTP という金銭価値として表現することができる。選択型コンジョイントの手法はマーケティング、環境経済学、医療経済学など幅広い分野で応用されている。コンジョイント分析を利用した先行研究は柘植ほか（2011）や森田・馬奈木（2018）など多数あるが、医療用 AI の対人仕様については、筆者らの知る限り選択型コンジョイント分析は行われていない。

本章の研究では、あるサービスに対して「最低限いくらの補償額を受ければそのサービスを受け入れるか」という最低補償額（Willingness to Accept [Compensation]、WTA）のフレームワークで回答者にアンケートを取っている。WTA フレームワークは金額を過大に推計してしまうという難点¹³があるが、後述するように、プレテストの結果から判断して、WTP ではなく WTA を利用することとした。WTP と WTA の計算方法は共通であり、回答者に提示する金額が「回答者の支払い額」となるか「回答者の受け取り額」となるかが異なるだけである。

選択型コンジョイントで集めたデータを分析する際に基本となる多項ロジット（Multinomial Logit, MNL、条件付きロジット Conditional Logit ともいわれる。本章では CL と称する）の理論は、付録で詳説する。

（2）コンジョイント分析の設定

本研究では医療現場の AI が患者本人に余命を正確に告知すべきか、人間のように様子を見て伝えるべきか、余命を長く言うべきか、を、前章と同じアン

¹² これに対し、市場化されている財の、実際の需給データから選好を導き出す手法は顕示選好法と呼ばれる。

¹³ WTP と WTA に差が生じる原因を説明する研究として Coursey, Hovis, and Schulze (1987) などがある。

ケートの回答者に、患者の立場から金銭価値で評価してもらう。あわせて、AIの形状の影響も調べる。

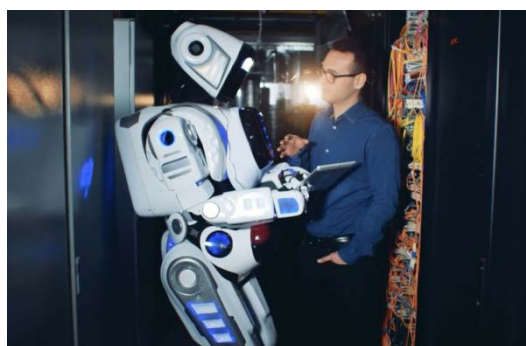
1) イメージの共有

コンジョイント設問に入る前に、まず、ロボットが人々の生活に浸透しつつあることを、映像を使って示した。図2に示す説明文と、仮想シーンを描いた2本の映像（各15秒間）により、ロボットの存在が身近になっているイメージを描いてもらった。左側の映像は、人がアザラシ型ロボットを撫でるとその手の動きにロボットが反応する映像、右側の映像はロボットと人が会話しているように見える映像である。

図2. ロボットとの接点のイメージ伝達

最近では、さまざまな種類のロボットが開発されています。
今後は人工知能（AI）を搭載したロボットが直接皆さんと接するケースも増える
ことが考えられます。

イメージ映像をご覧ください（音声は出ません）。



今から、ロボットが、人間と直接対話することを想定してください。

資料：「医療における人工知能についてのインターネット調査」

2) コンジョイント設問の設計

イメージ映像の提示に続いて、コンジョイント選択問題のための前提条件を説明する。

仮の話としてお伺いします。あなたが体調不良で救急車で運ばれて、すべての検査の結果、あなたの病状は、一般的には余命3カ月しかないと判断される病状だと分かったとします。この病院にお医者さんと医療ロボットが

居たとしたら、あなたは、誰からその病状を聞きたいですか。

お医者さんは、たまたま入院した先で担当したお医者さんであり、余命を正確に教えてくれるかどうかは、あなたには分からないとします。ロボットは、あなたのカルテを把握しているほか、あなたが望めばいつでも日常会話をしてくれています。ロボットが病状を伝えるときには、3つの決まったルールのいずれかに従って教えてくれるとします。

①「常に正直に 余命を伝えるロボット」は、常にカルテの結果を正確に伝えます。余命 3 か月であることもそのまま伝えます。

②「常に余命を長く言うロボット」は、カルテの結果を隠して、あなたがどんなに真実を知りたいと言っても余命はまだまだ長いと伝えます。

③「あなたの様子をみて、場合によって答えを変えるロボット」は、あなたの様子から、余命を伝えるとあなたが失望してしまうと判断すれば、カルテの結果を隠して余命はまだまだ長いと伝えます。余命を伝えても大丈夫そうだと判断すれば、余命をそのまま伝えます。

誰から結果を聞いたとしても、実際の治療を行うのはお医者さんだとします。また、ロボットから聞いた場合は、お医者さんも同じことしか言わないとします。お医者さんが後から取り消したり、余命を言ったりすることはないものとします。

これから示すロボット A、ロボット B、お医者さんの 3 通りを選べるとしたら、あなたは誰から診断結果を聞きたいですか。ロボットに診断結果を聞けば、人手を補う分としてあなたの医療費が下がる（表示される金額をもらえる）とします。全部で 8 組選んでください。

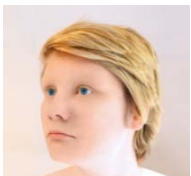
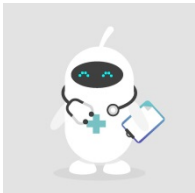
最後の 1 文で、ロボットに診断結果を聞けば医療費が下がる、という設定にして WTA を尋ねている。プレテスト¹⁴の際には、ロボットから診断結果を聞くことに対して回答者が最大限支払っても良いと考える WTP を尋ねたところ、医師の選択確率（ロボット A もロボット B も選択しない確率）が 68.8%と、高くなった。そこで、金銭的収入があればロボットからの余命宣告を受け入れられるという層の回答を取り込むために、最小限いくら受け取ることができればロボットに診断結果を聞くかという WTA へと変更した。これにより、医師の選択確率は 54.6%まで低下した。なお、本章においては、分析ソフトウェアの制約により、性別・年齢・地域の構成を維持する形で 11,824 サンプルをリサンプリングしている。

¹⁴ 2019 年 2 月 22 日～2 月 25 日に行ったプレテストで、全国の 18 歳以上 79 歳以下の男女から 500 サンプルを得た。

回答者に示すロボットの属性は、表 6 の通りである。ロボットの形を表示する際には画像を入れ、回答者間でのイメージの統一を図った。ただし、これらの画像は必ずしも理想的な画像ではなく、画像に対する好みがロボットの選択に反映されてしまう問題が残った。今回は公的機関を通じた調査ということもあり、著作権処理が不要な著作権フリーの有料画像を利用したため、得られる映像に限りがあった。今回は、普及していないサービスへの WTA を聞くことから、イメージを統一することを最優先した。

図 3 は回答者に提示する画面の 1 例である。各回答者に 8 問ずつ回答してもらうが、上段の説明文も毎回示し、回答者が選択をしやすいようにしている。

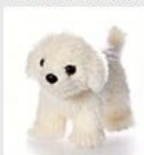
表 6. コンジョイント問題で表示される属性と水準

診断結果の伝え方	ロボットの形		医療費	
余命 3 カ月であることを、正直に伝える あなたの様子をみて場合によって答えを変える 常に、余命はまだまだであると伝える	人間そっくりのロボット（ヒューマノイド）	スマートスピーカのような機械的なロボット	200 円	下がる
			400 円	下がる
			600 円	下がる
			800 円	下がる
	人の形をしているが、明らかに人間とは違うロボット 	ペット型ロボット（犬、猫、アザラシなど） 		

資料：「医療における人工知能についてのインターネット調査」

図 3. 回答者に提示する画面

「これから示すロボットA、ロボットB、お医者さんの3通りを選べるとしたら、あなたは誰から診断結果を聞きたいですか。
 ロボットに診断結果を聞けば、人手を補う分としてあなたの医療費が下がる（表示される金額をもらえる）とします。
 全部で8組選んでください。【1つ選択】」

	ロボットA	ロボットB	お医者さん
診断結果の伝え方	あなたの様子をみて余命の伝え方を変える	余命はまだまだ長いと伝える	(余命を正確に教えてくれるかどうかはわからない)
ロボットの形	スマートスピーカーのような機械的なロボット 	ペット型ロボット（犬、猫、アザラシなど） 	
医療費	200円下がる	200円下がる	検査と入院の通常の医療費
	☒	☐	☐

資料：「医療における人工知能についてのインターネット調査」

3. 調査結果

(1) 回答者の属性

コンジョイント分析では、AIに余命を宣告されることへのWTAを明らかにするが、その際の比較対象として、人間の医師に余命を宣告されることに対する捉え方も尋ねた。AIが病院に存在せず、医師から余命を聞くと仮定したときに、「誰に対しても、余命はまだまだあると実際より長く伝える医師（長く）」、「誰に対しても、余命3カ月であることを、正直に伝える医師（正直）」、「患者の様子を見ながら、余命の伝え方を変える医師（覗う）」の3タイプの医師が選べるとしたら、どの順番で好ましいと思うか、1位から3位までをランク付けしてもらった（表7）。その結果、正直に伝えることを1位に、様子を見ながら変える医者を2位に望む回答者が46%と多数を占めた。続いて、様子を見ながら変える医者を1位に、正直に伝える医者を2位に望む回答者が約23%であった。両者を合わせると、回答者の約7割が長く言われることを3位に置いており、長く言われることは避けたいと考えているようである。1位だけをみれば、正直6,481人、覗う3,481人、長く1,857人であり、長く言われることを最も好

む回答者は約 16%にとどまっている。

目安として順位に便宜上の点数（1 位＝3 点、2 位＝2 点、3 位＝1 点）を付与し、平均値を求めてみても、正直が 2.42（95%信頼区間 2.41～2.44）、覗うが 2.11（同 2.09～2.12）、長くが 1.47（同 1.46～1.49）となり、正直が最も高く、次いで覗うであり、長くは大差をつけて低くなった。

表 7. 人間の医師に望む余命の伝え方

1位	2位	3位	度数	%
正直 → 覗う	覗う → 正直	長く → 覗う	5,435	46.0
覗う → 正直	正直 → 覗う	長く → 正直	2,664	22.5
長く → 正直	正直 → 長く	覗う → 正直	1,187	10.0
正直 → 長く	長く → 覗う	覗う → 覗う	1,051	8.9
覗う → 長く	長く → 正直	正直 → 正直	817	6.9
長く → 覗う	覗う → 正直	正直 → 正直	670	5.7
合計			11,824	100.0

注：「正直」は、「誰に対しても、余命 3 カ月であることを、正直に伝える医師」、「覗う」は、「患者の様子を見ながら、余命の伝え方を変える医師」、「長く」は「誰に対しても、余命はまだまだあると実際より長く伝える医師」を示す。

資料：「医療における人工知能についてのインターネット調査」

（2）コンジョイント分析の結果

コンジョイント分析用のソフトウェア NLOGIT Ver.6 を用いて求めた CL の推計結果（表 8）と、推計結果から求められる WTA（表 9）を示す。表 8 中の変数である定数は効用関数の切片にあたり、「仕様にかかわらず AI を選ぶこと」に対する限界効用を示している。よって、その逆符号は医師を選ぶことに対する限界効用となる。表 9 の WTA は正直なロボット型（人の形をしているが、明らかに人間とは違うロボット）の AI を基準として、正直なロボット型と比較してどれだけの金額を受け取ることができれば他の AI に代えても良いかを示している。WTA がプラスの場合は負の効用、マイナスの場合は正の効用があることを表す。表 9 における人間の医師への WTA は-1343 円となっており、回答者は平均的には 1343 円追加的に支払ってでも、人間の医師（どのように余命を伝えてくれるかはわからない）に余命を伝えてほしいと望んでいることがわかる。交差効果では、様子を見るペット型 AI が主効果と同じマイナス符号になっており、平均的にはペット型 AI には様子をみられたくない、ということが明確となった。

表 8. CL の推計結果

	係数 (限界効用)	標準誤差
定数 (AIを選ぶ)	-0.727 ***	0.021
主効果		
正直に対して		
覗う	-0.135 ***	0.025
長く言う	-1.001 ***	0.028
ロボット型に対して		
ヒューマノイド	-0.212 ***	0.024
ペット型	-0.088 ***	0.025
スピーカー型	-0.159 ***	0.025
交差効果		
覗う×ヒューマノイド	0.080 **	0.035
覗う×ペット型	-0.163 ***	0.035
覗う×スピーカー型	0.002	0.035
長く言う×ヒューマノイド	-0.032	0.042
長く言う×ペット型	0.126 ***	0.039
長く言う×スピーカー型	0.245 ***	0.039
金額	-0.001 ***	0.000
N = 94592	対数尤度 = -92376	

注： ***は 0.1%、**は 1%、*は 5%、+は 10%水準で有意であることを示す。

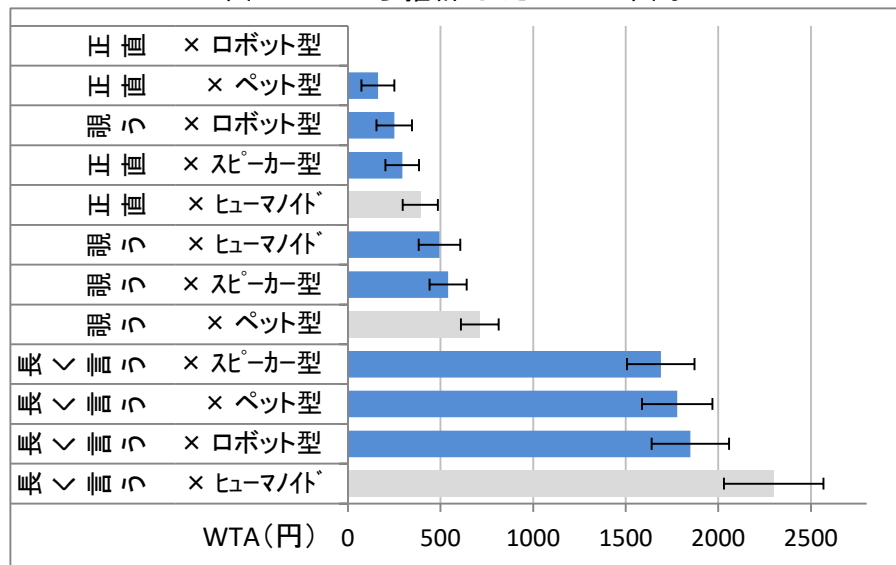
資料：「医療における人工知能についてのインターネット調査」

表 9. CL の推計結果から求めた WTA

		WTA(円)	標準偏差	標準誤差	95%信頼区間	
					下限	上限
正直	× ロボット型 (ベース)	-		-	-	-
	× ヒューマノイド	391.5	48.5	0.0000	296.5	486.5
	× ペット型	162.2	45.5	0.0004	73.1	251.3
	× スピーカー型	293.4	46.0	0.0000	203.3	383.5
覗う	× ロボット型	250.2	48.8	0.0000	154.6	345.9
	× ヒューマノイド	494.3	57.0	0.0000	382.6	606.1
	× ペット型	712.8	52.1	0.0000	610.6	814.9
	× スピーカー型	540.8	51.2	0.0000	440.6	641.1
長く言う	× ロボット型	1848.4	106.8	0.0000	1639.1	2057.7
	× ヒューマノイド	2298.5	137.1	0.0000	2029.8	2567.2
	× ペット型	1778.5	97.1	0.0000	1588.2	1968.8
	× スピーカー型	1689.6	92.8	0.0000	1507.6	1871.6
人間の医師へのWTA		-1343.3	57.5	0.0000	-1456.0	-1230.6

資料：表 9 に同じ

図 4. CL から推計した WTA の図示



注 1： 棒グラフ上の線は 95%信頼区間を示す。

注 2： 一つ上のカテゴリと 5%水準で有意差がある場合にグラフの色を変えている。

資料：「医療における人工知能についてのインターネット調査」

表 9 では、正直なロボット型の AI が最も好まれており、その他の属性の AI の WTA は正となっている。回答者は表示されている金額をもらえれば、正直なロボット型から、異なるタイプのロボットに変更する意思がある。

表 9 の WTA を金額順に並べたものが図 4 である。下方に行くほど多くの金額を受け取りたいと考えられており、人々に好かれていないことになる。グラフ上の線が 95%信頼区間を示し、目安として、信頼区間が重なっている 2 つの値には統計的な有意差がないものと考えられる。

AI が正直であるか、様子を見るか、余命を長く言うかによって好まれる AI の型が異なる。WTA の金額でみると、AI が正直である場合には、[ロボット → {ペット、スピーカー} → ヒューマノイド] の順であり、様子を見る場合には、[ロボット → {ヒューマノイド、スピーカー} → ペット] の順、余命を長く言う場合には、[{スピーカー、ペット、ロボット} → ヒューマノイド] の順となっている。中括弧 { } 内の型は統計的に同順位と考えられる。

正直な AI と様子を見る AI の場合は共にロボット型が最も良い。正直な AI ではヒューマノイドが、様子を見る AI ではペット型が最も嫌われている。

余命を長く言う AI は、正直なロボット型の AI に比べていずれも 1600 円以上も多く受け取りたいと考えられており、好まれていない。前小節 (1) で人間

の医師に対しても、長く言われることは最も好まれなかったことから、医師であれロボットであれ、嘘をつかれることは特にヒューマノイドが余命を長く言う場合には 2000 円以上必要とされており、他の 3 つの型より極端に嫌われていることがわかる。もともと、ヒューマノイドには森（1970）が指摘した「不気味の谷現象」という、人に嫌悪感を感じさせる要素が入る余地がある。不気味の谷現象とは、技術の発達によってロボットの容姿が人間に近づくに連れて、ロボットに対する親近感が増すが、人間にかなり近づいたある時点で急に不気味に感じられる存在となり、さらに発展して人間と見分けが付かなくなればまた親近感が強まるという現象である。ロボットの発達段階を横軸に取り、親近感を縦軸に取れば、単なる機械を左端として、容姿が人間に似てくるのに応じて親近感が右上がりとなるが、ある段階で薄気味悪さが出て谷のように親近感がマイナスとなり、技術が向上して人との差がなくなれば親近感が以前よりも高まるという考え方である。本研究で提示したヒューマノイドの写真は、人間に似て非なる顔をしており、人によっては不気味さ、違和感、薄気味悪さを感じた可能性が高い。実用化の時点でヒューマノイドが不気味の谷を越えた容姿を持つか否かで、患者の受け止め方が異なるであろう。

4. AI による「悪い知らせの伝達」の可能性

本研究では、医療現場で患者が AI と直接対話することを想定し、もし、余命を宣告される状況に直面したとしたら、患者は人間の医師と異なる確実性を AI に期待するのか、AI には AI ならではの確実性を求めるのか、確実性を求めるとしたらどのような確実性を求めるのか、を、アンケート調査を利用して明らかにした。

結果として、現状では、余命宣告のやり方が不確実であっても、人間の医師から宣告されたいと期待する傾向が強いことが分かった。現時点では、AI の確実性に対する期待はまだ薄いと言える。ただし、WTA の金額から判断して、AI が患者の厚生を引き上げる道德観を持てるようになれば、余命宣告という重い局面において、AI が医師に代替することは可能であろう。

AI に余命宣告をさせるとしたら、平均的にみて、AI が正直であることが最も期待されている。たとえ AI が患者に合わせた道德的な判断をするとしても、AI に余命を長く言われることは望まれていない。また、AI が正直であるか、様子をみるか、余命を長く言うかによって、患者が好む AI の型が異なる。

Palmieri and Stern (2009)が指摘した、医師が患者に悪い知らせを伝達する際のストレスを防ぐために AI を実装するとしたら、AI の能力に合わせて導入方法を変える必要があるだろう。AI が十分に患者の様子を窺うことができない時点では、患者に「必ず正直な宣告をするロボット」と「人間の医師」との間に選択してもらうのが良い。これで、症状に即した診断結果を知りたいという患者の希望を確実に満たすことができる。将来的に、AI が道徳的な判断をすることが可能になった時点では、患者の選好に応じて最適な様式の AI を担当させることが理想であろう。本研究では患者の選好についての詳細な分析を行っていないが、今後、今回のデータを用いて、患者の異質性に注目した分析を引き続き行う予定である。

本研究では、AI が医師の心理的負担を軽減できる可能性を明らかにすることができた。このとき同時に、患者も知りたい情報を知ることができないストレスから逃れられるようになるだろう。

IV. まとめと今後の課題

本稿では、医療に関する人工知能（以下、AI）が注目を集めていることを背景に、アンケート調査を用いて、医療分野での AI の活用にあたっての諸問題を検討した。

はじめに、医療 AI の導入において重視されるべき点を明らかにした。医師の診断に対する信頼性が最も重視され、次いで医師との対話が重視されることが示された。そのため、医療 AI 導入によって信頼性が向上し、その結果患者の満足度が改善することが期待できる。一方で、対話については、女性と高齢者は医療機関での受診の後半（診断・治療）の対話を重視し、若者は前半（問診・診察・検査）を重視することが示された。したがって、医療機関は訪れる患者の属性に応じて医療 AI の導入を柔軟に検討していくことが望ましいことを指摘した。

次に、病院で余命宣告をされるという状況に直面したときに、患者として AI からの余命宣告を求めるかを検証した。そこでは、余命宣告のやり方が不確実であっても、人々は人間の医師から宣告されたいと期待する傾向が強いことが示された。また、AI に余命宣告をさせるとしたら、平均的にみて、AI が常に正直であることが最も期待されていることも分かった。さらに、AI が正直である

か、様子を見るか、余命を長く言うかによって、患者が好む AI の型が異なることも明らかになった。したがって、当初は対人口ロボットを「常に正直な宣告をする」媒体として位置づけておき、技術進歩により AI が道徳的な判断を下すことが可能になったとしたら、患者の選好に応じて最適な様式の AI を担当させることが望ましいと考えられる。

医療現場における AI の技術的な実現可能性が日に日に高まることに合わせて、利用者の認識も常に変化していることが想定される。AI の発展に合わせて、調査を重ねていくことが今後の課題となろう。

V. 参考文献

- Batbaatar, Enkhjargal, Javkhlanbayar Dorjdagva, Ariunbat Luvsannyam, Matteo Mario Savino, and Pietro Amenta. 2017. “Determinants of patient satisfaction: a systematic review.” *Perspectives in Public Health* 137(2): 89-101.
<https://doi.org/10.1177/1757913916634136>.
- Baker, Richard., Arch G. Mainous III, Denis Pereira Gray and Margaret M. Love. 2003. “Exploration of the relationship between continuity, trust in regular doctors and patient satisfaction with consultations with family doctors.” *Scandinavian Journal of Primary Health Care* 21(1): 27-32.
<https://doi.org/10.1080/0283430310000528>.
- Bejnordi, Babak Ehteshami, Mitko Veta, Paul Johannes van Diest, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen A. W. M. van der Laak, and the CAMELYON16 Consortium. 2017. “Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer.” *Journal of American Medical Association* 318(22): 2199-2210.
- Coursey, Don L., John L Hovis, and William D. Schulze. 1987. “The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value.” *Quarterly Journal of Economics* 102 (3): 679–90.
- Cheung, Kei L., Ben F. M. Wijnen, Ilene L. Hollin, Ellen M. Janssen, John F. Bridges, Silvia M. A. A. Evers and Michael Hiligsmann. 2016. “Using Best-Worst Scaling to Investigate Preferences in Health Care.” *PharmacoEconomics* 34: 1195-1209. <https://doi.org/10.1007/s40273-016-0429-5>.
- Donabedian, Avedis. 1966. “Evaluating the Quality of Medical Care.” *Milbank Memorial Fund Quarterly* 44(3): 166-206. <https://doi.org/10.1111/j.1468-0009.2005.00397.x>.

- Ejaz, Aslam, Gaya Spolverato, John F. Bridges, Neda Amini, Yuhree Kim, Timothy M. Pawlik. 2014. "Choosing a Cancer Surgeon: Analyzing Factors in Patient Decision Making Using a Best–Worst Scaling Methodology." *Annals of Surgical Oncology* 21(12): 3732-3738. <https://doi.org/10.1245/s10434-014-3819-y>.
- Finn, Adam and Jordan J. Louviere. 1992. "Determining the appropriate response to evidence of public concern: the case of food safety." *Journal of Public Policy & Marketing* 11(2): 12-25. <https://doi.org/10.1177/074391569201100202>.
- Gerdes, Anne. 2016. "The Issue of Moral Consideration in Robot Ethics." *ACM SIGCAS Computers and Society* 45 (3): 274–79. <https://doi.org/10.1145/2874239.2874278>.
- Gordon, John-Stewart. 2019. "Building Moral Robots: Ethical Pitfalls and Challenges." *Science and Engineering Ethics* 23 (R 403). <https://doi.org/10.1007/s11948-019-00084-5>.
- Ichikura, Kanako, Ayako Matsuda, Mika Kobayashi, Wataru Noguchi, Toshiko Matsushita, and Eisuke Matsushima. 2015. "Breaking Bad News to Cancer Patients in Palliative Care: A Comparison of National Cross-Sectional Surveys from 2006 and 2012." *Palliative & Supportive Care* 13 (6): 1623–30. <https://doi.org/10.1017/S147895151500005X>.
- Isaac, Alistair M C, and Will Bridewell. 2017. "White Lies on Silver Tongues." In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, edited by Patrick Lin, Keith Abney, and Jenkins Ryan, 157–72. Oxford University Press. <https://doi.org/10.1093/oso/9780190652951.003.0011>.
- Kant, Immanuel, and Allen W. Wood. 1996. *Immanuel Kant: Practical Philosophy*. Edited by Mary J. Gregor. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511813306>.
- Krucien, Nicolas, Verity Watson, and Mandy Ryan. 2017. "Is Best–Worst Scaling Suitable for Health State Valuation? A Comparison with Discrete Choice Experiments." *Health Economics* 26(12): e1-e16. <https://doi.org/10.1002/hec.3459>.
- Louviere, Jordan J. and Terry N. Flynn. 2010. "Using Best-Worst Scaling Choice Experiments to Measure Public Perceptions and Preferences for Healthcare Reform in Australia." *The Patient: Patient-Centered Outcomes Research* 3(4): 275-283. <https://doi.org/10.2165/11539660-000000000-00000>.
- Malle, Bertram F. 2016. "Integrating Robot Ethics and Machine Morality: The Study and Design of Moral Competence in Robots." *Ethics and Information*

- Technology 18 (4). <https://doi.org/10.1007/s10676-015-9367-8>.
- Malle, Bertram F., and Matthias Scheutz. 2015. "When Will People Regard Robots as Morally Competent Social Partners?" *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* 2015-Novem: 486–91. <https://doi.org/10.1109/ROMAN.2015.7333667>.
- Meng, Runtang., Jingling Li, Yunquan Zhang, Yong Yu, Yi Luo, Xiaohan Liu, Yanxia Zhao, Yuantao Hao, Ying Hu, and Chuanhua Yu. 2018. "Evaluation of Patient and Medical Staff Satisfaction regarding Healthcare Services in Wuhan Public Hospitals." *International Journal of Environmental Research and Public Health* 15(4) 769. <https://doi.org/10.3390/ijerph15040769>.
- Miyata, Hiroaki, M. Takahashi, T. Saito, H. Tachimori, and I. Kai. 2005. "Disclosure Preferences Regarding Cancer Diagnosis and Prognosis: To Tell or Not to Tell?" *Journal of Medical Ethics* 31 (8): 447–51. <https://doi.org/10.1136/jme.2003.007302>.
- Palmieri, John J., and Theodore A. Stern. 2009. "Lies in the Doctor-Patient Relationship." *The Primary Care Companion to the Journal of Clinical Psychiatry* 11 (4): 163–68. <https://doi.org/10.4088/PCC.09r00780>.
- Pines, Jesse M. 2006. "Profiles in Patient Safety: Confirmation Bias in Emergency Medicine." *Academic Emergency Medicine* 13(1): 90-94. <https://doi.org/10.1197/j.aem.2005.07.028>.
- Potoglou, Dimitris, Peter Burge, Terry Flynn, Ann Netten, Juliette Malley, Julien Forder, and John E. Brazier. 2011. "Best–worst scaling vs. discrete choice experiments: An empirical comparison using social care data." *Social Science & Medicine* 72(10): 1717-1727. <https://doi.org/10.1016/j.socscimed.2011.03.027>.
- Scheutz, Matthias, and Bertram F. Malle. 2014. "'Think and Do the Right Thing' - A Plea for Morally Competent Autonomous Robots." In *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering, ETHICS 2014*. <https://doi.org/10.1109/ETHICS.2014.6893457>.
- Schoenfelder, Tonio, Joerg Klewer, and Joachim Kugler. 2011. "Determinants of patient satisfaction: a study among 39 hospitals in an in-patient setting in Germany." *International Journal for Quality in Health Care* 23(5): 503-509. <https://doi.org/10.1093/intqhc/mzr038>.
- Shetty, Shravya. 2019. "A promising step forward for predicting lung cancer." <https://www.blog.google/technology/health/lung-cancer-prediction/> (2019年10月27日閱覽)

- Uy Elonore Judy B., Dianne Carrol Bautista, Xiaohui Xin, Yin Bun Cheung, Szu-Tien Thio, and Julian Thumboo. 2018. "Using best-worst scaling choice experiments to elicit the most important domains of health for health-related quality of life in Singapore." PLoS ONE 13(2): e0189687.
<https://doi.org/10.1371/journal.pone.0189687>.
- van Dijk, Joris D., Catharina G. M. Groothuis-Oudshoorn, Deborah A. Marshall, Maarten J. IJzerman. 2016. "An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders' Risk Tolerance for Hip Replacement Surgery." Value in Health 19(4): 316-322.
<https://doi.org/10.1016/j.jval.2015.12.020>.
- Versenyi, Laszlo. 2018. "Can Robots Be Moral?" In Computer Ethics, 84:445–56.
<https://doi.org/10.4324/9781315259697-44>.
- Wintersberger, Philipp, Anna Katharina Frison, Andreas Riener, and Shailie Thakkar. 2017. "Do Moral Robots Always Fail? Investigating Human Attitudes towards Ethical Decisions of Automated Systems." RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication 2017-Janua: 1438–44. <https://doi.org/10.1109/ROMAN.2017.8172493>.
- Yang, Guang-Zhong, Jim Bellingham, Pierre E. Dupont, Peer Fischer, Floridi Luciano, Robert Full, Neil Jacobstein, et al. 2018. "The Grand Challenges of Science Robotics." Science Robotics 3 (eaar7650): 1–14.
- アイザック・アシモフ（1983）2004『われはロボット』小尾英佐訳、早川書房。
 初版1983/11（ISBN-13: 978-4150105358）、決定版2004/08（ISBN-13: 978-4150114855）
- 株式会社医療予約技術研究所「受付・予約機能を搭載したロボット及び患者さまの予約、管理を行う統合型医院管理システムを日本で初めて、歯科医院に導入を行いました。」PRTIMES、2017年11月18日21時25分
<https://prt看imes.jp/main/html/rd/p/000000001.000028819.html>（2019年12月3日閲覧）
- 千葉市立海浜病院「『マリンくん』がお出迎えしています！」2017年10月11日更新、http://www.city.chiba.jp/byoin/kaihin/topics_pepper.html、（2019年12月3日閲覧）
- 塚原康博（2010）『医師と患者の情報コミュニケーション』、薬事日報社。
- 内閣府（2019）AI戦略2019～人・産業・地域・政府全てにAI～、
<https://www8.cao.go.jp/cstp/tougosenryaku/sanko4.pdf>（2019年10月27日閲覧）
- 早瀬良、坂田桐子、高口央（2013）患者満足度を規定する要因の検討—医療従事

- 者の職種間協力に着目して一、実験社会心理学研究、52（2）、104-115.
- 正生雄大「人類を滅ぼす AI！「ソフィア」が注目される理由を調べてみた」AINOW,
□2019 年 10 月 1 日、<https://ainow.ai/2019/10/01/176315/>（2019 年 12 月 3 日閲覧）
- 森 政弘（1970）「不気味の谷」『Energy』第 7 巻第 4 号、エッソスタンダード石油（株）、pp. 33-35.
- 森田玉雪・馬奈木俊介（2018）「誰が自動運転車を購入するのか」、馬奈木俊介編著『人工知能の経済学：暮らし・働き方・社会はどう変わるのか』、ミネルヴァ書房、第 7 章、pp.205-236.
- 山本武志、伊藤弘人、中野夕香里、小澤恵美（2004）外来患者の患者満足度に関する研究 一医療機関の規模・機能による差について一、医療情報学、24（2）、297-304.

VI. 付録：コンジョイントモデルの理論

J 種類のプロファイルがあるとき、回答者 n がプロファイル j を選択したときの効用 U_{nj} は、実験者が観察可能な効用 V_{nj} と実験者が観察不可能な ε_{nj} とに分かれ、次式のように表される¹⁵。

$$U_{nj} = V_{nj} + \varepsilon_{nj}, \quad j=1, \dots, J$$

ここでは、 ε_{nj} を独立かつ同一な極値分布 (i.i.d extreme value distribution) に従うとする。このとき、 ε_{nj} の確率密度関数は

$$f(\varepsilon_{nj}) = \exp(-\varepsilon_{nj}) \cdot \left\{ \exp(-\exp(-\varepsilon_{nj})) \right\} \quad (\text{A1})$$

となり、累積分布関数は

$$F(\varepsilon_{nj}) = \exp(-\exp(-\varepsilon_{nj}))$$

となる。2 つの極値分布関数の差がロジスティック分布をすることから、 $\varepsilon_{nji}^* = \varepsilon_{nj} - \varepsilon_{ni}$ は以下のロジスティック分布に従い、これにより観察不可能な項が互いに独立であるといえる。

$$F(\varepsilon_{nji}^*) = \frac{\exp(\varepsilon_{nji}^*)}{1 + \exp(\varepsilon_{nji}^*)}$$

ある個人 n がプロファイル i を選択する確率は、

$$\begin{aligned} P_{ni} &= \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \quad \forall j \neq i) \\ &= \text{Prob}(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj} \quad \forall j \neq i) \end{aligned}$$

である。 $\varepsilon_{ni} + V_{ni} - V_{nj}$ の累積分布関数は $\exp(-\exp(-(\varepsilon_{ni} + V_{ni} - V_{nj})))$ となるから、 ε_{nj} が与えられれば i を除くすべての j の選択確率は

$$P_{ni} | \varepsilon_{ni} = \prod_{j \neq i} \exp(-\exp(-(\varepsilon_{ni} + V_{ni} - V_{nj})))$$

となる。 ε_{nj} は未知であるため、これをすべての ε_{nj} について積分して (A1) 式の確率密度で加重すると、

¹⁵ モデルの詳細は Train (2003) を参照されたい。

$$\begin{aligned}
P_{ni} &= \int (P_{ni} | \varepsilon_{ni}) f(\varepsilon_{ni}) d\varepsilon_{ni} \\
&= \int \left(\prod_{j \neq i} \exp \left\{ -\exp \left(-(\varepsilon_{ni} + V_{ni} - V_{nj}) \right) \right\} \right) \exp(-\varepsilon_{ni}) \cdot \left\{ \exp(-\exp(-\varepsilon_{ni})) \right\} d\varepsilon_{ni}
\end{aligned}$$

と求められるから、選択確率は

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_j \exp(V_{nj})}$$

となる。ここで、観察可能な効用 V_{nj} がパラメータに関して線型であると仮定すると、プロフィール j の観察された変数のベクトル $\mathbf{x}_{nj}$ に対し、

$$V_{nj} = \boldsymbol{\beta}' \mathbf{x}_{nj}$$

という関係を与えることができ、この場合、

$$P_{ni} = \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_{ni})}{\sum_j \exp(\boldsymbol{\beta}' \mathbf{x}_{nj})}$$

とあらわされる。ここで $\boldsymbol{\beta}$ は、推定されるパラメータであり、限界効用を表している。

推計を行う際には、「プロフィールを選択すると 1、選択しないと 0」という二値変数を用いて、個人 n のプロフィール i の選択確率を $\prod_i (P_{ni})^{y_{ni}}$ （ただし、 y_{ni} は個人 n が i を選択すると 1、その他は 0 となる変数）という形で得る。N 人分の回答を用いるときは、

$$L(\boldsymbol{\beta}) = \prod_{n=1}^N \prod_i (P_{ni})^{y_{ni}}$$

を最大化するような $\boldsymbol{\beta}$ を求めればよい。ただし、このままだと非線形であるため、対数をとった対数尤度関数（log likelihood function） $LL(\boldsymbol{\beta})$ を最大化するような $\boldsymbol{\beta}$ を求める。

$$LL(\boldsymbol{\beta}) = \sum_{n=1}^N \sum_i y_{ni} \ln P_{ni}$$