

2021 年度厚生労働行政推進調査事業費補助金（政策科学総合研究事業）
「医薬品・医療機器等の費用対効果評価における公的分析と公的意思決定方法に関する研究」

分担研究報告書

医療経済評価人材育成プログラムについて

後藤 励

慶應義塾大学 大学院経営管理研究科/大学院健康マネジメント研究科

1. はじめに

近年、医療資源の効率的な配分についての社会的要請が高まっている。保健医療の効率性を分析する学問分野として、医療経済学特に医療経済評価が各国で政策的な医療資源の配分決定に用いられている。日本では、2019 年度より医薬品・医療機器等の費用対効果評価が始まり、費用対効果評価を行う人材に対するニーズが高まっている。一方、医療経済評価は生物統計学や疫学等を基に治療アウトカムを定量的に分析する方法、心理学等を基に QOL を評価する方法、経済学等を基に費用や効率性を評価する方法についての理解が必要となる。また、実際の政策的な評価では、政策当局や医薬品・医療機器製造販売業者、臨床専門家との幅広いコミュニケーションも必要となる。こうした多彩な能力や経験を持った専門家はまだまだ少なく、費用対効果評価を進めるために必要な人材が常に不足しているというのが現状である。

慶應義塾大学大学院健康マネジメント研究では、2019 年度より国立保健医療科学院より「医薬品及び医療機器等の費用対効果評価のための人材育成プログラム開発事業」を受託して、「医療経済評価人材育成プログラム」を開始した。本稿では、プログラム内容とこれまでの修了者人数等の結果などを報告する。

2. 医療経済評価人材育成プログラムについて

本プログラムが設置されている慶應義塾大学大学院健康マネジメント研究科は、保健・医療・福祉・看護・公衆衛生に関わる幅広い領域に展開する大学院である。公衆衛生・スポーツ健康科学専攻（募集人員 30 名）と看護学専攻（募集人員 10 名）の 2 専攻で構成される。修士課程の修業年限は 2 年であるが、修士（公衆衛生学、医療マネジメント学）に限り、保健・医療・福祉の分野で一定の専門性と実務経験を有する者を対象に、早期修了制度を設けて在学 1 年での学位取得が可能となっている。それぞれの専攻に後期博士課程

(修業年限3年)が設けられている。

専任として教育を担当する研究科委員(教員)は22名であるが、本研究科を本務とする教員と他学部・研究科を本務としている教員が含まれる。研究科委員の構成は医学部、経済学部、商学部、経営管理研究科など多岐にわたっており、慶應義塾全体で学際的な教育・研究を担っている。

本プログラムでは、医薬品・医療機器等の費用対効果評価を行う人材を育成することを目的としているが、より幅広い人材に医療経済評価に興味を持ち学習と研究を行ってもらうために、ミッションとして「持続可能な保健・医療制度を支えるために、費用対効果評価を中心とした医療経済分析を、様々な視点を持って行うことができる人材を育成する」ことを掲げている。

本プログラムでは健康マネジメント研究科で行われる医療経済評価に関する授業を医療経済評価コース(HTAコース)と呼び、所定の単位を履修したものに対してサーティフィケート(修了証)を授与している。健康マネジメント研究科以外の慶應義塾大学の大学院学生の履修が可能である。さらに、学位課程に入学しない場合でも慶應義塾特別学生(科目等履修生)として履修することが可能である。

また、慶應義塾大学グローバルリサーチインスティテュート(KGRI)に医療経済・医療技術評価研究センターを2019年10月1日に設置し、有給の研究員として学生やポスドクが研究プロジェクトに対して参加することが出来るように整備している。さらに、2020年5月より慶應義塾大学医学部において医薬品・医療機器等の費用対効果評価にかかる公的分析班の業務を担うこととなった。HTAコース履修生、修了生等は、研究員や教員として実際の公的分析に参画する機会もある。

このように、教育(HTAコース)と研究(KGRI)、実務(公的分析班)が一体となり、学生は費用対効果評価研究プロジェクトのメンバーとして参加することで、費用対効果評価の専門家となる意義を学び、研究プロジェクトへの参加により得られた経験がキャリア形成の一助となることが想定される。

本プログラムでは、本学専任教員に加え、2022年4月現在、特任教授3名(非常勤)、特任講師2名(常勤)、特任助教1名(常勤)が健康マネジメント研究科で教育・研究指導を担っている。KGRIでは、特任講師1名(常勤)、特任研究員3名(常勤2名、非常勤2名)が研究を担っている。

HTAコース修了のための必修科目を図1に示す。全11科目20単位が必修となっており、標準的には1年半で20単位を履修することを想定している。疫学統計科目も医療経済評価科目も基礎から応用まで段階的に学習するようになっている。

疫学統計科目では、基礎疫学で観察研究および介入研究などの様々な研究方法についての基礎を学ぶ。生物統計学についても、基礎2科目、応用1科目の計3科目あり、確率と確率分布、統計量、推定と検定(点推定と区間推定や仮説検定の基礎)、相関と回帰、カテゴリカルデータの解析、ロジスティック回帰分析、生存時間分析、検出力・例数設計、

多変数モデル(multivariable model)や多変量データの取り扱い、線形混合モデル等を学ぶ
医療経済評価科目は、まず1学期目で 医療経済評価の考え方の概論を学ぶ「医薬経済学」があり、医療経済評価の関連学会である ISPOR 日本部会の理事・評議員を務めている講師に講義を依頼している。カリキュラム全体として1学期目は疫学統計学を集中的に2学期目以降は医療経済評価を中心に学ぶ。医療経済評価科目には、医療経済学全体やQOLと費用のそれぞれ個別のトピックを学ぶ科目がそれぞれ一科目ある。さらに、医療制度とレギュラトリーサイエンスでは、医療経済評価と大きく関連する、診療報酬制度や薬事制度、さらには日本の財政全体について学ぶ。医療経済評価特論は、医療経済評価に関する学術・政策の両面での最新のトピックについて、学生のディスカッションを基にしたワークショップ形式で行う。3学期目は、自ら医療経済評価ができるように演習科目を提供している。経済評価研究を批判的に吟味する費用対効果評価演習とモデル分析をワークショップ形式で学ぶ経済評価モデル分析演習が配置されている。

3. HTA コース修了者とその進路

2019年度から一部の科目を開講し、2020年度より本格的に開講した本コースでは、2020年度に18名、2021年度に23名が全20単位を取得しコース修了証を取得した。修了者41名の背景を表1に示す。医師や看護師等の医療資格者が全体の7割近くを占めるが、医療職以外の修了者も3割程度いた。

また本プログラムでは、医薬品・医療機器の製造販売業者の方々に対するHTA人材育成も重要だと考えているが、現職で勤務の方が4名、過去勤務経験がある方が2名修了されていた。製造販売業者に勤務経験がある修了者は薬剤師資格を持つものが6名中5名と多かった。

修了後の進路としては、社会人学生も多いため就学中の勤務先への勤務を継続するものや復職するものが最も多かった。ついで、就職(社会人学生の場合は転職)するものが多かった。博士課程に進学するものも11名と医療経済評価の研究を継続するものも一定程度は育成されていることがわかった。なお、慶應義塾大学医学部の公的分析部門には、HTA修了者やHTA科目履修者が2022年4月現在で6名(常勤の教員2名、非常勤の研究員4名)在籍している。国立保健医療科学院や他の公的分析機関に在籍しているものはまだでていないが、公的分析に関わるものの育成も行われつつある。

修了時に別途行ったアンケート調査によると、修了後の医療経済評価研究への参加意向として、強く希望としたものが約4分の1、どちらかといえば希望としたものが約2分の1、どちらともいえないとしたものが約4分の1であった。修了者の中でも医療経済評価への修了後の関わりについての希望は温度差があることがわかった。

4. 考察

本稿では、2019年度より国立保健医療科学院の「医薬品及び医療機器等の費用対効果評価のための人材育成プログラム開発事業」を受託して発足した、慶應義塾大学大学院健康マネジメント研究科の医療経済評価人材育成プログラムについて、その概要と現状をまとめた。

コース自体には、毎年40名程度の学生が入ってきており、すべての科目を取るコース修了者は毎年20名程度である。つまり「医療制度、医療経済、医療政策」というトピックの中で、医療経済評価に深く興味があるものは半分程度である。その中で、医療経済評価に関する研究を続けていくものの人数はその半分程度であり、現在、修士11名、博士9名が医療経済評価研究を学位論文のテーマにして研究を続けている。修了者へのアンケート調査から、医療経済評価研究への修了後の関わりについて機会があれば続けたいというものが半分程度いることがわかっているため、職場にもどった修了生についても、修了後にセミナー等の生涯学習の機会や研究の機会を提供することが医療経済評価研究の興味を保つために重要だと考えられる。

人材育成ではキャリア形成が非常に重要である。医療経済評価への関わり方は、学術研究、大学での公的分析、企業、臨床と様々であるため、まずは研究と公的分析に積極的に関わる機会を学生のうちに設け、経験の中から自分に向けたキャリアを見つけてもらうのが自然だと考えられる。現状では、公的分析の品目数も徐々に増加しており、公的分析に学術面から関与する人材ニーズも増加している。人材育成が進むにつれて、コース修了後も就業継続や公的分析と関わらない職への就職という進路から、公的分析を職として選ぶものが増えることが今後期待される。

特許・論文等なし

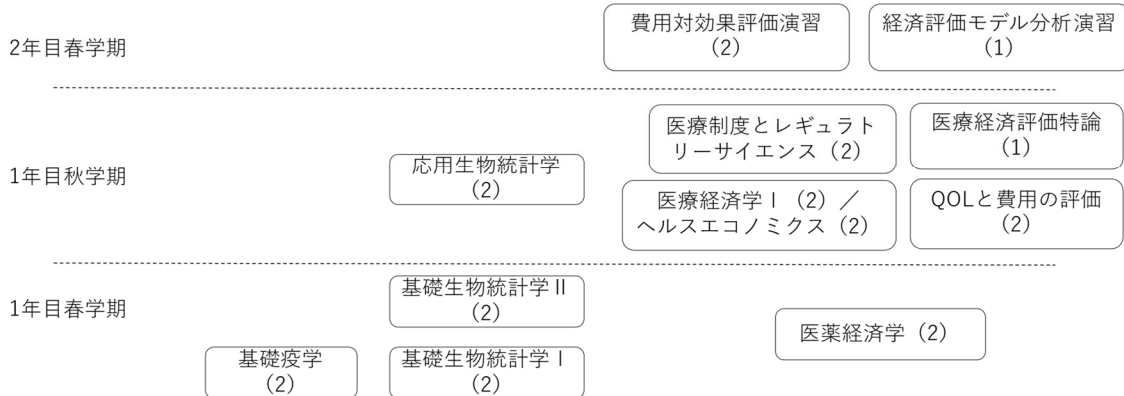
学会発表

後藤励（2022）“人材育成の視点から” ISPOR 日本部会 第16回学術集会 シンポジウム テーマ：費用対効果評価精度の現状と展望 2022年3月31日

図1：HTA コースカリキュラム

疫学・統計科目

医療経済評価科目



カッコ内数字は単位数

表1：HTA コース修了者41名の背景と進路

		人数
医療系資格	医師	9名
	薬剤師	6名
	看護師	10名
	助産師	1名
	医療系資格なし	15名
製造販売業者勤務	現職	4名
	退職	2名
修了後の進路	就業継続・復職	15名
	就職・就職活動継続	12名
	進学	11名
	起業・個人事業主	3名

QALY で活用する QOL 値に関する研究

能登真一（新潟医療福祉大学）

I. Patient Preference について

1) 背景

近年、医療製品の意思決定に患者さんの選好を反映させることが増えている。例えば、Medical Device Innovation Consortium (MDIC)¹⁾ は 2015 年に Patient Centered Benefit-Risk Project Report をまとめ、Patient Preference の重要性に言及している。具体的には、医療技術の規制上のベネフィット・リスク評価における Patient Preference 情報の潜在的価値として図 1 のようにまとめ、とくに High Benefit/High Risk の製品に対する Patient Preference の有用性を指摘している。製品開発ライフサイクルにおける Patient Preference 情報の活用と価値の可能性については、図 2 のように整理し、Discovery and Ideation の段階、Invention and Prototyping の段階から Product Launch and Reimbursement、Post-market Monitoring の段階までにおける Patient Preference の価値について整理している。また、FDA²⁾ も 2016 年に Patient Preference に関するガイダンスをまとめ、その定義や測定方法、患者アウトカムとの違いなどをまとめている。とくに、意思決定に Patient Preference 情報を含める必要性について、Patient Preference は個人差があり、患者が医療従事者、家族、規制当局、あるいは他の個人と同じ価値を様々なリスクとベネフィットに見出すとは限らないことを認識することが重要であるとし、さらに、Patient Preference はベネフィットとリスクに関する考え方と、治療／診断方法の好みの両方において異なる場合があると述べ Patient Preference の多様性について言及している。



図 1. ベネフィットとリスクの関数としての Patient Preference の価値

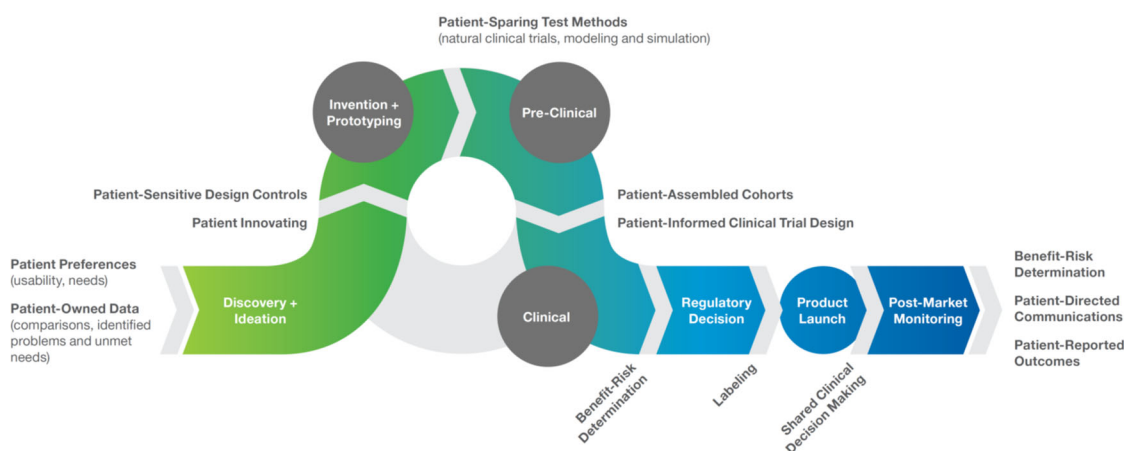


図 2. 製品開発ライフサイクルにおける Patient Preference 情報の活用と価値の可能性

2) 目的

Patient Preference に関する HTA 機関や企業の取り組み状況を把握することと、Patient Preference の測定方法を整理し、日本での適用の可能性を検討する。

3) 方法

欧州を中心とした Patient Preference への取り組み状況を NICE と PREFER, さらに ISPOR の活動内容をもとにまとめる。また、Patient Preference の測定方法については、

関連する論文をレビューする。

4) 結果

(1) NICE の取り組み

NICE は 2019 年に Myeloma UK という患者団体と協力のうえ骨髄腫患者を対象とした Patient Preference に関するプロジェクトを実施し、An exploratory study to determine how patient preferences data could be used in health technology assessment (HTA)³⁾ と題したレポートとしてまとめている。

このプロジェクトは Patient Preference を測定する方法論を HTA においてどのように活用できるかということについて探ろうとしたものである。

研究の具体的な目的は、骨髄腫患者の観点から、NICE のような機関が骨髄腫の治療に関する評価やガイダンスを作成する際に考慮すべき Patient Preference は何か、ということであった。

研究には骨髄腫患者を対象としたアンケート調査と、患者と介護者が同数で構成されるフォーカス・グループという、2つの調査方法が用いられた。前者では回答者の72%は56歳から75歳であり、回答者のほぼ半数は、調査時点で積極的に治療を受けていた。以下に示すとおり、「骨髄腫が回答者の生活に与える影響」「回答者が最も望んでいる治療効果」「回答者が最も望まない治療効果」についてたずねた。

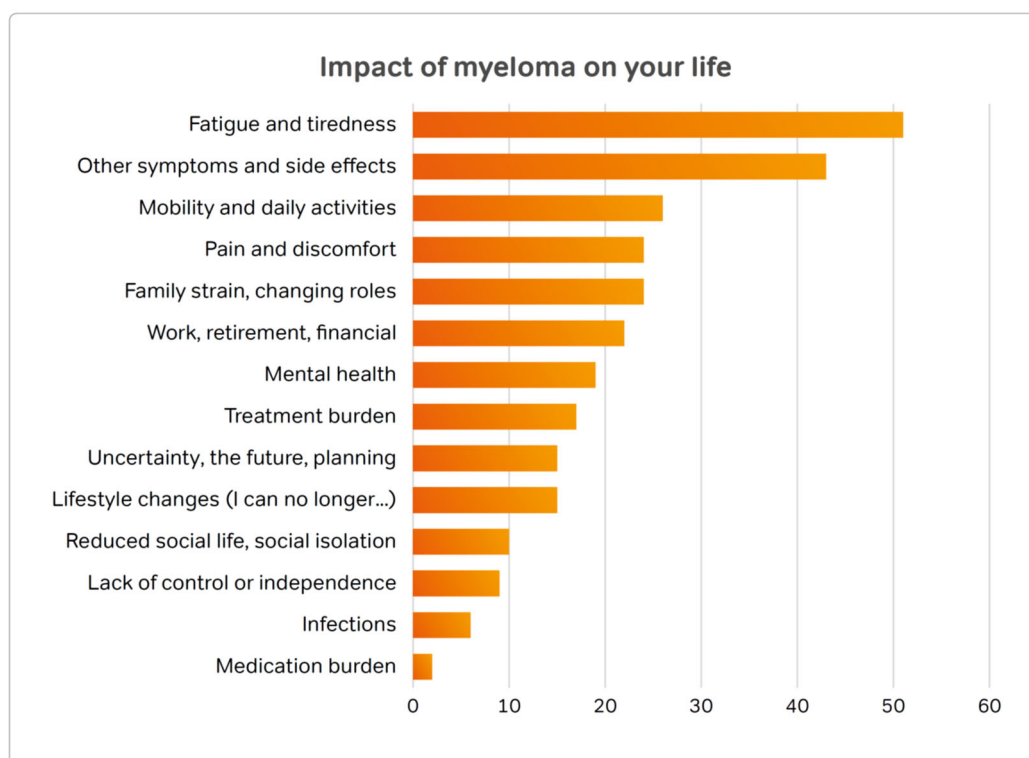


図3. 骨髄腫が回答者の生活に与える影響

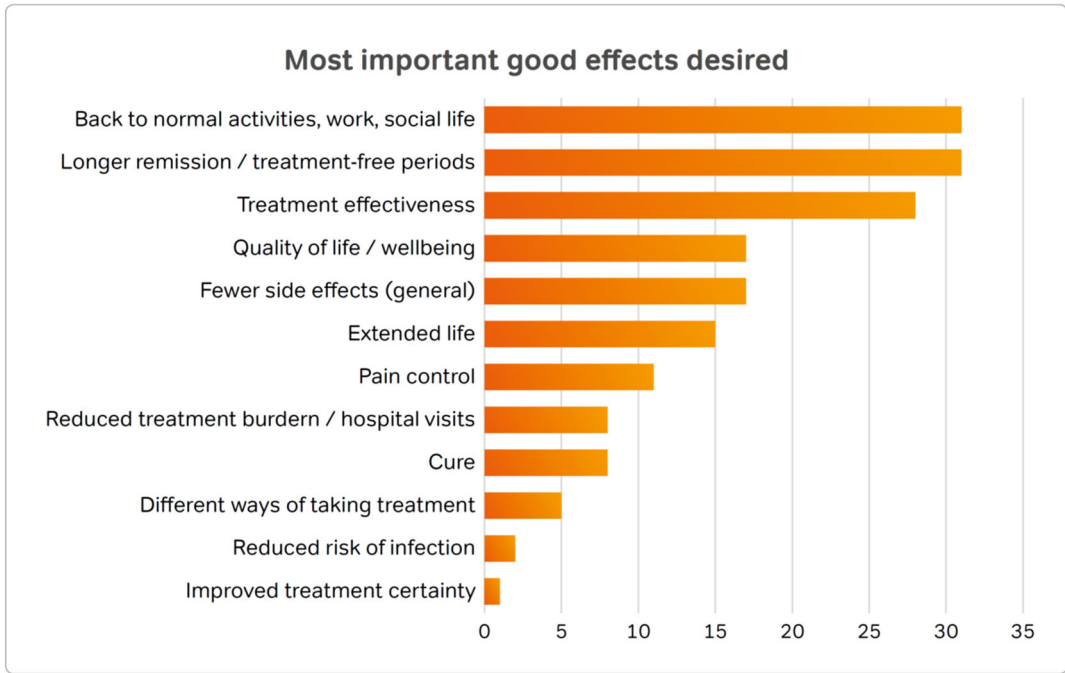


図4. 回答者が最も望んでいる治療効果

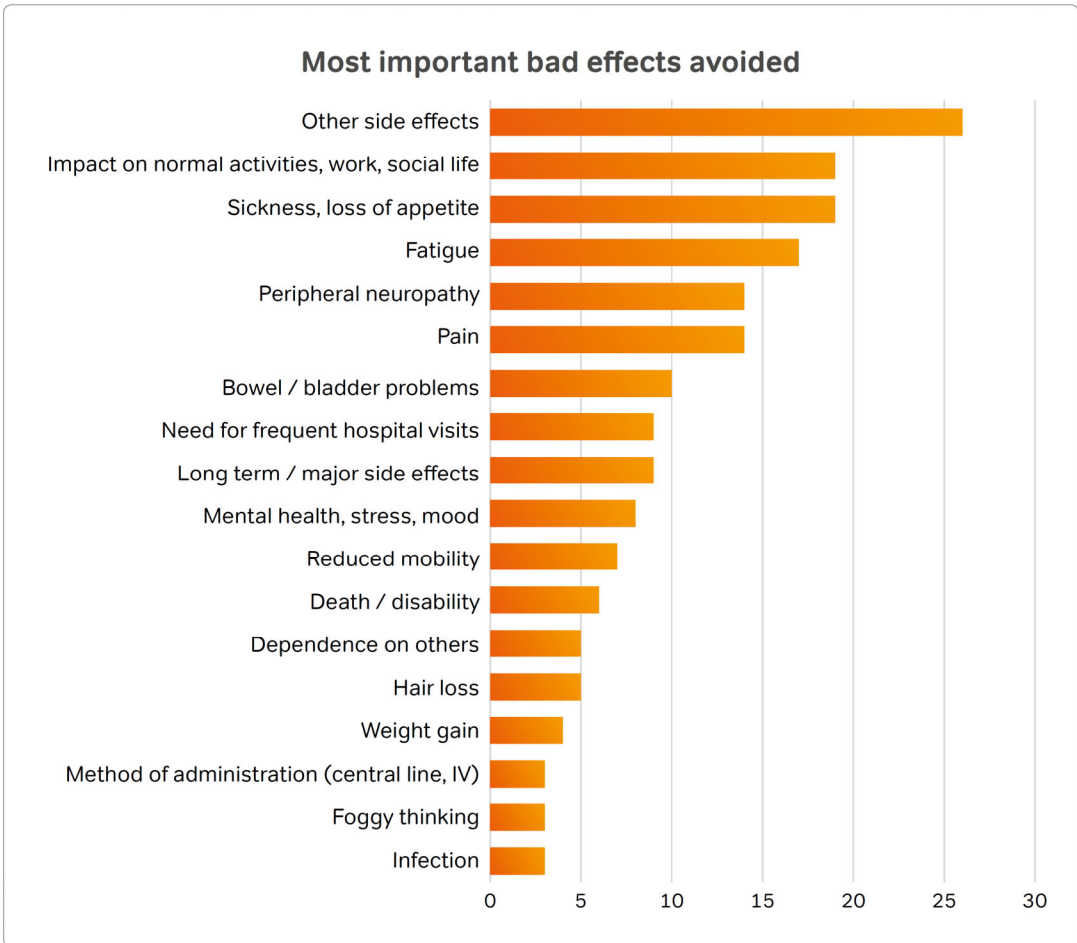


図 5. 回答者が最も望まない治療効果

骨髄腫が回答者の生活に与えた影響として最も多く報告されたのは「疲労・倦怠感」であり（図 3），次いで一般的な経験カテゴリーとしての「副作用」であった。骨髄腫の治療で最も望むプラス効果について，最も多くあげられたテーマは「治療効果」であり（図 4），「寛解・無治療期間の延長」というテーマも重要であることが指摘された。骨髄腫の治療で避けたいと思う最も重要な悪影響については，「重大な副作用」がもっとも多く指摘された（図 5）。これらのことから，DCE を設計する場合には，特定の副作用を別の属性として含めるのではなく，副作用の重篤度を属性レベルとして含めることが必要となる可能性がある」と述べている。

フォーカス・グループに対する調査では，患者と介護者の間の関係性の問題が様々な場面で議論された。それは，依存・非依存，仕事・退職・経済的問題や介護者への影響というテーマであった。さらに，Patient Preference を引き出すための ranking, point allocation, DCE, BWS という 4 つの手法について議論された。Patient Preference を測定する方法の中では，ranking と point allocation がもっともシンプルであるということと，DCE（図 6）と BWS（図 7）については長時間の集中力と選択肢を正確に識別し続ける能力が，一部の人々にとって問題になるかもしれないとの意見が出された。

最終的に，Patient Preference を測定するための Golden Standard は存在しないとしたうえで，DCE が異なる治療属性の相対的重要性を測定する方法として優れているとした。その一方で，HTA 機関が勧告を行ううえにおいて，Patient Preference から必ずしも有用な利益を得るとは限らず，さらなる研究が必要であると結論付けている。

Attribute	Option A	Option B
Effectiveness	50% reduction of symptoms	80% reduction of symptoms
Treatment duration	1 week	2 weeks
Dosage regimen	Daily pill	Pill twice a day
Side effects	Mild headache	Moderate headache

図 6. DCE のデザイン

Least important	Attribute	Most important
	Side effects	●
●	Treatment frequency	
	Treatment mode of delivery	
	Location of treatment delivery	
	Duration of treatment	

An example of a fictional BWS question - attribute case (type 1)

Least important	Attribute-level	Most important
	Side effects: Severe	●
	Treatment frequency: Weekly	
	Treatment mode: IV injection	
●	Location of treatment: Hospital	
	Duration of treatment: 10 weeks	

An example of a fictional BWS question - profile case (type 2)

Attributes	Treatment A	Treatment B	Treatment C
Side effects	Severe	Moderate	Moderate
Treatment frequency	Monthly	Monthly	Daily
Treatment mode	Oral	Injection	Oral
Treatment location	Home	Hospital	Home
Treatment duration	24 weeks	36 weeks	40 weeks
Best treatment	●		
Worst treatment		●	

図 7. BEST-WORST SCALING (BWS)のデザイン

(2) PREFER の活動

PREFER は、2016年に発足した Innovative Medicines Initiative (IMI) の下での官民共同研究プロジェクトであり、医薬品のベネフィットとリスクに関する患者の視点をいつ、どのように取り入れるべきかについて、企業、規制当局、HTA 機関向けのガイドラインの作成を支援するための提言策定を目的に組織された。メンバーは学術組織 (Uppsala University, Erasmus University Rotterdam など)、患者団体 (European Cancer Patients Coalition, European Patient Forum など)、HTA 機関 (Belgian Health Care Knowledge Centre (KCE)), さまざまな製薬企業などである。

なお、IMI は、欧州におけるバイオ医薬品の技術革新を促進し、患者さんにとってより良い、より安全な医薬品の開発を加速させることを目的に活動している組織であり、独自の革新的な資金提供スキームにより、安全性と有効性、知識管理、教育・訓練などの分野における研究プロジェクトを支援している。このプロジェクトは、欧州連合の Horizon 2020 プログラムと欧州製薬団体連合会 European Federation of Pharmaceutical Industries and Associations (EFPIA) の共同事業であり、双方から 600 万ユーロが拠出されている。

活動は「研究課題の策定」「ケーススタディ」「勧告作業」「全体の管理」という大きく 4 つのパッケージに分けられている。

「研究課題の策定」⁴⁾では、以下のようなスキームによって、第 1 層として、異なる手法、属性の調整、異なるサンプルを用いたプリファレンス・スタディー間の一貫性など、プリファレンス手法の妥当性と信頼性に関連するものが抽出された。第 2 層としては、プリファレンス研究の実施に関連する問題が指摘され、第 3 層として、ケーススタディでは適切に対処できない、Patient Preference 調査の結果の利用や解釈に関連した問題が指摘された。

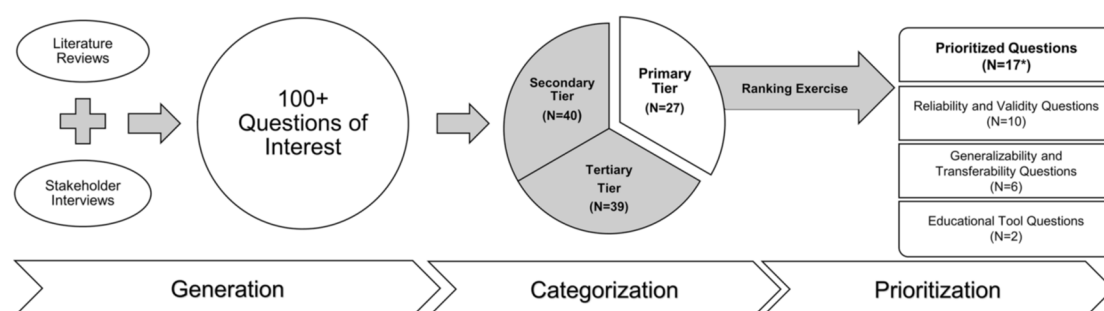


図 8. PREFER 研究のスキーム

「ケーススタディ」ではさまざまな疾患領域で実際の調査が実施された。神経筋疾患患者に対しては、BWS, Q-Methodology, DCE という 3 つの手法をランダムに割り付ける方法を用いて Patient Preference 評価が実施された (図 9) ⁵⁾。

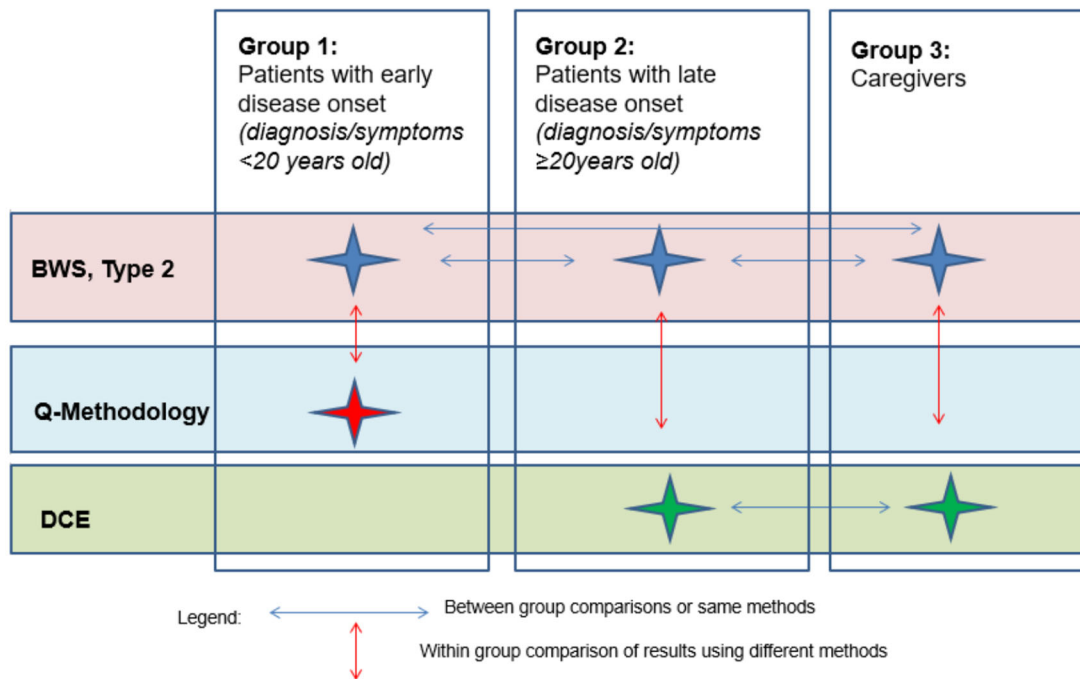


図9. 神経筋疾患患者に対するスタディデザイン

スウェーデンでは、リウマチ患者に図10のようなDCEを用いた調査を実施し、最も重要な属性は治療効果であり、次いで投与経路、使用頻度、重篤な副作用などが抽出されたと報告している⁶⁾。

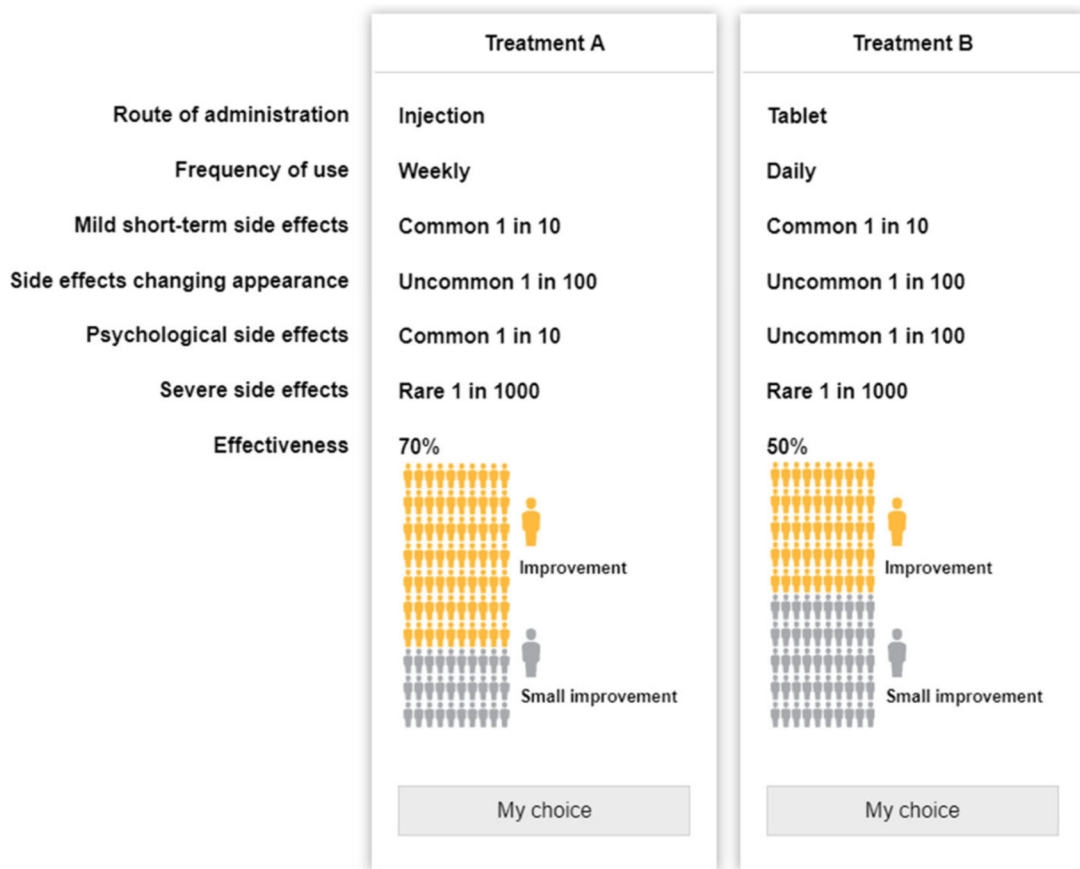


図 10. RA 患者に対する DCE デザイン

また、イタリアとベルギーでは、肺がん患者にフォーカス・グループ・ディスカッションを実施し、1) 治療によって期待されるプラス効果、例えば寿命の延長や日常機能の維持、2) 治療に関連するマイナス効果、例えば疲労など患者の日常機能に悪影響を与える有害事象、3) 治療効果の持続期間や種類に関する不確実性が抽出されている⁷⁾。

さらに、遺伝子治療について、ベルギーの成人血友病 A および B 患者に実施する調査のプロトコルとして、表 1 の属性を特定したと報告している⁸⁾。

「勧告作業」に関しては、HTA 機関へのアプローチを実施している。具体的には、カナダ、ドイツ、ベルギーの HTA 機関担当者に対して、フォーカス・グループ・ディスカッションを実施し、Patient Preference を HTA に活用する可能性について議論している。それによると、以下の属性をピックアップし Patient Preference を科学的助言と価値評価に利用することへの関心が見られたが、QALY や multi-criteria decision analysis に取り入れることはなかったと述べている⁹⁾。

このように、PREFER では、さまざまなプロジェクトをとおして、Patient Preference の調査とその HTA 機関での活用の可能性を探っている。

(3) ISPOR の活動

ISPOR ではこれまで、Task Force, Webinar 等で Patient Preference に関する話題を取り上げてきている。

まず、Task Force としては、2011 年に “Conjoint Analysis Applications in Health—a Checklist”¹¹⁾、2013 年に “Constructing Experimental Designs for Discrete-Choice Experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force”,¹²⁾そして 2016 年に “Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force”¹³⁾ と題したレポートがそれぞれ報告されている。

2011 年に発表された Checklist は表 1 に示すような内容で、Conjoint Analysis を使用した研究のガイドラインを示している。

表 1. A checklist for conjoint analysis applications in health care

具体的な項目	
1.	<p>明確なリサーチクエスチョンが示され、コンジョイント分析はそれに答えるための適切な方法であったか？</p> <p>1.1 定義された研究課題と検証可能な仮説が明示されていたか？</p> <p>1.2 研究の視点は説明され、研究は特定の意思決定や政策の文脈に位置づけられたか？</p> <p>1.3 リサーチクエスチョンに答えるためにコンジョイント分析を用いる根拠は何か？</p>
2.	<p>属性とレベルの選択はエビデンスに裏付けられていたか？</p> <p>2.1 属性の特定はエビデンス（文献調査、フォーカス・グループ、その他の科学的方法）によって裏付けられていたか？</p> <p>2.2 属性の選択は正当化され、理論と整合していたか？</p> <p>2.3 各属性に対するレベルの選択は、エビデンスによって正当化され、研究の視点や仮説と整合していたか？</p>
3.	<p>タスクの構成は適切であったか？</p> <p>3.1 コンジョイント・タスクの属性数は妥当だったか（つまり、完全プロファイルか部分プロファイルか）？</p> <p>3.2 各コンジョイント・タスクのプロファイルの数は妥当だったか？</p> <p>3.3 オプトアウトや現状維持の選択肢は含まれていたか（含まれるべきか）？</p>
4.	<p>実験デザインの選択は正当化され、評価されたか？</p> <p>4.1 実験デザインの選択は正当であったか？代替となる実験デザインは検討されたか？</p> <p>4.2 実験デザインの特性は評価されたか？</p> <p>4.3 データ収集装置に含まれるコンジョイント課題の数は適切であったか？</p>
5.	<p>研究課題から、嗜好は適切に引き出されたか？</p>

	<p>5.1 共役作業の動機づけと説明は十分であったか？</p> <p>5.2 適切な聴取形式（評価、順位付け、選択など）が用いられたか？無関心を許容する形式であったか（あるべきか）？</p> <p>5.3 コンジョイント課題では、嗜好の抽出に加えて、他の修飾的な質問（例えば、嗜好の強さ、嗜好の強さ、嗜好の強さ、嗜好の強さ）が行われたか？</p>
6.	<p>データ収集手段は適切に設計されていたか？</p> <p>6.1 適切な回答者情報（社会人口統計学的、態度、健康歴や状態、治療法など）が収集されたか？</p> <p>6.2 属性と水準は定義されていたか、また、文脈的な情報は提供されていたか？</p> <p>6.3 データ収集手段の負担の程度は適切であったか。回答者は励まされ、動機づけられたか？</p>
7.	<p>データ収集計画は適切だったか？</p> <p>7.1 サンプリング戦略は妥当であったか（例えば、サンプルサイズ、層別化、募集など）？</p> <p>7.2 実施形態は妥当か、適切か（対面式、紙とペン、ウェブベースなど）？</p> <p>7.3 倫理的配慮はなされていたか（例えば、募集、情報および/または同意、補償）？</p>
8.	<p>統計解析やモデルの推定は適切だったか？</p> <p>8.1 回答者の特性は調査・検証されたか？</p> <p>8.2 回答の質は調査されたか（例えば、合理性、妥当性、信頼性）？</p> <p>8.3 モデルの推定は適切に行われたか？クラスタリングやサブグループの問題は適切に処理されたか？</p>
9.	<p>結果や結論は妥当か？</p> <p>9.1 試験結果は検証可能な仮説を反映し、統計的不確実性を考慮したものであったか？</p> <p>9.2 研究の結論は証拠によって裏付けられ、文献にある既存の知見と比較されたか？</p> <p>9.3 研究の限界と一般化可能性について適切に議論されていたか？</p>
10.	<p>研究は明確、簡潔、かつ完全であったか？</p> <p>10.1 研究の重要性と研究背景は適切に説明されていたか？</p> <p>10.2 研究データ収集の手段・方法は記述されていたか？</p> <p>10.3 研究の意義は明確に述べられ、多くの読者に理解可能であったか？</p>

また、2016年に発表された推定値に関するは以下の表2のようなものである。

表2. The ESTIMATE checklist

項目	推奨される内容
推定値	<p>モデルから得られるパラメータ推定値の選択について、以下を含め、適切かつ完全に記述すること</p> <ul style="list-style-type: none"> 各変数が各レベルにおいて、効果コード化された水準に対応するのか、

	<p>ダミーコード化された水準に対応するのか、連続的な変化に対応するのか</p> <ul style="list-style-type: none"> 各変数が主効果に対応するか，交互作用に対応するか 連続変数が線形であるか，別の関数形式であるか
確率的な内容	<p>以下のような解析の確率的特性を記述する</p> <ul style="list-style-type: none"> パラメータ推定値の統計的分布 サンプル間のパラメータ推定値の分布（選好の不均質性） 推定関数の分散（オブザベーション間の分散の系統的な違いを含む）（スケール異質性）
トレードオフ	<p>モデルから推測されるトレードオフについて，以下のように説明する</p> <ul style="list-style-type: none"> 属性レベルの係数の大きさと方向性 実験に含まれるレベルの範囲における各属性の相対的重要度 回答者が属性間のトレードオフを希望する割合（限界代替率）
解釈	<p>統計モデルの特性を考慮し，以下のような結果の解釈を提供する</p> <ul style="list-style-type: none"> 結果から直接導き出せる結論 サブグループまたはセグメントを含むサンプルの対象母集団への適用性 結果の限界
方法	<p>分析に使用した統計分析方法を選択した理由を，以下を含めて記述する</p> <ul style="list-style-type: none"> その方法が実験によって得られたデータを分析するのに適している理由 その方法が根本的なりサーチクエスチョンに取り組むのに適切である理由 代替手法の中からその手法を選択した理由
前提条件	<p>モデルの前提条件とその前提条件が結果の解釈に与える影響を説明する</p> <ul style="list-style-type: none"> 誤差分布の仮定 観測値の独立性の仮定 価値関数の関数形式に関する仮定
再現性	<p>以下の説明を含め，再現を保証するために十分透明性のある方法で研究を記述する</p> <ul style="list-style-type: none"> データ設定（欠損データの取り扱いを含む） 価値関数や統計解析方法などの推定機能 推計に使用したソフトウェア
評価	<p>研究課題を解決するための統計解析手法の適切性を評価する</p> <ul style="list-style-type: none"> モデルの適合性 モデル仕様の感度分析

- 異なる手法で推定された結果の整合性

次に、Webinar では、2020 年に“Patient-Preference Information – What It Is and What It Is Not” と題して、Patient-Preference Information の測定方法や研究方法の紹介をしている。ここでは、Patient-Preference Information について、以下のような定義をしている。

表 3. Patient-Preference Information と PRO の比較

手法	特徴
Patient-Preference Information	特定の代替案や代替の医療介入で異なる結果や他の属性の選択肢に対する、患者にとって相対的な望ましさや受容性に関する定性的または定量的評価であり、患者がどのような治療や診断方法を好むかを測定するために設計されている。
Patient-Reported Outcome	患者の健康状態について、臨床医や他の誰によっても解釈されることなく、患者から直接もたらされるあらゆる報告のことであり、治療前、治療中、治療後の健康状態に関する患者の認識を測定するために設計されている。

さらに、ISPOR では、“A Framework for Studying Patient Preferences to Inform Decision Making in Healthcare: An ISPOR Good Practices Task Force Report” と題した追加の Task Force を実施しており、その成果が待たれている。

(4) Patient Preference の測定方法

Patient Preference の測定とその HTA への活用の機運が高まっている背景の 1 つに EQ-5D などの一般的な Preference-Based Measure (PBM) では測定できない、魅力があると考えられる。それは、測定方法の違いからはっきりと確認することができる。

そもそも、Patient Preference を測定するアプローチ方法は表 4 に示すように、顕示選好 (revealed preference) と表明選好 (stated preference) がある。それぞれの特徴としては、前者があらかじめ定められた状態に対する現実世界における選択を回答者に求めるのに対し、後者はあくまで仮想の、プログラムごとに属性やそのレベルなどが設定され、それに対する優先順位や選好といった相対的な望ましさを評価するものであるということになる。つまり、これまで議論されてきた Patient Preference は後者の表明選好 (stated preference) によって測定される価値ということになる。言い換えれば、顕示選好 (revealed preference) は利用可能な既存のサービスに対してその価値を問うのに対して、表明選好 (stated preference) は新たな技術の価値を問うのに適している可能性がある。

表 4. 顕示選好 (revealed preference) と表明選好 (stated preference) の比較

手法	特徴
顕示選好 (revealed preference)	現実世界における実際の選択行動の観察に依拠し、人々が実際に行った選択やトレードオフを検証するもの
表明選好 (stated preference)	回答者が仮想の実験に参加し、優先順位や選好を決定するもの

あらためて、表 5 に Patient Preference と健康関連 QOL の違いを整理してみる。Patient Preference の測定方法には大きく分けて 2 種類の方法がある。Patient Preference はあくまで、研究目的に応じて選択された成果・属性に対する選好を測定するものだということがわかる。

表 5. Patient Preference と健康関連 QOL の比較

	Patient Preference	健康関連 QOL
目的	健康状態, アウトカム, または治療の属性の評価	患者さんの健康状態の測定
時間軸	選好は比較的永続的であるが, 時間の経過とともに変化する可能性がある	リアルタイム, 介入前/介入後
領域	研究目的に応じて選択された成果・属性	あらかじめ定義された領域

表明選好 (stated preference) を測定する手法を表 6 に整理する。いずれも特徴があるが、これらはあくまで、属性間の相対的な選好や属性レベルの変化について、プロファイルを用いて引き出すものである。ただし、これらの方法で得られた選好は基数尺度には置き換えられず、あくまで序列を基準とした相対的な選好であるということを理解しておかなければならない。

表 6. Stated-Preference Methods の手法

手法	特徴
Conjoint Analysis, Discrete-Choice Experiments	各医療行為の属性に異なるレベルを割り当ててプロファイルを作成し、プロファイルを組み合わせて選択セットと呼ばれるプロファイル群を作成する。 プロファイルと選択のセットは、実験デザインによって決定される。患者には一連の選択セットが提示され、各選択セットから 1 つのプロファイルを選択するよう

	<p>求められる。</p> <p>回答のパターンを分析することで、研究に含まれる属性や属性レベルの変化に対して、患者がトレードオフを選択する割合が推定され、その結果、属性や属性レベルの変化の相対的な重要性、属性や属性レベル間のトレードオフの割合の測定が可能となる。</p>
Best-Worst Scaling	<p>オブジェクトケース、シングルプロファイルケース、マルチプロファイルケースの3種類がある。</p> <p>患者には一連の選択肢が提示され、最良または最も重要な選択肢と、最悪または最も重要でない選択肢を特定するよう求められる。</p> <p>複数プロファイルの場合は、DCEに非常によく似ている。一連のセットのそれぞれで、患者は、どのプロファイルが最高かまたは最も望ましいか、どのプロファイルが最悪かまたは最も望ましくないかを示すよう求められる。3種類の最良-最悪のスケールリングすべてにおいて、回答のパターンが分析され、各属性または属性レベルの相対的重要性が推定される。</p>
Direct-Assessment Question	<p>状態または相対的重要性の質問に対する直接的な回答を患者に求めるものである。</p> <p>状態または質問は、患者がある属性またはプロファイルに関する声明にどの程度同意するか、または他のすべての利用可能な属性またはプロファイルよりもある属性またはプロファイルを好むか、または選択するかどうかを患者に問うものである。</p> <p>プロファイルの直接評価のほとんどは、患者に提示されたプロファイルの中から優先されるプロファイルまたはプロファイルの序列を提供するだけであり、重みをもたらすことはない。</p>
Threshold Technique	<p>患者に基準プロファイルと代替プロファイルのどちらかを選択するよう求める方法である。</p> <p>基準プロファイルと代替プロファイルは、共通の属性セットで定義されるが、各属性のレベルは2つの代替間で異なることがある。この方法では、1つの属性が調査対象であるとみなされる。基準プロファイルを選択した場合、患者が基準プロファイルから代替プロファイルへ</p>

の選択を変更するまで、代替プロファイルの調査対象のレベルが向上する。代替プロファイルの場合、患者が代替プロファイルから基準プロファイルに選択を変更するまで、代替プロファイルの学習対象を悪化させる。患者が選択を変更した時点が閾値となる。

5) 考察

Patient Preference の重要性が高まっている現状について、その背景から現在までの各団体の取り組みについてまとめた。

Patient Preference はこれまでに盛んに測定されてきた PRO や健康関連 QOL (QOL 値を含む) とは異なり、新たな医療技術に対する患者の相対的な望ましさを測定できる概念、手法として注目されてきている。とくに、単なる治療効果だけではなく、製品の特徴やリスクに関する患者の望ましさを簡便に測定し、表すことができる手法として多く用いられてきている。これは、薬剤よりも医療機器におけるイノベーションをアピールする目的において、有用かもしれない。

しかしながら、現在のところ、Patient Preference そのものを QALY に反映させたり、それを HTA に用いたりしている機関はどこにもないのが現状である。

【文献】

1. Medical Device Innovation Consortium (MDIC) Patient Centered Benefit-Risk Project Report: A Framework for Incorporating Information on Patient Preferences regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. Medical Device Innovation Consortium; 2015.
2. Patient Preference Information—Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling: Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health and Center for Biologics Evaluation and Research; 2016.
3. NICE, Myeloma UK. Measuring Patient Preferences: An exploratory study to determine how patient preferences data could be used in health technology assessment (HTA) - Project report 2019.
4. Smith, Ian P.; DiSantostefano, Rachael L.; de Bekker-Grob, Esther W.; Levitan, Bennet et al. Methodological Priorities for Patient Preferences Research: Stakeholder Input to the PREFER Public–Private Project. The Patient - Patient-

- Centered Outcomes Research volume 14, pages449–453 (2021).
5. Jimenez-Moreno AC, Pinto CA, Levitan B et al. A study protocol for quantifying patient preferences in neuromuscular disorders: a case study of the IMI PREFER Project [version 1; peer review: 3 approved]. *Wellcome Open Res* 2020, 5:253.
 6. Schölin Bywall, Karin; Kihlbom, Ulrik; Hansson, Mats; Falahee, Marie et al. Patient preferences on rheumatoid arthritis second-line treatment: a discrete choice experiment of Swedish patients *Arthritis Research & Therapy* volume 22, Article number: 288 (2020).
 7. Petrocchi, Serena; Janssens, Rosanne; Oliveri, Serena; Arnou, Reinhard et al. What Matters Most to Lung Cancer Patients? A Qualitative Study in Italy and Belgium to Investigate Patient Preferences. *Front Pharmacol.* 2021 Mar 4; 12:602112.
 8. van Overbeeke, Eline; Hauber, Brett; Michelsen, Sissel; Peerlinck, Kathelijne et al.
 9. Patient preferences for gene therapy in haemophilia: Results from the PAVING threshold technique survey. *Haemophilia*: 27, Issue6, 2021, Pages 957-966.
 10. van Overbeeke, Eline; Forrester, Valérie; Simoens, Steven; Huys, Isabelle. Use of Patient Preferences in Health Technology Assessment: Perspectives of Canadian, Belgian and German HTA Representatives. *The Patient - Patient-Centered Outcomes Research* volume 14, pages119–128 (2021).
 11. Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, Bresnahan BW, Kanninen B, Bridges JF. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value Health.* Jan-Feb 2013;16(1):3-13.
 12. Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, IJzerman MJ, Bridges JF. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health.* 2016 Jun;19(4):300-15.
 13. Prosser LA. Statistical Methods for the Analysis of Discrete-Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health.* 2016 Jun;19(4):298-9.

II. QOL 値データベースのアップデート

1) 目的

平成 26 年（1999 年）度福田班において，国内で調査された QOL 値に関する論文のデータベースを構築した．これは医療経済評価研究に欠かせないデータのひとつである QOL 値について，日本国内で調査された研究論文を集めたものである．

このデータベースをアップデートし，来るべき費用対効果評価の施策導入に関係する方面で広く役立ててもらおうということが本研究の目的である．

2) 方法

データベースは 1 年に 2 回の割合で，各年度の当初にアップデートしている．対象は MEDLINE と医中誌であり，それぞれ国内で調査された QOL 値が掲載されている全論文をレビューした．

3) 結果

2021 年度に新たに公表された論文は 101 本であり，1999 年からの累計は 396 本となった（図 11）．

全体では，図 12 に示す通り，英文が 76%，和文が 24%となった．QOL 調査に使用された PBM ごとの内訳を図 13 に示す．EQ-5D-5L を使用した論文が 338 本と全体の 88%を占めた．

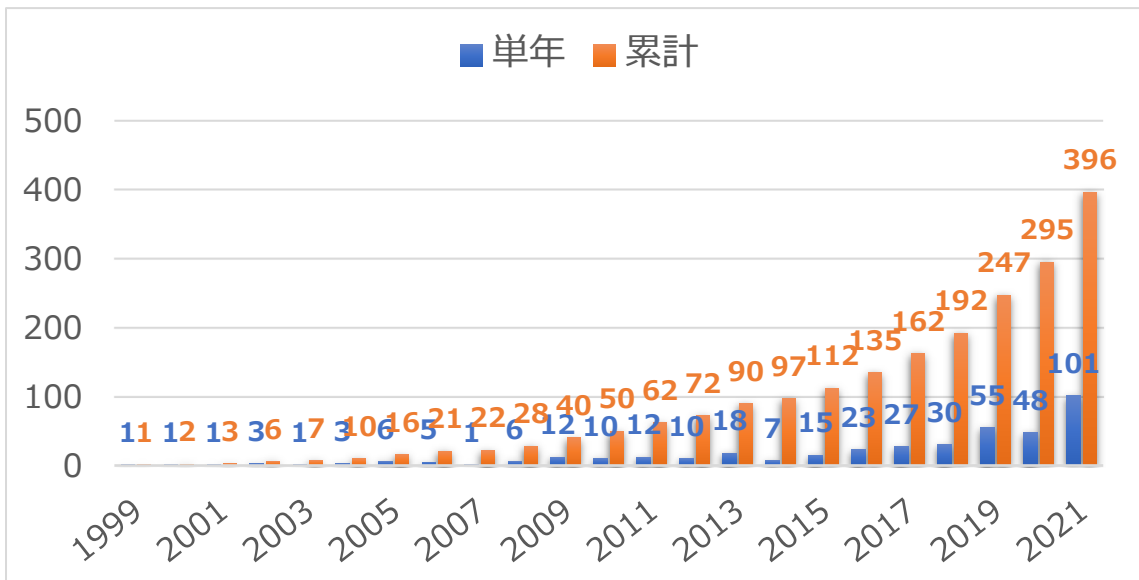


図 11. QOL データベースに掲載された論文の推移

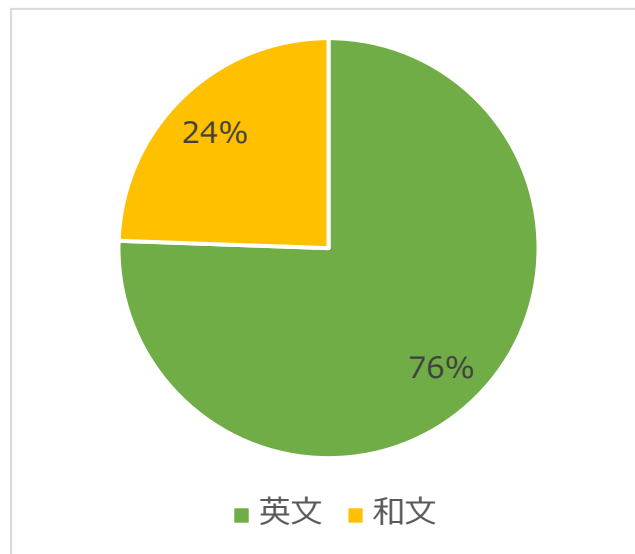


図 12. データベースに登録された論文の数

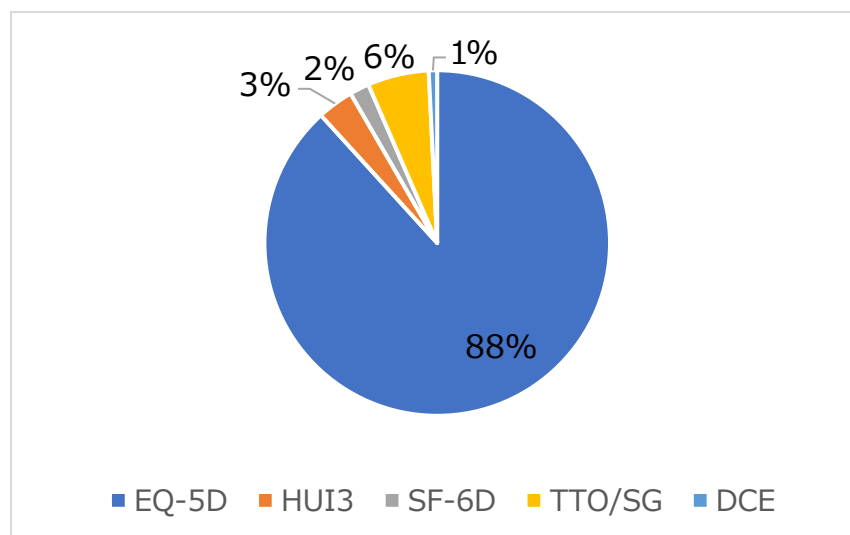


図 13. QOL 調査に使用された PBM の比較

EQ-5D について、その内訳を図 14 に示す。もっとも多かったのは、EQ-5D-3L であり、次いで EQ-5D-5L が続いた。これらの推移を図 9 に見ると、EQ-5D-5L は 2017 年以降に増え始め、その傾向は年々強まっているのがわかる。2021 年には EQ-5D-3L との報告数が逆転している。一方で、3L もしくは 5L の表記がなく、引用文献を調べてもそれが判明しない論文が一定数存在した。なお、ガイドラインで EQ-5D の使用を推奨している英国の NICE では、以下に示すように EQ-5D-3L の使用を position statement として発表している。

Position statement on use of the EQ-5D-5L value set for England (updated October 2019)

1. Since 2008, our preferred measure of health-related quality of life in adults has been EQ-5D. For most of that time, there has been only 1 version of the descriptive system, EQ-5D-3L. In this version, respondents rate their degree of impairment in different health domains using 3 response levels (no problems, some problems and extreme problems). The EuroQol Group has created a new descriptive system, EQ-5D-5L, with 5 response levels (no problems, slight problems, moderate problems, severe problems and extreme problems). EQ-5D-5L was designed to be more sensitive than EQ-5D-3L.
2. Our current guide to the methods of technology appraisal 2013 states that data collected using the EQ-5D-5L descriptive system may be used for reference-case analyses. When the guide was written, there was no value set for EQ-5D-5L from which to derive utilities. Our methods guide (section 5.3.12) states that: 'Until an acceptable valuation set for the EQ-5D-5L is available, the validated mapping function to derive utility values for the EQ-5D-5L from the existing EQ-5D (-3L) may be used (available from <http://www.euroqol.org>)'.
3. An EQ-5D-5L value set for England has been published (Devlin et al. 2018). Independent quality assurance (Hernandez Alava et al. 2018) and reports from 4 independent experts raised concerns about the quality and reliability of the data collected in the valuation study, and the methods used to model these data. Accordingly, we've chosen not to use the EQ-5D-5L value set for England published by Devlin et al (2018). The English valuation study was one of the first to be carried out anywhere in the world. Since then the international standard valuation study protocol has been improved and additional quality control methods introduced.
4. The following statement applies to all guidance-producing programmes at NICE that use cost-utility analyses.
5. We do not recommend using the EQ-5D-5L value set for England published by Devlin et al. (2018). Companies, academic groups and others preparing evidence submissions for NICE should use the 3L value set for reference-case analyses.
6. If data were gathered using the EQ-5D-5L descriptive system, utility values in reference-case analyses should be calculated by mapping the 5L descriptive system data onto the 3L value set. If analyses use data gathered using both EQ-5D-3L and EQ-5D-5L descriptive systems, the 3L value set should be used to derive all utility values, with 5L mapped onto 3L where needed.
7. The mapping function developed by van Hout et al. (2012) should be used for reference-case analyses, for consistency with the current guide to the methods of technology appraisal (even though several mapping functions are available; Hernandez Alava et al. 2017).

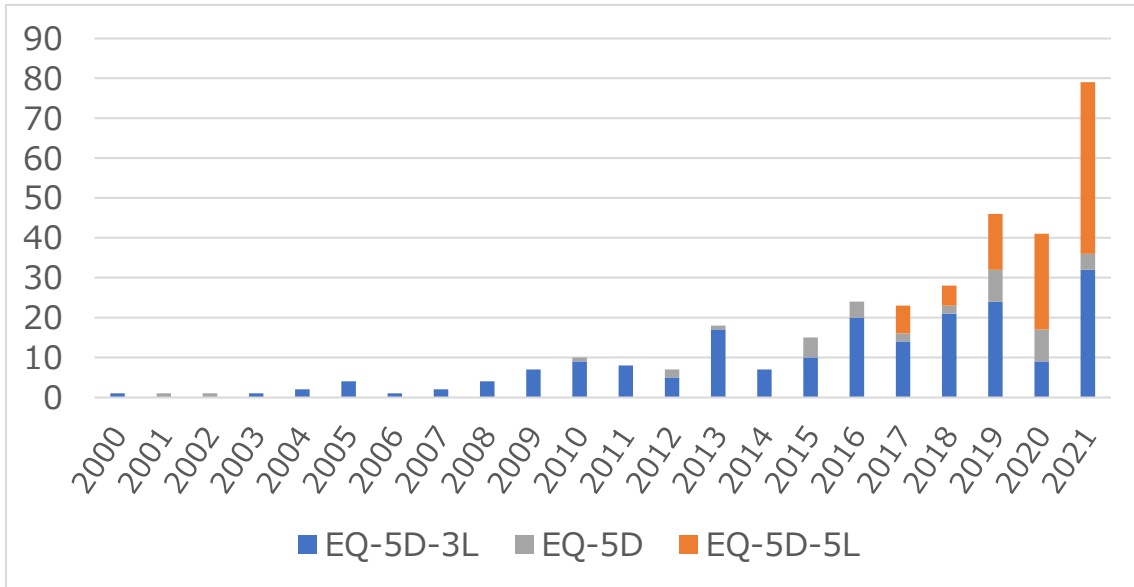


図 14. EQ-5D の内訳

QOL 値が測定された疾患の内訳を図 15 に示す。もっとも多くの QOL 値の報告があるのは整形疾患であり、次いで、がん、脳卒中、骨粗しょう症、腎疾患、RA 認知症などが続いた。

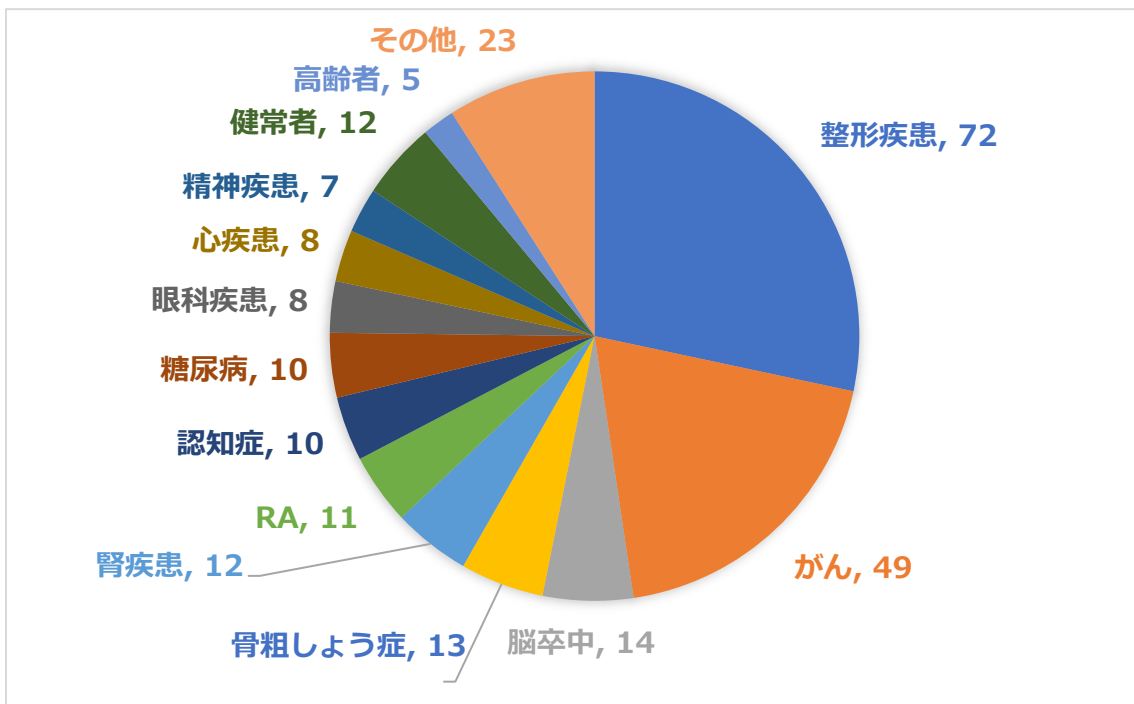


図 15. QOL 値が測定された疾患の内訳

また、研究の質を判断する1つの指標である症例数について見てみると、図16のように200例以上300例未満の論文がもっとも多く、100例以上200例未満の論文の数がその後続いた。

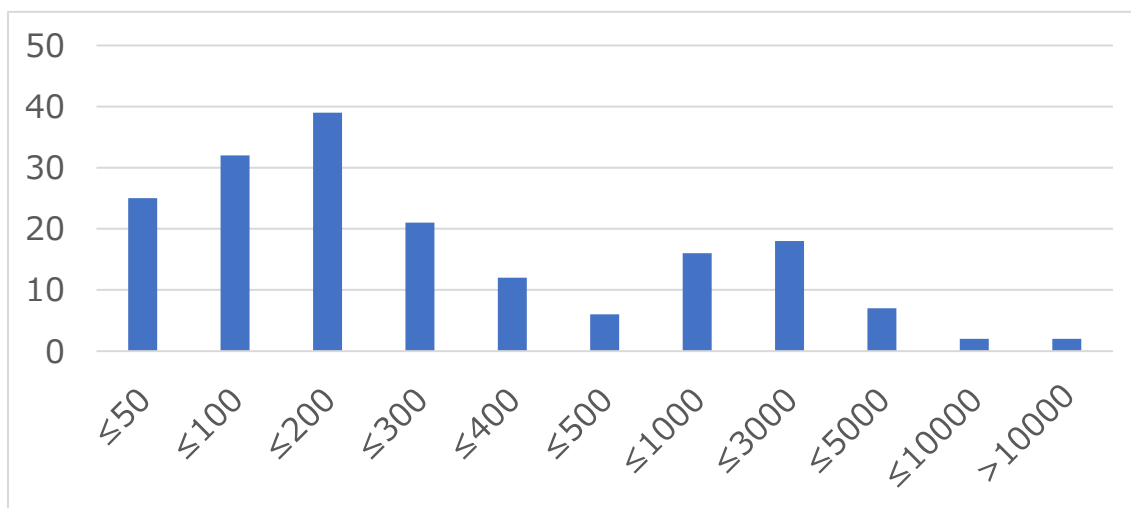


図16. 症例数による論文の比較

4) 考察

国内でQOL値を調査した論文の公表は年々増加傾向にあることに変わりはないが、疾患領域を含めて大きな変化はなかった。EQ-5Dを用いた研究が多いことは、費用対効果評価研究のガイドラインにも沿っており、良い傾向ととらえることができる。今後ますます国内のQOL値データの必要性が高まってくることが予想されるため、さらなる研究の発展が望まれる。

*QOL値データベースのURLは以下の通りである。

<http://chegol.com/database/>

5) Appendix

文献番号 Article No.	年 Year (a)	評価尺度 Instrument (a)	疾患分類 Disease Classification	症例数 n	状態分類 State category	状態 State	効用値 Utility	回答者 Response(a) (Self or Proxy)	代理人の立場 Relationship to proxy	介入方法 Intervention (a)	研究デザイン Study design	著者 Author (a)	雑誌 Journal	論文タイトル Title	巻 (号) Vol	ページ Page
E165	2021	EQ-5D	low back pain (LBP) and depression 腰痛 うつ病	504	Study	Low Back Pain Study	0.772	Self	-	-	randomized controlled trial	van Dongen JM	Med Care	Assessing the Impact of EQ-5D Country-specific Value Sets on Cost-utility Outcomes	59(1)	82-90
						Depression Study	0.762									
E176	2021	EQ-5D-3L	osteoporotic vertebral fracture 骨粗鬆症性椎体骨折	166	acute osteoporotic vertebral fracture	0 week	0.25	Self	-	-	randomized study	Inose H	Qual Life Res	Predictors for quality of life improvement after acute osteoporotic vertebral fracture: results of post hoc analysis of a prospective randomized study	30(1)	129-135
						12 weeks	0.71									
						48 weeks	0.75									
E196	2021	EQ-5D	chronic low back pain 慢性腰痛	602	chronic low back pain	-	0.65	Self	-	-	observational study	Imagama S	J Orthop Sci	Impact of background factors on outcomes of pharmacological therapy for chronic low back pain: A nationwide multicenter prospective study	26(1)	92-102
E203	2021	EQ-5D-5L	general population 一般住民	362	16-19	Male	0.978	self	-	-	cross-sectional study	Shiroiwa T	Value Health	Japanese Population Norms of EQ-5D-5L and Health Utilities Index Mark 3: Disability Catalog by Disease and Symptom in Community Settings	24(8)	1193-1202
				381		Female	0.967									
				637	20-29	Male	0.951									
				642		Female	0.953									
				633	30-39	Male	0.953									
				632		Female	0.944									
				640	40-49	Male	0.947									
				652		Female	0.945									

				615	50-59	Male	0.931								
				687		Female	0.925								
				682	60-69	Male	0.93								
				688		Female	0.927								
				723	70-79	Male	0.889								
				710		Female	0.876								
				765	80-89	Male	0.821								
				734		Female	0.774								
		HUI3		356	16-19	Male	0.892								
				369		Female	0.888								