

人間ドックにおける基本検査項目等の デルファイ法による社会経済的研究

学術委員会 基準検査検討小委員会

田倉智之¹⁾ 杉森裕樹^{2)*} 佐藤敏彦³⁾ 中山健夫⁴⁾ 高谷典秀⁵⁾ 渡辺清明⁶⁾ 山門 實^{7)**}

要 約

背景: 人間ドックは、疾病の危険因子を管理することで、一次予防や二次予防に貢献していると考えられる。一方で、人間ドックを含む予防的な健診は、EBMの観点からエビデンスが十分とは言えない状況にある。

目的: 本研究は、人間ドックにおける検査項目の社会経済的な意義を、専門職集団の知見等を定量化するデルファイ法にて評価し、検査・指導等の適切な普及に資することを目的とする。

方法: 対象検査は、基本検査項目等を中心に70項目とした。最終的なデルファイ調査の回答者は、人間ドック健診専門医の資格を有する106名(従事年数:17.80±7.40年、内科一般:72.6%)とした。本研究は、2段階のデルファイ調査を含む6つのステップから構成した。評価指標は、米国予防医学タスクフォース(USPSTF)の手法に準じて「期待される効果」、「避けがたい被害」、「必要となる費用」、「検査実践の状況」の4つの個別評価と1つの総合評価とした。評価の定量化は、アンケート方式でデータを収集し、5つのランクをもとにスコア化を実施した。

結果: 評価指標間の相関分析の結果、個別評価の「期待される効果」と「避けがたい被害」および「検査実践の状況」や総合評価の間には、有意な相関関係が認められた($p<0.01$)。第二段階のデルファイ調査の結果、基本検査項目等のうち88.6%で、結果を改善する方向に意見が収斂した($p<0.05$)。意見が収斂した後の「期待される効果」で評価が最も高いのは、血圧測定 4.14 ± 2.28 (スコア)であった。また「必要となる費用」の評価は、BMIで -2.51 ± 2.28 (スコア)と最も良かった。なお多くの基本検査項目等は、費用効果的となり、特にBMIや血圧測定のパフォーマンスが高かった。一方、意見が収斂した後の総合評価は、血圧測定が 1.41 ± 0.31 (スコア)と最上位に位置付けられ、次いでHbA1cが 1.35 ± 0.35 (スコア)となった。

結論: 本研究の結果、基本検査項目等は、医学的效果が期待され健康関連の被害も少なく、関わる検査費用や検査の普及状況も良好であった。また「費用対効果」分析でも、大多数の項目は費用効果的であった。基本検査項目等は、すべての項目で一様ではなく、人間ドック健診の目的に応じて、重みに差違があることが示唆された。

キーワード 基本検査項目、社会経済的評価、デルファイ法、米国予防医学タスクフォース、EBM

諸 言

人間ドックは、日本が長年培ってきた総合的な健康診断システムである。主に、壮年期や老年期において日常生活を支障なく送っている人を対象に、短期間に全身の総合的な健康診断を行うとともに本人の健康状態を評価・把握し、生活等における健康面での指導を行う事業である。人間ドック(二日ドック、一日ドック等)の総受診者数は、2012年において319万人となっている¹⁾。

この人間ドックの主な目的として、自覚されていない各種の慢性疾患を早期に発見する一次予防が挙げられる²⁾。すなわち、肥満、高血圧、高血糖、脂質異常等の危険因子をコントロールすることにより、脳血管疾患および心血管疾患等の一次予防に貢献していると考えられる。また、早期がん等の疾患を発見することにより、二次予防もさらに期待される。

予防医学に基づく健康管理指導の一環として発

1) 大阪大学大学院医学系研究科(医療経済産業政策学) 2) 大東文化大学大学院スポーツ・健康科学部健康科学科

3) 青山学院大学大学院社会情報学研究科 4) 京都大学大学院医学研究科社会健康医学系専攻健康情報学分野

5) 医療法人社団 同友会 6) 東京臨床検査医学センター 7) 足利工業大学看護学部

*: 実行委員長 **: 学術委員会委員長

展してきた人間ドックだが、事業のすそ野の拡がりや昨今の疾病予防への関心の高まり等を背景に、明確な根拠に基づく活動がさらに希求されている。特に、新たな理論や技術革新の応用機会も増し、また多くの医療資源を消費する傾向にあるため、国民の健康福祉の向上の観点からも、その社会的意義についてより一層の説明が望まれる。

目的

日本独自の健康診断システムである人間ドック健診はすでに50年の歴史を持ち、これまでも任意型健診として対策型健(検)診とともに我が国の予防医学を担ってきている。しかしながら、この任意型健診としての人間ドック健診の社会的な有用性や経済性については、これまで、科学的な評価が十二分になされてきたとは言い難い状況にある。

一方で、不確実性を伴う活動も内在し、健康改善の機序も複雑なうえ、エンドポイントが長期に渡る予防医学の領域では、強固なエビデンスの構築に一定の限界があるのも事実である。他の科学分野に目を向けると、事業等に占める技術要素の割合が高く、社会制度に及ぼす影響も大きい領域は、専門職集団の合議による評価と説明の仕組みも普及している。

そこで本研究は、検査や指導等の適切な普及を促すことを目的に、エキスパートの多様な知見等を収斂させる理論と手法を応用して、人間ドックにおける基本検査項目等の社会経済的な意義について、定量的な整理を試みる。

方法

(1)理論と手法

エキスパート・オピニオンで高度な専門技術を評価する方法は、医療分野以外で盛んに検討され推進がなされている。このテクノロジー・アセスメント(Technology Assessment: TA)は、技術のもたらす正や負の副次的な影響を総合的・包括的に予見・分析することで、社会的な課題設定や政策的な意思決定の方向性を、蓋然性を有して提示する理念や活動を指すものである。この検討過程においては、経済・環境・倫理・法・社会・文化

等の要因も広く含まれる。

TAの代表的な手法の1つとして、アメリカのシンクタンクであるランド研究所が開発した「デルファイ法(Delphi Method)」³⁾が挙げられる。この手法は、多方面から専門家の見解を集め、目的に沿ってある1つの意見に収斂をさせる方法である。そのため、現時点におけるエビデンスやノウハウを背景に、特定領域のエキスパートの集団による専門性の高い意見・知見に基づき、精度の良い将来予測や確度の高い評価に繋げることができる。

このデルファイ法を応用することで、医師の提供する医療技術の評価を試みた有名なケースに、1980年代のRBRVS(Resource-Based Relative Value Scale)⁴⁾のプロジェクトがある。専門医(約20領域)の集団が、検査の判断や画像の診断、患者の指導等の医療技術を提供する難易度を相対的に評価し、無形の医療技術の有用性について対外的な説明を進めている。その成果は、米国の支払制度にも応用されており、今日も米国のMedicare制度等で利用されている。

一般的にTAの手順は、医療分野では十分に標準化されておらず、評価の対象となる技術や領域によってその選択方法は様々である。しかし、デルファイ法については、近年、幾つかの報告が散見され知見の蓄積が進んでいる⁵⁻⁷⁾。

(2)研究の構成

本研究は、デルファイ法の先行研究⁸⁾を参考にしつつ研究方法等を検討し、6つの調査ステップを構成した(図1)。

最初に、評価対象の選定と評価デザインの確定を行い、その内容に基づいて、さらに調査シートの設計と作成を行った。続いて、第一段階のデルファイ調査を人間ドックの専門医グループに実施した。その調査データの解析後、得られた結果をフィードバックする方法(検査項目ごとに回答集団の平均値を情報提供)で、さらに同一グループへ同様に第二段階のデルファイ調査を実施し、意見の集約を行った。

最後に、この2つのデルファイ調査で得られたデータから、各パラメータの回答傾向や意見の収斂状況等の分析を行い、パネル会議におけるコン

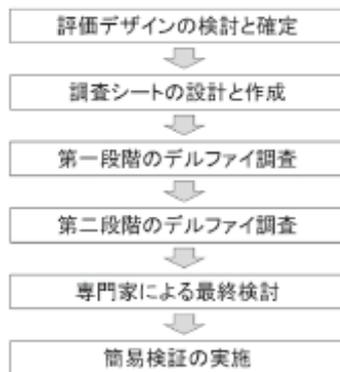


図1 研究のフローチャート

研究のステップは、6段階から構成されており、第一段階の調査は186名の専門医、第二段階の調査は106名の専門医からデータを収集し、その解析結果を7名の有識者で最終検討した。なお、いくつかの検査項目については、予防医学等の先行研究と照らし合わせを実施した。

センサスを経て、基本検査項目等に対するエキスパートの評価を取り纏めた。なお参考までに、評価が最上位の検査項目および“費用対効果⁽²⁾(期待される効果と必要となる費用から算出)”が最上位の検査項目について、主な臨床ガイドラインにおける評価状況や過去の研究報告の動向を精査し、結果の妥当性に関する検討を補足した。

(3)対象の集団

第一段階および第二段階のデルファイ調査の回答者は、日本人間ドック学会に所属の「人間ドック健診専門医」の資格を有する専門職であり、任意に本研究に対して協力を得た。

アンケート調査票の郵送方式で行った第一段階のデルファイ調査は、2012年2月に実施し、人間ドック健診専門医186名の回答を得た(全専門医に対して回収率30.0%)。同様にアンケート調査票の郵送方式で行った第二段階のデルファイ調査は、2013年7月に実施し、第一段階の調査の回答者のうち106名(回収率56.9%)から回答を得た。

なお、分析結果について最終的な精査を行ったパネル会議は、日本人間ドック学会の学術委員会メンバー7名(人間ドック健診専門医や公衆衛生学、医療経済学等の有識者)から構成され、会合は2013年10月に開催された。

(4)評価の指標

本研究では、基本検査項目等について、複数の次元と尺度による個別評価と総合評価を同時

に行った。

個別評価は、米国予防医学タスクフォース(USPSTF)の“insufficient evidence(不十分な科学的根拠)”に対応したupdated methodに準じた⁹⁾、「期待される効果」、「避けがたい被害」、「必要となる費用」、「検査実践の状況」の4つの指標を用いて、それぞれ5ランクからスコア化を実施した(4つの評価指標から専門医が考える意見を集約した)(図2)。すなわち、個々のパラメータは、ゼロを基準(標準的/一般的)に、ランクのマイナス2は“極端に低い/小さい”、ランクのプラス2は“極端に高い/大きい”とした。

また個別評価については、回答者よりランク間の重み付けに関する判断情報も収集し、ランク間のレベルについて適宜、計量的な補正を行った。つまり、基準となるゼロとプラス1、マイナス1の間隔に対して、プラス2およびマイナス2が何倍の意味を持つかたずね、その値で各回答値を補正した。

総合評価は、個別評価のみでは論じられない全ての社会的な意義に対する専門医の総合的な認識を、各基本検査について5ランクで整理を行った(ゼロを基準にプラスとマイナスにそれぞれ2ランクとした)。

(5)検査の項目

本研究では、基本検査項目を中心に、人間ドックの検査70項目について評価を行った(表1)。その主な大分類は、身体計測、生理、X線・超音波、生化学、血液学、血清学、尿、便、病理、問診・診察およびオプション検査となった。なお、一部の項目は重複しセット検査群(ex, 腫瘍マーカー)とした(以下、基本検査項目等とする)。

(6)統計と解析

前述の個別評価と総合評価について、おのおの基本統計量を算出した。

第一段階と第二段階のデルファイ調査の間における回答内容の推移は、マンホイットニー検定(ノンパラメトリックによる母平均の差の検定)で平均値の比較を行い、統計学的な有意差を確認した。さらに、第一段階の回答値のばらつきが、第二段階のデルファイ調査の施行によって収斂するかど

・【個別評価の解説】	I. 期待される効果 疾患の発見や病態の把握に有効であり対象者の健康改善や社会に寄与する程度 II. 避けたい被害 一定の確率で生じる顕在的な合併症や潜在的な副作用の影響（検査に直接関係） III. 必用になる費用 検査の提供時に必用となる直接的な費用または間接的な費用の大きさ（提供側） IV. 検査実践の状況 現在における検査の選択や普及の状況、または検査量（間隔・回数）の適正性
・【評価スコアの定義】	「重要レベル」 （ランク0が中間） ●ランク-2：かなり低い/かなり小さい ●ランク-1：やや低い / やや小さい ●ランク 0：どちらでもない（標準的、一般的） ●ランク+1：やや高い / やや大きい ●ランク+2：かなり高い/かなり大きい
・【総合評価の解説】	V. 検査の総合評価 上記の4つの評価軸を踏まえた、検査の社会的・経済的意義に関する総合評価
・【評価スコアの定義】	「重要性レベル」 （ランク0が中間） ●ランク-2：いらない（不要） ●ランク-1：なくても良い ●ランク 0：どちらでもない（あっても良い）← 中間、基準 ●ランク+1：重要（必用） ●ランク+2：大変重要（絶対必要）

図2 評価パラメータと評価スコア

各検査項目の評価は、効果・被害・費用・実践の4つの個別パラメータと総合パラメータの計5つから構成した。また、その評価スコアは、個別パラメータでゼロを基準(標準的/一般的)とする5ランクとした(ランク-2は極端に低い/小さい、ランク+2は極端に高い/大きい)。総合評価も同様にゼロを基準にプラス・マイナスの2ランクとした。

うかについて、第二段階の回答者106名の母集団で等分散分析(F検定)を行い検証した。

また、回答者の人間ドックに対する経験量がエキスパート・オピニオンに与える影響(バイアス)を検討するために、4つの個別評価および総合評価と人間ドックの従事歴の相互関係を、第一段階のデルファイ調査のデータを用いたスピアマン順位相関の分析で整理した。

診療サービスに対する医療経済評価の1つに、費用負担と診療成果から算出される費用効果(パフォーマンス)の概念があり、海外では社会政策の議論等へ応用されてきた。そこで、第二段階のデルファイ調査のデータを利用し、「期待される効果」と「必要になる費用」の二軸から、基本検査項目等の「費用対効果」分析を行った。

さらに、選択バイアス等の考察のために、総合評価と「費用対効果」の相互関係を、スピアマン順位相関分析で検証した(第一段階のデルファイ調査のデータを利用し、各検査項目を単位に実施)。

なお、統計学的な検定の有意水準を5%とし、

表1 対象とした検査名(70の基本検査項目等)

番号	検査	番号	検査
1	身長	36	ヘマトクリット
2	体重	37	血小板数
3	肥満度	38	MCV
4	BMI	39	MCH
5	腹囲	40	MCHC
6	血圧測定	41	CRP
7	心電図	42	血液型(ABO)
8	心拍数	43	血液型(Rh)
9	尿潜血検査	44	梅毒反応
10	尿圧検査	45	HBs抗原
11	視力検査	46	尿蛋白
12	聴力検査	47	PH
13	呼吸機能検査	48	尿糖
14	胸部X線	49	尿酸値
15	上部消化管X線	50	尿潜血
16	腹部超音波	51	尿比重
17	総蛋白	52	便潜血
18	アルブミン	53	子宮頸細胞診
19	クレアチニン	54	内科
20	尿酸	55	婦人科
21	総コレステロール	56	特定健診22項目問診票
22	HDLコレステロール	57	乳房触診
23	LDLコレステロール	58	マンモグラフィ
24	中性脂肪	59	乳房超音波
25	総ビリルビン	60	HCV抗体
26	AST(GOT)	61	薬師 MRI/MRA
27	ALT(GPT)	62	PSA
28	γ-GTP	63	腫瘍マーカーセット
29	ALP	64	胸部CT
30	血糖(空腹時)	65	内臓脂肪CT
31	血糖(75gブドウ糖負荷試験)	66	経腸超音波
32	HbA1c	67	骨密度
33	赤血球	68	甲状腺検査
34	白血球	69	頸動脈超音波
35	血色素	70	ヘリコバクター・ピロリ

(注)番号57~70はオプション検査項目

結果の数値はすべてmean±SDで示した。また、統計解析のソフトは、SPSSを利用した。

結果

(1)対象者群と評価指標

第一段階のデルファイ調査の回答者186名の人間ドックの従事年数は、17.06±8.55(年)であり、専門領域は、内科一般が72.5(%)と最も多く、次いで消化器内科が30.1(%)、循環器内科が16.1(%)となった(表2)。第二段階のデルファイ調査の回答者106名の人間ドックの従事年数は、17.80±7.40(年)であり、専門領域は、内科一般が72.6(%)と最も多く、次いで消化器内科が31.1(%)、循環器内科が17.9(%)であった。以上から、サンプル数が異なるものの、第一段階および第二段階のデルファイ調査の母集団の間に、構造的な差がないと考えられた。

人間ドックの従事年数と各種評価の指標に関わる相関分析の結果、従事年数は4つの個別評価や総合評価との間に相互関係が認められなかった

($p>0.05$)(表3)。すなわち、人間ドックの従事年数に関する選択バイアスは存在せず、本研究の結果に影響を与えないと推察された。

また評価パラメータ間の相関分析の結果、個別評価の「期待される効果」と「避けがたい被害」および「検査実践の状況」や総合評価の間には、有意な相関関係が認められた($R_s=0.3396$ $p<0.01$, $R_s=0.4763$ $p<0.01$, $R_s=0.3182$ $p<0.01$)。さらに「避けがたい被害」は、「必要となる費用」「検査実践の状況」の間で、有意な相関関係が認められた($R_s=0.2762$ $p<0.01$, $R_s=0.2923$ $p<0.01$)(表3)。

なお総合評価は、4つの個別評価では不十分な幅広い評価を目的とした指標であるが、個別評価の半数(「期待される効果」および「検査実践の状況」)と統計学的に有意な相関関係にあるため、個別評価の内容も含む包括的なパラメータとして論じることは妥当と推察された。

(2)意見収斂と評価結果

第一段階および第二段階のデルファイ調査の間の回答者の意見収斂の動向について、等分散分析

表2 回答者の基本属性

	第一段階調査	第二段階調査
回答者数(人)	186	106
人間ドック従事年数(年)	17.06±8.55	17.80±7.40
専門領域構成(上位5科-重複有)(%)		
1位	内科一般 72.6	内科一般 72.6
2位	消化器内科 30.1	消化器内科 31.1
3位	循環器内科 16.1	循環器内科 17.9
4位	糖尿病内科 13.4	内視鏡内科 12.3
5位	呼吸器外科 9.1	腫瘍内科 10.4

第一段階および第二段階のデルファイ調査の回答者は、日本人間ドック学会に所属の「人間ドック検診専門医」の資格を有する専門職であり、任意に本研究に対する協力を得た。

表3 回答者が有するバイアスの検証と各評価指標間の関係

	相関係数					
	人間ドックの従事年数	期待される効果	避けがたい被害	必要となる費用	検査実践の状況	検査の総合評価
人間ドックの従事年数		-0.0066	0.0149	-0.0336	-0.0591	-0.0073
期待される効果	0.9320		-0.3396	0.0201	0.4763	-0.3182
避けがたい被害	0.8466	0.0000**		0.2762	-0.2923	-0.0342
必要となる費用	0.6621	0.7938	0.0003**		-0.1439	0.0046
検査実践の状況	0.4422	0.0000**	0.0001**	0.0614		-0.2780
検査の総合評価	0.9246	0.0000**	0.6563	0.9524	0.0003**	

スピアマンの相関分析 **： $p<0.01$

第一段階のデルファイ調査の回答値を用いて、回答者の人間ドック従事年数と個別評価の4パラメータおよび総合評価の間で相関分析を行い、回答者が保有するバイアスの検証とともにパラメータ間の関係を整理した。

表4 第一段階から第二段階における意見の収斂状況

ケース	検査項目数	割合	P値
分散に差あり			
(第一段階>第二段階)	18	25.7%	**
(第一段階<第二段階)	37	52.9%	*
分散に差なし(第一段階=第二段階の可能性あり)	0	0.0%	
合計	15	21.4%	
合計	70	100.0%	

等分散分析(F検定) * : $p < 0.05$, ** : $p < 0.01$
 第一段階および第二段階のデルファイ調査の周の回答者の意見収斂の動向について、等分散分析(F検定)で解析を行った。70検査項目について、統計学的に有意な差が認められるか、さらにどちらのデルファイ調査の分散が大きいか、という観点から整理を行った。

表5 第一段階から第二段階へのデルファイ調査結果の推移

	第一段階調査	第二段階調査	95%CI	P値
ランク間の重み付け(差)の変位(単位:倍)				
ランク0と+1の差に対して「ランク+2」	3.10 ± 1.40	⇒ 2.70 ± 1.30	-0.80, -0.10	**
ランク0と-1の差に対して「ランク-2」	2.90 ± 2.10	⇒ 2.40 ± 1.00	-1.00, -0.10	**
各評価パラメータの変位(単位:スコア)				
期待される効果	1.50 ± 0.68	⇒ 2.30 ± 1.29	0.65, 0.96	**
避けがたい被害	-1.71 ± 0.55	⇒ -2.55 ± 0.85	-0.93, -0.75	**
必要となる費用	-0.40 ± 0.67	⇒ -0.59 ± 0.76	-0.28, -0.11	*
検査実践の状況	1.13 ± 0.93	⇒ 1.45 ± 1.18	0.22, 0.41	**
検査の総合評価	0.83 ± 0.49	⇒ 0.98 ± 0.32	0.10, 0.20	**

母平均の差の検定 * : $p < 0.05$, ** : $p < 0.01$
 第一段階および第二段階のデルファイ調査について、個別評価および総合評価のパラメータの回答値(70検査項目の平均値)を整理した。また、個別評価で実施したランク間の重み付けについても、第一段階および第二段階のデルファイ調査の平均値を整理した。

(F検定)で解析を行ったところ、基本検査項目等において、第二段階のデルファイ調査の分散が第一段階のそれよりも大きい項目はなかった(表4)。また、第二段階のデルファイ調査の分散が第一段階のそれよりも有意に小さかったのは、基本検査項目等のうち88.6%であった($p < 0.05$)。すなわち、専門医のグループの意見の多くは、デルファイ調査により収斂したと考えられた。

個別評価の回答値は、「期待される効果」において基本検査項目等全体の平均値が、第一段階のデルファイ調査の1.50 ± 0.68(スコア)から第二段階の2.30 ± 1.29(スコア)へ上昇した($p < 0.01$) (表5)。「避けがたい被害」では、同様に平均値が-1.71 ± 0.55(スコア)から-2.55 ± 0.85(スコア)へ改善した($p < 0.01$)。「必要となる費用」では、その平均値が-0.40 ± 0.67(スコア)から-0.59 ± 0.76(スコア)へ改善した($p < 0.05$)。「検査実践の状況」では、その平均値が1.13 ± 0.93(スコア)か

ら1.45 ± 1.18(スコア)へ上昇した($p < 0.01$)。また、総合評価の回答値は、基本検査項目等全体の平均値が、第一段階のデルファイ調査の0.83 ± 0.49(スコア)から第二段階の0.98 ± 0.32(スコア)へ上昇した($p < 0.01$)。

なお、第二段階のデルファイ調査の個別評価において「期待される効果」で評価が最も高かったのは、血圧測定で4.14 ± 2.28(スコア)であった(表6)。また「避けがたい被害」の評価は、BMIで-3.70 ± 1.85(スコア)と最も良かった。さらに「必要となる費用」の評価も、BMIで-2.51 ± 2.28(スコア)と最も小さかった。続いて「検査実践の状況」で評価が最も高かったのは、血圧測定で4.14 ± 2.28(スコア)であった(図3)。一方、意見が収斂した第二段階のデルファイ調査における総合評価でも、血圧測定が1.41 ± 0.31(スコア)と最上位に位置付けられ、次いでHbA1cが1.35 ± 0.35(スコア)、ALTが1.34 ± 0.39(スコア)となった。

表6 各個別評価において上位の検査項目の評価結果

上位10検査項目	第一段階		第二段階	
	検査	平均 標準偏差	検査	平均 標準偏差
効果	検査	平均 標準偏差	検査	平均 標準偏差
	血圧測定	2.56 ± 1.39	血圧測定	4.13 ± 2.28
	HbA1c	2.51 ± 1.51	HbA1c	4.08 ± 2.37
	子宮頸部細胞診	2.43 ± 1.44	子宮頸部細胞診	3.99 ± 2.44
	便潜血	2.39 ± 1.43	マンモグラフィ	3.99 ± 2.50
	血糖(空腹時)	2.37 ± 1.57	HCV抗体	3.92 ± 2.52
	マンモグラフィ	2.25 ± 1.37	LDLコレステロール	3.89 ± 2.49
	LDLコレステロール	2.22 ± 1.47	血糖(空腹時)	3.80 ± 2.32
	BMI	2.22 ± 1.39	血色素	3.78 ± 2.22
	血色素	2.19 ± 1.55	γ-GTP	3.78 ± 2.58
	腹部超音波	2.19 ± 1.44	尿酸	3.75 ± 2.67
	検査	平均 標準偏差	検査	平均 標準偏差
被害	検査	平均 標準偏差	検査	平均 標準偏差
	身長	-2.70 ± 2.15	BMI	-3.70 ± 1.85
	体重	-2.63 ± 2.18	心拍数	-3.68 ± 1.87
	肥満度	-2.61 ± 2.19	血圧測定	-3.65 ± 1.87
	BMI	-2.59 ± 2.20	身長	-3.57 ± 1.82
	心拍数	-2.51 ± 2.18	体重	-3.55 ± 1.90
	視力検査	-2.43 ± 2.17	視力検査	-3.54 ± 2.02
	血圧測定	-2.42 ± 2.20	聴力検査	-3.49 ± 2.00
	聴力検査	-2.41 ± 2.22	心電図	-3.49 ± 1.94
	心電図	-2.41 ± 2.21	腹部超音波	-3.48 ± 2.25
	腰痛	-2.33 ± 2.23	特定健診22項目問診票	-3.45 ± 2.01
	検査	平均 標準偏差	検査	平均 標準偏差
費用	検査	平均 標準偏差	検査	平均 標準偏差
	BMI	-1.52 ± 1.71	BMI	-2.51 ± 2.28
	肥満度	-1.51 ± 1.71	肥満度	-2.39 ± 2.25
	心拍数	-1.49 ± 1.61	腰痛	-2.32 ± 2.15
	腰痛	-1.44 ± 1.75	身長	-2.32 ± 2.27
	体重	-1.34 ± 1.59	心拍数	-2.32 ± 2.18
	血圧測定	-1.24 ± 1.71	体重	-2.16 ± 2.19
	身長	-1.22 ± 1.52	血圧測定	-2.01 ± 2.15
	視力検査	-1.02 ± 1.74	視力検査	-1.36 ± 1.69
	尿酸	-1.01 ± 1.64	内科	-1.25 ± 2.01
	尿比重	-1.00 ± 1.52	尿蛋白	-1.13 ± 1.65
	検査	平均 標準偏差	検査	平均 標準偏差
実践	検査	平均 標準偏差	検査	平均 標準偏差
	血圧測定	2.27 ± 1.85	血圧測定	3.24 ± 2.56
	血糖(空腹時)	2.20 ± 1.84	ALT(GPT)	3.11 ± 2.53
	ALT(GPT)	2.19 ± 1.84	γ-GTP	3.10 ± 2.55
	AST(GOT)	2.17 ± 1.84	内科	3.08 ± 2.59
	血色素	2.16 ± 1.85	AST(GOT)	3.04 ± 2.52
	胸部X線	2.15 ± 1.81	尿蛋白	3.01 ± 2.52
	尿蛋白	2.15 ± 1.88	血糖(空腹時)	2.99 ± 2.48
	中性脂肪	2.14 ± 1.85	白血球	2.98 ± 2.56
	LDLコレステロール	2.10 ± 1.83	血色素	2.94 ± 2.48
	赤血球	2.10 ± 1.86	中性脂肪	2.92 ± 2.41

(3) “費用対効果”(パフォーマンス)と他の結果

「期待される効果」と「必要になる費用」の調査で得られた結果から、基本検査項目等の費用対効果(パフォーマンス)を整理した(図4)。多くの基本検査項目等は、費用効果的(cost-effective)なゾーン(効果がプラスでかつ費用がマイナス)にあった。特にBMIは、期待される効果が3.41±2.47(スコア)、必要となる費用が-2.51±2.28(スコア)と高

いパフォーマンスが伺えた。同様に血圧測定も、期待される効果が4.14±2.28(スコア)、必要となる費用が-2.01±2.15(スコア)とパフォーマンスが高かった。

なお、「期待される効果」や総合評価に対して「必要になる費用」は、相関分析において統計学的に有意な関係になく、相互に独立変数と見なされた(表3)。また、参考までに、“費用対効果”の分析

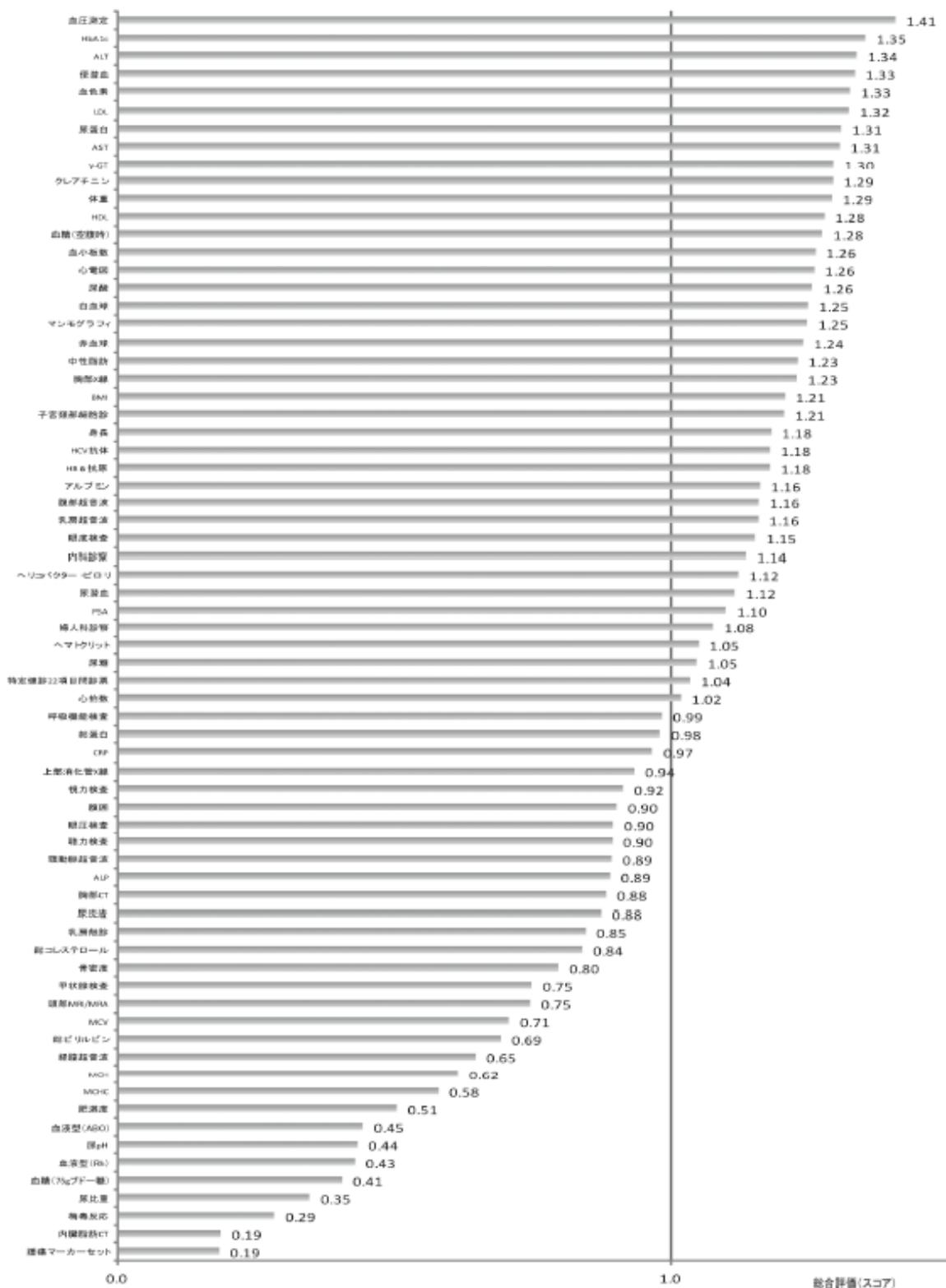


図3 基本検査項目等の総合評価の結果
70項目について総合評価(ゼロが基準で数値が大きいものが高い評価)の順位を整理した。

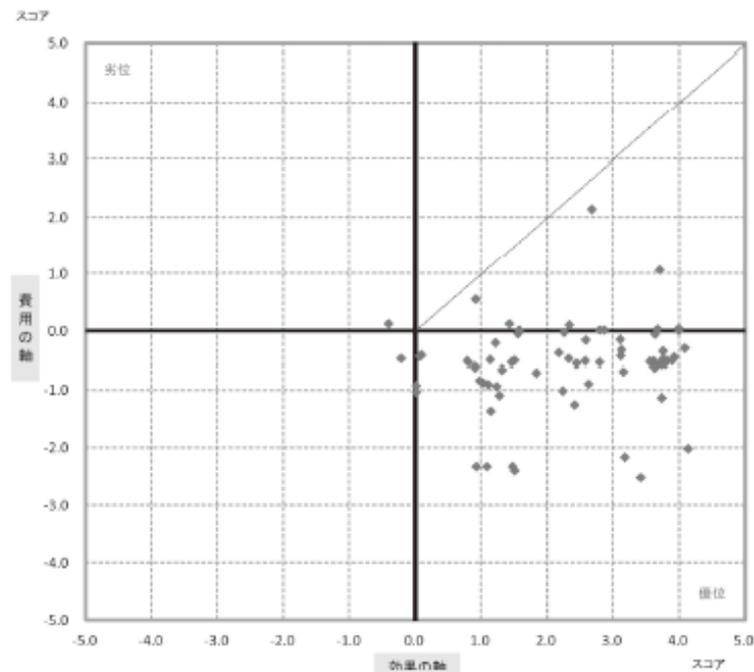


図4 基本検査項目等の社会経済性(費用対効果)の分析結果

費用のパラメータを縦軸に、効果のパラメータを横軸に設定し、各検査項目に対する評価結果をプロットした。右下の面は、費用対効果(パフォーマンス)が優位なゾーンとなり、左上の面は、費用対効果(パフォーマンス)が劣位なゾーンとなる。

結果と総合評価への回答の相互関係について整理を行ったところ、統計学的な相関関係にないことが明らかとなった($R_s=0.017$ $p=0.882$)。

以上から、さらに精緻な分析を要するものの、基本検査項目等に対するエキスパートの総合評価において、「費用対効果」や社会経済性の要素は大きな比重を占めないと考えられた。つまり総合評価は、施設経営等の観点とは独立して回答がなされており、関連するバイアスは小さいと理解された。

考 察

本研究では、人間ドックの基本検査項目の社会的意義に関する評価として、デルファイ法によりエキスパート・オピニオンの意見集約を試みた結果、以下の示唆を得ている。

(1)基本検査項目等の評価

「期待される効果」、「避けがたい被害」、「必要となる費用」、「検査実践の状況」の4つのパラメータを用いた個別評価によると、70の基本検査項目等の大部分は、医学的效果が期待できるうえ、

健康関連の被害も少なく、さらに関わる検査費用や検査の普及状況も適正の範囲にあると推察される。また、「期待される効果」と「必要となる費用」による「費用対効果」の整理から、多くの検査項目が費用効果的(cost-effective)な状況にあると考えられる。

5つのランクによる総合評価では、70の基本検査項目等のうち、55.7%の項目がスコア1(重要/必要:ランク+1)よりも上の評価を平均的に得ており、多くの検査項目がエキスパートから比較的良好に認識されていると考えられる。この結果の解釈は、第一段階のデルファイ調査から第二段階にかけて、9割弱の基本検査項目等において回答が統計学的に収斂していた点、および前述の個別評価の分析結果からも支持されると推察される。

なお本研究では、デルファイ調査の過程で意見が収斂しなかった基本検査項目等が1割ほど散見される。研究の質を向上させるには、この実態が調査手法上の条件から生じるものなのか、または予防医学の領域特性によるものなのか、その背景

や理由について注意深く検討を進め、得られた知見を結果の解釈等に反映することが望まれる。そのためには、回答者に対する確認プロセスと要因の構造化等が不可欠であり、今後の研究課題に位置付けられる。

(2)回答者のバイアス

デルファイ法においては、調査手法の課題として、有意抽出法(Non-random sampling)が議論になる。一方で、本研究のような主旨における対象者の選定や属性について、最も望ましい抽出法に関する指針は明確ではない¹⁰⁾。本研究では、日本人間ドック学会が規定する専門医を対象としたため、回答者の基本検査項目等に関する専門性は担保がなされているものの、デルファイ調査に対する協力が任意であったため、一定の参加バイアスが潜在すると考えられる。

例えば、エキスパート・オピニオンといえども、本研究のような主観的なデータをもとにした結果については、一般に、専門資格や所属団体、就労環境等の回答者が有する社会的な属性がバイアス要因として挙げられる。

本研究では、専門資格や各種経験のバイアスに関しては、人間ドックの従事歴が各種評価に及ぼす影響を検証することで、専門医グループ自体の均質性を示すと同時に、サンプルの構造が妥当であることを明らかにしている。

さらに、基本検査項目等に対する総合評価において、“費用対効果”や社会経済性の要素が占める比重は、おおむね小さいことが統計的に示唆されているため、施設経営等の観点とは独立した回答行動がなされており、本研究の結果に対する就労環境等のバイアスは低いと理解される。

なお本研究では、第一段階から第二段階のデルファイ調査に移行する過程で、約4割のサンプルが母集団から脱落している。これが結果にどのような影響を与えているのか、今回の分析において十分な配慮を行っていないが、研究自体に消極的(理解不足)な専門医とともに、基本検査項目等に対する否定的な意見を持つ群が相対的に多いと予想される。今後は、本研究の意義をさらに高めるためにも、示されたこの論点について引き続き探

究が望まれる。

(3)分析結果の検証方法

本研究で得られた結果をより頑強にするためには、アプローチの異なる研究成果と照合することが重要となる。その手法の1つに、予防医学や公衆衛生学、医療経済学に関する先行研究(大規模試験やリアルワールド)の報告について、システマティックレビューを行い、メタアナリシス等を展開することが挙げられる。

例えば、一次予防を主眼に予防的な健診の有効性を評価したランダム化比較試験(RCT)がある¹¹⁾。壮年期の約1万人を対象とした評価では、総死亡率や休業回数に差は認められていないが、一部の原疾患について、死亡率で統計的に有意な減少が認められている。

ただし、このような研究成果は多くないため、人間ドックを含む予防的な健診は、EBMの観点からエビデンスレベルや研究報告の数量で十分とは言えない状況にある。以上を踏まえつつ、参考までに、本研究の「期待される効果」の評価が最上位の血圧測定および“費用対効果”が最上位のBMIについて、主な臨床ガイドラインにおける評価状況や先行する研究報告の動向について整理を行い、それら他の研究成果との照合により、本研究の簡易検証を試みる(表7)。

脳卒中や末期腎不全、心筋梗塞等の危険因子をコントロール(早期診断と検査管理)する予防医学の意義は、臨床現場において広く認知されており、それらを背景に血圧測定は本研究においても高く評価されたと推察される¹²⁻¹⁷⁾。BMIを予防医学に位置付けた費用対効果は、残念ながら研究報告¹⁸⁻²⁰⁾は少ないが、小児の肥満防止に対して、BMIをインジケータとする健康管理ツールの“費用対効果”は高い(AU\$376/ Δ BMI)という報告がある¹⁸⁾。このように、間接的な有効性も含む蓋然的な整理の範囲ではあるものの、両者の解釈が概ね一致するため、本研究の結果は一定の信頼性があると示唆される。

本結果が恒常的な最終結果ではなく、医療技術進歩や価値観の変遷により、時代とともに変化することが推察され、継続的に検討を重ねてその時

表7 有効性や費用効果に関わる先行研究の概要

検査項目 (期待される効果)	イベントリスク	エビデンスレベル	出典 (引用文献番号)
	脳卒中の危険因子 ・収縮期血圧160mmHg以上の患者の脳梗塞の発症リスクは3.46倍、拡張期血圧95mmHg以上では3.18倍	II b (日本高血圧学会高血圧治療ガイドライン2013)	12), 13)
血圧測定	末期腎不全の発症リスク ・収縮期血圧が10 mmHg 上昇することに末期腎不全の発症は20～30%上昇	I (エビデンスに基づくCKD診療ガイドライン2009)	14)～16)
(早期診断と検査管理の意義)	心筋梗塞の発症リスク ・収縮期血圧値1標準偏差(19mmHg)の増加に対する心筋梗塞の発症リスクは、男性で1.2倍、女性で1.3倍と高い	II a (心筋梗塞二次予防に関するガイドライン2011)	17)
検査項目 (費用対効果)	健康改善効果	研究デザイン	出典 (引用文献番号)
	小児の肥満防止 ・BMIをインジケータとする健康管理ツール(healthy beginnings)の費用対効果は高い(AU\$376/ΔBMI)	RCT	18)
BMI	成人病の肥満抑制 ・BMIをインジケータとする生活改善活動(practice nurses)の費用対効果は高い(US\$536/ΔBMI)	66症例、12カ月観察	19)
(検査管理と指導介入の意義)	学童の肥満抑制 ・BMIをインジケータとする生活指導(医師等)や学習トレーニングの経済性は高い可能性あり(£1,500/ΔBMI/ids)	143症例、12カ月観察	20)

参考までに、「期待される効果」の評価が最上位の血圧、および「費用対効果」が最上位のBMIについて、主な臨床ガイドラインにおける評価状況や研究報告の動向を精査した。

代ごとに目的に応じて結果も変遷していくことが考えられる。また、人間ドックは内科医中心としてさまざまなバックグラウンドの医師が参画しているが、ある意味で幅広い視点に立って順位付けが可能であったとも考えられる一方、なじみの薄い項目は評価にバラツキがある状況も想像される。この点も考慮が必要である。また、サービス提供サイドの医師の視点だけでなく、サービス受益者サイドの人間ドック受診者や患者の視点(効用値を含む)でも、今後分析が必要である。また、すべての人間ドックの項目が本検討で用いた4つの評価指標だけで評価出来るとは考えていない。

以上から、基本検査項目等の社会経済的な評価は、それを取り巻く社会環境や研究実態を踏まえつつ、アウトカム系の指標によるTA等も視野に入れながら、将来、さらに多面的なエビデンスの蓄積と検証が望まれる。さらに、本検討をもとに、人間ドックに携わる医師と人間ドック受診者間の双方向コミュニケーションが活性化され、適正な人間ドックの基本検査項目選択に必要な健全な意思決定(Decision making)の環境が育まれることが期待される。

結 論

70の基本検査項目等の社会的意義を、専門医グループによるデルファイ法にて評価したところ、第二段階の調査において意見の収斂が認められた。また、各項目は、医学的効果が期待され健康関連の被害も少なく、関わる検査費用や検査の普及状況も概ね適正範囲にあることが示された。費用対効果の分析においても、多くの項目は費用対効果が良い結果となった。以上から、基本検査項目は、社会的な意義が高いと示唆された。

謝 辞

ご多忙のおり、本研究のデルファイ調査にご協力頂いた人間ドック健診専門医の方々には、厚く御礼を申し上げます。

利益相反

すべての著者について、本研究に関しての利益相反はありません。

(注)：本論文で用いる「費用対効果」は、医療経済領域で生存年や血圧等の検査値や質調整生存年(QALY)を効果指標として用いる「費用効果(cost-effectiveness)分析や費用効用(cost-utility)分析」と厳密には区別されるもので、社会経済研究手法にもとづく「費用対効果」という意味で用いた。また、本研究では人間ドック専門医の視点で健康改善や社会貢献を「期待される効果」としている。

文献

- 1) Yamakado M: Taking stock of the Ningen Dock. *Highlighting Japan* 2010; 4(2): 10-11.
- 2) 日本人間ドック学会：2012年人間ドックの現況, http://www.ningen-dock.jp/wp/common/data/other/release/dock-genkyou_h24.pdf [2014.6.4]
- 3) Norman Dalkey, Olaf Helmer: An Experimental Application of the Delphi Method to the use of experts. *Management Science* 1963; 9: 458-467.
- 4) Becker ER, Dunn D, Braun P, et al: Refinement and expansion of the Harvard Resource-Based Relative Value Scale: the second phase. *Am J Public Health* 1990; 80: 799-803.
- 5) Michels ME, Evans DE, Blok GA: What is a clinical skill? Searching for order in chaos through a modified Delphi process. *Med Teach* 2012; 34: e573-581.
- 6) Maarsingh OR, Dros J, van Weert HC, et al: Development of a diagnostic protocol for dizziness in elderly patients in general practice: a Delphi procedure. *BMC Fam Pract* 2009; 10: 12.
- 7) Cook C, Brismée JM, Fleming R, et al: Identifiers suggestive of clinical cervical spine instability: a Delphi study of physical therapists. *Phys Ther* 2005; 85: 895-906.
- 8) 田倉智之：人間ドック健診と医療経済. *人間ドック* 2012; 26: 775-785
- 9) Petitti DB, Teutsch SM, Barton MB, et al; U.S. Preventive Services Task Force: Update on the methods of the U.S. Preventive Services Task Force: insufficient evidence. *Ann Intern Med* 2009; 150: 199-205.
- 10) Murphy MK, Black NA, Lamping DL, et al: Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998; 2: i-iv, 1-88.
- 11) Friedman GD, Collen MF, Fireman BH: Multiphasic Health Checkup Evaluation: a 16-year follow-up. *J Chronic Dis* 1986; 39: 453-463.
- 12) Kannel WB, Wolf PA, McGee DL, et al: Systolic blood pressure, arterial rigidity, and risk of stroke. The Framingham study. *JAMA* 1981; 245: 1225-1229.
- 13) Tanaka H, Ueda Y, Hayashi M, et al: Risk factors for cerebral hemorrhage and cerebral infarction in a Japanese rural community. *Stroke* 1982; 13: 62-73.
- 14) Tozawa M, Iseki K, Iseki C, et al: Blood pressure predicts risk of developing end-stage renal disease in men and women. *Hypertension* 2003; 41: 1341-1345.
- 15) Shulman NB, Ford CE, Hall WD, et al: Prognostic value of serum creatinine and effect of treatment of hypertension on renal function. Results from the hypertension detection and follow-up program. The Hypertension Detection and Follow-up Program Cooperative Group. *Hypertension* 1989; 13(5 Suppl): 180-93.
- 16) Walker WG, Neaton JD, Cutler JA, et al: Renal function change in hypertensive members of the Multiple Risk Factor Intervention Trial. Racial and treatment effects. The MRFIT Research Group. *JAMA* 1992; 268: 3085-3091.
- 17) Miura K, Nakagawa H, Ohashi Y, et al: Four blood pressure indexes and the risk of stroke and myocardial infarction in Japanese men and women: a meta-analysis of 16 cohort studies. *Circulation* 2009; 119: 1892-1898.
- 18) Hayes A, Lung T, Wen LM, et al: Economic evaluation of "healthy beginnings" an early childhood intervention to prevent obesity. *Obesity (Silver Spring)* 2014; doi: 10.1002/oby.20747. [Epub ahead of print]
- 19) Karnon J, Ali Afzali HH, Gray J, et al: A risk adjusted cost-effectiveness analysis of alternative models of nurse involvement in obesity management in primary care. *Obesity (Silver Spring)* 2013; 21: 472-479.
- 20) Hollinghurst S, Hunt LP, Banks J, et al: Cost and effectiveness of treatment options for childhood obesity. *Pediatr Obes* 2014; 9: e26-34.

Socioeconomic Evaluation Regarding The Standard Testing Items in The Multiphasic Health Check-ups (Ningen Dock): a Delphi Survey

Toshiyuki Takura, Hiroki Sugimori, Toshihiko Sato, Takeo Nakayama, Norihide Takaya,
Kiyooki Watanabe, Minoru Yamakado:

Academic Committee, Subcommittee for The Standard Testing Items in Ningen Dock Health Check-ups

Abstract

We evaluated the socioeconomic significance regarding the standard testing items in the multiphasic health check-ups (Ningen Dock) by the Delphi survey among the Ningen Dock specialists. We surveyed by reference to the methods of the U.S. Preventive Services Task Force, which tackled with the issue of "insufficient evidence". The 4 evaluation domains were "potential preventable burden", "potential harm of the intervention", "costs (both monetary and opportunity)", and "current practice". This study will contribute to the appropriate testing in Ningen Dock.

Keywords: standard testing item, socio-economical study, Delphi survey, USPSTF, EBM
