

マス・スクリーニングシステムのテクノロジー・アセスメント
－評価枠組および現在の動向－

久繁哲徳

要約：わが国で実施されている新生児の先天代謝異常・内分泌疾患のマススクリーニング・システムは、他の先進諸国でも類を見ない広範囲な内容を有している。しかしながら、現在の効果評価・効率評価では、世界的には先天性甲状腺機能低下症およびフェニルケトン尿症の2疾患のみ明確な根拠が確立している。その意味で、わが国のスクリーニング・システムに関する総合的な評価が求められている。その具体的な評価方法として、テクノロジー・アセスメントの枠組みを導入し、新生児マス・スクリーニングの評価課題を設定した。第一段階として、米国特別委員会の最良根拠の方法により、各疾患別にスクリーニングの効果に関する批判的吟味を実施することとした。第二段階として、この効果評価を基礎に、各疾患毎に費用－便益分析を実施するとともに、その結果を総合的に評価し、スクリーニング・システムの効率を検討する。

見出し語：テクノロジー・アセスメント、効果、効率

はじめに

わが国における疾病予防対策において、二次予防（マス・スクリーニング）は極めて重要な位置を占めている。乳幼児に対する先天代謝異常・内分泌疾患健診、学童に対する定期健診、成人に対する循環器・がん検診など、全国におよぶ多様な検診システムが確立されている¹⁻³⁾。

しかしながら、こうした検診システムについては、その効果が明確に確立

されていない段階で導入・普及されていることが大きな問題点であることが指摘されている¹⁻³⁾。とくに、検診の効果に関しては、単に疾患の早期発見に留まるのではなく早期治療による最終的な健康結果の改善が達成されるべきであること、また有害性と得失から評価する必要があることに注意すべきである。

こうした観点から、世界的には1960年代以降、健康診断のあり方について急速に見直しが進められてきた¹⁻³⁾。その中でも、1979年のカナダ定期健康診断特別委員会（Task force）⁴⁾、19

88年の米国予防医学特別委員会⁵⁾の評価は、現在までに報告された世界中の全文献を総合的に評価した記念碑的な成果であった。

これらの評価の結論としては、明確な効果の根拠が認められる健康診断の数は極めて限られていること、さらに効果的な予防対策を実施するためには、効果の認められた健康診断を系統的に積み上げて行くことが重要であることが指摘されている^{4, 5)}。

また、検診を健康政策として実施する場合は、社会資源に限りがあるため、たとえ効果的な検診であっても、その優先順位を決定し実行することが求められる。その場合は、検診に要する費用と得られる利益を結びつける経済的効率を評価する必要がある²⁾。

こうした臨床的な効果と経済的効率を総合的に評価する方法として、近年、医療のテクノロジー・アセスメントが注目されている^{6, 7)}。わが国においても、近年、学校健診において適用が試みられるようになった⁸⁾。そこで、二次予防の中でも重要な位置を占めている先天代謝異常・内分泌疾患のスクリーニングシステムについて、テクノロジー・アセスメントの枠組みに基づき、臨床的效果とともに経済的効率の評価を実施したいと考えた。

今回は、その評価に向けて、テクノロジー・アセスメントの評価枠組みを設定するとともに、実施計画の策定、および現在の国際的な動向について検討を行うこととした。

1. テクノロジー・アセスメントの評価枠組み

1) 評価方法

医療のテクノロジー・アセスメントは、保健医療政策の総合的な研究方法であり^{6, 7, 9)}、「技術の適用にともなう、技術的・経済的・社会的結果を検討する、包括的な政策研究の一分野」を医療に適用したものである。

医療のテクノロジー・アセスメントの具体的な評価過程を、医療技術の普及過程と対比させて示したのが、図1・図2である。

医療技術の開発は、通常、実験による技術革新に始る。そして動物実験により効果が期待でき、安全性が保証された段階で、臨床での評価が開始される。

まず最初は、小数の患者を対象とした事例研究が行われる。そこで有効性が確認され安全性に問題がなければ、具体的な医療技術の<効能>(efficacy)と<安全性>(safety)を評価する。<効能>とは、理想的条件下で特定の対象者に対する臨床の有効性が認められる場合をさす。

スクリーニングでは、特定の対象あるいは一部の地域での有効性を意味する。この時、疾患の早期診断に対する有効性が先行するが、その後の早期治療による健康結果の改善の評価が求められる。また同時に、医療技術は、つねに危険と背中合せであるため、安全かどうかを並行して評価する³⁾。ただし、スクリーニングの場合は、検査の危険性は一般的に少ない。

こうした評価の関門をくぐることでできた医療技術については、さらに一般的な条件下で多様な対象者についても同じように有効性が認められるか、つ

まり＜効果＞（effectiveness）の評価を行う。スクリーニングでは、広範な対象および地域でも、効能と同様な有効性が認められるかどうかを評価する。また、早期診断とともに健康改善での有効性評価が必要とされる。

それと平行して、医療技術が必要とする人に利用できるような状態となっているかどうか、＜利用度＞（availability）の評価を行っておかなければならない。スクリーニングでは、全員を対象として実施するか、または特定の高危険群を対象として実施するか評価する必要がある。

ところが、われわれの社会では資源に限りがあるため、効果が認められた医療技術をすべて思うままに利用することはできない。医療技術の利用によって、それに費やす社会的資源から最大限の利益が得られるよう評価する必要がある。このように費用と健康結果とを結びつけて評価するのが、＜効率＞（efficiency）すなわち医療経済的評価である。経済的評価は、医療の領域でもとくにスクリーニングで精力的に実施されてきており、他の保健医療技術との比較、あるいはスクリーニングの実施内容・程度の比較など重要な役割を果たしている。

また臓器移植、脳死、遺伝子治療、胎児医療などのような医療技術の種類によっては、それと平行して、生命倫理および社会・文化的影響についても評価を行う必要がある。スクリーニングでも出生前診断では、これらの評価を総合的に実施することが求められる。

これらの評価によって保証された医療技術は一般の臨床に普及し、確立さ

れた技術として利用されることになる。ただし、こうした評価過程の間でも、有効性が否定されるなど問題となる医療技術は放棄される。また、新しい代替技術が開発され旧式化した技術も放棄される。

ただし、有効な医療技術であっても、適正な対象に適正な利用が行われているか、＜監視＞（monitoring）の評価を継続して行うことが重要な課題となる。スクリーニングにおいても、広範囲な対象・地域で実施されている段階では、定期的に追跡調査として早期診断および健康改善を実施する。

この一連の評価を総合的に実施するのが医療のテクノロジー・アセスメントであり、もっとも重要な決め手は効果と効率の評価である。

2) 評価指標

実際に行われているテクノロジー・アセスメントの評価範囲と実施状況を表1に示した⁷⁾。医療技術は、薬、医療機器、治療手技など狭義の医療技術にとどまらず、支援システム、組織・管理システムなど広義の技術までを含む。こうした評価は、多くの場合、効果に効率の評価を組み合わせる形で行われる。ただし先にも述べたように、臓器移植や脳死、遺伝子操作など社会におよぼす影響の大きい問題については、倫理的、法的、あるいは社会文化的な評価まで範囲を広げることが必要となるが、その頻度はあまり大きなものではない。

ただし、テクノロジー・アセスメントの内容は1970年代から80年代にかけて徐々に変化してきている（表2）¹⁰⁾

。たとえば、視野についてみると、当初は安全性と効能を評価範囲であったのが、近年では効果、効率、生命倫理へと範囲が広がっている。こうした評価の焦点が広がりや深まりを見せるにつれて、評価実施者は、科学者・医者から経済学・疫学・統計学などの学際的な集団へと変化してきた。また一方で、評価利用者も、それに対応して、医者から一般大衆まで含むさまざまな関連集団へと広がっている。

このようにテクノロジー・アセスメントの内容が量・質ともに変化して行く中で、とくに注目されるのは評価の指標である。その指標は、図3に示すように歴史的に変化してきた¹¹⁾。

当初は、医療施設、機器設備、医療従事者数、従事者資格など<構造> (structure) の評価を行っていた。これは医療に投入する資源そのものを意味している。スクリーニングでは、検査機器、実施頻度などがこれに該当する。それに引続き、<過程> (process) に評価の重点が移った。これは、医者と患者の間で行われる医療活動の内容を指している。したがって、検査や診断、治療の内容が、時代の要求する適切な基準に合致しているか否かを評価する。具体的にはカルテの審査などがこれに該当する。スクリーニングでは、早期診断が行われているか、検査の有効性評価が焦点となる。また、早期診断後の早期治療へのコンプライアンスも評価の対象として重要な意味を持つ。

これら2つの指標は比較的測定し易いため、評価によく利用されてきた。もちろん、優れた<構造>であれば、

適切なく<過程>につながり、ひいては最終的なく<健康結果>の改善につながると考えられる。しかしながら、こうした指標が、目的とする健康改善に本当に関連しているかどうかは不明である。そこで、近年は、医療の最終的な産出となる<健康結果> (health outcome) の改善に焦点をあてるように変化してきた。具体的には、有病率、死亡率、合併症発生率、医療過誤、在院日数などが用いられている。スクリーニングでは、対象疾患による死亡率、精神的・身体的障害発生率などが健康結果の指標となる。

2. 新生児マススクリーニングの評価方法

テクノロジー・アセスメントの枠組みにしたがって、新生児の先天異常・内分泌疾患のマス・スクリーニングの評価を実施することを計画した。

評価の過程は図4に示すように、4つの段階に分けられる。第1段階は、疾患の負担の評価である。疾患が、個人あるいは社会に対してどの程度の負担を与えているかが問題となる。したがって、発生頻度、死亡率、有病率、障害の程度、治療費用の項目について評価する。

第2段階は、スクリーニング検査の有効性の評価である。第一の基準は、検査の安全性、簡便性、費用、患者の受入れ易さ、および検査手技の測定可能性である。つまり検査の技術的有効性の評価である。第二の基準は、スクリーニング検査の感度、特異度、適中率である。この基準では、スクリーニングによる疾患の早期発見の有効性を

評価する。通常、スクリーニングの有効性評価は、この段階の評価に留まるものが多いが、図5に示すように、これはあくまでスクリーニングの中間的な有効性である。早期発見が早期治療につながり、最終的に健康結果の改善が認められない限り、スクリーニングを実施する意味はない。

第3段階は、スクリーニングの効果評価である。効果については、適切な健康結果の指標を選択し、その改善が認められるかどうか検討する必要がある。その基準としては、カナダおよび米国の特別委員会の基準を採用する(表3)^{4, 5)}。この基準は、効果評価に用いられた初めての明示的な基準であり、研究設計の効力に基づいて判定が行われる。最も効力の高いのは無作為臨床試験(randomized controlled trial)であり、分析疫学的分析がそれに次ぎ、観察疫学的分析、専門家の意見と効力は低下する。とくに、スクリーニングの効果評価を行う上では、表4に示す4つの偏り(bias)を十分に吟味することが求められる。

第4段階は、経済的評価である。複数の代替案の中かから、最も費用-効果的、費用-便益的なスクリーニングを評価・選択する。経済的評価には、臨床的効果における無作為臨床試験のような黄金律が存在しないため、表5に示すような、経済的評価の適切性(質)を保証する基準を用いて評価する^{12, 13)}。ただ注意すべき点は、経済的評価は、明確な効果を前提とする。したがって、効果が認められない場合は実施する意味がない。

以上のような枠組みで評価を行うが、

実際の評価に際しての情報源が問題となる。情報源から見た評価方法は2つに大別される(表6)。一つは、既存の情報を分析し統合評価する方法である。テクノロジー・アセスメントは、前に述べたように、多角的総合的に医療技術を評価するため、情報源の多くの部分を既存の情報から得ることが多い。例えば、コンピュータ文献検索を行い批判的吟味を行う。この場合、短時間で比較的少ない費用により評価が可能となる。しかしながら、目的とする分析に完全に適合する情報は限られているため、質の面で問題が残される。

もうひとつは、新たに評価のために、調査・研究を実施し必要な情報を入手する方法である。この方法では、分析にとって質の高い情報が利用可能となる。しかしながら、長時間を要し、費用も高い。その意味では、実際の分析では、これらの方法を組み合わせて情報を入手することになる。今回の評価でも、前者を中心として情報を入手し評価を行う。

統合的な評価の方法としては、判断分析、メタ・アナリシス、合意形成、最良根拠などの方法があるが²⁾、今回は、効果については米国の特別委員会の最良根拠の方法を、また効率については判断分析の方法を用いる。

3. マスクリーニング評価の動向

新生児のマスクリーニングに関する評価については、個別の研究が実施されているが、総合的なテクノロジー・アセスメントは実施されていない。そうした試みの一つが、カナダおよび米国の特別委員会の評価である。わが

国で実施されている新生児の先天代謝異常・内分泌疾患のスクリーニングの項目について、カナダの特別委員会の評価結果を対比してみよう⁴⁾。表7に示すように、検討されているのは先天性甲状腺機能低下症とフェニールケトン尿症の2疾患のみである。

先天性甲状腺機能低下症およびフェニールケトン尿症はいずれも、根拠の質としては、無作為臨床試験は実施されておらず分析疫学的研究に基づくものであるが、明確な効果が認められるため、勧告では実施すべき優れた根拠があるとされている。

カナダの委員会の評価結果から10年を経過した後、米国の特別委員会でも同様の評価が行われたが、対応する疾患としてはカナダと同様2疾患であり、評価結果は一致している⁵⁾。

一方、他の国ではこうした明確な基準に基づく系統的な評価が行われていないが、英国では疫学者の Holland¹⁾がスクリーニング全般の評価を行っており、その中で Hall 委員会の評価を紹介している。その評価結果では、カナダ・米国同様に先天性甲状腺機能低下症およびフェニールケトン尿症については明確な根拠があるととして実施を勧めているが、その他の先天代謝異常疾患に関しては今後の検討が必要であると指摘している。

これら2疾患に対する効率の評価も種々行われており、とくにフェニールケトン尿症のスクリーニングは、最も費用-効果的な保健医療プログラムであることが認められている¹⁶⁾。

しかしながら、わが国のように新生児の先天代謝異常・内分泌疾患に対す

る広範囲なスクリーニング・システムを確立している国はまれであり、その意味では上記の2疾患以外のスクリーニングの効果評価は現在も確立されていないし、効率評価もほとんど実施されていない。したがって、わが国の実際の研究結果に基づき、スクリーニング・システムの効果評価・効率評価を総合的に実施すべき必要があると考えられる。

まとめ

わが国で実施されている新生児の先天代謝異常・内分泌疾患のマススクリーニング・システムは、他の先進諸国でも類を見ない広範囲な内容を有している。しかしながら、現在の効果評価・効率評価を見る限り、世界的には先天性甲状腺機能低下症およびフェニールケトン尿症の2疾患が明確な根拠が確立している状況である。その意味では、わが国のスクリーニング・システムに関する総合的な評価が求められているといえる。その具体的な評価方法として、テクノロジー・アセスメントの枠組みの下で、評価課題を設定した。今後、この課題に沿ってマススクリーニング・システムの効果と効率の評価を実施するが、今年度は各疾患別に、米国特別委員会の最良根拠の方法により、スクリーニングの効果に関する批判的吟味を実施することとした。次年度は、この効果評価を基礎に、各疾患毎に費用-便益分析を実施するとともに、その結果を総合的に評価し、スクリーニング・システムの効率を検討する。

文献

- 1) 久繁哲徳：臨床の指針（その1），健康診断・人間ドックをめぐる，日放技会誌，39:479-483,1992
- 2) 久繁哲徳：疾病予防対策の評価と実行に向けての合意形成，成人病の二次予防と一次予防の効果をめぐる，日本医事新報，3516:48-52,1991
- 3) 久繁哲徳：効果的・効率的医療の検討，上・下，社会保険旬報，1751:10-13, 1752:16-21,1992
- 4) Canadian Task Force on the Periodic Health Examination: Task force report, Can Med Assoc J, 121 : 1193-1254, 1979
- 5) US Preventive Services Task Force: Guide to Clinical Preventive Services, William & Wilkins, New York, 1989
- 6) Banta HD & Behney CJ:Technology assessment, Last JH ed, Public Health and Preventive Medicine, 12th ed, ACC, 1986
- 7) Committee for Evaluating Medical Technologies in Clinical Use: Assessing Medical Technologies, Nation Acad Press, 1985
- 8) Hisashige A: Technology assessment of periodic health examinations for school children in Japan, Int J Technol Assess Health Care, 8:219-233,1992
- 9) Office of Technology Assessment : Medical Technology and costs of the Medical Program, GPO, 1984
- 10) Fuchs VR & Garber AM: The new technology assessment, New Engl J Med, 323:673-677, 1990
- 11) 久繁哲徳：看護の質とクウォリ-，教育と医学，40(5):461-470,1992
- 12) 久繁哲徳 編：臨床判断学，篠原出版，1990
- 13) 久繁哲徳，西村周三 監訳：臨床経済学，篠原出版，1989
- 14) Holland WW & Stewart S: Screening in health care, Nuffield Provincial Hospital Trust, 1990
- 15) Torrance GW: Measurement of health state utilities for economic appraisal, J Health Econom, 5:1-30,1986

図1 テクノロジー・アセスメントの評価過程

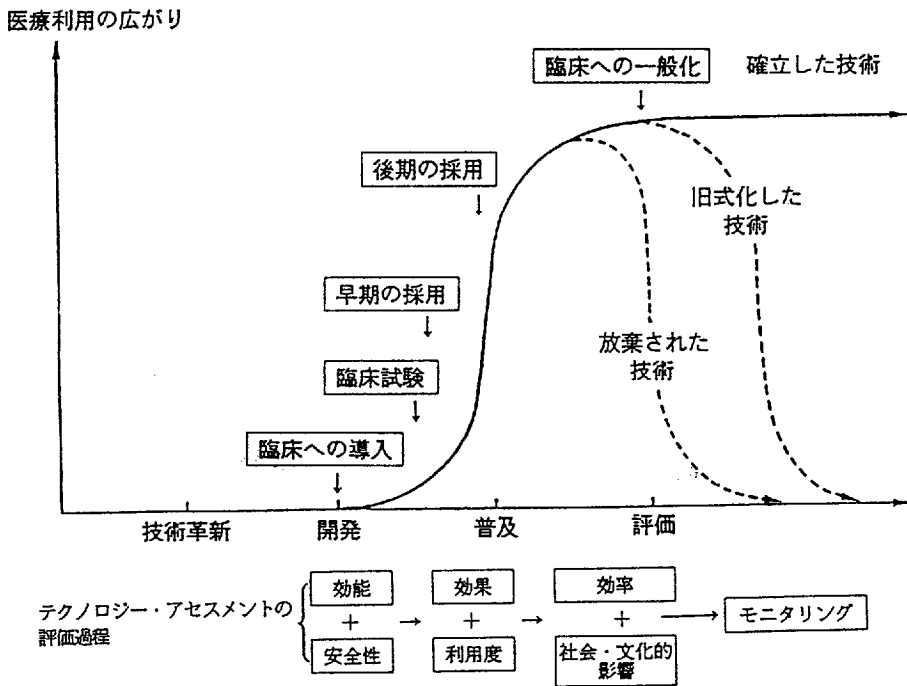
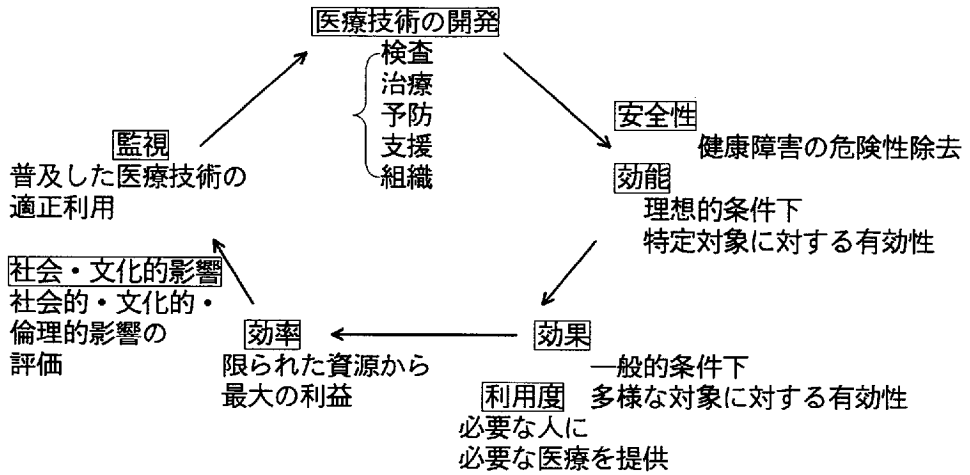


図2 医療技術の普及と評価の過程

図3 医療の質の評価指標

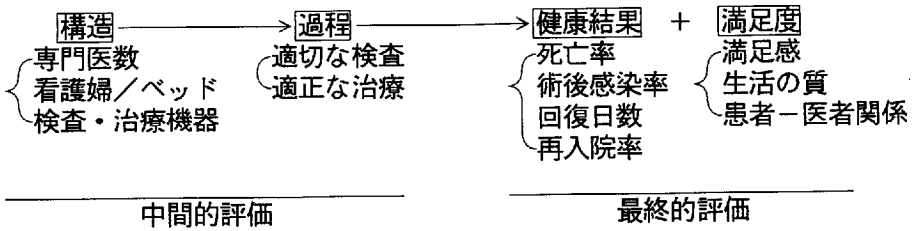


図4 マス・スクリーニングのテクノロジー・アセスメントの評価過程

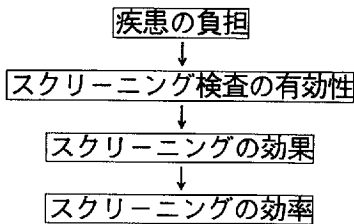


図5 スクリーニングの過程とその評価

2次予防

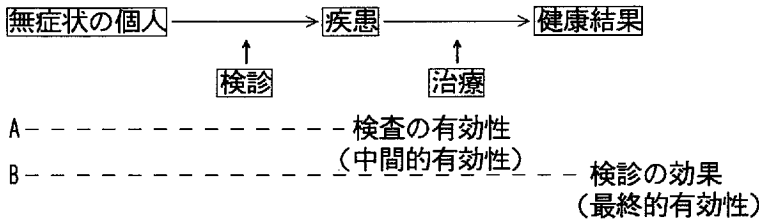


表1 テクノロジー・アセスメントの評価範囲と実施状況

| 技術 | 評価範囲 | | | |
|--------|------|----------|----|-------------|
| | 安全 | 効能 効果 | 効率 | 倫理・法・ 社会 |
| 薬 | ◎ | ◎ | ○ | △ |
| 医療機器 | ◎ | ◎ | ○ | △ |
| 治療手技 | ○ | ○ | ○ | △ |
| 支援システム | ○ | ○ | ○ | △ |
| 組織・管理 | - | ○ | ○ | △ |

◎：よく実施 ○：ときどき実施 △：まれに実施

表2 医療のテクノロジー・アセスメントの内容

| 視野 | | 評価利用者 | |
|----------------------------------------------------|-------|-----------------------------------------|-------|
| 安全性 効能 | 当初の項目 | 医者 | 当初の項目 |
| 効果 費用と便益 生命倫理 社会文化的影響 | 近年の焦点 | 健康保険会社 雇用者 政府諸機関 政策決定者 一般大衆 | 近年の焦点 |
| 評価実施者 | | (Fuchs VRら1990より, 久繁まとめ) | |
| 科学者 医者 | 当初の項目 | | |
| 生命統計学者 経済学者 疫学者 オペレーションズ・リサーチ研究者 心理測定者 | 近年の焦点 | | |

表3 効果評価の基準と勧告の分類 (米国予防医学TF)

介入の効果

- I. 適切に設計された無作為臨床試験 (RCT) による証拠が少なくとも一つ
- II-1. 優れた設計の無作為化の無い臨床試験 (CT) による根拠
 - 2. 優れた設計のコホート研究, 症例-対照研究による根拠 (複数の研究機関, 研究集団によることが望ましい)
 - 3. 介入の有無を問わない多様な時系列研究による根拠 (意図しない実験による劇的な結果もこうした根拠に含まれる)
- III. 権威者の意見 (臨床経験, 記述研究, 専門委員会の報告に基づく)

勧告の分類

- A. 定期健診の実施を特に推薦すべき条件を満たす優れた根拠がある
- B. 定期健診の実施を特に推薦すべき条件を満たす正当な根拠がある
- C. 定期健診の実施を特に推薦すべき条件を満たす根拠が乏しい (しかし, 他の根拠から推薦する)
- D. 定期健診の実施を除外すべき条件を満たす正当な根拠がある
- E. 定期健診の実施を除外すべき条件を満たす優れた根拠がある

表4 スクリーニングにおいて吟味すべき偏り

- 1 罹り易さの偏り (susceptibility bias)
- 2 先行期間の偏り (lead time bias)
- 3 罹間期間の偏り (length time bias)
- 4 過剰診断の偏り (overdiagnosis bias)

表5 経済的評価を吟味するためのチェックリスト

1. 明確に定義した問題を回答可能な形で提出しているか？
2. 競合する代替案を包括的に記述しているか？
3. プログラムの効果を確認した根拠があるのか？
4. それぞれの代替案について重要かつ関連する費用と結果をすべて明らかにしているか？
5. 費用と結果を適切な物理的単位で正確に測定しているか？
6. 費用と結果の価値付けは信頼できるか？
7. 費用と結果について時期の違いを調整しているか？
8. 代替案について費用と結果の増分分析を行っているか？
9. 感度分析を行っているか？
10. 研究結果の提示と考察は利用者にとって重要な問題を全て含んでいるか？

表6 テクノロジー・アセスメントにおける情報の収集と評価

| 方法 | 時間 | 費用 | 質 |
|-----------------------------|----|----|---|
| 既存の情報の分析・評価 (文献検索と批判的吟味) | 短 | 少 | 低 |
| 新しい情報の作成・収集 (調査・研究の実施) | 長 | 多 | 高 |

表7 カナダの Task force による乳幼児のスクリーニングの評価

先天性甲状腺機能低下症

効果 : 症例発見に対するスクリーニングの効能は極めて優れている。乳児期の診断は精神遅滞を改善する。治療は極めて効果的。

根拠の質 : II-1*

勧告・意見 : A 必要ならば補助手段として血清サイロキシン検査を再実施してもよい。高危険群：乳児の母親がヨウ化物・¹³¹I-チオチラシル・放射性ヨウ素を摂取した場合

フェニールケトン尿症

効果 : スクリーニングと治療は極めて効果的（ただし、生後4-21日までに実施）。

根拠の質 : II-1*

勧告・意見 : A 生後4日未満の新生児ではスクリーニングは不備となるため、2度目の検査が必要となる。

*カナダの特別委員会のII-1は、米国の委員会の分類（表3）ではII-2となる。



検索用テキスト OCR(光学的文字認識)ソフト使用

論文の一部ですが、認識率の関係で誤字が含まれる場合があります



要約:わが国で実施されている新生児の先天代謝異常・内分泌疾患のマススクリーニング・システムは,他の先進諸国でも類を見ない広範囲な内容を有している。しかしながら,現在の効果評価・効率評価では,世界的には先天性甲状腺機能低下症およびフェニルケトン尿症の2疾患のみ明確な根拠が確立している。その意味で,わが国のスクリーニング・システムに関する総合的な評価が求められている。その具体的な評価方法として,テクノロジー・アセスメントの枠組みを導入し,新生児マス・スクリーニングの評価課題を設定した。第一段階として,米国特別委員会の最良根拠の方法により,各疾患別にスクリーニングの効果に関する批判的吟味を実施することとした。第二段階として,この効果評価を基礎に,各疾患毎に費用-便益分析を実施するとともに,その結果を総合的に評価し,スクリーニング・システムの効率を検討する。