

Treatment Panel)」である。この専門家パネルは、治療へのコンプライアンス、治療継続率、患者のコンプライアンス、結核管理、結核根絶、直接観察療法を考慮しつつ、結核治療法に対する現在の根拠を評価することに焦点を当てている。このパネルは、こうした疑問のそれぞれに答えるための科学的根拠は十分あり、結核の治療と管理のための科学に立脚した指針を作成できるという結論に達した。しかし、このパネルは、この問題のほとんどの側面は、臨床的視点ではなく、公衆衛生の視点からとらえる必要があると感じていた。別の言い方をすれば、集団ベースの戦略を学的研究（たとえば、コホート研究や症例対照研究）が含まれそうである。したがって、公衆衛生上の問題に適用する際には、科学的な問題の再定義が必要になる。

5. 職場の保健とEBM

合衆国では毎年、職場における外傷や疾病によって、6500人が外傷によって亡くなり、1320万人が致命的でない外傷を負い、60,300人が病気で死亡し、86万7000人が病気にかかっている（1997年）。直接費（650億ドル）と間接費（1060億ドル）を合わせて、総コストは1710億ドルと見積られている。この推定値はおそらく過小評価されている。なぜなら、痛みや苦痛に関連するコストや、家族が提供する在宅ケアのコストを無視しており、また職業性の外傷や疾病の数は過小報告されている可能性が高いからである。

健康増進、健康損失の予防、財政的危険の回避といったものがすべて、雇用前、あるいは入所前の医学的検査を推進させる原動力となっているようだ。van Dijkら（1993年）はそうした評価に対するエビデンスに基づく方法論を記述している。彼らは、予測される最大の予防効果と労働者個人が能力低下に陥る可能性とを検証した。彼らが用いたパラメータには、検査の妥当性、関心のあつたアウトカムについての疫学データ、およびリスク要因が含まれている。彼らのアプローチは、一般的な問題にも適用できる。たとえば、動物実験施設の技師の仕事に応募する人のアトピーの問題を考えてみる。

A. 用いられる検査の特性：検査陽性適中度（PPV）

$$PPV = \frac{a-p}{a-p+(1-b)*(1-p)} \dots [1]$$

[a=検査の敏感度、b=検査の特異度、p=リスク要因の保有割合]

実験動物施設の環境では、アトピー性反応をもつ人はきわめてアレルギー反応を起こすリスクが高いため、この施設の技師の仕事に応募する人の過敏反応の可能性、つまり「アトピー」を確かめる必要があるとしよう。この場合の採用基準は、候補者をアトピーの人と非アトピーの人とに識別することになるだろう。こうした人の識別するのに用いられる検査法—すなわち、血中 IgE 濃度の評価と質問票の組合せ—の敏感度と特異度は、最適条件下ではいずれも 90 % である。一般集団におけるアトピーの有病割合はおおよそ 5 % である。検査で陽性となった場合、その陽性者がアトピーである可能性を示す。こうした条件を上記の公式 [1] に代入すると、それによるアウトカムは、検査陽性と判定された人のうち、32 % が正しくアトピーと判定されて、職場への配置や採用から除外されることになる。別の見方をすれば、検査陽性者のうちの 68 % は誤ってアトピーと判定されるが、それにもかかわらず採用から除外されることになる。

6. 関心のある影響と関連したリスク要因の疫学：集団寄与危険度

上述のアトピーの例のように、リスク要因に関する医学的検査によって候補者を除外することによって、職業性疾患の患者数を減少させられるものと期待できる。理論的には、この方法による選別を行うと、職業関連疾患の発生率は、最大でそのリスク要因に帰する割合だけ減少することになる。議論を簡単にするため、競合したり、相互作用をしたりするようなリスク要因は存在しないものと仮定すれば、そのリスク要因に帰することができる職業関連疾患の割合、すなわち集団寄与危険度 (PAR) を算出する以下の公式を導くことができる。

$$PAR = \frac{p * (RR - 1)}{1 + p * (RR - 1)} \dots [2]$$

[p=リスク要因の保有割合、RR=相対危険度]

もし、アトピーの人は、正常と考えられる人よりも実験動物に対してアレルギー反応を起こすリスクが 3 倍高いとしたら、アトピーであること (集団の保有割合 = 5 %) に基づいて採用を決定すれば、職業関連動物アレルギーの新規患者数は、最大で 9.1 % 減少させることができる。検査法の敏感度が 90 % であれば、実際の減少率は 10 % 少なく、おおよそ 8.2 % になる。

7. 予防的措置の有効性

雇用前医学検査に基づく採用が、それだけの貢献度をもつかを評価する際には、そうしたプロセスの結果を測定することが有用かもしれない。具体的には、選別による健康便益を、投入された努力やコストに対して評価すべきである。これを行うためにはさまざまな方法が考えられる。たとえば、特定の職業性疾患の新規患者発生を1人減少させるために実行すべき医学検査の回数を計算することもできる。あるいは、職業関連疾患の新規患者数を減少させるのに必要な採用拒否候補者数を計算することもできる。Van Dijk と de Kort (1995年) がこの目的のために導いた公式は

$$K = \frac{1}{a \cdot \text{PAR} \cdot I} \quad \dots [3]$$

であり、採用拒否候補者数 K_{rej} は、

$$K_{rej} = \frac{p}{\text{PPV} \cdot \text{PAR} \cdot I} \quad \dots [4]$$

となる。ただし、 a =検査法の敏感度、 PAR =集団寄与危険度、 I =疾病もしくは有害アウトカム（累積）発生率、 PPV =検査陽性適中度である。

したがって、アトピーと動物技師の例で言えば、実験動物に対するアレルギーを獲得する累積発生率 (I) は、動物技師 1000 人当たり約 100 人になる。 p 、 PAR 、 PPV の値を用いれば（上記の式 [3]、[4] を用いて）、122 回の医学検査を行い、17 人の候補者を拒否しなくてはならない、と計算される。この 17 人の候補者のうちの 12 人 (71 %) は、誤ってアトピーと判定されたことになる。

8. 結論

以上より、以下のことが結論される。(1)臨床サービス指針 (the Guide to Clinical Services) の第2版は、プライマリケア医師にとって貴重な参照情報である。(2)タスクフォースはまた、現在の私たちの知識が不完全で、さらに研究が必要な領域が数多く存在することも示した。さまざまな章で、数多くの重

要な研究課題が示されている。(3) 予防医学システムの創設には、エビデンスに基づく道具に加えて、明確に定義された目標をもつ体系化された計画と、そうした目標に向かって成功しているかどうかを評価する計画が必要である。(4) タスクフォースの様々なアプローチ法のなかには、ランダム化臨床試験のように、公衆衛生（合衆国におけるプライマリケア）の実践には不適當なものもある。この点については、因果関係を示す疫学的根拠があれば、適切な介入方法を実行するのに十分な根拠が提供されるだろう。(5) 職場の保健については、雇用前医学検査の例を示した。科学的に精密に検討すると、そうした検査は疾病を減少させるのに有効性が低かったり、コスト対有効性が低かったりするかもしれない。産業医学にはバランスのとれた取り組みを追求しなくてはならない。そうすれば、エビデンスに基づく実践が可能になるだろう。

V. わが国の健診の現状

1. 予防は治療に優る？

わが国では「予防は治療に優る」ということが良く言われる。一見正しそうな言葉ではあるが、この言葉と「予防といえば健康診断」という風潮とあわせると、必ずしも正しくはなくなる。健康診断の目的には表V-1のようなことが考えられるが、その中心は表の①および②である。したがって、自覚症状が出る以前の早期に疾病（や不健康状態）を発見し、それが放置されていた場合に辿る自然経過を、治療や保健指導により改善することができて初めて有効となる。

表V-1 健康診断の主な目的

①疾病の早期発見・早期治療

自覚症状が出る以前の早期に疾病を発見し、それが放置されていた場合に辿る自然経過を、治療や保健指導により改善すること。ただしこれが成立するためには、目的とする疾患が、自覚症状が出る以前に健診で発見できる前臨床期がある程度長く持続すること、早期に発見することでその自然経過を改善しうること、等の条件を満たすことが必要である。

②不健康状態の把握とその進行の防止

明らかな疾病には至っていない場合でも、発症以前の不健康状態について、その存否や程度を知り、その自然経過を改善させること。

③個別の健康教育

健診という機会を利用して、直接には検査内容とは別に、や職業や生活習慣等の個人毎の状況から、起こりうる疾病や不健康状態について情報提供を受け、予防のための方策を実行する動機付けや、実行の援助を受けること。

④職業への適性判断や将来計画のための情報入手

職業の継続や変更、その他の個人の将来の計画を立てるに当たって、健康状態が適するか否かの情報を得ること。

⑤不安の解消

特にドックなどを受診する動機として、本人もしくは家族のために健康についての不安を解消するということがある。実際、健診の結果何も異常がな

いといわれ、安心することも健康診断の効用として挙げることができよう。ただし、これは文化・習慣に根ざしたもので、本来の効用ではなく、また健康診断の内容がすべての健康障害の有無を検出するかのような誤解を伴っていることが多い。

⑥集団としての問題の発見

以上①から⑤の被験者個人にとっての目的に加え、健診がある集団を対象として行われるとき、集団の検査項目毎の異常割合や疾病の頻度等は、集団を取り巻く環境の健康に対する問題点の発見・改善に役立つ。近年WHOでは個人を対象とするスクリーニングに対して、集団を対象にするサーベイランスということを強調している。

現在行われている健康診断項目は、必ずしも全て自覚症状が出る前の早期に疾病を発見できるものではない。そもそもどういう疾病もしくは健康状態を発見しようとしているのか、必ずしも明確でない場合すらある。まず、健診で発見すべき疾患としては一定の条件がある（表V-2）。そして、その疾患を発見すべき検査法の備えるべき条件を考慮する必要がある。その主なものを表V-3に示した。また、疾病発見後に行われる治療や保健指導が、確かに疾病の自然経過を変更し、改善しているとの証明は必ずしも多くない。加えて、健診は有益どころか害をなす場合もある。レントゲン被曝をはじめ健診にともなう身体的侵襲、採血に伴う感染等医療事故の危険という直接的な害はいうまでもない。これに加え、見落とし（スクリーニングの偽陰性）による対処の遅れ、誤った診断（スクリーニングの偽陽性）による無用の精神的負担や精密検査に伴う負荷も考える必要がある。さらに被験者の健診時間中の労働時間の損失や、健診にかかわる医療関係職員の賃金や直接の検査費用も、それらがもしより有益に用いる資源を圧迫しているなら無視できない。従って、果たしてわが国の予防は治療に優っていると言えるだけの根拠があるのかどうか、検討してみる必要がある。

表V-2 健診の備えるべき条件

(1) 慢性疾患であること

健診が年1回もしくはたかだか2回であることを考えると、定期健診で発見することを期待する疾患は、慢性疾患に限られる。急性疾患は健診までに治癒してしまうか、はるかに進行・悪化してしまい、二次予防の効果

が期待できない。

(2) 頻度がある程度高い疾患であること

全員を対象にその検査を受けさせる以上、ある程度高率にその疾患が見つかるものでなければならない。極めて稀な疾患は、そのために多くの被検者への負担を合理化しない。

(3) 予後が不良であること

健診の身体的・経済的その他の負担を考えるなら、その疾患または不健康状態の予後が、生命もしくは QOL に深刻な影響を及ぼすものでなければ、その疾患をや状態を発見することが最終的な利益につながらない。

(4) 介入により自然経過を改善しうること

診断がついても全く対処の方法がない場合、健診で疾患を発見することによる最終的な利益は発生しない。臨床の場合は他の疾患との鑑別や同じ疾患でも程度による予後予測等の目的で、治療の見込みのない疾患を目的として確定診断のための検査をすることがあるが、集団全員を対象とする健康診断の目的にはなじまない。

表V-3 検査自身の備えるべき要件

(1) 低侵襲

職域定期健康診断の検査項目を侵襲の程度で段階分けしてみると、問診、身体計測および尿検査、採血を伴う検査、そして胸部エックス線検査の3段階程度に分けることができよう。このうち胸部エックス線検査は、学校保健では中学までは間接検査は全廃、高校1年では一度行うが、それ以外、直接撮影や必要なときに行うという体制に既に移行している。むしろ妊娠のチャンスのある世代に間接撮影が強制される制度は検討の必要がある。

(2) 低費用・簡便

臨床医はしばしば検査の精度や妥当性のみに着目し、最新・高度医療を健診の場にも持ち込みたがる。一方、現場では直接の費用や検査に要する時間が非常に重要であり、それに見合う利益が説明されなければ受け入れられない。

(3) 受け容れられ易さ

かつて染料工場で、膀胱ガンの一次検査に膀胱鏡が用いられたことがあったが、翌年からは検査を従業員に拒否されたという。医学的な有効性だけでは、現在疾病を持たない健診受診者に受け入れられない。

(4) 実施可能な確定診断法の存在

健診で異常所見を指摘された場合、検査が最終的には介入を目的としている以上、介入の可否とその内容を決定するための確定診断が必須である。確定診断なしでは次に述べる、検査の特異度や鋭敏度を調べることもできない。例えば尿蛋白検査は健診の検査法としては簡便でもっとも受け入れられやすいが、尿蛋白が陽性であっても、確定診断の腎生検はあまりに危険性が大きく、健診集団の確定診断検査としては現実的でない。従って、毎年膨大な数の尿蛋白有所見者が発見されているが、そのうちどの程度が介入を必要とする腎疾患であるのかはわからない。

(5) 検査の妥当性

適当なレベルの特異度と鋭敏度が検査の妥当性の指標となり、それを検査の有効性とも呼ぶ。ROC 曲線も含め、この詳細はVI章で述べる。

2. わが国の健診の現状

上述したように、わが国では一般国民から専門家である医師まで、予防というところから健診という発想があるため、一年間に全国で健康診断を受診する人は大変な数に上る。実際に日本の健康診断の現状を見てみると、地域では老人保健法で、1996年の実績で40歳以上の住民が受診する基本審査で1,000万人、近年いろいろな議論が起こっているがん検診は、肺がんの690万人を筆頭に大腸、胃、子宮、乳房のそれぞれがん検診があり、延べ数で2,000万件となっている。母子保健では1996年の出生数は120万人であるが、妊産婦健診や乳児検診、1歳6ヶ月健診、3才児健診と多岐にわたり、何回も検診が行われる。

一方、労働者は労働安全衛生法や人事院規則等にわかれ、制度上複雑であるが、いずれにしろ、被雇用者は年に一回以上、定期の健康診断を受けることが義務づけられている。その結果、被雇用者の総数約5,200万人が、年に一回以上健康診断を受けているはずである。同様に学校では学校保健法により児童・生徒・学生と教職員あわせて2,400万人が健康診断の受診を義務づけられている。

この他、わが国では自発的な受診による人間ドックが盛んである。総合健診学会は人間ドックを、(1)総合健診、(2)病院宿泊ドック(1～7日の宿泊ドック)、(3)日帰り型ドックに分けているが、その会員施設に対する調査で、1997年の(1)の受診者は155万に達し、その他のドックも80万人の受診者があった。

3. 健康診断の内容

前述のように、わが国の健診の中で一番受診者や件数が多いものは職域における健診であるが、職域健診は通常、業務別の特殊健康診断と一般定期健康診断に分けられ、それぞれ表V-4（労働衛生のしおり、p.171-178）に示す健診が含まれる。これらのうち1997年の受診者数は労働省の統計の範囲だけでも特殊健診が158万であるのに対し、一般定期健康診断は1155万人の受診者があり、後者には公務員などが含まれていないことを考えると、わが国で最も受診者が多い健診は、職域の一般定期健康診断ということになる。一般定期健康診断についての規定は労働安全衛生規則第44条にあり、表V-5に示す項目を年に1回以上行うことが定められている。地域や学校などその他の制度による定期健診も、これとほぼ同様な内容である。ここにある多数の項目のうち現場では、血液検査による肝機能と高脂血症の検査が実は、一番健康指導などの事後措置に利用されているように思われる。

表V-4 職域の健康診断

イ 一般健康診断

- (イ) 雇い入れ時の健康診断（労働安全衛生規則第43条）
- (ロ) 定期健康診断（労働安全衛生規則第44条）
- (ハ) 特定業務（深夜業、坑内労働等）従事者の健康診断（労働安全衛生規則第45条）
- (ニ) 海外派遣労働者の健康診断（労働安全衛生規則第45条の2）
- (ホ) 結核健康診断（労働安全衛生規則第46条）

ロ 業務別特殊健康診断

- (イ) 法令で義務づけられているもの
 - じん肺法
 - 労働安全衛生法
 - 高気圧作業安全衛生規則
 - 電離放射線障害防止規則
 - 鉛中毒予防規則
 - 四アルキル鉛中毒予防規則
 - 有機溶剤中毒予防規則
 - 特定化学物質等障害予防規則
 - 労働安全衛生規則（給食従業員の検便）

労働安全衛生規則（歯科医師による健康診断）

（ロ） 通達で示されているもの

紫外線，赤外線にさらされる業務

マンガン化合物を取り扱う業務

亜硫酸ガスにさらされる業務

砒素を取り扱う業務

その他

キーパンチャーの業務

振動工具の取扱い業務

重量物取扱い作業

金銭登録の業務

引金付工具を取り扱う業務

VDT作業

レーザー機器を取り扱う業務

半導体製造工程

騒音作業

表V-5 一般定期健康診断の検査項目

問診

身体計測 身長、体重

視機能 視力、色覚検査

聴力 1,000Hz および 4,000Hz の測定

胸部 胸部X線検査

循環器機能 血圧、心電図（安静時）

赤血球 赤血球数、血色素量

代謝 尿糖、血糖値、総コレステロール、トリグリセリド、HDL コレステロール

肝機能 GOT、GPT、 γ GPT

腎機能 尿蛋白*

*地域では腎機能にクレアチニンが加わる。

4. わが国の健康診断の全体的な特徴

本研究で検討するわが国の一般定期健診の特徴として、対象を一律に検査すること、全ての検査項目について基本的に年1回の間隔で行われること、等の特徴を有する。その個々について以下に論じる。

(1) 一律の対象者

一般に、定期健康診断は一度対象を設定すると一律に検査することを原則としており、性・年齢、既往歴、喫煙歴その他の曝露からくるリスクや必要度に基づいての考慮や選択に乏しい。そのことは、検査が発見を目指す疾患の有病率が非常に低い集団も同様に検査することになり、著しく検査の効率を悪くする場合が考えられる。特に、職域健診ではそれが最も顕著である。

(2) 一律の検査項目

上の有病率が低い場合の逆に、疾患があることがわかっており、そのため現在通院中で、より詳しい検査がすでに行われていたり、他の検査項目の目的疾患についてもその存否が明らかになっているような場合、本来、その検査を省くべきであろう。しかし事務的な簡便さのため、重複がわかっていながら一律に検査することが、現場では珍しくない。また前年の検査結果や、健診時の自覚的な訴えなどで検査項目の加除が望まれるときも、それには対応しにくいシステムとなっている。

(3) 一律の検査間隔

現在の定期健診は、通常の場合は1年以内ごとに1回行うことが多い。しかし多岐にわたる検査項目の目的疾患がすべて、一年間隔に検査することが最も適当であるとする根拠はない。進行の遅い疾患については、より長い間隔をおいて経過を観察し、進行の早い疾患については、より頻回に検査した方がよいのかもしれない。

(4) 毎回の断面検査

健康診断は同じ集団に繰り返し行われる場合、同じ対象を継続的に観察できることがある。一回の検査では原因の分からない異常な検査値も、経過を追う中で解釈がつくことは多い。また、いわゆる正常値の範囲内の値の動きであっても、個人ごとの例年の値を知っていれば、早期に異常を発見する可能性がある。そこで健診を請け負う検査機関によっては、結果報告に際して前回以前の

結果を対照しやすい表の形式を工夫したり、グラフで経過を示しているところもある。また、医師や保健婦が個々人の検査結果を頭に入れ、それに基づいて判断する事により、データを継続的に取り扱うことができている場合も多い。しかし、そういう個別の努力がなければ、制度としては、健診は毎年の検査が独立した断面検査の繰り返しであり、昨年以前の結果を生かす構造になっていない。その結果、職域健診の場合労働省の発表する健診有所見者の割合が、すべての項目について、年々一律に増加していくのみとなっている。

(5) 非患者が対象

健康診断の別の特徴として、対象が自覚症状などの異常に既に気づいて検査を受ける病院などの患者と異なり、少なくとも通常の生活を送れる程度に健康で、基本的に無症状・無徴候の非患者であるということがある。既に診断がついた疾患がある場合は言うまでもなく、一定の異常に気づいている場合一律の検査では効率が悪く、見落としの危険性も高い。したがってこのような場合、定期健診とは異なる、目的に即した検査を行うべきである。スクリーニング検査では、このようにあらかじめ対象集団が病院の患者とは異なるので、その検査結果の分析においては次章で数量的に詳細に述べるように、患者に対する臨床検査とは異なる取り扱いが必要である。しかし、現実にはそのような考慮がなされている場合は多くない。

(6) 拡張主義・楽観主義

上のような問題の背景には、予防医学の独自の技法への理解が欠如し、予防医学は臨床医学を薄めたものというような意識が臨床医の間に広範に存在することと関係していると思われる。臨床医学では、情報は多ければ多い程良く、たくさん検査することが対象にとって利益になるという考えが多い。すなわち病院などの臨床場面では、学会で症例報告できるような稀有なる一例を見つけることに大きな価値がおかれており、一つの検査値の異常に対する鑑別診断のリストは多ければ多いほどよいとする傾向がある。逆に臨床医にとって、鑑別診断に入れるべき疾患名を落としたことによる見落としは最も恥ずべきこととされる。そこには、有病率を想定し検査の異常値が真に疾患のためである蓋然性を考えるという配慮は乏しく、あらゆる可能性を考えての検査の追加がなされる。この臨床医学としても必ずしも適当ではない考え方が、そのまま予防医学に持ち込まれている。その結果、健診制度の改変のたびに健診項目が増えることはあっても減ることはなく、終始一貫拡充を続け、健診に伴うマイナス側面についての考慮が全くというほど見られない。

(7) 無評価

医療に限らず一般に、わが国では制度、施策について実施後の評価が欠如しているといわれる。健診については、学校健診で胸部エックス線検査や胸囲・色覚検査等について見直しが行われたことがあるが、成人の健診では見直しといえば追加・拡張ばかりであった。その原因は社会全体の傾向にもあるが同時に、実証的に有効性や効果の測定を行った評価研究が極めて少ないことにも関係している。

5. 諸外国の健診体制

(1) 米国予防医療研究班報告

1922年米国医師会は、健康な人を対象にする年1回の定期健康診断を初めて提案した。しかし1970年代頃より、全ての人々に、同じ検査を等しく毎年1回という頻度で行うことの有用性に英国で疑問が提出され(Cochrane Holland)、カナダでは1976年より政府が体系的に定期健康診断の有用性について見直すための研究班を召集した。このカナダの研究班の最初の報告は1979年に発表されたが、そこでは判断に用いた根拠(Evidence)の質について、ランク付けを行い、評価の結果としての予防活動を実行すべきか否かの勧告も、推奨の強さ(または否定)のランクをつけたものであった。その後、カナダの研究班報告は何度か改訂を行っているが、これに刺激を受け米国でも同様の、予防医学に関する研究班報告が作られ、1989年、その最初の報告が発表された。更に1996年にはその第2判も発表されているが、1989年の初版については医学書院から翻訳がでており、わが国でも広く知られている。

(2) 米国研究班報告とわが国の定期健康診断項目

米国研究班報告でもカナダに習い、Evidenceの質を分類し、Evidenceが確かなものかどうかによって、ある検査を定期健康診断に含めるべきとするものを2段階、含めるべきでないとするものもやはり2段階にわけ、これにどちら向きのEvidenceも不十分を加えて、勧告を5種類に分けている(表IV-1参照)。そこでわが国の健康診断項目をこの報告書での評価に当てはめたものが表V-6である。

この表でみると、わが国の地域や職域で広範に行われている検査項目の中で、確かなEvidenceに基づき実施勧告のAにランクされているのは、血圧の測定だけである。このほか一段低い質のEvidenceであるが、実施が勧告されているBランクのものは、身長・体重そして血清の総コレステロールである。しかし視

力・聴力・尿糖をはじめ大部分が、Cランクの勧告、すなわち、どちらともいえないという評価を受け、肺癌についてはやらない方が良いDランクという結論となっている。なおここでアルファベットが入っていない尿蛋白などの項目は、米国研究班のリストに無いものであるが、特に注目に値するのは、わが国の定期健診のいわば中心となる肝機能検査が、検討の対象にも入っていなかったということである。すなわち米国研究班が予防活動についてそれまでの種々の研究を詳細にレビューした結果では、わが国で毎年義務的に行われている定期検診項目のほとんどが、実施すべきか否か議論が分かれるか、そもそも検討の対象にもならない項目であるということである。それにもかかわらず、わが国でこれらの検査が毎年行われているのは、米国研究班を上回る調査研究と検討がわが国で行われたからではない。先年のがん検診を巡る議論で明らかになったように、わが国は健診の評価ということにおいて欧米より遙かに遅れていると言わざるを得ない。

表V-6 日本の健診項目と米国研究班報告の勧告

米国研究班報告		
	Evidence の質	勧告ランク
① 既往歴・業務歴	—	—
② 自覚症状・他覚症状	—	—
③ 身長・体重	I, II-2, II-3	B
視力	III	C
聴力	III	C
④ 胸部X線, 喀痰検査		
(肺癌)	I, II-1, II-2	D
(結核)	—	—
⑤ 血圧	I	A
⑥ 尿検査		
尿糖	II-2	C
尿蛋白	—	—
⑦ 貧血検査		
(鉄欠乏性貧血)	I, II-1, II-2	C
(多血症)	—	—

⑧ 肝機能検査 (AST、ALT、 γ -GTP)	—	—
⑨ 血中脂質検査 (T. Chol)	I, II-2	B
(中性脂肪)	II-2	C
(HDL-C)	II-2, III	C
⑩ 血糖	II-2	C
⑪ 心電図	II-2	C

しかし、健診は対象とする集団の疾病構造によっても有効性が異なる。先の米国研究班の翻訳者、福井次也氏（京都大学総合診療部）も、米国のレポートを丸呑みするのではなく、わが国の状況の中で、独自に評価することが必要と述べている。

(3)健診評価の視点 (Evidence Based Medicine による健康診断)

わが国独自の状況という意味でまず考えなければならないのは、疾病構造である。職域健診のいくつかの項目の直接の目的となる冠動脈疾患の死亡率は、欧米に比べわが国ははるかに低い。このことは、すでに行われてきた健診があったからこその結果である、という可能性は十分検討しなければならないが、果たしてそれだけであろうか。いずれにしろ、死亡率や有病率の低い疾患についての検査は、その検査で疾患を発見する事の利益が低いということの意味する。その他、遺伝的な背景や文化・生活習慣等は疾病の原因を考える上でも、検査で疾病を発見した後保健指導などの介入を行う上でも、重要な因子となる。そこでそういう背景を持つ日本人の集団、日本の事業所で行われた研究は、その結果を日本人集団に適応しやすく、わが国での有用性を判断する上で重要な証拠となる。しかし残念ながら、評価ということがそもそも盛んでないわが国で、そうした研究は多くない。そこでいくつか我々自身で、評価研究を試みることになった。

これを踏まえた上で、以下は先の米国予防医療研究班の報告等と同様な手法をとることになる。すなわち、理想的な状況下での検査の有効性 (efficacy) の評価とその実際の適用場面での有用性 (effectiveness) を、基礎になる情報の質をランク付けした上で、検討していくことである。

検査の有効性のEvidenceとして、最も高い質の証拠とみなされているのは、無作為割付比較試験である。これは、疫学の分類では介入試験に属するが、そ

のデザインにおいて、動物実験に極めて近い。そもそもこのような研究に対しては、原理原則として許されないと主張する意見すらまだあるが、その点においても、試験はその実施状況の影響を受けるため、動物実験的な理想化された条件下でなされた結果をそのまま、一般化できるとは限らない。その具体的な例としては、スクリーニングの有効性における有病率の影響があげられる。

最近、G. Roseの「予防医学のストラテジー」がわが国で翻訳され注目されている。Roseは予防活動にはハイリスクストラテジーと集団ストラテジーがあり、後者の重要性を強調しているが、健診を評価するに当たっても個人のみならず、集団としての有用性を評価する必要がある。基本的にスクリーニングは、ハイリスク者を発見し、そのものに介入を行うハイリスクストラテジーの手段である。しかし、個人はその属する集団を取り巻く環境に規定されており、その全体的な環境に介入を行わなければ、個人のリスクの軽減も困難なことがある。上記書でRoseが示した例を用いるなら、日本人に比べとフィンランド南部の集団では全体のコレステロール値の分布が高い値を示し、個人ではなく、社会全体の食習慣等を変えていく必要性を示している。予防医学におけるこのふたつのストラテジーの対比はWHOも重視しており、個人に対するスクリーニングに対して、集団へのアプローチとしてサーベイランスの言葉を当てている。特に、有害業務の検診に当たっては、騒音性難聴など不可逆性の病態に対し、検診受診者個人に対しては利益がなくても、集団としてみれば問題発見のきっかけとなり、防音対策が進むということなどは考え得る。

VI. スクリーニング

スクリーニングは、1951年にUS Commission on Chronic Illnessにより「迅速に実施可能な問診、試験、臨床検査、その他の手技によって無自覚の疾病やリスクファクターを識別することである」と定義されて以来、疾病の早期発見を目的として行われてきた。スクリーニングの主な目的は、先に述べた疾病の早期発見の他に、健康であることの確認、心身の異常の早期発見と言える。その他、スクリーニングの波及効果として、(1)検診結果を知らせるときに教育・指導を併せて行えば、禁煙や過度の飲酒の防止など、ライフスタイルの改善が可能となる、(2)そのような活動を通して健康管理活動の点検や立案に繋がる、(3)健康情報の継続的な収集による個人の正常範囲の同定が可能となる、などが期待される。しかし、スクリーニングは、健康な者あるいは医療機関を受診するほどの自覚症状のない者を対象に労力と費用を掛けて行うため、その有用性について慎重に検討する必要がある。さらに、スクリーニングに適した疾患とそうでない疾患があることも念頭に入れる必要がある。一般的に言えば、患うと重い病気であるが早期の治療効果がよく、臨床診断前(preclinical)の有病率が高い疾患が適した疾患といえよう。例えば高血圧性疾患などがそのよい例であり、早期発見による治療効果は高く、生命を脅かす脳卒中等の合併症を予防することができる。さらに、有病率は高い。このような疾患に対してスクリーニングは有用な方法である。

ここではスクリーニングでの目的疾患の選択、評価法、および地域・職域での検査結果の有効利用を中心に整理した。

1. 目的疾患の選択

ここでは労働安全衛生規則第44条にある検査項目(表V-6)を取り上げる。これらの検診項目のうち、果たしてスクリーニングの目的疾患が明確に認識されている検査項目はどれほどあるだろうか。検査法(スクリーニングテスト)の有効性を検討するときには検査目的疾患を‘gold standard’として有効性の検討を行う。しかし、実際に検査を実施する際には必ずしもその目的疾患のみを対象としていない場合が少なくない。産業医として職域でスクリーニングを行う際には、その検査により何を明らかにしようとしているのか、どのような疾患を発見するために行っているのかということについて、改めて見直す

必要があろう。職域での健康管理体制では、定期健康診断により高血圧、糖尿病、結核、心臓病、腎臓病、などの疾病管理を行っている（スクリーニングマニュアル）が、個々の検査項目についてこれらの疾病とスクリーニングとの関わりということに焦点を当てて検討する必要がある。

例えば、肝機能検査項目の GOT を例に取ってみよう。GOT は肝機能検査の 1 項目としてほとんどの地域・職域健診で行われている。ところが、この検査によってどのような「疾患」あるいは「異常な症状」を発見しようとしているのであろうか。その有用性はどの程度あるのであろうか。米国予防医療研究班では、一つ一つの検査法について目的疾患を明確にし、その評価を行い、必要に応じて提言を行っている。日本では諸外国にはないスクリーニング検査も多々行われており、それらの評価は独自に行わねばならない。さらに、検査法やその有用性は国による違いもあり、諸外国での有用性の評価を参考にしながら、日本でも目的疾患を明らかにした上で独自に評価を行う必要がある。

2. 正常値 (Normal value)

スクリーニングでは 'gold standard' として疾病の有無や正常・異常を取り扱う。そのため、ここでは正常値あるいは正常範囲について述べておく。

「正常」は単純には定義できず、Sackett 教授らは「正常」には 6 つの定義（ガウス分布、パーセンタイル、文化的望ましさ、危険因子、診断的、治療的）があるとしている。「ガウス分布」や「パーセンタイル」は「平均的」という意味であり、「正常範囲」は健常者の検査値の分布から求めということになる。

“文化的望ましさ”は、文化・社会的な「望ましさ」という基準であり、正常範囲は健康状態を示す範囲である。しかし「健康」と「不健康」の間にはシャープな境界線があるわけでない。“危険因子”は何らかの危険因子の程度または有無で分けることになろう。“診断的”は「標的障害の可能性が高率になる範囲」としているが、カットオフ値の問題等が絡んでくる。“治療的”は「措置の必要度」といったもので、何らかの措置が必要かどうかを判断するための範囲であり、この意味での「正常」では単に「健康—疾病」の情報の他に「介入の便益」が関係してくる。

このように正常には、絶対的な基準があるわけではなく、取り決めとして定義しているのである。検査値の場合には一般的には第 1 または第 2 の定義を用いることが多いので、この分布からの定義に基づく「正常範囲」について概説する。この場合には「健康な人の 95 % が入る範囲を正常とみなす」という基準を与えて、その分布を考慮して正常範囲を与えているのが一般的である。一

般に、検査値の分布は正規分布か対数正規分布をするものが多いことが知られている。そこで正常範囲は、正規分布の場合には（平均値±2標準偏差）の間に、対数正規分布の場合には対数変換後のデータから（平均値±2標準偏差）を求め、これを逆変換して求めたり、パーセント法で2.5パーセンタイルと97.5パーセンタイルとして正常範囲を求める。その他、確率紙を利用して求めたり、モデルを利用して求めたりと様々な方法が提案されている。

この「平均的」という意味での正常範囲を取り扱うときに留意したいことは、健常人の検査値の分布の「はずれた」ところは偶然による変動以上の数値と見なすため、健常人の5%が「異常」として分類されてしまうことである。つまり、健常人であっても20回に1回は「異常」と分類されてしまい、極論すれば、20種類の検査を同時に行った場合には1つくらいは「異常」と分類されてしまう可能性があるわけである。

さらに、正常範囲は健常な集団のデータを基に推定され、一個人の健常なときのデータを基に推定される値ではないことに留意する必要がある。一般に検査値は、性、年齢、その他の要因によっても異なる。性別に正常範囲が設定されていることは多いが、年齢別については、必要に応じた年齢別正常範囲を設定し、それを‘gold standard’として評価することも望まれる。

もう一つ大切なことは、検査値の個人間変動は個人内変動よりも大きいということである。したがって、個人の正常範囲は集団のそれよりもずっと狭い範囲で、しかも必ずしも集団のデータから求めた正常範囲のデータとは一致するとは限らない。例えば、総コレステロール値を例にとると、家族性高脂血症の人はいわゆる病気でないという意味で“健康”であっても、総コレステロール値が230以上あることが少なくない。ところが、集団の正常範囲では常に異常と診断されてしまい、要精検の指示が出てしまうことになる。そのような場合には、その人にとってのスクリーニング検査結果は意味が薄れてしまう。その点、個人の経時データを集積しておけば、個人の正常範囲の推定にも利用でき、異常を発見するのにも有効である。

3. スクリーニング評価の理論

(1) スクリーニングテストとスクリーニングプログラム

スクリーニングの有用性を検討する時、スクリーニング検査法自体の有効性と実際にスクリーニングを行うときのスクリーニングプログラムの有用性とは質を異にするため、分けて考える必要がある。スクリーニングテストは検査法のみに関わる問題であり、その評価では精度と正確性を問題とする。他方、ス

クリーニングプログラムは実際にスクリーニングを実施するときの様々な問題点を含む。その中でスクリーニングテスト自体の有効性も1つの判断要素となるが、適中度や侵襲性、費用、その他の要素に左右されることになり、多分に政策的である。幾つかの基準を満たす中での最適なものをスクリーニングプログラムでは有用と見なす。

(2) スクリーニングテストの有効性

スクリーニングテスト自体の有効性の評価では、精度 (precision) と正確性 (accuracy) が問題となる。精度は信頼性 (reliability) ともいい、検査値自体の変動が少ないことや再現性 (repeatability) が高いことをいう。まず精度について検討し、精度が高い検査法であれば、次に正確性について検討することになる。なぜなら、精度が悪く検査の都度異なった結果が出てしまうような検査では使いものにならない。他方、正確性は妥当性 (validity) の1つの基準であり、スクリーニングテストで実際に 'gold standard' としての疾病あるいは異常をどの程度正確に判別できるかをみるもので、その評価指標には敏感度と特異度がある。このように正確性を検討するときに gold standard が必要となるのである。

'gold standard' は疾病の有無として臨床診断の結果を用いることが多いが、分類は曖昧なこともある。疾病の種類や重症度などによっても gold standard の決め易いものと決め難いものがあり、さらに、現場で検査を行う場合にはその検査法本来の目的疾患とは別の目的で利用することも少なくない。このような場合には、再度目的疾患でのスクリーニングテストの有効性の検討を行う。換言するならば、スクリーニングテストを行う際、目的疾患を明確にした上で検査法の選択や評価を行うべきである。

(3) 敏感度と特異度

多くの検査診断では疾患ごとにカットオフ値 (cut-off value/point) と呼ばれる臨床的な判別点が定められ、検査値がそれを越えるときに陽性、越えないときに陰性と判定される。しかし現実には、本当は陰性 (疾患ではない)なのに検査では陽性と判断されてしまう (偽陽性) 確率や、逆に、本当は陽性 (疾患である)なのに検査では陰性と判断されてしまう (偽陰性) 確率が存在する。これに対し、疾患を有する人 (真の陽性者) が陽性と判断される (真陽性) 確率や、疾患を有しない人 (真の陰性者) が陰性と診断される (真陰性) 確率もある。特に、検査法の有効性には敏感度と特異度が問題となる (表VI-1)。

検査の評価に用いられる敏感度と特異度を表VI-1のような標本での分類を用いて表現する。