

ヘルスクレームの科学的評価

(独)国立健康・栄養研究所 栄養所要量策定企画・運営担当リーダー 佐々木 敏

はじめに

ヘルスクレーム(health claim:健康強調表示)とは、ある食品あるいはその食品が含む栄養素・機能性物質が健康に効果のあることを示す表示、あるいは、これらと疾病との関係、疾病の予防・治療のできることを明確に主張したり、間接的に主張したりすることとされている。

ヘルスクレームに要求される評価が、どのような科学的な手続きによって行われ、その結果はどのように解釈されるべきであるかについて基本的な点についてまとめてみたい。

1. 評価対象データの特定と収集

評価対象として扱いうるデータを得るために、最初に個別試験の評価を集中して行う。先ず意味のある良質の試験を特定し、その試験の強みと弱みを判定する。その際に重要なことは、①ある特定の試験のみを重要視するのではなく、試験の科学的妥当性の基準をあらかじめ設けておき、それにしたがって評価すること、②ひとつではなく、可能な限り多数の試験を収集し、総合的に評価すること、③収集に当たって、公に入手可能な試験に限り、それらの証拠を評価することである。評価対象となるデータは、食品物質と疾病の関係を説明するために必要な疑問に対して確実に解答を与える『信頼度の高い試験デザインで実施及び計測がなされ、適切な意味が与えられている』試験である。

一般的には、ヒト試験がもっとも重点的に評価の対象とされる。動物モデル試験及びインビトロ(研究室)試験など、いわゆる非ヒト試験で得られたデータも利用可能であるが、非ヒト試験のみではヘルスクレームを裏付けるには十分でない。それは、非ヒト試験の結果には

ヒトの生理的効果に外挿する際の不確実性に難があることが主な理由である。

非ヒト試験に比べるとヒト試験の実施は困難かつ高価であり、その評価も複雑かつ困難である。一方、ヒト試験の最大の利点は、その食品が含む栄養素・機能性物質の健康への効果を直接にヒトで検討できることにある。そして、ヒト試験の目的は、栄養素・機能性物質が健康に及ぼすそのメカニズムを明らかにすることではなく、『ヒトにおいて有効か否か、また、その量的な意味はどの程度かを明らかにすること』にある。

2. ヒト試験における評価方法

ヒト試験(ヒト研究)は、介入試験(介入研究)(interventional studies)と観察試験(観察研究)(observational studies)に大別される。それぞれはさらに細分類され(表1)、それぞれの評価方法を用いて得られた結果を評価するうえでのエビデンス・レベル(level of evidences)が提唱されている。しかし、この基準は、薬剤を中心にする医学的治療法を科学的に判定するための方法論として発展してきたため、食品を対象とする場合に、必ずしも、単純にこの基準に従うことはできず、食品におけ

る特有の問題や特徴をじゅうぶんに考慮することが重要である。

介入試験の『黄金基準(gold standard)』と言えるものはランダム化割付比較試験(randomized controlled trial、略称RCT)である。しかし、食品に関する介入試験は薬剤に関する介入試験と異なる点が数多くあることを理解しておかななくてはならない。介入として食品物質を使用するために、二次的な混在要因(confounders、confounding factors、交絡要因、交絡因子とも呼ぶ)の混入を避けたい場合があるということである。また、食品試験ではプラセボ(偽薬)を用いる対照群の設置は不可能な場合があり、被験者は介入に関して盲検化されないことになる。これらの問題のために、食品に関する介入試験の結果は医薬品を扱った介入試験に比して確実性に乏しいデータとなることが多い。介入試験は目的とする食品や栄養素が目的とする健康状態に与える影響を検討するもっとも信頼度の高い試験方法であると一般に認められているが、信頼度の高い結果を得るために非常にたくさんの人数をその試験に投入する必要がある場合や、結果を得るまでに時間がかかる(時には数十年

表1 疫学研究(試験)の種類と長所・短所

	研究方法	長所	短所
①介入研究	介入を行い、その効果を検討する。	交絡因子の排除が可能。	費用がかかる。 時間がかかる。 代理指標が用いられることがある。
	②コホート研究	原因を調査し、その後に結果を観察する。	交絡因子の調整がある程度可能。多数の原因について一度に調査が可能。
観察研究	③症例対照研究	結果を調査し、その後に、過去における原因を調査する。	比較的安価。 時間がかからない。 少数の対象者で可能。
	④横断(断面)研究	原因と結果を同時に調査する。	
	⑤生態学的研究、経時的変化	集団を単位として、原因と結果を観察する。	大きな集団に関する調査が可能。 比較的安価。 時間がかからない。

※①～⑤はエビデンスレベル。

注意:複数の種類の疫学研究の結果を個々に評した上で、全体について総合的に評価することが重要である。

の年月を要すると考えられる場合もある)が多い。そのため、次に述べる観察試験の全体もしくは一部に科学的証拠を求めなければならないことも多い。なお、観察試験は観察研究と呼ばれることのほうが多いため、以下では試験ではなく、研究と呼ぶことにする。

観察研究の範疇に入る研究の重要性を測る普遍的に妥当な方法は存在しない。しかし、一般的にみて、観察研究の種類を説得力の強さの順(降順)に挙げると、コホート研究(cohort or longitudinal study)、症例対照研究(ケース・コントロール研究)(case-control study)、横断研究(断面研究)(cross-sectional study)、時系列試験(time-series study)、生態学的研究(エコロジカル試験)(ecological study)、記述疫学(descriptive epidemiology)及び症例報告(case report)となるであろう。

実際にはそれぞれの観察研究の実施方法を個別に評価し、その説得力を評価しなくてはならず、介入研究であればよいとか、エコロジカル研究からの結果の信頼度は低く、コホート研究からのものは高いというような一元的なものではない。そして、ひとつの研究手法でも、実際にはさまざまな異なる問題をはらんでいる。そのため、研究成果を正しく理解するためには、報告内容(特に方法)を丁寧に読み、どのような目的のために実施された研究であるか、どの目的にこの結果は使えるかをじっくりと考えて、研究成果を適切に活用することが重要である。

3.食品及び食品成分の特定ならびに計測

その食品及び食品成分を特定することとはヘルスクレームを承認することにおいて考慮すべき重要な問題となる。食事全体もしくは食事内の特定の食品類ではなく、その食品もしくは食品成分(栄養素やある特定の物質)によって有用性がもたらされていることを示す証拠がない限り、当該食品や食品成分と疾病との関係付けは確立されない。

例えば、食物繊維と心疾患との関係についてのヘルスクレームが評価された時に指摘されたように、多くの試験が、食物繊維に富む比較的大量の単一の食品もしくは食物繊維源(例えば調理された豆)の効果を評価することを目的と

していて、食物繊維総量としての摂取を検討するものではなかった。そのため、そこで得られた効果が食物繊維によるものであると結論することができなかった。これは、試験食が食物繊維摂取量を増加させたことは明らかであっても、『他の栄養素摂取量は変化しなかった』という保証を与えることができず、得られた結果の原因を食物繊維に特定することができなかったという問題のためである。そのために、試験食(検討の対象としている食品や栄養素)を中心に、試験食以外の摂取食品に由来する栄養素摂取量も測定し、考慮することが必要となる。

このように、摂取量の測定方法、すなわち、食事調査の方法、食事アセスメントの方法を厳密に評価することが重要である。摂取量の評価方法としては、食事記録法、24時間思い出し法、食事歴法、食物摂取頻度法が知られている。

短日間の食事記録法や24時間思い出し法は個人の栄養素もしくはその他の食品の通常摂取を評価する方法としては適切ではないと考えられている。調査必要日数は対象者の特性(特に年齢)によっても、また、栄養素によっても異なる。そのため、実際には相当の配慮が必要である。さらに、エネルギーや3大栄養素に比べると、ミネラルやビタミンなどの微量栄養素に要求される調査必要日数は長いという特徴がある。現在、いわゆる生活習慣病に関連し、ヘルスクレームの検討対象となる栄養素は、ほとんどの場合、後者であることは特に注意したい点である。

一方、食事歴法や食物摂取頻度法で使用される食物摂取頻度質問票の食品リストに、試験の対象とされる集団におけるその栄養素の主食品源が含まれていないような場合、この方法はその栄養素の摂取量を評価する方法としては不適切なものになってしまう。食物摂取頻度質問票は、大規模な疫学研究で利用することを目的として開発されたものがほとんどであるため、そのままではヘルスクレームの評価試験には必ずしも適当でないものが多く、注意を要する。いずれにしても、既存の食事評価法を利用するに当たっては、利用目的や利用可能性、利用限界についてじっくりと検討することが重要である。

日常生活である程度摂取されている

栄養素を補足する目的で用いる補助食品や強化食品の場合は、対象としている栄養素を対象者が日常的にどの程度を摂取しており、強化することによってどの程度摂取量が増加し、その結果、どの程度の健康影響を期待しうるのかを評価しなくてはならない。つまり、評価試験中(正しくは、試験前、試験中、試験後)に、評価対象としている食品から摂取される分だけではなく、すべての食品から摂取される分を評価することが必要となる。

加えて、いずれの方法でも、栄養素摂取量を計算するために使用される食品成分データベースが信頼に足るものであることが前提となる。食品成分データベースが完備していない栄養素を対象とするような場合にはこれは無視できない問題となる。

4.個別試験の評価基準

個別試験(個別研究)の評価は、試験(研究)デザイン・実施・解析・解釈のそれぞれについてなされる。

(1)試験デザインの妥当性及び明瞭性

試験デザインの妥当性及び明瞭性は具体的には次のような観点から評価される。

- ①当該試験で解決すべき疑問点は試験デザイン設定時に明確に記載されているか。
- ②試験に適用される方法論が明確に記載され、試験によって提起される疑問点に答えるのに適切か。
- ③介入試験では期間もしくは追跡調査期間が問題の転帰に関連した効果を検出するのにじゅうぶんか。
- ④潜在的な混在要因の特定、評価及びその管理はおこなわれているか。
- ⑤被験者脱落(試験完了前に試験を中止する被験者)は評価され、説明されているか、そしてそれらは妥当であるか。
- ⑥被験者数は有意な効果を認めるのにじゅうぶんで、統計学的検出力が得られる程度か。
- ⑦被験集団は、ヘルスクレームの対象となる集団(たとえば年齢、性別分布、人種、経済状態、地理的要因、家族歴、健康状態及び試験参加動機等)を代表しているか。
- ⑧試験の被験者採用基準及び除外基準は明確に述べられ、適切であるか。

⑨被験者選択時の偏りを最小限に抑えるような募集方法が用いられているか。

⑩比較対照介入試験では被験者はランダムに割り付けられているか。

(2)介入もしくは暴露とその転帰の評価
介入もしくは暴露とその転帰の評価は次の観点からなされる。

①食事摂取評価の際の解析方法及び質的管理手法は適切であるか。

②介入試験の場合、問題の食品物質の適切な摂取量(即ち、有効と仮定する量)が計画されモニターされ、確保されているか。

③試験物質を添加した基礎食事、あるいは対照群及び介入群の食事は正しく計測され、記述されているか。

④介入の方式及び設定は『実生活』を代表しているか。

⑤試験の結果に影響を与えるような、試験期間中の食事もしくは健康関連生活習慣(体重減少、運動、飲酒、禁煙)に随伴する変化が特定され、評価、管理されているか。

⑥疾病関連転帰の内容は明確化され、適切に計測されているか。バイオマーカーが計測されている場合には、疾病転帰との関連性について検証されているか。

⑦有用効果と同様に有害効果についても検出しようとする努力が為されているか。

(3)統計解析法を評価するポイント
統計解析法を評価するポイントは次のように考えることができる。

①用いられた統計解析方法は適切であるか。

②『統計学的有意』の解釈は適切か。
例えば、統計学的に有意ではない差に対しては、傾向があるという表現ではなく、差はなかったと表現すべきである。

5. 証拠の総括表の作成と総合評価

評価過程の一部として、科学的証拠を包括的に評価し、評価の方向付けを行うために、評価対象となった試験の総括表を作成する。つまり、個別に評価された結果をもとに、一定の基準でそれぞれの試験に『グレード』を与えた表を作成する。そして、証拠の総合評価を行うが、栄養的暴露と健康関連転帰を関連づける証拠の説得力の強さは、個々の試験の質のみではなく、全体的

な『グレード』、即ち、試験の数、結果の一貫性及び効果の程度等を勘案した証拠全体としての評価に基づく。

具体的な原則は以下のように考えることができるであろう。なお、このうち②から⑦までと⑨は、ヒト試験における因果関係を示すために必要な基準、いわゆるHillの基準として知られている(ただし、ここにある②と③はHillの基準では『一致性』としてひとつにまとめられている)。

①証拠の多くが説得力のある試験デザインによって得られていること。

②関連性を裏付ける試験の数がじゅうぶんに多いこと。

③状況及び性質が異なる集団間で結果が一貫していること(関連の一貫性)。

④効果発現の程度の信頼度を示す統計学的有意性が高いこと。

⑤関連の独立性が保障されること(暴露もしくは介入以外の他の要素が係わる関連の程度と、試験の対象とされる暴露もしくは介入が係わる関連の程度の対比として測られ、前者に比べて後者が係わる程度が大きいこと)。

⑥量-反応関係が示されていること(食品物質の暴露が強くなるほど発現する効果も大きくなること)。

⑦時間的関係が成立していること(転帰の発現に対して暴露が常に先行していること)。

⑧誘発除去の効果が観察されていること(介入を中止した被験者において、関連する転帰に反転がみられることをいう)。ただし、必須ではない。

⑨生物学的信憑性があること(動物試験など、メカニズム研究の結果と整合性があり、生物学に信憑性が高いことを指す)。これは因果関係があると推論するための必要条件ではないが、因果関係の成立の可能性を高めることになる。

上記の過程を経て証拠の総合性の判定がなされ、『有意な科学的同意の成立』に至る。その際に重要なことは、個々に状況が異なるために、ケースバイケースで検討されることが必要であり、そして、当該ヘルスクレームに係わるデータ及び情報は関連の学会に公開されなければならないということである。『当該ヘルスクレームに係わるデータ及び情報』は有資格専門家が入手できる証拠であるという点は特に重要である。これは、『デー

タや情報が受理審査のある学術誌に公開されている』ことを意味する。一般的に抄録では審査者にとって情報が不十分である。専門家が証拠の全てにアクセスできる状態でなければ、その専門家の意見の価値は限定的なものとなり、偏ったものとなる危険性があるからである。

おわりに

健康強調表示ヘルスクレームは、健康・栄養食品の利用者にその効果の有無や程度、信頼度に関する情報を与える重要なものである。したがって、その内容はじゅうぶんな科学的根拠に基づいていなければならない。しかし、これらの情報が正しく活用されるためには、健康・栄養食品の開発者側及び評価者側だけでなく、利用者側も含めた三者すべてが、ここで紹介した考え方や評価手法を理解していなくてはならないと考えられる。

なお、本稿は、「佐々木敏. 2章7.健康強調表示health claimの科学的評価. よくわかるサプリメント 医者と患者のための完全マニュアル. 信川益明監修、三宝社発行(共著) 2003: 96-114」を基にして作成した。

〈著者略歴〉

佐々木 敏(ささき さとし)

94年3月 大阪大学医学部大学院博士課程修了、ルーベン大学医学部大学院(ベルギー)博士課程修了

95年5月～96年9月 名古屋市立大学医学部公衆衛生学教室助手

96年10月～02年3月 国立がんセンター研究所支所臨床疫学研究部室長

02年1月から現在に至る 独立行政法人国立健康・栄養研究所 栄養所要量策定企画・運営担当リーダー

04年4月から現在に至る お茶の水女子大学大学院人間文化研究科人間環境科学専攻食環境科学講座 教授(客員)

〈学位〉

医学博士(疫学) ルーベン大学(平成6(1994)年3月)

医学博士(公衆衛生学) 大阪大学(平成6(1994)年3月)

〈研究テーマ〉

人間栄養学、栄養疫学

〈主要著書〉

1) 徳留信寛、佐々木敏 監・編訳. 食事評価法マニュアル 医歯薬出版. 1997.

2) 佐々木敏、等々力英美 編著. EBN入門 ミ生活習慣病を理解するために― 第一出版. 2000.

3) 佐々木敏. Evidence-based Nutrition: EBN 栄養調査・栄養指導 医歯薬出版. 2001.

