

遡及的診療録調査における有害事象把握の方法論に関する研究

医師判定の信頼性に関する検討

| | | | | | |
|-----|--------------------|------|-----------------------|------|------------------|
| イケダ | シュンヤ | コバヤシ | ミア | サカグチ | ミサ |
| 池田 | 俊也 ^{*、2*} | 小林 | 美亜 ^{*、2*、3*} | 坂口 | 美佐 ^{4*} |
| カネコ | トシヒロ | ヒロセ | マサヒロ | サカイ | ヒデト |
| 兼児 | 敏浩 ^{5*} | 廣瀬 | 昌博 ^{6*} | 塚 | 秀人 ^{7*} |

目的 医療安全対策を推進する上で、その基礎情報となる全国的な有害事象の発生頻度の把握が必要と考えられている。諸外国では、有害事象の頻度や種類を判定するため、看護師による第一次レビューと医師による第二次レビューの二段階による遡及的診療録レビューが相次いで実施されているが、この際には、レビュー者の判定の信頼性を確保することが重要な課題と認識されている。

そこで本研究では、看護師による第一次レビューにおいて「有害事象の可能性あり」とスクリーニングされた診療録100冊を対象とし、3人の医師レビュー者がマニュアルに基づき独立して有害事象に関する判定を行い、その結果を比較することによって、医師判定の信頼性の検証を行うことを目的とした。

方法 某病院において、平成14年度に退院した精神科以外の入院患者の全診療録から無作為抽出した250冊のうち、看護師による第一次レビューで基準該当(+)あるいは要検討として判定された159冊の診療録から100冊を無作為抽出し、3人の医師レビュー者が独立して有害事象の判定を行った。その後、評価マニュアル作成者であるWGの医師を加えた4人で、各レビュー者の判定結果について討議を行い、レビュー者全員の合意が得られた判定を最終判定とした。

医師レビュー者の判定結果の信頼性を検証するために、各医師2人間の判定結果の一致率および、各医師の判定結果と最終判定結果の適中率を求めるとともに、 κ 統計量を算出した。

結果 4人の医師が討議を行う前に、各医師が独立して100症例のうち有害事象(+)と判定した症例数は、18症例～27症例であった。各医師間の一致率は83.0～90.0% ($\kappa=0.52\sim0.70$)であった。

4人の医師の討議によって16例が有害事象(+)と最終判定され、7例は判定保留となった。有害事象の有無に関する各医師の適中率は、医師86.0～96.8% ($\kappa=0.56\sim0.88$)であった。

医療との因果関係に関する判定ならびに予防可能性に関する判定については、各医師の適中率は必ずしも良好ではなかった。

結論 有害事象の有無については個々の医師の判定の信頼性は良好な水準であることが確認された。一方、医療との因果関係・予防可能性に関する判定については必ずしも良好ではなく、様々な専門領域の医師が討議の上で判定することが望ましいと考えられた。

Key words : 有害事象, 診療録, 遡及的レビュー, 信頼性, 医療安全

* 慶應義塾大学医学部医療政策・管理学教室

2* 財団法人医療科学研究所

3* ニューヨーク大学教育学系大学院

4* 九州大学大学院医学系学府医療経営・管理学専攻
専門職大学院

5* 三重大学医学部附属病院安全管理室

6* 京都大学医学部附属病院安全管理室

7* 東海大学医学部付属病院

連絡先: 〒160-8582 東京都新宿区信濃町35番地

慶應義塾大学医学部医療政策・管理学教室

池田俊也

I 緒 言

医療事故を予防し、安全な医療環境を実現するための具体的な対策を整備・推進するためには、医療事故ならびに有害事象の実態を把握することが重要である。米国科学アカデミー医学研究所 (Institute of Medicine) が2001年に出版した「To Err is Human」では、医療事故により死亡する年間患者数を44,000人～98,000人と推計している¹⁾が、この根拠となったのは、米国ハーバード大学のグループにより1984年にニューヨーク州²⁾で、また1992年にユタ・コロラド両州³⁾で実施された遡及的診療録調査による有害事象の把握結果であった。その後、オーストラリア⁴⁾、英国⁵⁾、ニュージーランド⁶⁾、デンマーク⁷⁾、カナダ⁸⁾などで同様の調査が相次いで実施されている。

各国で実施された遡及的診療録レビューは、看護師による第一次レビューと医師による第二次レビューの二段階で構成される。第一次レビューでは、看護師が診療録から有害事象の可能性のある症例のスクリーニングを行う。第二次レビューでは、医師が第一次レビューでスクリーニングされた症例に対して有害事象の有無ならびに種類の判定などを評価する。

諸外国と同様、日本においても第一次レビューの信頼性・妥当性は検証されている^{9,10)}。一方、第二次レビューの信頼性においては、諸外国では、医師レビュー者間もしくはレビューに精通した上級医師レビュー者と医師レビュー者間における有害事象の有無ならびに種類の判定の一致率によって検証されている^{2~4,6,8)}。しかし、これらの有害事象に関する判定は、医師による「暗黙的规则 (implicit criteria)」による判断がなされており、その信頼性については十分な水準にないとの指摘もある¹¹⁾。

そこで本調査研究では、医師による暗黙的规则による判断を可能な限り回避するために、パイロットスタディを通して、有害事象を判定するためのマニュアルを作成し、そのマニュアルに基づき第二次レビューを実施している。

本調査研究の目的は、看護師による第一次レビューにおいて「有害事象の可能性あり」とスクリーニングされた診療録100冊を対象とし、3人の医師レビュー者がマニュアルに基づき独立して有害

事象に関する判定を行い、その結果を比較することによって、医師判定の信頼性の検証を行うことである。

II 研究方法

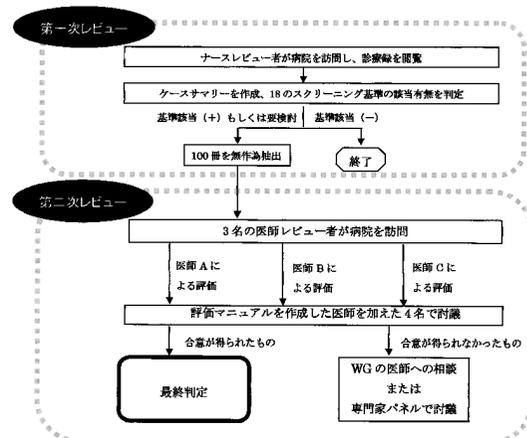
1. 遡及的診療録レビュー

病床数約600床を有する特定機能病院において、平成14年4月から平成15年3月の間に退院した精神科以外の入院患者の全診療録から無作為抽出した250冊を調査対象とし、平成15年4月～6月に遡及的診療録レビューを行った。本調査研究の手順は図1に示した。なお医師判定の信頼性の検証において対象とした診療録は、上記の250冊のうち、基準該当 (+) もしくは要検討と判定された159冊から、今回の調査のために無作為抽出した100冊である。

まず看護師のレビュー者が有害事象の可能性のある症例をスクリーニングするための18の基準を活用して第一次レビューを実施した¹²⁾。第一次レビューにおいては、看護師レビュー者がケースサマリーの作成および「1.基準該当 (+): 基準に該当する」、「2.要検討: 基準に該当するか否か判定に迷う」、「3.基準該当 (-): 基準に該当しない」という3段階で評価を実施し、「評価シート1」に記載した。

つぎに、第二次レビューとして、第一次レビューで基準該当 (+) または要検討として判定された診療録から100症例を抽出し、評価マニュアル¹³⁾に従って、14年以上の臨床経験を有する3人

図1 本調査研究の手順



の医師レビュー者（医師A：日本麻酔科学会認定麻酔専門医，医師B：日本内科学会認定内科専門医，医師C：日本外科学会指導医）が独立して有害事象の判定を行った（図2）。その3人の医師は，本調査研究のために組織されたワーキンググループ（WG）のメンバーであり，前年度の予備調査¹⁴⁾に参加しレビューを実施している。

3人の医師レビュー者は，有害事象の判定に関するマニュアルを熟読し，そのマニュアルに記載された判定基準や判定方法に従って第二次レビューを実施した。このマニュアルは，パイロットスタディにおいて把握された実際の有害事象の事例を通し，有害事象の判定に至るまでの経緯や有害事象として判定した理由を詳細に記したものである。

有害事象の事例は，典型的な有害事象症例および判定に困難を要した症例から構成されている。事例のカテゴリーとしては，院内感染，褥瘡，外来治療から病状悪化による入院，医療行為による副作用・合併症，搬送された重症患者の転帰，誤嚥性肺炎，イレウスの再発，診断・治療上の判断に関する問題，管理上の問題があげられている。

有害事象の判定は，本調査研究の有害事象の把握範囲である「患者への意図せぬ傷害（injury）や合併症（complication）で，一時的または恒久的な障害（disability）を生じ，疾病の経過でなく医療との因果関係（causation）が認められるもの」

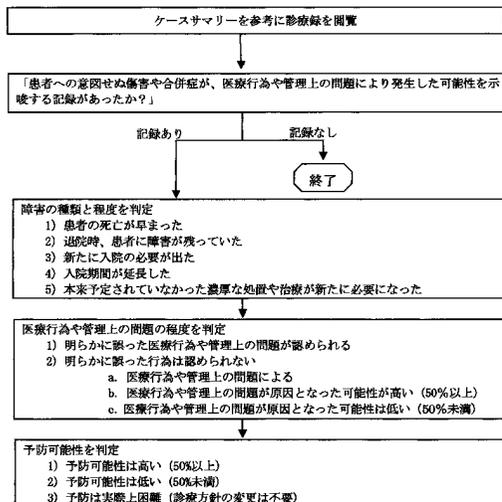
に従って行った。具体的には，①患者の死亡が早まった，②退院時，患者に障害が残っていた，③新たに入院の必要が出た，④入院期間が延長した，⑤①～④のいずれにも該当しないが，本来予定されていなかった濃厚な処置や治療が新たに必要になった，という障害の種類いずれかに該当するか否かの評価によって，有害事象の有無の判定が行われた。なお「重篤な原疾患・病状に合併しうる障害で，自然経過・予後に影響がなかった」症例については，有害事象の把握範囲から除外した¹⁾。

遡及的診療録レビューでは，「有害事象に関する記録が存在する」ということが有害事象の判定にあたっての前提条件となる。そこで診療録の閲覧を行う際に，まず「患者への意図せぬ傷害や合併症が，医療行為や管理上の問題により発生した可能性を示唆する記録があったか？」という記録の有無について確認を行った。そして「記録あり」と確認された場合に，有害事象の該当基準番号，有害事象の発生日およびその具体的な内容，有害事象の障害の種類を「評価シート2」に記載した。

その後，医療行為や管理上の問題の程度，予防可能性についても判定を行い，その判定結果も「評価シート2」に記入した。医療行為や管理上の問題の程度に関する判定においては，①「明らかに誤った医療行為や管理上の問題が認められる」，②「明らかに誤った行為は認められないが，医療行為や管理上の問題による」，③「明らかに誤った行為は認められないが，医療行為や管理上の問題が原因となった可能性が高い（50%以上）」，④「明らかに誤った行為は認められず，医療行為や管理上の問題が原因となった可能性は低い（50%未満）」の4段階で判定した。予防可能性については，①「予防可能性は高い（50%以上）」，②「予防可能性は低い（50%未満）」，③「予防は実際上困難（診療方針の変更は不要）」の3段階で判定した。

このような3人の医師による独立した有害事象の判定後に，評価マニュアル作成者であるWGの医師を加えた4人で，各レビュー者の判定結果について討議を行い，レビュー者全員の合意が得られた判定を最終判定とした。なお，評価マニュアル作成者であるWGの医師は，前年度の予備調査終了後に，米国，豪州，カナダにおける先行

図2 医師による有害事象の判定



研究を実施した研究者と意見交換を行うことによって、判定の適切性ならびに諸外国における判定基準との整合性について確認を行っている。

なお、レビュー者全員の討議において、合意が得られなかった、もしくは判定が困難であった症例については、WGに属する眼科医、産婦人科医等に照会し確認を行う、もしくは各専門領域の学会から推薦された医師からなる専門家パネルで討議に基づいて、有害事象の最終判定を行っているが、今回の検討においては、このような照会しなければならない症例については除外した。

2. 信頼性の検証方法

医師レビュー者の判定結果の信頼性を検証するために、各医師2人間の判定の一致率、および、最終判定を正答とした場合の各医師の適中率を求めるとともに、 κ 統計量¹⁵⁾を算出した。また医師の専門性が有害事象の判定に影響するか否かを検証するために、各医師レビュー者における外科系症例、内科系症例での適中率を算出した。

なお、一致率は、2人とも同一回答をした症例数を、全症例数で除して算出した。また、適中率は、真陽性数と真陰性数の和を、真陽性数、偽陽性数、真陰性数、偽陰性数の和で除して算出した。

3. 倫理的配慮

本研究は、調査対象病院における倫理委員会において承認を受け実施した。なお本研究は、人の疾病の成因および病態の解明並びに予防および治療の方法の確立を目的とする疫学研究ではないため、「疫学研究に関する倫理指針」の適用対象外である。同指針は、既存資料のみを用いる観察研究ではインフォームドコンセントを省略でき、その条件として研究実施の公開を義務づける（指針7(2)②イ）が、有害事象を把握するという本研究の特殊性から鑑み、研究実施を非公開とすること、調査対象となる患者からのインフォームドコンセントを省略すること、そして原則として研究に支障がない範囲で患者個人情報の保護を厳守するとともに、調査結果から病院を特定できるような形での公表は行わないことについて、調査病院の倫理委員会から承認を得ている。

III 研究結果

1. 対象患者の属性

第一次レビューにおいて、基準該当（+）もし

くは要検討と判定された159件中から抽出された100件における患者の基本的属性を表1に、診療録の記載状況については表2に示した。

患者の属性として、自院の外来、救急部・救急外来・時間外外来から、緊急な入院が必要になった場合、他院や他施設の外来からの紹介、もしくは搬送され、緊急に入院になったというような緊急入院の割合は39%、予定入院は61%であった。

表1 調査対象症例の基本的属性

| 項目 | | n=100 |
|-------------|-----------------------|-------------|
| 年齢 | 平均値±標準偏差 | 54.1±27.0 |
| | 中央値（最小値～最大値） | 65（0～95） |
| 入院日数 | 平均値±標準偏差 | 30.4±39.4 |
| | 中央値（最小値～最大値） | 18.5（2～254） |
| | | No. |
| 年齢階級 | | |
| | 0～24歳 | 20 |
| | 25～44歳 | 11 |
| | 45～64歳 | 18 |
| | 65歳以上 | 51 |
| 性別 | | |
| | 男性 | 54 |
| | 女性 | 46 |
| 入院状況 | | |
| | 緊急入院 | 39 |
| | 予定入院 | 61 |
| 退院先 | | |
| | 自宅 | 76 |
| | 他施設へ転院 | 20 |
| | 死亡 | 4 |
| 診療科 | | |
| | 内科* | 37 |
| | 外科** | 18 |
| | 産婦人科 | 6 |
| | その他 | 39 |
| 入院時の患者の精神状態 | | |
| | 痴呆・意識障害なし | 77 |
| | 痴呆あり | 5 |
| | 意識障害あり | 10 |
| | 不明 | 8 |
| | 手術・侵襲的処置あり | 72 |
| | 入院診療計画書に入院予定期間が記されている | 58 |

* 循環器内科、呼吸器内科、消化器内科、代謝内科、血液科、膠原病科、神経内科、腎臓内科が含まれる。

** 消化器外科、胸部外科、脳神経外科が含まれる。

入院時記録（病歴，診察，身体所見，入院理由，検査・処置・治療などの予定）が適切に記載されていた割合は93%であった。医師経過記録（48時間ごとに記載されている）は34%，看護経過記録（毎日記載されており，患者の状態変化あるいは処置が行われた場合は詳細な記録が必要）は96%

表2 記録が適切に記入されていると判断した割合

| 診療録の種類 | | n = 100 No. |
|---------|------|----------------|
| 入院時記録 | 適切 | 93 |
| | 不適切 | 7 |
| 医師記録 | 適切 | 34 |
| | 不適切 | 66 |
| 看護経過記録 | 適切 | 96 |
| | 不適切 | 4 |
| 処置記録 | 適切 | 73 |
| | 不適切 | 1 |
| | 該当なし | 26 |
| 病理報告書 | 適切 | 34 |
| | 不適切 | 0 |
| | 該当なし | 66 |
| 退院時サマリー | 適切 | 100 |
| | 不適切 | 0 |

が適切に記載されていた。処置記録（手術・麻酔記録，侵襲的処置の記録）が必要とされる74件中，73件が適切に記載されていた。

有害事象の有無に関する評価が不可能と判断された診療録はなかった。

2. 有害事象の有無に関する判定

各基準について基準該当(+)となった症例数，ならびに，基準該当(+)とはならなかったが要検討となった症例数を表3に示した。

4人の医師が討議を行う前に，各医師が独立して100症例のうち有害事象(+)と判定した症例数は，医師Aは18症例，医師Bは24症例，医師Cは27症例であった。医師AとBの一致率は90.0% ($\kappa=0.70$)，医師BとCの一致率は83.0% ($\kappa=0.55$)，医師AとCの一致率は83.0% ($\kappa=0.52$)であった。

4人の医師の討議によって16例が有害事象(+)と最終判定された。全100症例のうち7例については，有害事象の有無や種類を判定するためには専門家の意見が必要として，判定保留となった。これらは，泌尿器科，皮膚科，眼科，産科の症例であった。

表4に，判定保留となった7例を除いた93例に

表3 第一次レビューで基準該当(+)ならびに要検討と判定された基準の事象数*

| 内 容 | 基準該当(+) | 要検討 |
|-------------------------------------------------|---------|-----|
| 基準1 調査対象入院前の診療・処置・ケア等の結果として生じた，予定外の入院 | 33 | 28 |
| 基準2 調査対象入院後12ヶ月以内の予定外の再入院 | 11 | 5 |
| 基準3 病院で生じた患者のアクシデントや傷害 | 7 | 0 |
| 基準4 薬剤副作用反応 | 6 | 3 |
| 基準5 集中治療室や医療依存度の高い部署への予定外の移送 | 4 | 0 |
| 基準6 別の急性期病院への予定外の転院 | 1 | 0 |
| 基準7 調査対象入院における予定外の再手術 | 3 | 1 |
| 基準8 手術中，侵襲的処置，経陰分娩（鉗子分娩も含む）における，予定外の臓器の除去，損傷，修復 | 3 | 0 |
| 基準9 その他の患者の合併症（例：急性心筋梗塞，脳血管障害，肺塞栓症など） | 8 | 3 |
| 基準10 入院時に認められなかった，新たに発症した神経障害 | 3 | 1 |
| 基準11 調査対象入院中における予測外の死亡 | 2 | 1 |
| 基準12 不適切な自宅への退院 | 3 | 4 |
| 基準13 心停止，呼吸停止，低アプガースコア | 0 | 0 |
| 基準14 中絶や分娩，出産に関連した傷害や合併症（新生児の合併症を含む） | 2 | 1 |
| 基準15 院内感染/敗血症 | 9 | 1 |
| 基準16 医療行為や管理上の問題に関連した患者や家族の不満 | 5 | 0 |
| 基準17 検討中もしくは係争中の，訴訟を示す文書（弁護士からの文書など） | 0 | 0 |
| 基準18 他の基準に当てはまらないその他の望ましくない転帰 | 5 | 8 |

* 1症例で複数の事象が該当した症例あり

表4 3人の医師の有害事象に関する判定の一致度

| 医師 A | 医師 B | 医師 C | 症例数 | 最終判定で有害事象 (+) |
|------|------|------|-----|---------------|
| + | + | + | 11 | 11 |
| + | + | - | 2 | 2 |
| + | - | + | 0 | 0 |
| - | + | + | 2 | 1 |
| + | - | - | 2 | 1 |
| - | + | - | 4 | 1 |
| - | - | + | 8 | 0 |
| - | - | - | 64 | 0 |

ついでに3人の判定結果の違いを示した。3人とも有害事象 (+) と判定した症例は11例 (11.8%) であり、これらの症例は討議の結果すべて有害事象と最終判定された。2人が有害事象 (+) と判定し残りの1人が有害事象 (-) と判定した症例は4例 (4.3%) で、このうち最終判定で有害事象 (+) と判定されたのは3例であった。1人が有害事象 (+) と判定し残りの2人が有害事象 (-) と判定したのは14例 (15.1%) であり、このうち最終判定で有害事象 (+) と判定されたのは2例であった。3人とも有害事象 (-) と判定したのは64例であり、これらは最終判定ですべて

表5 事例 (一部改変)

事例1 80代男性

「入院前の経過」

8年前より、肺気腫の診断にて近医 (開業医) で外来フォロー。

本年1月、左肺自然気胸にて当院入院、胸腔ドレーン挿入し、経過良好にて在院約2週間で退院。

「調査対象入院」

本年9月、呼吸困難と左胸痛が出現。当院救急外来受診し、左肺自然気胸にて緊急入院。胸腔ドレーン挿入。入院約1週間目に胸膜癒着術施行。経過良好にて在院約2週間で退院。

「退院後の経過」

本年4月、呼吸困難と右胸痛が出現。当院救急外来受診し、右肺自然気胸にて緊急入院。胸腔ドレーン挿入。入院約1週間目に胸膜癒着術施行。経過良好にて在院約2週間で退院。

「看護師のスクリーニング結果」

基準1: 調査対象入院前の診療・処置・ケア等の結果として生じた、予定外の入院

基準2: 調査対象入院後12ヶ月以内の予定外の再入院

「医師の判定」

3名とも、疾病の自然経過による入院と判断し、「有害事象の記録なし」と判定。

事例2 50代女性

「調査対象入院前の経過」

本年1月、婦人科癌に対して手術を施行。左リンパ節は血管を巻き込んでいたため郭清できず、外来でフォローとなった。本年5月、CT上、左リンパ節の増大がさらに認められ、入院となった。

「調査対象入院」

入院後、炎症所見が認められ、化学療法を施行。化学療法1クール施行後、炎症所見は陰性化した。肝機能障害が出現したため、腹痛時に内服していた非ステロイド性消炎鎮痛剤を中止するとともに、肝保護剤投与にて対応。肝機能の正常化を待って、約90日後に2クール目を施行した。

「看護師のスクリーニング結果」

基準4: 薬剤副作用 (非ステロイド性消炎鎮痛剤による肝障害)

「医師の判定」

有害事象の記録の有無については、3名とも、薬剤性肝障害による「有害事象の記録あり」と判定。ただし、原因薬剤は不明と判定した。

障害の種類と程度は、3名とも「入院期間が延長した」と判定。

医療行為や管理上の問題の程度は、2名は「医療行為や管理上の問題による」、1名は「医療行為や管理上の問題が原因となった可能性が高い」と判定した。討議の結果、「医療行為や管理上の問題による」と最終判定された。

予防可能性は、2名は「予防は実際上困難」、1名は「予防可能性は低い」と判定した。討議の結果、「予防は実際上困難」と最終判定された。

有害事象(−)と判定された。

有害事象の有無に関する最終判定を基準とした場合の各医師の適中率は、医師Aが96.8% ($\kappa = 0.88$)、医師Bが94.6% ($\kappa = 0.82$)、医師Cが86.0% ($\kappa = 0.56$)であった。

外科系症例における有害事象の有無判定の適中率は、医師A、医師B、医師Cの順に、100%、93.2%、86.3%であった。内科系症例における有害事象の有無判定の適中率は、医師A、医師B、医師Cの順に93.6%、95.7%、85.1%であった。

3. 医療との因果関係に関する判定

有害事象(+)症例の「医療との因果関係」に関する最終判定は、「明らかに誤った行為は認められないが、医療行為や管理上の問題による」が9例、「明らかに誤った行為は認められないが、医療行為や管理上の問題が原因となった可能性が高い(50%以上)」が4例、「明らかに誤った行為は認められず、医療行為や管理上の問題が原因となった可能性は低い(50%未満)」が3例であった。最終判定を基準とした場合の適中率は、医師Aが78.6% ($\kappa = 0.65$)、医師Bが71.7% ($\kappa = 0.49$)、医師Cが41.7% ($\kappa = 0.24$)であった。

4. 予防可能性に関する判定

有害事象(+)症例の「予防可能性」に関する最終判定は、「50%以上」が1例「50%未満」が5例「実際上困難」が10例であった。最終判定を基準とした場合の適中率は、医師Aが64.3% ($\kappa = 0.40$)、医師Bが78.6% ($\kappa = 0.56$)、医師Cが66.7% ($\kappa = 0.13$)であった。

表5に、有害事象の内容について判定が一致した事例(事例1)と、判断が一部食い違った事例(事例2)を示した。

IV 考 察

1. 各医師の判定結果間の一致率

廻及的診療録調査における有害事象の有無の判定について、複数医師間の判定の信頼性に関する先行研究の結果は、米国ユタ・コロラド州における研究では一致率80.4~82.9% ($\kappa = 0.40 \sim 0.41$)¹¹⁾、オーストラリアの研究では一致率80% ($\kappa = 0.55$)⁴⁾であった。米国ユタ・コロラド州の研究では医師レビュー者の属性、人数などは明らかではないが、豪州では臨床経験10年以上であり、内科9人、麻酔科5人、産科4人、外科2

人、小児科1人の合計21人の医師がレビューを実施している。

各国の判定基準に若干の相違があることから今回の結果との単純な比較はできないが、今回の研究における医師AとBの間の一致率については、先行研究の結果を上回る水準であった。

マニュアルに類似ケースについての記載があった症例、あるいは、記載がなくても臨床の見地から自明の症例については、有害事象の有無ならびにその内容に関して3名の医師の判断がほぼ一致していた。

一方、有害事象の有無の判定に関して医師の意見が分かれた症例について、その主な要因を分析すると、①専門領域の判断による違い、②有害事象が複数ある場合において、主たる有害事象の選択における優先順位の判断が困難、③他の医療施設での処置・治療・ケアが原因となって入院した場合における情報量が不十分、④障害の程度の捉え方の違い、という4つが考えられた。

第一に、今回の医師の判定結果の信頼性検証では、各医師の専門領域とは異なった診療科の診療録レビューも実施しており、専門的な知識が要求されるような複雑な症例では、各医師が専門とする領域によって解釈や判断が異なり、有害事象の判定結果に影響を与えていた。

第二に、本調査研究では、1症例に対して「主な有害事象」として1つの有害事象を判定しているが、1症例に対して複数の事象が生じ、事象の優先順位の判断に困難を要する症例においては、医師間の判定に影響を与えていた。本調査研究では、複数の有害事象が認められる場合を想定し、「主な有害事象」の選択基準を設定している¹¹⁾。しかし、複数の有害事象が生じた場合で、障害のレベルや医療行為や管理上の問題の程度がほぼ同程度であった場合には、選択基準に従って選択することは困難であり、「主な有害事象」として、とるべき事象に相違が生じることがあった。このような症例に関しては、討議によって検討を行うことで、とるべき「主な有害事象」の合意を得ていくことが必要であると考えられた。

第三に、本調査研究では、他の医療施設の処置・治療・ケアが原因となって当該病院に入院となった症例も有害事象として把握しているが、把握できた他の医療施設からの情報量によって、こ

の有害事象の判定結果が左右された。たとえば、他の医療施設で発生した事象については、情報が不十分であることが多かった。したがって、限られた情報の中から有害事象を判定しなければならないような症例では、「病態の自然経過」で生じた入院であるか、それとも「他の医療施設において不適切な診療・処置・ケアが行われた」ことにより生じた入院であるかの判断が困難であり、医師により評価が異なる場合がみられた。適及的診療録レビューでは、記録に十分な記載がない場合には、有害事象の判定を行うことは難しく、適及的診療録レビューの限界でもある。したがって、このような症例に関しては、各医師の専門的知識と経験的な判断に基づいた討議によって、多角的な視点から検討し、有害事象の判定を行っていくことの必要性が示唆された。

第四に、障害の種類である「退院時、患者に障害が残っていた」を判定する場合に、医師間でどの程度の障害を有害事象とするかで、障害の捉え方の見解が異なっていた。たとえば、聴神経腫瘍術後の顔面神経不全麻痺は、仕事・家事に影響のないレベルの障害とするか否かで意見の相違が生じていた。したがって、医師間で、有害事象としてみなす障害の程度の捉え方に相違が生じた場合には、討議によって合意形成を図り、有害事象とする障害の程度について明確にし、マニュアルに反映させていくことが必要と考えられた。

2. 各医師の判定結果の適中率

諸外国において、第二次レビューの監督者に該当する医師の有害事象の判定を基準として、医師レビュー者の判定結果の適中率を検証した結果、 κ 統計量は0.47~0.61であった^{2,6,8)}。本調査研究における各医師の適中率は、諸外国と比較高い水準であった。

内科症例については、内科医Bの適中率が最も高かった。また、麻酔科医Aは手術を有する全般的な症例に関わっていることから、術中や集中治療室で生じた有害事象をはじめとして、外科系の様々な症例に関する判定に優れていた。

一方、外科医Cの有害事象の判定については、偽陽性症例が9例あり、適中率を下げる原因となっていた。偽陽性症例が多かった理由として、判定に困難を有する症例や有害事象の判定に関して新たな基準や解釈が必要となる場合に、有

害事象(+)として判定する傾向が認められた。たとえば、本調査研究では「濃厚な処置や治療」を一般に入院管理が必要な処置・治療と見なしているが、処置や治療を外来で行うことも可能であるが入院管理によっても行われているというような境界にある症例においては、偽陽性として判定される傾向にあった。また入院中に患者の自己管理による処置(例:自己導尿)が原因となり、感染などを生じた場合には、医療管理上の問題とするか否かによる判定基準が確立していなかったため、医師Cは患者の自己管理上の問題も医療管理の中に含め、有害事象(+)としていた。また医師Cが臨床を積んできた病院においては、高い医療水準にあったため、その水準を基準として厳しく有害事象の判定を行ったことが偽陽性症例の増加につながったことも考えられる。

こうした状況に対応するため、今後実施される本調査研究の通常の第二次レビューにおいては、有害事象の判定に関する新たな基準や解釈が必要となる場合や、判断に迷うような症例については、「陽性」として取るのではなく「要検討」とし、医師レビュー者間で討議を行う方式とした。要検討症例を複数医師によって討議することにより、合意形成をはかりマニュアルを随時改訂することもでき、有害事象の判定の精度を高めることに繋げることができると考えられる。

3. 医療との因果関係の程度や予防可能性に関する判定

カナダ調査における医療との因果関係の判定一致率の κ 統計量は0.45、予防可能性に関する判定一致率の κ 統計量は0.69と報告されている⁸⁾。本調査研究の医療との因果関係の程度や予防可能性に関する判定の最終判定を基準とした場合の適中率は、医師間で大きなばらつきが認められた。

医療との因果関係の程度や予防可能性に関する最終判定は、各専門領域における調査対象時点での医療水準を踏まえた高度な判断が要求され、判定には困難を伴う場合も多かった。本調査研究では、各医師が独立してレビューを実施した後に、評価マニュアル作成者の医師が確認を行い、両者の合意が得られないものについては全体で討議を行っている。そして、合意が得られないものについては、本調査研究のワーキンググループの各診療科の医師から助言を求めたり、また専門家パネ

ルを開催し各領域の専門医から意見を聞く機会を設けている。したがって、各医師の適中率自体は必ずしも良好ではなかったが、最終的には妥当性ある判定に繋げることができると考えられる。

4. 本研究における限界

今回の研究結果を解釈する際の限界としては、次のような点があげられる。

第一に、今回の調査は1病院のみで行われているため、結果が他の病院にも普遍的に当てはまるかどうか問題となる。今回の調査では調査対象病院名を公表しない取り決めとなっているため病院の特性を詳細に述べることはできないが、地域の中核病院であり他院からの紹介患者や重症例の受け入れも多かったことから、診療録レビューの際に専門的な判断を要する症例が多数含まれていた。このため、一般的な病院における調査よりも適中率の低下につながった可能性がある。一方で、診療情報管理が整備されており診療録の記載様式が整っていたことから、診療録の記載内容から適切な状況把握が可能であり、個々の医師の判断の妥当性が高まったとも考えられる。他の調査対象病院における医師レビューに際して同様の適中率が得られるかどうかは、今後の検証が必要である。

第二に、評価結果の妥当性に関する評価が困難な点が挙げられる。今回の調査では入院診療録を調査対象としており、必要に応じて当該病院の外来診療録を閲覧することができた。しかし、インシデントレポートやアクシデントレポートを参照することが認められなかったこと、患者や担当医ら関係者へのヒアリングは行わない取り決めとなっていたこと、別の医療機関が原因となった有害事象については紹介状を除き他院からの情報は得られなかったこと、などの制約があり、外的基準による妥当性の評価は行うことができなかった。今後、インシデントレポートやアクシデントレポートの参照などが可能な状況において、信頼性とともな妥当性の検証を行うことを計画したい。

V 結 語

有害事象の有無については個々の医師の判定の信頼性が確認された。一方、医療との因果関係・予防可能性に関する判定については、様々な専門領域の医師が討議の上で判定することが望ましい

と考えられた。

本研究は、平成16年度厚生労働科学研究「医療事故の全国的発生頻度に関する研究」（主任研究者 東海大学医学部付属病院 堺秀人）の研究成果の一部である。

本研究の実施に当たっては、倫理委員会における審査、診療録の抽出、調査場所の確保など、調査対象病院の関係者の皆様に多大なご協力をいただきました。心より感謝申し上げます。

(受付 2004.10.19)
(採用 2005. 3.16)

文 献

- 1) Kohn LT, Corrigan JM, Donaldson MS. To Err is Human: Building a Safer Health System. Washington DC: National Academies Press, 2000; 26-48.
- 2) Brennan TA, Leape LL, Laird NM, 他. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard medical practice study T. New England Journal of Medicine 1991; 324: 370-376.
- 3) Thomas EJ, Studdert DM, Burstin HR, 他. Incidence and types of adverse events and negligent care in Utah and Colorado. Medical Care 2000; 38: 261-271.
- 4) Wilson RM, Runciman WB, Gibberd RW, 他. The quality in Australian health care study. The Medical Journal of Australia 1995; 163: 458-471.
- 5) Vincent C, Neale G, Woloshynowych M, 他. Adverse events in British hospitals: Preliminary retrospective record review. BMJ 2001; 322: 517-519.
- 6) Davis P, Lay-Yee R, Briant R, 他. Adverse events in New Zealand public hospitals I: Occurrence and impact. The New Zealand Medical Journal 2002; 115: U271.
- 7) Schioler T, Lipczak H, Pedersen BL, 他. Danish Adverse Event Study, Incidence of adverse events in hospitals. A retrospective study of medical records. Ugeskrift for laeger 2001; 163: 5370-5378.
- 8) Baker GR, Norton PG, Flintoft V, 他. The Canadian Adverse Events Study: The incidence of adverse events among hospital patients in Canada. Canadian Medical Association Journal 2004; 170: 1679-1686.
- 9) 北沢直美, 池田俊也, 兼児敏浩, 他. 有害事象把握のための診療録レビューにおける看護師スクリーニングの信頼性の検証. 病院管理 42, 1-13, 2005.
- 10) 小林美亜, 池田俊也, 堺 秀人. 診療録調査により有害事象を把握する試み—看護師による遡及的診療録レビューの妥当性検証—. 日本看護管理学会誌 8, 12-20, 2005.

- 11) Thomas EJ, Lipsitz SR, Studdert DM, 他. The reliability of medical record review for estimating adverse event rates. *Annals of Internal Medicine* 2002; 136: 812-816.
 - 12) 小林美亜, 鈴木琴江, 中田かおり, 他. 遡及的診療録調査により有害事象を把握するための方法: 看護師による有害事象把握のための基準作成. *看護管理* 2004; 14: 756-762.
 - 13) 中田かおり, 池田俊也, 小林美亜, 他. 医療事故の全国的発生頻度に関する研究における有害事象の判定基準について. *病院管理* 2004; 41: 111-123.
 - 14) 小林美亜, 池田俊也, 兼児敏浩, 他. 有害事象把握のための遡及的診療録レビュー手法に関する検討. *医療と社会* 2004; 14(2), 55-69.
 - 15) Agresti A: *An Introduction to Categorical Data Analysis*, 1st edition. NewYork: Wiley-Interscience, 1996.
-

METHODOLOGY OF RETROSPECTIVE CHART REVIEW TO DETECT ADVERSE EVENTS: THE RELIABILITY OF PHYSICIAN JUDGMENT

Shunya IKEDA^{*、2*}, Mia KOBAYASHI^{*、2*、3*}, Misa SAKAGUCHI^{4*},
Toshihiro KANEKO^{5*}, Masahiro HIROSE^{6*}, and Hideto SAKAI^{7*}

Key words : adverse events, medical record, retrospective review, reliability, patient safety

Purpose To establish a scientific basis for promoting patient safety, basic information related to the incidence of adverse events (AEs) is needed. In studies in several other countries, trained nurses screened for potential AEs using explicit criteria in the first stage, and physicians reviewed selected charts in the second stage. To assure the accuracy of retrospective chart review, it is important to verify the reliability of AE judgments by physician reviewers. The purpose of this study was to test this reliability of judgment of AEs (their presence, causation of healthcare management and preventability) by three physician reviewers.

Methods This study used 100 selected charts of non-psychiatric inpatients in an acute care hospital. Three physicians independently assessed AEs and discussed their judgments with the physician who created the manual for judging AEs. We considered judgments of the AEs agreed on by the four physicians to be final AE judgments and compared the reliability of each measure related to AE judgments among the physician reviewers using the kappa statistic.

Results The number of AE cases each physician reviewer judged ranged from 18 to 27. Agreement on the presence of an AE ranged from 83.0% to 90.0% ($\kappa=0.52\sim0.70$). Ultimately, AEs were judged to have occurred in 16 cases while 7 cases were deferred. The agreement on the presence of an AE between the physician's and the final judgment ranged from 86.0% to 96.8% ($\kappa=0.56\sim0.88$). However, agreement on the causation of healthcare management and preventability between the physician's and the final judgment was not in the acceptable range.

Conclusion The reliability of each physician's judgments regarding the presence of an AE was satisfactory. However, the reliability of judgments related to the causation of health care management and preventability was not necessarily satisfactory. Therefore, it is considered important to judge causation and preventability based on discussion with clinical experts in the relevant field.

* Department of Health Policy and Management, Keio University School of Medicine

2* The Health Care Science Institute

3* Division of Nursing, The Steinhardt School of Education, New York University

4* Health Care Administration and Management, Graduate School of Medical Sciences, Kyushu University

5* Patient Safety Division, Mie University Hospital

6* Patient Safety Division, Kyoto University Hospital

7* Tokai University Hospital