

Matthias Schwenkglens, Prof. Thomas D. Szucs
European Center of Pharmaceutical Medicine
Universität Basel

3. August 2005

Überlegungen in Zusammenhang mit dem Ärzterating der santésuisse in der Deutschschweiz

speziell

- **Machbarkeitsüberlegungen betreffend die wissenschaftliche Untersuchung eines möglichen Zusammenhangs zwischen dem Ärzterating der santésuisse und dem Erreichen von Zielen der Hyperlipidämie-Behandlung in der ärztlichen Grundversorgung**
- **Überlegungen zu alternativen Studiendesigns**
- **Ergebnisse einer in diesem Zusammenhang durchgeführten Literaturrecherche**

im Auftrag von Dr. Micheal Romanens, Belchenstrasse 18, 4600 Olten

Kontaktadresse

Matthias Schwenkglens, M.A. M.P.H.
c/o ECPM Executive Office
University Hospital
CH-4031 Basel
Switzerland

phone +41 61 261 45 83

fax +41 61 261 45 84

email m.schwenkglens@unibas.ch

Inhaltsverzeichnis

Zusammenfassung der Ergebnisse.....	3
Hintergrund und Auftrag	4
Machbarkeitsüberlegungen betreffend eine Studie zur Frage eines Zusammenhangs zwischen dem Ärzterating der santésuisse und dem Erreichen von Zielen der Hyperlipidämie-Behandlung in der Grundversorgung	5
1. Methodische Schwierigkeiten und mögliche Studientypen	5
2. Designaspekte einer retrospektiven Kohortenstudie	13
3. Zusammenfassung und Beurteilung	17
Alternative Vorgehensweisen bzw. Studiendesigns	19
Durchgeführte Literaturrecherchen und Resultate	22
Literaturangaben	25
Literaturangaben - Abstracts (alphabetisch nach Autoren)	28

Zusammenfassung der Ergebnisse

Die wissenschaftliche Untersuchung eines möglichen Zusammenhangs zwischen dem Ärzt rating der santésuisse in der Deutschschweiz und dem Erreichen von Zielen der Hyperlipidämie-Behandlung in der ärztlichen Grundversorgung wäre möglich, aber mit starken Unsicherheiten verknüpft und mit erheblichen methodischen und praktischen Problemen verbunden. Die bestmögliche Form wäre die einer retrospektiven Kohortenstudie. Die Autoren raten jedoch von einer Durchführung eher ab.

Alternative Studienmöglichkeiten beinhalten im wesentlichen (1) die Durchführung einer rein deskriptiven Studie zur 'Epidemiologie' des Ärzt ratings in der Deutschschweiz und (2) die Durchführung einer ebenfalls primär deskriptiven Studie, die aber so angelegt wird, dass gleichzeitig die Erfordernisse des ersten Teils der retrospektiven Kohortenstudie erfüllt werden. Auf dieser Basis könnte dann ein (fundierterer, aber immer noch mit sehr grossen Unsicherheiten behafteter) Entscheid über die letztendliche Durchführung der vollen retrospektiven Kohortenstudie getroffen werden. Weiter bestünde (3) die Möglichkeit, für ein Sample von Arztpraxen ein adäquates *casemix adjustment* durchzuführen und auf dieser Basis die Angemessenheit der Ratingresultate der santésuisse zu überprüfen. Eine Anschluss-Studie mit Bezug zur Hyperlipidämie wäre auch hier denkbar.

Literaturangaben, die im Rahmen der durchgeführten Abklärungen identifiziert wurden, sind am Schluss des Dokuments kurz kommentiert und die zugehörigen Abstracts sind als Materialien zur Verfügung gestellt. Die Aussage der santésuisse, die Korrektur der verursachten Kosten auf Basis der Variablen Spezialisierung des Arztes, Region sowie Alter und Geschlecht der Patienten sei ausreichend, steht in scharfem Kontrast steht zur durchweg in der Literatur festgestellten Notwendigkeit eines *casemix adjustments*, das der unterschiedlichen Morbidität verschiedener Patientenkollektive adäquat Rechnung trägt.

Hintergrund und Auftrag

Das Ärzt rating der santésuisse erfolgt gestützt auf das Krankenpflegeversicherungsgesetz (KVG) und besteht inhaltlich in einem Kostenvergleich zwischen Ärzten der gleichen Spezialisierung, die in derselben Region tätig sind [1,2]. Weiterhin erfolgt eine Adjustierung der Kosten nach Alter und Geschlecht des Patientenkollektivs. Ärzte mit, ausgehend von dieser Berechnungsgrundlage, deutlich überhöhten Kosten (üblicherweise mindestens 130% des jeweiligen Referenzwertes) können von santésuisse kontaktiert werden. In einem mehrstufigen Prozess, so santésuisse, wird dann versucht, eine Klärung bzw. eine Kostensenkung zu erreichen. Als letzte Möglichkeit erhebt die santésuisse Regressforderungen, wobei diese Praxis durch die schweizer Gerichte im wesentlichen gestützt wird.

Nach Auffassung des Auftraggebers ist die Praxis der santésuisse ungerecht, erstens weil keine ausreichende Berücksichtigung der Patientenzusammensetzung der einzelnen Praxen erfolgt, und zweitens wegen einer uneinheitlichen Vorgehensweise. So sei die Auswahl der gemahnten etc. Ärzte willkürlich und Regressforderungen beschränkten sich weitgehend auf die Kantone Aargau, Bern und Solothurn. Der Auftraggeber geht davon aus, dass die Qualität der ärztlichen Grundversorgung in der Schweiz durch die gegenwärtige Praxis des Ärzt ratings beeinträchtigt wird und dass die unter anderem in einer Beeinträchtigung der Qualität der Hyperlipidämie-Behandlung zum Ausdruck kommt.

Die Autoren wurden beauftragt, die Machbarkeit einer wissenschaftlichen Studie zu prüfen, die geeignet wäre, diesen vermuteten Zusammenhang wissenschaftlich valide zu untersuchen. Die nachfolgenden Teile dieses Dokuments widmen sich dieser Fragestellung. Weiterhin werden Möglichkeiten für alternative Studienansätze skizziert. Abschliessend werden die Ergebnisse von Literaturrecherchen, die im Rahmen der Machbarkeitsstudie erforderlich waren, kurz vorgestellt.

Die Autoren sind Mitarbeiter einer universitären Einrichtung und vertritt eine neutrale Position. Es bestehen keinerlei Interessenkonflikte bezüglich der Fragestellung dieses Auftrags.

Machbarkeitsüberlegungen betreffend die wissenschaftliche Untersuchung eines möglichen Zusammenhangs zwischen dem Ärzterating der santésuisse in der Deutschschweiz und dem Erreichen von Zielen der Hyperlipidämie-Behandlung in der ärztlichen Grundversorgung

1. Methodische Schwierigkeiten und mögliche Studientypen

Im folgenden wird versucht, anhand von neun Kernfragen zu erörtern, ob die Untersuchung eines möglichen Zusammenhangs zwischen dem Ärzterating (*physician profiling*) der santésuisse in der Deutschschweiz und dem Erreichen von Zielen der Hyperlipidämie-Behandlung in der ärztlichen Grundversorgung wissenschaftlich sinnvoll wäre und welche methodischen und praktischen Probleme mit der Übersetzung dieser Fragestellung in ein konkretes Studiendesign verbunden wären.

Terminologie:

- Der Begriff 'Exposition' steht für den Grad der Betroffenheit des einzelnen Arztes durch das Ärzterating.
- Der Begriff '*outcome*' steht die Qualität der Hyperlipidämie-Behandlung.
- Der Begriff '*confounders*' steht für 'verzerrende' Variablen, d.h. für Variablen, die sowohl mit der Exposition als auch mit dem *outcome* in Zusammenhang stehen. Um ein im wesentlichen unverzerrtes Studienresultat zu erhalten, müssen alle relevanten *confounders* erhoben und in der statistischen Analyse berücksichtigt werden.

Kernfrage 1: Erscheint ein Zusammenhang zwischen dem in der Deutschschweiz durchgeführten Ärzterating und der Qualität der Hypercholesterinämie-Behandlung in der ärztlichen Grundversorgung plausibel?

Für das Vorliegen eines solchen Zusammenhangs sprechen folgende Argumente:

- Die Prävalenz der Hyperlipidämie bzw. von Risikokonstellationen, die laut den gültigen Guidelines einer Behandlung bedürfen, ist hoch. Daher besteht bei Behandlungskosten von einigen hundert Schweizer Franken pro Patient und Jahr ein erhebliches kurzfristiges Sparpotential.
- Es ist aus diversen Studien bekannt, dass der Behandlungsgrad der Hyperlipidämie und die Qualität der lipidsenkenden Behandlung im Vergleich zu den Vorgaben der gültigen Guidelines unzureichend sind [3-6]. Deshalb handelt es sich hier um ein relativ 'weiches' Behandlungsgebiet, dessen Vernachlässigung typischerweise keine Kunstfehlerprozesse

nach sich ziehen dürfte. (Dies gilt zumindest für eine abgrenzbare Teilmenge der betreffenden Patienten.)

Argumente gegen das Vorliegen des vom Auftraggeber vermuteten Zusammenhangs fehlen. Allerdings wurde dieser Zusammenhang bisher in der Literatur nicht beschrieben und entsprechende Studien sind nicht verfügbar. Nur eine Arbeit stellte einen Zusammenhang zwischen ökonomischen Gegebenheiten und der Qualität der Hyperlipidämie-Behandlung dar, wobei die untersuchte Exposition eine limitierende Vergütungsregelung war [7]. Dies ist nicht direkt vergleichbar.

Fazit Kernfrage 1: Das Vorliegen eines Zusammenhangs zwischen dem in der Deutschschweiz durchgeführten Ärzterating und der Qualität der Hypercholesterinämie-Behandlung erscheint grundsätzlich plausibel, ist aber bisher nirgendwo belegt. Auch auf internationaler Ebene fehlen entsprechende Studien.

Kernfrage 2: Sofern ein Zusammenhang zwischen dem in der Deutschschweiz durchgeführten Ärzterating und der Qualität der Hypercholesterinämie-Behandlung besteht, ist das Ausmass dieses Effekts von Bedeutung. Mit anderen Worten: Ist der Effekt praktisch von Bedeutung, ist er 'klinisch relevant'?

Die Qualität der Hypercholesterinämie-Behandlung (mit den Elementen (1) Identifikation der Risikopatienten, (2) Initiierung der Behandlung und (3) Erreichen der Behandlungsziele) wird durch eine Vielzahl arztseitiger (z.B. Erfahrung, Interesse, zeitliche Verfügbarkeit, Patientenmix) und patientenseitiger (z.B. *compliance*, finanzielle Aspekte wie Franchise und Selbstbehalt) Faktoren beeinflusst. Ob der Einfluss des Ärzteratings relativ zu diesen Faktoren eine substantielle Rolle spielt, ist derzeit nicht abzuschätzen.

Auf der analytischen Ebene bedeutet dies, dass der Effekt, auch wenn vorhanden, in der unadjustierten (nicht für andere Faktoren korrigierenden) Analyse möglicherweise unsichtbar wäre. Seine Sichtbarmachung durch Regressionsverfahren würde eine adäquate Messung aller relevanten Einflussfaktoren (*confounders*) voraussetzen [8].

Fazit Kernfrage 2: Das Ausmass des vermuteten Effekts und damit seine praktische Bedeutung sind derzeit nicht einschätzbar.

Kernfrage 3: Wäre ein prospektives oder ein retrospektives Studiendesign vorzuziehen?

Zur Untersuchung des vorliegenden Sachverhalts kommt realistischlicherweise nur eine Beobachtungsstudie in Frage. Ein experimentelles Design ist zumindest unter den Bedingungen des schweizerischen Gesundheitswesens nicht möglich. Ein prospektives Vorgehen im Rahmen einer Beobachtungsstudie würde, da eine Information der Zielpopulation (Ärzte und eventuell Patienten) unvermeidbar wäre, das Verhalten der

Akteure beeinflussen und dadurch die Ergebnisse verzerren, möglicherweise gar ins Gegenteil verkehren. Es kommt daher nur ein retrospektives Vorgehen in Frage, eventuell in Verbindung mit einer Lipidmessung am Ende der retrospektiven Beobachtungsperiode. Die Nachteile des retrospektiven Ansatzes liegen vor allem im Bereich der Datenverfügbarkeit und der Datenqualität.

Fazit Kernfrage 3: Unter den gegebenen Umständen käme nur eine retrospektive Studie, eventuell mit einer Lipidbestimmung am Ende der Beobachtungsperiode, in Frage.

Kernfrage 4: Ist mit Problemen aufgrund von Selektionseffekten zu rechnen?

Verzerrende Selektionseffekte könnten erstens auf der Ebene der teilnehmenden Ärzte auftreten. Sofern ein Studiendesign gewählt würde, das ein Patienteneinverständnis erfordert, wären zweitens verzerrende Selektionseffekte auf der Patientenebene möglich. Das Problem der Teilnahmebereitschaft der Ärzte ist als das kritischere einzustufen. Zum Beispiel wäre das Beteiligungsverhalten derjenigen Ärzte, die ihr Behandlungsverhalten aufgrund des Ärzteratings verändert haben, möglicherweise stark atypisch, wobei die Abweichung in beide Richtungen gehen könnte. Dies könnte die Studienergebnisse stark und unkontrollierbar verzerren.

Fazit Kernfrage 4: Eine sehr hohe Beteiligungsrate der angefragten Ärzte wäre von zentraler Bedeutung. Mechanismen, um eine solche sicherzustellen, müssten im Studiendesign implementiert und gegebenenfalls im Rahmen einer Pilotstudie bzw. einer ersten Studienphase getestet werden.

Kernfrage 5: Wie könnte ein Studiendesign zur Beantwortung der Forschungsfrage aussehen?

Grundsätzlich bestünde die Möglichkeit, entweder eine Fall-Kontroll-Studie oder eine retrospektive Kohortenstudie durchzuführen.

- **Möglichkeit 1: Fall-Kontroll-Studie**

- o **Beobachtungseinheit:** Personen der Risikopopulation.
- o **Rekrutierung der Teilnehmer:** Gewinnung einer Zufallsstichprobe der Risikopopulation. Personen, die die Einschlusskriterien erfüllen, könnten z.B. über *computer-assisted telephone interviews* (CATIs) zufällig ausgewählt werden. Bei Teilnahmebereitschaft und nach Erteilen des *informed consent* würden mögliche *confounders* erhoben sowie der behandelnde Arzt identifiziert.
- o **Messung des outcome:** Entweder auf Basis der Krankenakten des behandelnden Arztes oder durch Aufbieten der Teilnehmer für eine Lipidbestimmung. Wäre das

Einverständnis zur Durchführung einer Lipidbestimmung Teilnahmevoraussetzung, wären *outcome*-Daten für die meisten Teilnehmer vorhanden.

- o **Messung der Exposition:** Angaben der teilnehmenden Ärzte. Gewinnung der Ratingresultate etc. vorzugsweise direkt von *santésuisse* nach Einverständnis der teilnehmenden Ärzte.
- o **Bemerkungen:**
 1. Es wäre nicht korrekt, nur Personen mit bekannter Hyperlipidämie einzuschliessen, da das Bekanntsein dieses Risikofaktors bereits durch das Ärzterating beeinflusst sein könnte.
 2. Eine Rekrutierung der Studienteilnehmer über Zeitungsannoncen wäre denkbar. Die Gefahr eines Selektionsbias wäre jedoch deutlich grösser als im Falle des CATI-Ansatzes. Ausserdem könnten durch die Zeitungsannoncen selbst Verzerrungen induziert werden.
- o **Nachteile:**
 1. Kosten der Telefoninterviews und der Lipidbestimmungen.
 2. Wäre die Teilnehmer bereit, sich einer Lipidbestimmung zu unterziehen?
 3. Gefahr eines Selektionsbias, insbesondere, wenn im Rahmen des *informed consent* über das Studienziel informiert würde. Andererseits wäre es vermutlich ethisch nicht akzeptabel, die Teilnehmer über das Studienziel im unklaren zu lassen. (Eventuell besteht hier eine abklärungsbedürftige Grauzone.)
 4. Eventuell bestünde das Problem einer 'seltenen Exposition': Möglicherweise hätten nur wenige der Teilnehmer einen Arzt, der durch das Ärzterating betroffen ist. Dies hätte negative Auswirkungen auf die erforderliche Fallzahl (siehe Kernfrage 9: Fallzahlschätzung, Beispiel 1).
 5. Es wäre nicht möglich, parallel die 'Epidemiologie' des Ärzteratings in der Schweiz zu erfassen, es sei denn, ein Fragebogen hierzu würde parallel oder vorgängig von allen potentiell betroffenen Ärzten ausgefüllt, oder entsprechende Daten würden von *santésuisse* zur Verfügung gestellt.
- **Möglichkeit 2: Retrospektive Kohorten-Studie**
 - o **Beobachtungseinheit:** Personen der Risikopopulation (wie Möglichkeit 1).
 - o **Rekrutierung der Teilnehmer:** Gewinnung einer Zufallsstichprobe der in Frage kommenden Ärzte. Innerhalb jeder teilnehmenden Praxis Bestimmung des Expositionsstatus. Bei gleich vielen exponierten und nicht exponierten Praxen Auswahl eines Zufallssamples der Patienten, die die Einschlusskriterien erfüllen. (Verbesserung der statistischen Macht durch dieses zweistufige Vorgehen.)

- o **Messung der Exposition:** Angaben der teilnehmenden Ärzte. Gewinnung der Ratingresultate etc. vorzugsweise direkt von santésuisse nach Einverständnis der teilnehmenden Ärzte.
- o **Messung des outcome:** Krankenakten der teilnehmenden Ärzte. Eventuell Aufbieten der eingeschlossenen Patienten für eine Lipidbestimmung. Es müsste jedoch sichergestellt werden, dass vor der Gewinnung des Lipidprofils keine Einflussnahme durch den behandelnden Arzt stattfinden kann. Probleme: Entstehung von Selektionsbias, patientenindividueller *informed consent* zwingend.
- o **Bemerkungen:**
 1. Es wäre nicht korrekt, nur Personen mit bekannter Hyperlipidämie einzuschliessen, da das Bekanntsein dieses Risikofaktors bereits durch das Ärzterating beeinflusst sein könnte. Allerdings bestünde die Möglichkeit, die Studie auf eine Hochrisikopopulation (z.B. Patienten mit bekannter koronarer Herzkrankheit) zu beschränken.
 2. Es stellt sich die ethische Frage, ob es akzeptabel wäre, Patientendaten anonym aus den Krankenakten zu extrahieren, ohne einen individuellen *informed consent* einzuholen.
- o **Vorteile:**
 1. Möglichkeit, durch eine vermehrte Auswahl (*oversampling*) exponierter Ärzte die erforderliche Fallzahl günstig zu beeinflussen (siehe auch Kernfrage 9: Erforderliche Fallzahl).
 2. Sicherstellung einer ausreichenden Zahl von Patienten pro Arzt, um den Effekt des *physician clustering* (das Phänomen einer relativen Häufung vergleichbarer Patienten bei einzelnen Ärzten) adäquat zu berücksichtigen [9].
 3. Möglichkeit einer Bestimmung der 'Epidemiologie' des Ärzteratings in der Deutschschweiz. (Allerdings erscheint es absurd, bei santésuisse weitgehend vorhandene Daten im Rahmen einer aufwendigen Studie von den Ärzten neu zu erheben.)
- o **Nachteile:**
 1. Möglichkeit von Selektionsbias, insbesondere bei unvollständiger Teilnahmebereitschaft der Ärzteschaft.
- o **Weitere Details siehe Abschnitt Designaspekte einer retrospektiven Kohortenstudie**

Fazit Kernfrage 5: Als Studiendesign zur Untersuchung des vermuteten Sachverhalts kämen grundsätzlich eine Fall-Kontroll-Studie oder eine retrospektive Kohortenstudie in Frage, wobei die letztere Möglichkeit deutlich vorzuziehen wäre. In beiden Studiendesigns wäre potentiell mit problematischen Selektionseffekten zu rechnen,

und Massnahmen zur Reduzierung dieser Effekte konfliktieren eventuell mit Erfordernissen der *Good Clinical Practice (GCP)* bzw. Ethik.

Kernfrage 6: Kann die Exposition adäquat bestimmt werden?

Folgende Aspekte sind hier von Bedeutung:

- Objektive Faktoren (Resultate des Ärztatings, bereits erfolgte Ermahnungen durch *santésuisse*, bereits erfolgte Konfrontation mit Regressforderungen) und psychologische Faktoren (subjektives Gefühl der Betroffenheit) interagieren in nicht bekannter Weise. Zum Beispiel könnte die Angst vor Regressforderungen möglicherweise auch 'präventiv' die Verhaltensweisen eines Teils der Ärzte beeinflussen, so dass bei diesen nie ein hohes Ratingresultat sichtbar würde.
- Alle identifizierten Studien, die sich unter verschiedenen Gesichtspunkten (meist unter Qualitätsverbesserungsaspekten) mit den Auswirkungen von Ärztatings auseinandersetzen, verwenden die Ratingresultate direkt. Dies würde jedoch in der vorliegenden Situation zu kurz greifen. Es existieren damit keine validierten Messinstrumente (kein Goldstandard) für die Situation in der Deutschweiz.
- Die objektiven Daten (d.h. Ratingresultate, Informationen über ermahnte bzw. in Regress genommene Ärzte) liegen bei *santésuisse* vor, sind jedoch aus Datenschutzgründen nicht zugänglich. Bei gegebener Kooperationsbereitschaft käme eine Datenübernahme von *santésuisse* auf Basis von Einverständniserklärungen eines grossen Teils der Ärzteschaft in Frage. Alternativ könnte dieser Teil der Daten direkt von den Ärzten erhoben werden, die bereit wären, an einer Studie teilzunehmen.
- Weiche Daten (z.B. subjektives Empfinden einer Bedrohung durch das Ärztating, subjektive Einschätzung einer Beeinflussung des eigenen Behandlungsverhaltens durch das Ärztating) könnten ebenfalls direkt von den Ärzten erhoben werden. Die Autoren gehen jedoch davon aus, dass solche weichen Indikatoren nicht geeignet sind, um die Exposition im Sinne der hier interessierenden wissenschaftlichen Fragestellung zu messen. Die Einflüsse, die das Antwortverhalten beeinflussen könnten, wären sehr vielfältig und die Richtung ihrer Auswirkungen wäre nicht einschätzbar.

Fazit Kernfrage 6: Zur Messung der Exposition müsste ein Indikator konstruiert werden und dessen Validität müsste im Rahmen einer Pilotstudie bzw. einer ersten Studienphase geprüft werden. Da jedoch kein Goldstandard vorhanden ist, wäre auch diese Prüfung nicht unproblematisch. (Detailvorschläge siehe Abschnitt 'Design einer retrospektiven Kohortenstudie'.)

Kernfrage 7: Kann das *outcome* adäquat gemessen werden?

Grundsätzlich kommen hier zwei Vorgehensweisen (einzeln oder kombiniert) in Frage, ein rein retrospektives, Krankenakten-basiertes Vorgehen oder eine Bestimmung des aktuellen Lipidstatus bei den eingeschlossenen Patienten.

Krankenakten-basiertes Vorgehen:

- In publizierten Studien wurden verschiedene Einzelindikatoren (durchgeführte Messungen, initiierte Behandlungen, erreichte Zielwerte) verwendet, um die Qualität der Hyperlipidämie-Behandlung zu messen [10].
- Im vorliegenden Fall wäre es vermutlich hilfreich für die Analyse, zusätzlich einen Indikator zu konstruieren, der die genannten Einzelindikatoren kombiniert. Gleichzeitig sollte dieser Indikator geeignet sein, Veränderungen der Behandlungsqualität über Zeit (konkret: eine Verschlechterung der Behandlungsqualität nach dem Eintreten einer 'Bedrohung' durch das Ärzterating) zu erfassen. Derartige kombinierte Indikatoren sind in der Literatur bisher nicht beschrieben.
- Bei einem retrospektiven, Krankenakten-basiertem Vorgehen ergeben sich immer Probleme der Datenverfügbarkeit. Verschiedene Studien haben Krankenakten einer Analyse von Mustern der Hyperlipidämie-Behandlung zugrunde gelegt, zumeist allerdings in den USA bzw. Kanada, wo die Datenverfügbarkeit tendenziell besser ist als in der Schweiz [6,10,11]. Da eine Vielzahl von Arztpraxen in die hier diskutierte Studie einbezogen würden, würden Unterschiede im Dokumentationsniveau unter Umständen zusätzliche Probleme aufwerfen.

Bestimmung des aktuellen Lipidstatus bei den eingeschlossenen Patienten:

- Vorteilhaft an dieser Lösung wäre, dass ein klarer Endpunkt zur Verfügung stünde.
- Es müsste sichergestellt werden, dass keine Beeinflussung der Messergebnisse durch die Studiensituation (konkret: durch eine kurzfristige Intervention des behandelnden Arztes) entstehen könnte. Wie dies zu realisieren wäre, ist unklar.
- Weitere Nachteile sind offensichtlich: Erhöhter Aufwand, zwingende Notwendigkeit eines *informed consent*, erhöhte Hürde bei der Ethikbeurteilung, Frage der Teilnahmebereitschaft (eventuell Notwendigkeit eines Anreizes).

Fazit Kernfrage 7: Eine Extraktion der notwendigen Daten aus den Krankenakten der teilnehmenden Ärzte wäre möglich, allerdings wäre mit Qualitätsunterschieden und Datenlücken zu rechnen. Die Konstruktion eines summarischen Indikators von Veränderungen der Behandlungsqualität über Zeit wäre möglich. Eine Prüfung der Machbarkeit und Validierung im Rahmen einer Pilotstudie wäre wünschenswert. Eine (zusätzliche oder alleinige) Bestimmung des aktuellen Lipidstatus bei allen ausgewählten Patienten wäre aus methodischer Sicht wünschenswert, aber mit erheblichen praktischen Problemen verbunden. Sie wäre nur sinnvoll, wenn eine hohe

Teilnahmebereitschaft der Patienten sichergestellt werden könnte. (Siehe auch Abschnitt 'Design einer retrospektiven Kohortenstudie'.)

Kernfrage 8: Können alle wichtigen potentiellen *confounders* adäquat gemessen werden?

Als potentielle *confounders* kommen unter anderem in Frage [9,12,13]:

- Auf der Patientenebene: Alter, Geschlecht, Morbidität, Bildungsstand, Zivilstand, Heimbewohnerschaft, Versicherungsdetails (Franchise, Zusatzversicherungen), *compliance* mit Behandlungsmassnahmen.

Diese *confounders* können durchweg gemessen. Allerdings sind einige der benötigten Angaben (*compliance*, fraglich Versicherungsdetails) nicht in den Krankenakten der Ärzte enthalten, so dass für eine vollständige Erfassung ein Einbezug der Patienten nötig wäre. (Eventuell ist die Nichterhebung einiger eher sekundärer Variablen vertretbar. Dies wäre zu prüfen.)

- Auf der Arzteebene: Alter, Geschlecht und berufliche Erfahrung des Arztes, Spezialisierung des Arztes, Arbeit im Rahmen einer Einzel- oder Gemeinschaftspraxis, Patientenzahl der Praxis, Struktur der Praxis im Hinblick auf die Franchisen und Zusatzversicherungen der Patienten, Schweregrad der Patientenkollektivs der Praxis, Belastung mit Notfällen, etc.

Einige dieser Variablen können nicht rasch und mit geringem Aufwand erhoben werden.

Speziell findet sich in der Literatur keine Möglichkeit, den Schweregrad des Patientenkollektivs einer Praxis rasch und mit geringem Aufwand zu messen. Möglich wären relativ aufwendige Vorgehensweisen, z.B. *Ambulatory Care Group*-Bestimmungen für alle Patienten der eingeschlossenen Praxen, oder die Verwendung von groben Indikatoren (z.B. Überweisungs- und Hospitalisationsraten [14]).

Fazit Kernfrage 8: Potentielle *confounders* auf der Patientenebene können entweder aus den Krankenakten der Ärzte oder direkt von den Patienten erhoben werden. Ob letzteres erforderlich ist, bleibt zu prüfen. Die Erhebung einiger sehr wichtiger *confounders* auf der Arzteebene ist problematisch bzw. mit erheblichem Aufwand verbunden.

Kernfrage 9: Erforderliche Fallzahl?

Die Fallzahl ist derzeit kaum zu schätzen, da sowohl die Häufigkeit der Exposition als auch das Ausmass des zu erwartenden Effekts (Unterschied zwischen den Patienten exponierter versus nicht exponierter Ärzte hinsichtlich der Qualität der Hyperlipidämie-Behandlung) unbekannt ist.

- **Beispiel 1:**

Annahme der Bestätigung einer korrekten Hyperlipidämie-Behandlung in 45% versus 50% der exponierten vs. nichtexponierten Patienten. Annahme, dass 10% der teilnehmenden Patienten exponiert sind. Fallzahl, um bei $\alpha = 0.05$ eine statistische Macht von 80% zu erreichen: etwa 9'000 Patienten.

- **Beispiel 2:**

Annahme der Bestätigung einer korrekten Hyperlipidämie-Behandlung in 45% versus 50% der exponierten vs. nichtexponierten Patienten. Annahme, dass 50% der teilnehmenden Patienten exponiert sind, was im Rahmen der retrospektiven Kohortenstudie erreichbar wäre, nicht jedoch in der Fall-Kontroll-Studie. Fallzahl, um bei $\alpha = 0.05$ eine statistische Macht von 80% zu erreichen: etwa 3'100 Patienten.

- **Zusätzliche Aspekte:**

- o Die gezeigten Fallzahlschätzungen setzen voraus, dass alle relevanten *confounders* adäquat gemessen werden.
- o Die Schätzungen setzen unabhängige Beobachtungen voraus, praktisch wird aber ein *clustering* innerhalb der Arztpraxen zu beobachten sein, was die erforderliche Fallzahl stark erhöhen kann [9]. Die Grösse dieses Effekts ist schwierig abzuschätzen.

Fazit Kernfrage 9: Wegen grosser Unsicherheiten im Bereich der nötigen Inputparameter ist eine solide Fallzahlschätzung kaum möglich. Die oben gezeigten Beispiele basieren auf rein hypothetischen Schätzungen des Effekts. Ein Design, das es erlaubt, die Zahl der exponierten Teilnehmer (Patienten) zu erhöhen (also das retrospektive Kohortendesign), wäre in jedem Fall vorzuziehen.

2. Designaspekte einer retrospektiven Kohortenstudie

Grundlegendes Design:

- Stufe 1: Gewinnung einer Zufallsstichprobe der in Frage kommenden Ärzte. Dabei Überrekrutierung in den verstärkt vom Ärzterating betroffenen Kantonen. Innerhalb jeder teilnehmenden Praxis Bestimmung des Expositionsstatus zu einem im der Vergangenheit liegenden Zeitpunkt (z.B. in den Jahren 2002-2003).
Alternativ könnte für diese Stufe eine Vollerhebung durchgeführt werden.
Alternativ könnte der Expositionsstatus im Einverständnis mit den teilnehmenden Ärzten direkt auf Basis der *santésuisse*-Daten ermittelt werden, was methodisch wegen geringerer Fehlerquellen vorzuziehen wäre.
- Stufe 2: Bestimmung von zwei gleich grossen Zufallsstichproben der exponierten und nicht exponierten Praxen. Auswahl von Zufallssamples der Patienten dieser Praxen, die

die Einschlusskriterien erfüllen. (Verbesserung der statistischen Macht durch dieses zweistufige Vorgehen.)

- Bestimmung des *outcome* zu einem Zeitpunkt nach Eintritt der Exposition (z.B. in den Jahren 2004-2005).
- Der Referenzzeitpunkt der Bestimmung der *confounders* wäre zu diskutieren (vor oder allenfalls während des Zeitpunkts der *outcome*-Messung, keinesfalls danach).

Pilot-Studie:

Es wäre notwendig, die Machbarkeit des Designs, die Teilnahmebereitschaft auf Ärzte- und Patientenebene sowie die Datenverfügbarkeit im Rahmen einer Pilotstudie zu überprüfen. Gleichzeitig müsste in diesem Rahmen die Eignung (Validität und Reliabilität) der gewählten Erhebungsinstrumente weitmöglichst überprüft werden. Ein Teil dieser Aufgaben könnte alternativ im Rahmen einer ersten Studienphase mit voller 'Fallzahl' auf der Ärzteebene adressiert werden (siehe oben, Grundlegendes Design Stufe 1).

Ethische Aspekte:

Ein Ethik-Votum wäre in jedem Fall erforderlich, ein *informed consent* der teilnehmenden Ärzte ebenfalls. Auf der Patientenebene sind zwei mögliche Szenarien zu unterscheiden.

- Bei einem ausschliesslich Krankenakten-basierten Vorgehen wäre die Datenextraktion eventuell ohne Einverständnis der betreffenden Patienten möglich, wenn nur anonyme Daten die Praxen verliessen. Kompliziert wird die Situation allerdings dadurch, dass die Datenextraktion möglichst nicht durch die Ärzte selbst, sondern durch neutrale, externe Personen (*study nurses* oder allenfalls geschulte MedizinstudentInnen) erfolgen sollte. Der letztendliche Entscheid könnte nur durch ein Ethikkomitee getroffen werden.
- Bei einer zusätzlichen Erhebung von Daten direkt von den Patienten bzw. im Falle einer Lipidbestimmung wäre mit Sicherheit ein individuelles Patienteneinverständnis erforderlich. Es wäre mit dem Ethik-Komitee abzuklären, wie genau der Studienzweck im *informed consent*-Formular beschrieben werden muss. Je genauer der Beschrieb, desto grösser wären möglicherweise die Verunsicherung der Patienten und resultierende Selektionseffekte.

Ein- und Ausschlusskriterien:

- Es wäre zu diskutieren, die Teilnahme auf solche Ärzte zu beschränken, die bereits über einen längeren Zeitraum in ihrer Praxis tätig sind und stabile Arzt-Patient-Beziehungen etabliert haben (weniger Varianz, weniger Störgrössen im Hinblick auf den vermuteten Effekt).
- Eine Möglichkeit, die Variabilität auf Patientenebene zu reduzieren, bestünde darin ausschliesslich eine relativ eng definierte Patientengruppe einzuschliessen, z.B. jüngere Patienten (40-65 Jahre?) mit nachgewiesener koronarer Herzkrankheit (angiographisch

oder durch durchgemachten Myokardinfarkt), aber ohne schwere sonstige Komorbiditäten.

Messung der Exposition:

- Möglichkeit 1: Entwicklung eines Score, der Informationen über die Ansässigkeit in einem exponierten Kanton und die Erfahrungen des einzelnen Arztes (Ratingresultate, bereits erfolgte Ermahnungen durch santésuisse, bereits erfolgte Konfrontation mit Regressforderungen) kombiniert.
Beurteilung: Problematisch und schwierig zu validieren, weil verschiedene objektive und subjektive Faktoren, deren Wechselwirkungen unklar sind, kombiniert werden müssten.
- Möglichkeit 2: Definition eines Arztes als exponiert, wenn er in (z.B.) 2002 oder 2003 ein Ratingresultat von mindestens (z.B.) 120% hatte und dieses in (z.B.) 2004-2005 um mindestens (z.B.) 10% reduzierte. Ein solches Mass könnte umschrieben werden als 'Anzeichen einer Reaktion des Arztes auf ein hohes Ratingresultat'. Es besitzt den Vorteil der Objektivierbarkeit. Die Validierung würde, in Abwesenheit eines Goldstandards, durch Vergleiche mit den potentiellen Einzelindikatoren einer Betroffenheit der Ärzte durch das Ärzterating erfolgen.
- Eventuell wäre es sinnvoll, anstelle einer binären Variable mehrere Graden der Betroffenheit durch das Ärzterating zu definieren. Dies wäre zu prüfen.
- Von der Verwendung subjektiver Aussagen zum 'Bedrohungsempfinden' wird, wie bereits ausgeführt, wegen Zweifeln an der Validität solcher Indikatoren abgeraten. (Eventuell Verwendung für Validierungszwecke.)

Messung des outcome:

Wenn bei den teilnehmenden Patienten eine aktuelle Lipidbestimmung durchgeführt wird, stellt die Erreichung der derzeit empfohlenen Zielwerte den primären Endpunkt dar.

Naheliegenderweise würden die von der Arbeitsgruppe Lipide (AGLA) der Schweizerischen Gesellschaft für Kardiologie vorgeschlagenen Zielwerte Verwendung finden (LDL < 2.6 mmol / l). Bei einem rein Krankenakten-basierten Vorgehen stellt sich die Situation komplexer dar:

- Möglichkeit 1: Verwendung der auch in der publizierten Literatur verwendeten Einzelendpunkte (unter anderem Anteil der Patienten, bei denen das Vorliegen einer Hyperlipidämie untersucht wurde; Anteil der Patienten mit bekannter Hyperlipidämie, die behandelt werden; Anteil der Patienten mit bekannter Hyperlipidämie, die die empfohlenen Zielwerte erreichen).
- Möglichkeit 2: Ergänzend Kombination der in der Literatur verwendeten Einzelendpunkte zu einem gemeinsamen Endpunkt, z.B. Anteil der Patienten, bei denen entweder
 - o der Lipidstatus bis (z.B.) 31.12.2003 unbekannt war, bei denen aber in (z.B.) 2004 oder 2005 einer Bestimmung erfolgte und bei Bedarf eine Behandlung initiiert und zu einem angemessenen Zeitpunkt nach der Initiierung kontrolliert wurde, oder

- o zum (z.B.) 31.12.2003 eine behandlungsbedürftige Hyperlipidämie bekannt war und bei denen in 2004 oder 2005 eine Behandlung erfolgte, in angemessenen Abständen kontrolliert wurde und die empfohlenen Zielwerte erreicht wurden.

Dieser Endpunkt könnte in etwa bezeichnet werden als 'Bestätigung korrekter Lipidwerte oder korrigierende Aktivität in (z.B.) 2004 oder 2005'.

Dieses Konzept sollte noch dahingehend ausgearbeitet werden, dass auch der Aspekt einer Veränderung der Behandlungsqualität im Zeitverlauf (z.B. nach dem 31.12.2003 gegenüber vor dem 31.12.2003; eventuell längere Zeitreihen) Berücksichtigung findet.

Messung der potentiellen *confounders*:

- Erhebung der potentiellen *confounders* auf der Patientenebene aus den Krankenakten der Ärzte, eventuell direkt von den Patienten (Bildungsstand, Versicherungsdetails wie Franchise etc.).
- Erhebung der potentiellen *confounders* auf der Arzzebene direkt von den Ärzten bzw. aus den Krankenakten. Teilweise problematisch bzw. mit erheblichem Aufwand verbunden. Ob der Einsatz von *proxy measures* adäquat wäre, bliebe zu prüfen (z.B. Anteil der Patienten, die im Referenzzeitraum eine Hospitalisation hatten, unter Nichtberücksichtigung der Unfälle, als grober Schätzer des Schweregrads der verschiedenen Patientenkollektive - siehe hierzu Roos *et al.* 1995 [14]).
- Subjektive Beurteilungen der Ärzte bezüglich der (arztbezogenen, patientenbezogenen, gesundheitssystembezogenen) Ursachen des Nichterreichens von Lipidzielen sollten zusätzlich erfasst werden, dürften aber als weiche Daten von limitierter Validität und damit für die Hauptanalyse nur fraglich relevant sein.
- Subjektive Beurteilungen der Ärzte bezüglich Ausweichstrategien (z.B. Überweisung an Spezialisten oder Klinikambulanzen) sollten ebenfalls erfasst werden.

Primäre Analysen:

Bestimmung des Zusammenhangs zwischen Exposition und *outcome*, dabei Kontrolle für die potentiellen *confounders*, vermutlich durch logistische Regression. Eventuell ergänzend *multilevel modelling*.

Sekundäre Analysen:

Eine direkte Aussage zur 'Epidemiologie' des Ärztteratings in der Deutschschweiz wäre auf Basis der in Stufe 1 gewonnenen Daten möglich. Je nach Gestaltung der Datenerhebung könnten weiche Daten (subjektives Bedrohungsempfinden der Ärzte, subjektive Arbeitsbelastung oder Belastung durch Notfälle etc.) miterfasst werden, die in Grenzen auch für die Hauptfragestellung von Interesse wären: Obwohl nicht direkt zur Bestimmung der Exposition geeignet, könnten sie für Validierungszwecke hilfreich sein.

Weitere Aspekte:

- Um die spätere Glaubwürdigkeit der Studie sicherzustellen, wäre es notwendig ein *steering board* oder *advisory committee* zu installieren, in dem alle Interessengruppen einschliesslich der Krankenversicherer oder, falls dies nicht möglich ist, mehrere neutrale Wissenschaftler vertreten wären. Die Hauptaufgabe dieses Gremiums wäre, strittige Fragen zum Vorgehen, die Auswirkungen auf das Studienresultat haben könnten, breit abzusichern.
- Der Aufwand für die teilnehmenden Ärzte und Patienten müsste möglichst gering gehalten werden. Wo diese Vorgabe nicht voll erfüllt werden könnte, wäre eine Schaffung von Teilnahmeanreizen (z.B. in Form von Aufwandsentschädigungen) zu diskutieren. Solche Anreize müssten vom zuständigen Ethikkomitee genehmigt werden.

3. Zusammenfassung und Beurteilung

- Das Vorhandensein des vermuteten Zusammenhangs erscheint grundsätzlich plausibel. Es finden sich jedoch keine entsprechenden Hinweise in der Literatur. Eine Erforschung fand bisher nicht statt.
- Das Ausmass des vermuteten Effekts und damit seine praktische Bedeutung sind derzeit nicht einschätzbar.
- Die Messung der Exposition (Betroffenheit der Ärzte durch das Ärztering) erscheint problematisch, da eine Vielzahl objektiver und subjektiver Faktoren in nicht bekannter Weise interagieren. Ein Indikator müsste konstruiert und dessen Validität müsste im Rahmen einer Pilotstudie geprüft werden. Da jedoch kein Goldstandard vorhanden ist, wäre auch diese Prüfung nicht unproblematisch.
- Die Messung des *outcome* und der *confounders* erscheint möglich, wäre aber mit erheblichem Aufwand verbunden. Praktische Probleme stellen sich unter anderem auf der Ebene der Datenverfügbarkeit aus den ärztlichen Krankenakten (Unvollständigkeit) beziehungsweise auf der Ebene der Teilnahmebereitschaft der Patienten (wenn diese direkt involviert oder gar für eine Blutabnahme aufgeboten werden sollen). Ethische Fragestellungen wären in beiden Fällen abzuklären.
- Als Studiendesign wäre am ehesten die Form einer retrospektiven Kohortenstudie geeignet. Auch bei Wahl dieses Designs blieben allerdings die oben genannten Probleme bestehen. Weiterhin könnte sowohl auf der Ärzteebene als auch auf der Patientenebene erheblicher Selektionsbias entstehen.
- Für keines der diskutierten Studiendesigns ist derzeit eine aussagekräftige Fallzahlschätzung möglich, da zentrale Grössen unbekannt sind.
- In der Summe verbleiben sehr erhebliche Unsicherheiten, ob der vermutete Zusammenhang wissenschaftlich valide nachgewiesen werden könnte. Die Autoren raten

deshalb von der Studiendurchführung eher ab, bei Kosten, die sicherlich in sechsstelliger Höhe lägen. Im Falle der Durchführung sollte eine adäquate Pilot-Phase vorgesehen werden. Es sollten *go / no go*-Entscheide auch noch zu späteren Zeitpunkten möglich sein (z.B. wenn Ethikentscheide einer optimalen Studienanlage entgegenstünden oder wenn sich eine unzureichende Teilnahmebereitschaft der Ärzteschaft ergäbe).

Alternative Vorgehensweisen bzw. Studiendesigns

Möglichkeit 1:

Bei ähnlichem Studienaufbau könnte ein andere 'Indikatorkrankheit' (z.B. Hypertonie [15]) anstelle der Hyperlipidämie gewählt werden.

Beurteilung: Es finden sich keine deutlichen Hinweise in der Literatur, dass dies vorteilhaft wäre. Ebenfalls sind in der Literatur keine einfachen Indikatoren beschrieben, anhand derer eine (milde) Unterbehandlung über verschiedene Krankheitsbilder hinweg ermittelt werden könnte. Die meisten der beschriebenen Probleme liessen sich durch eine Änderung auf dieser Ebene nicht beheben. Der Indikator Hyperlipidämie scheint damit gut gewählt.

Möglichkeit 2:

Es könnte eine rein deskriptive Studie zur 'Epidemiologie' des Ärzteratings in der Schweiz durchgeführt werden, auf Basis einer Zufallsstichprobe der in Frage kommenden Ärzte oder im Rahmen einer Vollerhebung. Zu erfragen wären Arzt- und Praxischarakteristika sowie die vermuteten subjektiven und objektiven Komponenten und Indikatoren der Auswirkungen des Ärzteratings in der Schweiz (Resultate des Ärzteratings, bereits erfolgte Ermahnungen durch santésuisse, bereits erfolgte Konfrontation mit Regressforderungen, subjektives Gefühl der Betroffenheit, subjektive Einschätzung eigener Reaktionen und Verhaltensänderungen). Vorzugsweise wären ergänzend zu den Angaben der Ärzte die Daten der santésuisse direkt zu verwenden.

In der wissenschaftlichen Literatur gibt es ein Beispiel für eine ähnliche Studie [16].

Beurteilung: Nach Auffassung der Autoren wäre eine derartige Studie aufgrund ihrer Anlage nicht geeignet, einen negativen Einfluss des Ärzteratings auf die Qualität der Patientenbehandlung zu belegen.

Die erhobenen objektiven Daten wären geeignet, einen Überblick über den aktuellen Stand des Ärzteratings zu geben. Unter Umständen könnte auch eine longitudinale Komponente implementiert werden (Erfragen von Verläufen über Zeit).

Die erhobenen subjektiven Daten wären ihrem Charakter nach eher von publizistischem denn von wissenschaftlichen Interesse. (Sie könnten allerdings für die Validierung des auf S. 15 angesprochenen Indikators einer 'Reaktion des Arztes auf ein hohes Ratingresultat' von Nutzen sein.)

Möglichkeit 3:

Die als Möglichkeit 2 beschriebene Studie könnte so durchgeführt werden, dass gleichzeitig alle Stufe 1-Elemente des oben beschriebenen retrospektiven Kohortendesigns (siehe S. 13) enthalten sind.

Beurteilung: Die Anforderungen an die Wissenschaftlichkeit und der Aufwand wären im Vergleich zu Möglichkeit 2 erhöht.

Die Nutzenseite, es wäre zusätzlich möglich, einen Teil der in Bezug auf die Kohortenstudie offenen Fragen zu klären und einen Teil der diesbezüglich anstehenden Aufgaben abzuarbeiten:

- Ermittlung der Teilnahmebereitschaft der Ärzteschaft, damit teilweise Abschätzung des Risikos von Selektionsbias auf dieser Ebene.
- Gewinnung der Daten zur Exposition und Validierung des auf S. 15 vorgeschlagenen Indikators, soweit ohne Verfügbarkeit eines Goldstandards möglich.
- Auf dieser Basis fundierterer Entscheid bezüglich der Durchführung des Lipidteils.

Verbleibende und neue Probleme:

- Fundamentale Probleme des Kohortendesigns blieben bestehen. Insbesondere bliebe das Ausmass des vermuteten Effekts uneinschätzbar, mit allen direkten und indirekten Implikationen. Es wäre weiterhin keine solide Fallzahlschätzung möglich. Alle Probleme auf der Ebene der Messung des *outcome* und der *confounders* blieben ebenfalls ungelöst.
- Es würde naheliegen, die teilnehmenden Ärzte hinsichtlich ihrer Teilnahmebereitschaft an einer Folgestudie (d.h. der Lipidstudie) zu befragen. Bereits dies könnte jedoch das Behandlungsverhalten ändern und Verzerrungen induzieren.

Möglichkeit 4:

Als weitere Alternative wäre eine Studie denkbar, die eine Neuberechnung der Resultate des Ärzt ratings durchführt, unter Berücksichtigung der Erfordernisse eines adäquaten *casemix adjustment*. Die Resultate würden dann mit den primären Ratingresultaten verglichen. In diesem Zusammenhang könnte analysiert werden, ob Ärzte, die sowohl vor als auch nach der Durchführung des *casemix adjustment* ein hohes Ratingresultat aufwiesen, eher in der Lage waren, ihre Kosten zu senken, als diejenigen, deren hohes Resultat primär durch den *casemix* bedingt war.

Bei denjenigen Ärzten, die primär ein hohes, nach dem *casemix adjustment* aber ein niedriges Ratingresultat aufwiesen, wäre von einem 'schlechten' *casemix* auszugehen, also von einem Patientenkollektiv mit überdurchschnittlicher Morbidität. In einem zweiten Schritt könnte die weiterführende Frage adressiert werden, ob bei diesen Ärzten unter dem Einfluss des Ärzt ratings eine Verschlechterung der Behandlungsqualität (z.B. im Bereich der Hyperlipidämie) erfolgte. (Diese Hypothese wäre eine Verfeinerung der Hypothese der oben diskutierten retrospektiven Kohortenstudie.)

Beurteilung: Deutliche Diskrepanzen zwischen den direkten Resultaten des Ärzt ratings der santésuisse und den korrigierten Resultaten, die sich aus einem adäquaten *casemix*

adjustment ergäben, wären nach der Literaturlage zu erwarten [17]. Eine wichtige Argumentationshilfe wäre damit verfügbar.

Nachteile:

- Im ersten Schritt würde das hier angedachte Studiendesign keinen formalen Beleg eines Zusammenhangs zwischen dem aktuell in der Deutschschweiz praktizierten Ärzterating und der Qualität der Patientenbehandlung liefern.
- Der Aufwand für die Auswahl, die Adaptation und den Einsatz eines geeigneten Patientenklassifikationssystems (z.B. Ambulatory Care Groups, kurz ACG [18]) wäre vermutlich erheblich.

Durchgeführte Literaturrecherchen und Resultate

In Zusammenhang mit der Frage der Machbarkeit der Untersuchung eines Zusammenhangs zwischen dem Ärzterating der santésuisse und dem Erreichen von Zielen der Hyperlipidämie-Behandlung in der Grundversorgung waren eine Reihe von Literaturrecherchen erforderlich. Deren Resultate sind für die Gesamtbeurteilung des in der Deutschschweiz durchgeführten Ärzteratings von Bedeutung und werden hier kurz vorgestellt.

Die Durchführung der Recherchen erfolgte in Medline unter dem Primat einer effizienten Nutzung begrenzter Ressourcen, weshalb die Suchstrategien keinen Anspruch auf methodische Rigorosität und die Resultate keinen Anspruch auf Vollständigkeit erheben. Alle identifizierten Forschungsarbeiten mit relevanter Fragestellung sind im Folgenden berücksichtigt. Es erfolgte keine Selektion nach den rapportierten Forschungsergebnissen.

Fragestellung 1: Gibt es publizierte Studien zu den Auswirkungen von Ärzteratings? Wie wurden die Betroffenheit von Ärzten durch Ärzteratings (verstanden als Exposition) gemessen?

- Verwendete Suchkriterien: '*physician profiling*' AND '*primary care*'; '*physician profiling*' AND '*economics*'; automatisierte Suche nach ähnlichen Artikeln.
- Hauptresultate:
 - o Es gibt eine Reihe von Studien zu den Auswirkungen von Ärzteratings. Die untersuchten Ratings wurden jedoch in fast allen Fällen mit dem expliziten Ziel einer Qualitätsbeurteilung und Qualitätsverbesserung (anstatt Kostenbeurteilung und Kostenersparnis) durchgeführt [12,19-22].
 - o Die Ratingresultate waren in allen Fällen direkt verfügbar und wurden als Expositionsvariable verwendet. Das Thema eines Einflusses unterschiedlicher 'Bedrohungsgrade' etc. stellte sich nicht.

Fragestellung 2: Wie wird die Qualität der Hyperlipidämie-Behandlung in der Literatur gemessen? Welche Resultate wurden ermittelt?

- Verwendete Suchkriterien: Diverse.
- Hauptresultate:
 - o Eine Unterbehandlung der Hyperlipidämie, gemessen an den jeweils gültigen Guidelines, wurde für diverse Länder konstatiert [3-6].
 - o Zur Beurteilung wurden verschiedene Endpunkte herangezogen (Anteil der Patienten, bei denen das Vorliegen einer Hyperlipidämie untersucht wurde; Anteil der

Patienten mit bekannter Hyperlipidämie, die behandelt wurden; Anteil der Patienten mit bekannter Hyperlipidämie, die die empfohlenen Zielwerte erreichten) [6,10,11,23]. Diese Einzelendpunkte wurden jedoch in keiner der ermittelten Arbeiten zu einem Gesamtscore zusammengefasst.

- o Krankenakten wurden für die Gewinnung der erforderlichen Daten verwendet [6,10,11].
- o Die Primärarztpraxis wurde als idealer 'Ort' der Sekundärprophylaxe der koronaren Herzkrankheit (inklusive Behandlung der Hyperlipidämie) beschrieben und Untersuchungen zur strukturellen Optimierung der Behandlungsqualität werden aktuell durchgeführt [23].

Fragestellung 3: Wie kann generell eine (milde) Unterbehandlung der Patienten einer Praxis gemessen werden?

- Verwendete Suchkriterien: 'under-treatment' AND '(indicator OR indicators)'; 'under-treatment' AND 'primary care'; automatisierte Suche nach ähnlichen Artikeln.
- Hauptresultate:
 - o Es wurden keine einfachen, krankheitsübergreifenden Indikatoren einer (milden) Unterversorgung in der ärztlichen Primärversorgung identifiziert.
 - o Eine kanadische Studie zu geographischen Unterschieden der Qualität der Primärversorgung wurde am Beispiel der Hypertonie durchgeführt [15].

Fragestellung 4: Wie kann der Schweregrad der Patienten einer Praxis (*casemix*) gemessen werden?

- Verwendete Suchkriterien: Diverse.
- Hauptresultat: Es existieren verschiedene Instrumente, jedoch keines, das geeignet wäre, den *casemix* einer Praxis rasch abzuschätzen (d.h. ohne jeden einzelnen Patienten klassifizieren zu müssen) [14,18,24,25].

Im vorliegenden Sachzusammenhang relevante Aussagen der gefundenen Arbeiten:

- Kosteneffekte von Ärzt ratings wurden beobachtet [26-28].
- Es wurde festgestellt, dass Ärzt ratings [ohne adäquates *case-mix adjustment*] Qualitätseffekte haben bzw. zur Abschiebung von schwer kranken Patienten führen. In einem Fall wurde die subjektive Perzeption der Ärzte erfragt [16], im anderen Fall handelte es sich eher um eine Feststellung als um eine Analyse [29].
- Ein korrektes *casemix adjustment* ist für ein aussagekräftiges Ärzt rating notwendig [14,22,26,30-36].

- Ein korrektes *casemix adjustment* ist für die Beurteilung der Behandlungsqualität notwendig [9,12,13,17,19-22,36].
- Für die Messung des *casemix* in der Primärversorgung wurden verschiedene Instrumente entwickelt: *Ambulatory Care Groups (ACG)* [17,18,37]; Duke Case-Mix System [24]; von der Art der verschriebenen Medikamente ausgehende Systeme [25,33]. Eine Studie kombinierte die Zahl der Überweisungen und Hospitalisationen mit Komorbiditätsdaten [14,38].
- Eine Studie weist auf das Phänomen des *physician clustering* (Ansammlung ähnlicher Patienten beim selben Arzt) hin und betont, dass dieser Effekt beim Ärzterating berücksichtigt werden muss [9].
- Die Aussage, dass es wichtig ist, die Ratings *durch* Ärzte durchzuführen, findet sich nur ein einmal [39].

Literaturangaben

1. Amstutz R. Wirtschaftlichkeitsverfahren: Das Ziel heisst Verhaltensänderung. *infosantésuisse* 2005, 3.
2. Kraft P. Wirtschaftlichkeitsverfahren bauen auf verlässliche Grundlagen. *infosantésuisse* 2005, 4-5.
3. Mantel-Teeuwisse AK, Verschuren WM, Klungel OH, de Boer A, Kromhout D. Recent trends in (under)treatment of hypercholesterolaemia in the Netherlands. *Br J Clin Pharmacol* 2004, 58, 310-6.
4. Garcia Ruiz FJ, Marin Ibanez A, Perez-Jimenez F, et al. Current lipid management and low cholesterol goal attainment in common daily practice in Spain. The REALITY Study. *Pharmacoeconomics* 2004, 22 Suppl 3, 1-12.
5. Ruof J, Klein G, Marz W, Wollschlager H, Neiss A, Wehling M. Lipid-lowering medication for secondary prevention of coronary heart disease in a German outpatient population: the gap between treatment guidelines and real life treatment patterns. *Prev Med* 2002, 35, 48-53.
6. Eaton CB, Monroe A, McQuade W, Eimer MJ. Cholesterol testing and management: a national comparison of family physicians, general internists, and cardiologists. *J Am Board Fam Pract* 1998, 11, 180-6.
7. Autier P, Creplet J, Vansant G, et al. The impact of reimbursement criteria on the appropriateness of 'statin' prescribing. *Eur J Cardiovasc Prev Rehabil* 2003, 10, 456-62.
8. Harnick DJ, Cohen JL, Schechter CB, Fuster V, Smith DA. Effects of practice setting on quality of lipid-lowering management in patients with coronary artery disease. *Am J Cardiol* 1998, 81, 1416-20.
9. Greenfield S, Kaplan SH, Kahn R, Ninomiya J, Griffith JL. Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results. *Ann Intern Med* 2002, 136, 111-21.
10. Camenga D, Gill JM. Treatment of hypercholesterolemia in patients with diabetes at a family practice office. *Del Med J* 2003, 75, 415-20.
11. Alzahrani T, Marrat S, Haider A. Management of dyslipidemia in primary care. *Can J Cardiol* 2003, 19, 1499-502.
12. Katon W, Rutter CM, Lin E, et al. Are there detectable differences in quality of care or outcome of depression across primary care providers? *Med Care* 2000, 38, 552-61.
13. Fiscella K, Franks P. Influence of patient education on profiles of physician practices. *Ann Intern Med* 1999, 131, 745-51.
14. Roos NP, Carriere KC, Friesen D. Factors influencing the frequency of visits by hypertensive patients to primary care physicians in Winnipeg. *Cmaj* 1998, 159, 777-83.
15. Veugelers PJ, Yip AM, Burge F. Inequalities in health and health services delivery: a multilevel study of primary care and hypertension control. *Chronic Dis Can* 2004, 25, 101-7.
16. Stoddard JJ, Reed M, Hadley J. Financial incentives and physicians' perceptions of conflict of interest and ability to arrange medically necessary services. *J Ambul Care Manage* 2003, 26, 39-50.

17. Tucker AM, Weiner JP, Honigfeld S, Parton RA. Profiling primary care physician resource use: examining the application of case mix adjustment. *J Ambul Care Manage* 1996, 19, 60-80.
18. Greene BR, Barlow J, Newman C. Ambulatory care groups and the profiling of primary care physician resource use: examining the application of case mix adjustments. *J Ambul Care Manage* 1996, 19, 86-9.
19. Garnick DW, Fowles J, Lawthers AG, Weiner JP, Parente ST, Palmer RH. Focus on quality: profiling physicians' practice patterns. *J Ambul Care Manage* 1994, 17, 44-75.
20. Goldfield N. Case mix, risk adjustment, reinsurance, and health reform. *Manag Care Q* 1994, 2, iv-viii.
21. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *Jama* 1999, 281, 2098-105.
22. Tucker JL, 3rd. The theory and methodology of provider profiling. *Int J Health Care Qual Assur Inc Leadersh Health Serv* 2000, 13, 316-21.
23. Murphy AW, Cupples ME, Smith SM, Byrne M, Leathem C, Byrne MC. The SPHERE Study. Secondary prevention of heart disease in general practice: a randomised controlled trial of tailored practice and patient care plans with parallel qualitative, economic and policy analyses. [ISRCTN24081411]. *Curr Control Trials Cardiovasc Med* 2005, 6, 11.
24. Parkerson GR, Jr., Michener JL, Yarnall KS, Hammond WE. Duke Case-Mix System (DUMIX) for ambulatory health care. *J Clin Epidemiol* 1997, 50, 1385-94.
25. Sloan KL, Sales AE, Liu CF, et al. Construction and characteristics of the RxRisk-V: a VA-adapted pharmacy-based case-mix instrument. *Med Care* 2003, 41, 761-74.
26. Rich EC, Kralewski J, Feldman R, Dowd B, Bernhardt TS. Variations in the management of primary care: effect on cost in an HMO network. *Arch Intern Med* 1998, 158, 2363-71.
27. Roblin DW. Applications of physician profiling in the management of primary care panels. *J Ambul Care Manage* 1996, 19, 59-74.
28. Ramsey R. IPA cites physician profiling as key capitation strategy, expands effort to specialists. *Capitation Manag Rep* 2003, 10, 39-41.
29. Reid RA. Using the tail to wag the dog: a consumer's perspective on management of the cost of cancer. *Cancer* 1998, 82, 2022-5.
30. Cowper PA, Peterson ED, DeLong ER, et al. The impact of statistical adjustment on economic profiles of interventional cardiologists. *J Am Coll Cardiol* 2001, 38, 1416-23.
31. Tufano JT, Conrad DA, Liang SY. Addressing physician compensation and practice productivity. *J Ambul Care Manage* 1999, 22, 47-57.
32. Nickerson C, Rutledge RW. A methodology for choosing a physician profiling system: the case of First Option Health Plan. *J Health Care Finance* 1999, 26, 5-13.
33. Roblin DW. Physician profiling using outpatient pharmacy data as a source for case mix measurement and risk adjustment. *J Ambul Care Manage* 1998, 21, 68-84.
34. Welch HG, Miller ME, Welch WP. Physician profiling. An analysis of inpatient practice patterns in Florida and Oregon. *N Engl J Med* 1994, 330, 607-12.
35. Salem-Schatz S, Moore G, Rucker M, Pearson SD. The case for case-mix adjustment in practice profiling. When good apples look bad. *Jama* 1994, 272, 871-4.
36. Chang W, McCracken SB. Applying case mix adjustment in profiling primary care physician performance. *J Health Care Finance* 1996, 22, 1-9.

37. Carlsson L, Borjesson U, Edgren L. Patient based 'burden-of-illness' in Swedish primary health care. Applying the Johns Hopkins ACG case-mix system in a retrospective study of electronic patient records. *Int J Health Plann Manage* 2002, 17, 269-82.
38. Pope GC. Medical conditions, health status, and health services utilization. *Health Serv Res* 1988, 22, 857-77.
39. Tompkins CP, Bhalotra S, Garnick DW, Chilingerian JA. Physician profiling in group practices. *J Ambul Care Manage* 1996, 19, 28-39.

Literaturangaben - Abstracts (alphabetisch nach Autoren)

Alzahrani, T., S. Marrat, et al. (2003). "Management of dyslipidemia in primary care." *Can J Cardiol* 19(13): 1499-502.

BACKGROUND: Cardiovascular disease remains the leading cause of mortality in Canada. The link between hyperlipidemia and coronary heart disease has been clearly established. There is overwhelming evidence for reductions in coronary events and cardiovascular mortality with lowering of low-density lipoprotein cholesterol (LDL-C). Despite the evidence, hyperlipidemia treatment remains suboptimal. **OBJECTIVE:** To evaluate compliance with published dyslipidemia guidelines in a primary care setting. The primary outcome measure was target LDL-C level. **METHODS:** Retrospective chart review of a random selection of 300 patients diagnosed with hyperlipidemia in a large academic family medicine clinic. The primary outcome measure was a target LDL-C level of less than 2.5 mmol/L for patients with diabetes or coronary heart disease. For patients without diabetes or coronary heart disease, Framingham risk assessment tables were used to determine ideal target LDL-C levels. **RESULTS:** Overall, 53% of patients achieved target LDL-C. Target LDL-C levels were achieved in 48% of patients with diabetes or coronary heart disease. Males were twice as likely to be prescribed lipid lowering therapy than females. Males on lipid lowering therapy were twice as likely as females on lipid lowering therapy to achieve target LDL-C levels. Males with diabetes or coronary heart disease were twice as likely as females with diabetes or coronary heart disease to achieve target LDL-C levels. Only 44% of patients with diabetes or coronary heart disease were prescribed lipid lowering therapy. **CONCLUSION:** Results from an academic family medicine clinic indicate suboptimal compliance with current dyslipidemia management guidelines. Primary care physicians need to continue to take an aggressive stance on lipid lowering strategies, especially in high-risk patients and females.

Amstutz, R. (2005). "Wirtschaftlichkeitsverfahren: Das Ziel heisst Verhaltensänderung." *infosantésuisse*(1-2): 3.

Autier, P., J. Creplet, et al. (2003). "The impact of reimbursement criteria on the appropriateness of 'statin' prescribing." *Eur J Cardiovasc Prev Rehabil* 10(6): 456-62. **BACKGROUND:** In Belgium, regulations restrict the reimbursement of statins to patients with total serum cholesterol above 250 mg/dl (6.41 mmol/l) after a three-month lipid-lowering diet. We investigated the possible impact of these regulations on characteristics of Belgian patients receiving a lipid-lowering drug. **DESIGN:** From 1998 to 2000, standard questionnaires on coronary risk factors and treatments with lipid-lowering drugs were completed by 301 GPs sampled at random in the 11 Belgian Provinces. Questionnaires had to be completed for 18 consecutive patients 35 years old or more attending GPs' practices, irrespective of the underlying motive for attendance. **RESULTS:** Of the 5511 patients included in the study, 1519 (28%) had established coronary disease or diabetes mellitus, or ≥ 2 non-cholesterol coronary risk factors. Most (70%) of these patients were not treated with a lipid-lowering drug. Only 22% of patients with established coronary disease, 10% of patients with diabetes mellitus and 9% of patients with ≥ 2 coronary risk factors were treated with a statin. Fifty-nine percent of fibrate users and 50% of statin users had a pre-treatment cholesterol level above 250 mg/dl, but had no or only one non-cholesterol coronary risk factor. **CONCLUSIONS:** In Belgium, the majority of patients at higher risk of coronary event do not benefit from lipid-lowering drugs, particularly the statins. In contrast, one of two statin users and three of five fibrate users should probably not receive the drug prescribed. Regulation based on blood cholesterol level

encourages the overlooking of other risk factors relevant for selecting patients having the greatest chance to benefit from statin treatment.

Camenga, D. and J. M. Gill (2003). "Treatment of hypercholesterolemia in patients with diabetes at a family practice office." *Del Med J* 75(11): 415-20.

BACKGROUND: Patients with diabetes have an increased risk of developing coronary heart disease (CHD) and therefore need optimal management of hypercholesterolemia. Previous studies have shown that management of these patients remains suboptimal. **OBJECTIVE:** Our study aimed to determine the extent to which patients with diabetes in a family practice setting in Delaware are being managed according to national guidelines for the management of hypercholesterolemia. **METHODS:** Using a retrospective medical chart review, we identified all active patients with diabetes in a family practice office. We determined the percentage of patients who had a low-density lipoprotein (LDL) documented in their chart and the percentage of patients who received LDL screening within the past year. We then stratified the sample into three different categories of LDL level (< 100 mg/dl, 100-129 mg/dl, and > or = 130 mg/dl) and determined the percentage of patients who were on lipid-lowering therapy within each stratum. **RESULTS:** According to the medical record review, 80.30% of the active patients with diabetes had an LDL recorded in their chart. Of the patients with diabetes, 57.99% had an LDL measurement taken within the past year. Of this group, 41.03% had an LDL < 100 mg/dl, 31.41% had an LDL between 100 and 129 mg/dl, and 25.00% had an LDL > or = 130 mg/dl. Of the total number of patients with an LDL < 100 mg/dl, 53.93% were on lipid-lowering agents, while 38.18% of the patients with an LDL > or = 130 mg/dl were on lipid-lowering agents. Of the patients with an LDL measurement taken in the past year, 67.19% of those with an LDL < 100 mg/dl were on lipid-lowering medications, while 43.59% of patients with an LDL > or = 130 mg/dl were on lipid-lowering medications. **CONCLUSIONS:** In a Delaware family practice office, the screening rate for hyperlipidemia in patients with diabetes is similar to that found in other studies, but still leaves room for improvement. Lipid control rates are somewhat better than the rates reported in previous studies, but they are still not optimal. While many of the patients requiring lipid-lowering therapy are on these medications, a substantial minority has an LDL of < 100 mg/dl without being on medications. Although these findings may reflect patient preference to defer medications, it suggests that there is room for improvement with medical therapy.

Carlsson, L., U. Borjesson, et al. (2002). "Patient based 'burden-of-illness' in Swedish primary health care. Applying the Johns Hopkins ACG case-mix system in a retrospective study of electronic patient records." *Int J Health Plann Manage* 17(3): 269-82.

Patients from one municipality in Sweden utilizing primary health care (PHC) during 1998 and 1999 have been categorized into 81 groups. The groups show each patient's own case-mix in terms of illness. Grouping was carried out using the case-mix instrument adjusted clinical groups (ACG), developed by the School of Hygiene and Public Health at Johns Hopkins University in Baltimore, USA. The resulting pattern provided a more adequate reflection of the scope of primary care's task than that yielded by diagnoses alone. Changes over time in terms of illness patterns for a population could be described, analysed and assessed from medical and health economic perspectives. One of the conclusions from this study was that the ACG instrument is a relevant tool in describing the outcome of work by the primary health care centre. The ACG is of interest in the improvement of the quality of primary care in Sweden. The ACG should be a driving force in the development of health indices in both national and international comparisons, as a result of its focus on the health status of patients and populations instead of on diagnoses and diseases.

- Chang, W. and S. B. McCracken (1996). "Applying case mix adjustment in profiling primary care physician performance." *J Health Care Finance* 22(4): 1-9.
This project was conducted with two objectives: developing a simple and meaningful performance efficiency index for profiling primary care physician (PCP) practice patterns and testing the explanatory power of the ambulatory care group case mix system in controlling for variation in patients' resource needs. The findings indicate that large administrative databases can provide valuable information for health services management. However, effort should be made to focus on methodological issues in analyzing those data sources. Conventional demographic risk adjustment factors (age and sex) are not adequate. Greater priority needs to be given to developing and applying additional risk adjustment systems to maximize the value of large administrative databases in profiling the relative efficiency of PCPs.
- Cohen, M. M. and L. MacWilliam (1995). "Measuring the health of the population." *Med Care* 33(12 Suppl): DS21-42.
A set of 102 population-based indicators was developed from multiple administrative data sources; these indicators were used to compare the health status of 1 million Manitoba residents across eight administrative regions for 1 year. Marked variations in health status were shown. Despite theoretically equal access to care in a universally insured system and high rates of utilization of hospital and physician services, residents of Manitoba's two northern, more remote regions scored most poorly--consistently and with statistical significance--across a variety of health status indicators. The strength of the various indicators was evaluated, and premature mortality emerged as the most useful "flagship" indicator for future analyses. Indicators that purport to be sensitive to how well a health care system is performing showed patterns similar to those derived from more classic measures (eg, mortality, low birth weight). Furthermore, the "system sensitive indicators" did not appear to be sufficiently independent of utilization biases.
- Cowper, P. A., E. D. Peterson, et al. (2001). "The impact of statistical adjustment on economic profiles of interventional cardiologists." *J Am Coll Cardiol* 38(5): 1416-23.
OBJECTIVES: The objective of this study was to identify preprocedure patient factors associated with percutaneous intervention costs and to examine the impact of these patient factors on economic profiles of interventional cardiologists. BACKGROUND: There is increasing demand for information about comparative resource use patterns of interventional cardiologists. Economic provider profiles, however, often fail to account for patient characteristics. METHODS: Data were obtained from Duke Medical Center cost and clinical information systems for 1,949 procedures performed by 13 providers between July 1, 1997, and December 31, 1998. Patient factors that influenced cost were identified using multiple regression analysis. After assessing interprovider variation in unadjusted cost, mixed linear models were used to examine how much cost variability was associated with the provider when patient characteristics were taken into account. RESULTS: Total hospital costs averaged \$15,643 (median, \$13,809), \$6,515 of which represented catheterization laboratory costs. Disease severity, acuity, comorbid illness and lesion type influenced total costs ($R^2 = 38\%$), whereas catheterization costs were affected by lesion type and acuity ($R^2 = 32\%$). Patient characteristics varied significantly among providers. Unadjusted total costs were weakly associated with provider, and this association disappeared after accounting for patient factors. The provider influence on catheterization costs persisted after adjusting for patient characteristics. Furthermore, the pattern of variation changed: the adjusted analysis identified three new outliers, and two providers lost their outlier status. Only one provider was consistently identified as an outlier in the unadjusted and adjusted analyses. CONCLUSIONS: Economic profiles of interventional cardiologists may be misleading if they do not adequately adjust for patient characteristics before procedure.

Eaton, C. B., A. Monroe, et al. (1998). "Cholesterol testing and management: a national comparison of family physicians, general internists, and cardiologists." *J Am Board Fam Pract* 11(3): 180-6.

BACKGROUND: We wanted to compare the frequency of cholesterol testing and treatment of hypercholesterolemia in patients cared for by family physicians, general internists, and cardiologists. **METHODS:** This study was a continuous cross-sectional survey of 1991 ambulatory office visits using a national probability sample of US physicians' office practices (National Ambulatory Care Survey). The physicians surveyed self-reported their specialty as family practice, internal medicine, or cardiology. Records of 33,795 patient visits to 1354 physicians were reviewed to find out whether the physicians reported cholesterol testing, cholesterol counseling, and charting of patient use of lipid-lowering medications. The results were compared among the three specialist groups. **RESULTS:** During an annual health examination (9.77 million office visits), a cholesterol test was reported by 23.5 percent of family physicians, 43.5 percent of internists, and 13.1 percent of cardiologists ($P < 0.01$). For all hypercholesterolemic patients (23.52 million office visits), the age- and sex-adjusted percentages of reported cholesterol-reduction counseling during office visits were 38.3 percent for family physicians, 42.4 percent for internists, and 36.5 percent by cardiologists (NS), and percentages of reported lipid-lowering medication prescriptions were 13.4 percent for family physicians, 25.1 percent for internists, and 28.4 percent for cardiologists ($P < 0.01$). In hypercholesterolemic patients with coronary heart disease (3.47 million office visits), the age- and sex-adjusted percentages of cholesterol reduction counseling reported during office visits were 64.4 percent for family physicians, 47.1 percent for internists, and 35.9 percent for cardiologists (NS) and the age- and sex-adjusted percentages of lipid-lowering medication prescriptions reported were 13.9 percent for family physicians, 62.5 percent for internists, and 34.7 percent for cardiologists ($P < 0.01$). **CONCLUSIONS:** Recommended goals regarding cholesterol testing and management were not reached by any physician group. Internists tested for hypercholesterolemia during an annual health examination more frequently and had more patients using lipid-lowering medications than did family physicians or cardiologists. Understanding the reasons for these specialty differences might lead to improvement in the diagnosis and management of hypercholesterolemia and therefore reduction in cardiovascular disease.

Fiscella, K. and P. Franks (1999). "Influence of patient education on profiles of physician practices." *Ann Intern Med* 131(10): 745-51.

BACKGROUND: Few data are available about the effect of patient socioeconomic status on profiles of physician practices. **OBJECTIVE:** To determine the ways in which adjustment for patients' level of education (as a measure of socioeconomic status) changes profiles of physician practices. **DESIGN:** Cross-sectional survey of patients in physician practices. **SETTING:** Managed care organization in western New York State. **PARTICIPANTS:** A random sample of 100 primary care physicians and 50 consecutive patients seen by each physician. **MEASUREMENTS:** Ranks of physicians for patient physical and mental health (Short Form 12-Item Health Survey) and satisfaction (Patient Satisfaction Questionnaire), adjusted for patient age, sex, morbidity, and education. **RESULTS:** Physicians whose patients had a lower mean level of education had significantly better ranks for patient physical and mental health status after adjustment for patients' level of education level than they did before adjustment ($P < 0.001$); this result was not seen for patient satisfaction. After adjustment for patients' level of education, each 1-year decrease in mean educational level was associated with a rank that improved by 8.1 (95% CI, 6.6 to 9.6) for patient physical health status and by 4.9 (CI, 3.9 to 5.9) for patient mental health status. Adjustment for education had similar effects for practices with more educated patients and those with less educated patients. **CONCLUSIONS:** Profiles of physician practices that base ratings of physician performance on patients' physical and mental

health status are substantially affected by patients' level of education. However, these results do not suggest that physicians who care for less educated patients provide worse care. Physician profiling should account for differences in patients' level of education.

Garcia Ruiz, F. J., A. Marin Ibanez, et al. (2004). "Current lipid management and low cholesterol goal attainment in common daily practice in Spain. The REALITY Study." *Pharmacoeconomics* 22 Suppl 3: 1-12.

OBJECTIVE: To evaluate prescribing patterns of lipid-lowering drugs used in management of patients at risk of coronary heart disease (CHD) in usual clinical practice in Spain and to assess low-density lipoprotein cholesterol (LDL-C) goal attainment among CHD and CHD equivalent patients (< 100 mg/dL) and non-CHD patients with two or more risk factors (< 130 mg/dL) who were prescribed lipid-lowering drugs. **METHODS:** Cohort study with retrospective chart review at 23 primary care centres and 16 lipid treatment centres across Spain (59% primary care; 41% outpatient lipid centres). Physicians consecutively identified eligible patients. Adults (aged > or = 18 years) with CHD/CHD equivalent or two or more major risk factors prior to first prescription of lipid-lowering drugs were eligible. Medical records were reviewed by physicians to collect patient characteristics, baseline and follow-up laboratory values and lipid-lowering drug treatment data. **RESULTS:** 619 patients (45.5% CHD and CHD equivalent patients and 54.5% non-CHD with two or more major risk factors) were included in the study with an average study follow-up of 3.6 years. Mean age was 60.1 years (SD 10.2), and 47.8% were female. Mean baseline LDL-C was 178 mg/dL (SD 45.0) for the CHD/CHD equivalent patients and 191 mg/dL (SD 56.95) for patients with two or more risk factors. Statins were the initial lipid-lowering drugs in 90.2% of patients; 52.5% of patients were initiated on low-dose (simvastatin 10mg or lower potency) statins. Overall 20.2% of CHD/CHD equivalent and 31.4% of patients with two or more risk factors attained LDL-C goal during the study period; of patients not attaining goal, 28.7% required an additional LDL-C reduction of > 30% to attain goal. In a logistic regression model for goal attainment, CHD/CHD equivalent patients (odds ratio [OR] 0.47; 95% confidence interval [CI] 0.31, 0.72) and patients with baseline LDL-C > 190 mg/dL (OR 0.53; 95% CI 0.35, 0.80) were least likely to reach cholesterol goal when compared with patients having baseline LDL-C > 100 mg/dL and < 130 mg/dL. **Conclusion:** Only 12.9% of patients attained LDL-C goal on their initial lipid-lowering drugs, and an additional 13.4% achieved goal after a change in their lipid-lowering therapy, resulting in 73.7% of patients not attaining goal after at least 3 years of follow-up, after initiation of lipid-lowering therapy. Patients who would gain the most from aggressive lipid lowering (CHD patients and patients with high baseline LDL-C) were least likely to achieve goal. More effective lipid management is needed to help these patients lower their cholesterol to goal levels or even lower.

Garnick, D. W., J. Fowles, et al. (1994). "Focus on quality: profiling physicians' practice patterns." *J Ambul Care Manage* 17(3): 44-75.

This article presents a physician practice profiling system developed using Medicare data to evaluate the quality of care provided by primary care physicians. We discuss four attributes to physicians' practice profiles that make them useful for quality improvement: flexibility, user involvement in developing profiles, explicit plans for evaluation, and fairness to groups of providers. This system serves as a model for physician profiling with a focus on quality of care measurement.

Goldfield, N. (1994). "Case mix, risk adjustment, reinsurance, and health reform." *Manag Care Q* 2(3): iv-viii.

Case mix measures represent the fundamental tool that measures clinical differences in groups of patients. It is important to be careful how these measures are used, as quality of care delivered to patients is at stake. Risk adjustment methodologies are

excellent for physician profiling if one adopts them as the first step in the quality improvement process. Differences in hospital or primary care practice profiles, adjusted for the case mix measures, often reveal quality improvement opportunities. This first step is the easy part. The hard part comes in working together with the health care team to determine the sources for these differences. After identifying and implementing solutions, the quality improvement process just continues. As benefits true commitment to quality improvement, risk assessment methodologies will only improve. In turn, this will provide the health care team with a greater ability to deal with "treasures" uncovered by applying case mix measures.

Greene, B. R., J. Barlow, et al. (1996). "Ambulatory care groups and the profiling of primary care physician resource use: examining the application of case mix adjustments." *J Ambul Care Manage* 19(1): 86-9.

A variety of profiling models and tools is utilized by payers, providers, and regulators to evaluate physician work, performance, and resource utilization. In physician profiling, the provider's pattern of practice is expressed as a rate of service or outcome. The article by Tucker, Weiner, Honigfeld, and Parton (this issue) compares the practice-based norms of primary care physicians by adjusting for case mix using ambulatory care groups (ACGs), a population-based classification method. Once the case mix is adjusted, the actual use of resources, as measured by overall charges, is compared with the expected value of resource use. In the Center for Research in Ambulatory Health Care Administration (CRAHCA) Physician Profiling Project, funded by The Robert Wood Johnson Foundation, physicians learn which services other physicians in their specialties perform. Physicians are able to compare their profiles with state and national level medians. The profiling project is one of the first demonstration projects in the field to profile ambulatory care practice patterns and collect patient demographics. An aspect of the project is to test the ACG classification system to data selected from 130 nonacademic practices representing over 5,000 physicians.

Greenfield, S., S. H. Kaplan, et al. (2002). "Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results." *Ann Intern Med* 136(2): 111-21.

BACKGROUND: Patient characteristics (case-mix bias) and physician-level variation (clustering) are often overlooked in profiling the quality of care provided by different groups of physicians, such as specialties. **OBJECTIVE:** To examine the effect of case-mix bias and physician-level clustering on differences in quality of diabetes care between specialty groups participating in the American Diabetes Association's Provider Recognition Program. **DESIGN:** Retrospective record review of both process and outcome measures over 1 year and a cross-sectional patient survey. The sample included 29 solo and group practice sites in diverse regions of the United States. Of the 29 sites, 15 were endocrinology sites and 14 were primary care sites. **PATIENTS:** 1750 adults with diabetes. **MEASUREMENTS:** Process measures included frequency of hemoglobin A(1c), lipid, and urine protein testing; blood pressure measurement; and foot and eye examinations. Outcome measures included A(1c)level, blood pressure, lipid levels, and patient satisfaction. Patient case-mix variables included age, sex, health status, level of education, ethnic minority status, and duration of diabetes. **RESULTS:** Unadjusted differences between endocrinologists and generalists were statistically significant for most process and outcome measures. Inclusion of patient case-mix variables reduced the statistical significance of specialty differences for some quality measures. After accounting for the substantial physician-level clustering, observed differences between specialties were no longer statistically significant for any of the quality measures except patient satisfaction. **CONCLUSIONS:** The findings underscore the importance of designing physician profiling studies with sufficient power to account for physician-level variation (clustering) as well as patient case-mix. Studies that are not designed with both

sufficient numbers of physicians and patients per physician may distort differences in quality of care between physician groups.

Harnick, D. J., J. L. Cohen, et al. (1998). "Effects of practice setting on quality of lipid-lowering management in patients with coronary artery disease." *Am J Cardiol* 81(12): 1416-20.

We undertook a study to determine whether there were differences in the quality of lipid management in patients with coronary artery disease (CAD) in 2 different practice settings (which represent different socioeconomic classes), and to determine the level of compliance with the National Cholesterol Education Program guidelines by academic physicians in managing patients with CAD. A retrospective cross-sectional study was performed using a systematic chart review of 270 medical records (131 from the cardiology clinic, 139 from the cardiology private practice) of patients with known CAD at an academic tertiary care center in New York City. The total proportion of patients with CAD having a lipid profile ordered in the clinic and private suite was 43%. Of these people, 22% had a low-density lipoprotein cholesterol (LDL) \leq 100 mg/dl and 54% had an LDL \leq 130 mg/dl (10% and 23% of the total population, respectively). The total proportion of patients taking lipid-lowering medications was 29%. When comparing the quality of treatment between the 2 settings, there were no statistically significant differences in the percentages of patients who had lipid profiles measured (40% clinic vs 47% private suite, $p > 0.10$), in the percentage of patients with LDL \leq 130 mg/dl (50% clinic vs 57% private suite, $p > 0.10$) or in the weighted percentage of patients taking lipid-lowering medications (29% clinic vs 48% private suite, $p = 0.099$). The performances of individual physicians, however, varied widely. The percentages of patients with lipid profiles measured by individual physicians ranged from 0% to 83%, while the percentages of patients on drug treatment by a physician ranged between 10% and 88%. These findings indicate that socioeconomic differences, represented by different practice settings, do not account for differences in the screening for, control of, or use of medications in managing hyperlipidemia. Rather, individual physicians are accountable for differences in lipid management.

Hofer, T. P., R. A. Hayward, et al. (1999). "The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease." *Jama* 281(22): 2098-105.

CONTEXT: Physician profiling is widely used by many health care systems, but little is known about the reliability of commonly used profiling systems. OBJECTIVES: To determine the reliability of a set of physician performance measures for diabetes care, one of the most common conditions in medical practice, and to examine whether physicians could substantially improve their profiles by preferential patient selection. DESIGN AND SETTING: Cohort study performed from 1990 to 1993 at 3 geographically and organizationally diverse sites, including a large staff-model health maintenance organization, an urban university teaching clinic, and a group of private-practice physicians in an urban area. PARTICIPANTS: A total of 3642 patients with type 2 diabetes cared for by 232 different physicians. MAIN OUTCOME MEASURES: Physician profiles for their patients' hospitalization and clinic visit rates, total laboratory resource utilization rate and level of glycemic control by average hemoglobin A1c level with and without detailed case-mix adjustment. RESULTS: For profiles based on hospitalization rates, visit rates, laboratory utilization rates, and glycemic control, 4% or less of the overall variance was attributable to differences in physician practice and the reliability of the median physician's case-mix-adjusted profile was never better than 0.40. At this low level of physician effect, a physician would need to have more than 100 patients with diabetes in a panel for profiles to have a reliability of 0.80 or better (while more than 90% of all primary care physicians at the health maintenance organization had fewer than 60 patients with diabetes). For profiles of glycemic control, high outlier physicians could dramatically improve their

physician profile simply by pruning from their panel the 1 to 3 patients with the highest hemoglobin A1c levels during the prior year. This advantage from gaming could not be prevented by even detailed case-mix adjustment. CONCLUSIONS: Physician "report cards" for diabetes, one of the highest-prevalence conditions in medical practice, were unable to detect reliably true practice differences within the 3 sites studied. Use of individual physician profiles may foster an environment in which physicians can most easily avoid being penalized by avoiding or deselecting patients with high prior cost, poor adherence, or response to treatments.

Katon, W., C. M. Rutter, et al. (2000). "Are there detectable differences in quality of care or outcome of depression across primary care providers?" *Med Care* 38(6): 552-61. OBJECTIVE: The objective of this work was to determine whether there are detectable differences among primary care physicians in measures of quality of care or clinical outcome for depressed patients during the first 2 months of treatment with antidepressant medication. METHODS: We studied 1,599 depressed primary care patients initiating antidepressant treatment from 63 family physicians in 4 primary care clinics of a staff-model health maintenance organization. Patients were interviewed 6 to 8 weeks after initiating antidepressant medication with a telephone structured interview that included the Structural Clinical Interview for DSM-IV Diagnoses (SCID). Automated databases of the HMO were used to examine 3 quality-of-care measures: (1) the percentage of patients who had a refill of their antidepressant by 6 weeks, (2) the percentage of patients who had a return visit by 3 weeks, and (3) the percentage of patients having a return visit by 6 weeks. The percentage of patients in each primary care physician panel who had > or =4 persistent DSM-IV depressive symptoms at 6 to 8 weeks was the main clinical outcome variable. To adjust for case-mix differences between physician panels, patient age, gender, and medical comorbidity were controlled for in the analyses. Two covariates were used to adjust for differences in patients' clinical severity: self-report of > or =2 prior depressive episodes and an SCID diagnosis of major depression during the patient's worst episode in the last 2 years. Physician age, gender, and part-time versus full-time practice were also used as covariates. RESULTS: The wide observed range of variability in quality-of-care and clinical outcome measures by physician practice decreased markedly in the statistical model that controlled for patient-level covariates and differences in the number of patients seen per provider. We did not detect differences in physician practice for the percentage of patients who had a return visit by 3 or 6 weeks, the percentage of patients who had a refill of their antidepressant prescription by 6 weeks, or the percentage of patients with an adverse clinical outcome of depression. CONCLUSIONS: We did not find important differences in measures of quality of care or patient outcomes by physician. These results may have implications for the use of physician profiling and other forms of physician report cards.

Kraft, P. (2005). "Wirtschaftlichkeitsverfahren bauen auf verlässliche Grundlagen." *infosantésuisse*(1-2): 4-5.

Mantel-Teeuwisse, A. K., W. M. Verschuren, et al. (2004). "Recent trends in (under)treatment of hypercholesterolaemia in the Netherlands." *Br J Clin Pharmacol* 58(3): 310-6. AIM: To assess recent trends in undertreatment of hypercholesterolaemia (1998-2002). METHODS: Data were obtained from the third cross-sectional examination of the Monitoring Project on Risk Factors for Chronic Diseases (n = 4878; age 31-70 years), conducted in the Netherlands. Treatment eligibility was established according to Dutch guidelines. Data from the second examination (1993-1997) were used to assess time trends. The association between demographic variables, cardiovascular disease risk factors, drug use and lipid-lowering medication was assessed using multivariable logistic regression. RESULTS: Overall, 45.9% [95% confidence interval (CI) 41.4, 50.4] of respondents eligible for treatment were treated, and 17.4% (95%

CI 13.9, 20.9) were both treated and controlled (1998-2002). Treatment increased significantly after 1995, showed a slight decrease in subsequent years until 2000, when treatment increased again. Subgroups less frequently treated for primary prevention included among others males [odds ratio (OR) = 0.08; 95% CI 0.03, 0.21], younger patients (OR = 0.93 per year; 95% CI 0.88, 0.98), diabetics (OR = 0.19; 95% CI 0.07, 0.56), untreated hypertensives (OR = 0.21; 95% CI 0.09, 0.49) and current smokers (OR = 0.09; 95% CI 0.03, 0.25). In secondary prevention, patients with a history of stroke were less likely to receive treatment (OR = 0.41; 95% CI 0.18, 0.94) compared with patients with a history of ischaemic heart disease. CONCLUSIONS: Treatment of hypercholesterolaemia has steadily increased over the past 10 years in the Netherlands. However, at present still less than one out of two eligible for treatment is treated, and only about one out of six is both treated and controlled.

Murphy, A. W., M. E. Cupples, et al. (2005). "The SPHERE Study. Secondary prevention of heart disease in general practice: a randomised controlled trial of tailored practice and patient care plans with parallel qualitative, economic and policy analyses. [ISRCTN24081411]." *Curr Control Trials Cardiovasc Med* 6(1): 11.

BACKGROUND: The aim of the SPHERE study is to design, implement and evaluate tailored practice and personal care plans to improve the process of care and objective clinical outcomes for patients with established coronary heart disease (CHD) in general practice across two different health systems on the island of Ireland. CHD is a common cause of death and a significant cause of morbidity in Ireland. Secondary prevention has been recommended as a key strategy for reducing levels of CHD mortality and general practice has been highlighted as an ideal setting for secondary prevention initiatives. Current indications suggest that there is considerable room for improvement in the provision of secondary prevention for patients with established heart disease on the island of Ireland. The review literature recommends structured programmes with continued support and follow-up of patients; the provision of training, tailored to practice needs of access to evidence of effectiveness of secondary prevention; structured recall programmes that also take account of individual practice needs; and patient-centred consultations accompanied by attention to disease management guidelines. METHOD: SPHERE is a cluster randomised controlled trial, with practice-level randomisation to intervention and control groups, recruiting 960 patients from 48 practices in three study centres (Belfast, Dublin and Galway). Primary outcomes are blood pressure, total cholesterol, physical and mental health status (SF-12) and hospital re-admissions. The intervention takes place over two years and data is collected at baseline, one-year and two-year follow-up. Data is obtained from medical charts, consultations with practitioners, and patient postal questionnaires. The SPHERE intervention involves the implementation of a structured systematic programme of care for patients with CHD attending general practice. It is a multi-faceted intervention that has been developed to respond to barriers and solutions to optimal secondary prevention identified in preliminary qualitative research with practitioners and patients. General practitioners and practice nurses attend training sessions in facilitating behaviour change and medication prescribing guidelines for secondary prevention of CHD. Patients are invited to attend regular four-monthly consultations over two years, during which targets and goals for secondary prevention are set and reviewed. The analysis will be strengthened by economic, policy and qualitative components.

Nickerson, C. and R. W. Rutledge (1999). "A methodology for choosing a physician profiling system: the case of First Option Health Plan." *J Health Care Finance* 26(2): 5-13.

The public continues to demand improved health care at affordable costs. It appears that this public pressure will force health maintenance organizations (HMOs) and other health care providers to face their toughest days in the years to come. They will need to find solutions to resolve this "quality and cost" problem. One method that is likely to have the ability to improve quality of care and control costs is physician

profiling. That is, data related to a physician's care or outcomes can be collected and compared with profiles of various populations, for example, other similar providers in a relatively close geographical proximity. This article examines the issues involved in physician profiling, including the characteristics of profiling systems and the direct and indirect benefits of having such a system. Further, the article uses an actual HMO example to discuss considerations that administrators should make when choosing a physician profiling system.

Parkerson, G. R., Jr., J. L. Michener, et al. (1997). "Duke Case-Mix System (DUMIX) for ambulatory health care." *J Clin Epidemiol* 50(12): 1385-94.

The Duke Case-Mix System (DUMIX), which combines age, gender, patient-reported perceived and physical health status, and provider-reported or auditor-reported severity of illness to classify patients by their risk of high future utilization, explained 17.1% of the variance in future clinic charges and 16.6% of the variance in return visits. When a random half of 413 ambulatory adults were classified into four risk classes by predictive regression coefficients from the other half, there was a stepwise increase in actual future utilization by risk class. The most accurate classification was for Class 4 (highest risk) patients, with a sensitivity of 40.8%, specificity of 82.1%, and likelihood ratio of 2.3. These 23.7% of patients accounted for 44.2% of charges for all patients. When predictive coefficients from this population were used to classify a different group of 206 ambulatory adults, past utilization also increased in stepwise order by case-mix class.

Pope, G. C. (1988). "Medical conditions, health status, and health services utilization." *Health Serv Res* 22(6): 857-77.

Using data from the 1980 National Medical Care Utilization and Expenditure Survey (N = 11,530), four commonly used health status indicators are interpreted in terms of the underlying medical conditions they reflect. It is found that self-rated health status, role limitations, restricted activity days, and functional limitations measure similar conditions. These conditions tend to be chronic and severe; heart and cerebrovascular disease are especially associated with poor health as measured by all of the variables. Disability days is most likely to reflect acute, transitory morbidity. Practical suggestions for the appropriate use of the four variables are made. In addition, the conditions associated with the most ambulatory utilization of health services are identified. Among these conditions, those which are and are not measured adequately by the health status indicators are disclosed. It is concluded that the health status variables, either individually or as a group, do not measure many variations in health that are strongly related to utilization.

Ramsey, R. (2003). "IPA cites physician profiling as key capitation strategy, expands effort to specialists." *Capitation Manag Rep* 10(3): 39-41.

Reid, R. A. (1998). "Using the tail to wag the dog: a consumer's perspective on management of the cost of cancer." *Cancer* 82(10 Suppl): 2022-5.

This article describes the health insurance policy used by one, small, fully insured company. The most important benefits for reducing the human and financial cost of cancer are benefits that improve the decision processes of the healthy person, and a benefit that promotes involvement of a primary care physician at the time of death. These benefits in our health insurance plan are designed to: 1) assure easy access to a primary care physician; 2) provide incentives to teach the beneficiary to use the knowledge of the primary care physician as he/she makes decisions regarding care; 3) provide claims monitoring to identify small problems before they become serious; and 4) eliminate factors such as capitation and physician profiling, which provide strong financial incentives for the primary care physician to disengage from the care of seriously ill patients. Once serious illness is identified, it is important to pay generously for the person's struggle to regain health. The strategy described in this

article does not rely on networks or negotiations with providers, and therefore can operate uniformly throughout the U.S. All large companies using this strategy to date have experienced a permanent drop in health care costs, largely due to a reduction in the incidence of expensive cases.

- Rich, E. C., J. Kralewski, et al. (1998). "Variations in the management of primary care: effect on cost in an HMO network." *Arch Intern Med* 158(21): 2363-71.
OBJECTIVE: To determine the relation to cost of different aspects of the management of primary care among group practices within a health maintenance organization network. MEASURES: A cross-sectional survey study of medical practices conducted with Blue Cross Blue Shield of Minnesota, St Paul. The subjects were group practices accepting financial and administrative responsibility for primary care services in the managed care plans of Blue Cross Blue Shield of Minnesota. One hundred twelve primary care practices and 153397 enrollees were included in this analysis. The principal resource use measure in this study was nonhospital cost per member per year estimated from payments to providers plus subscriber-eligible liability. RESULTS: The medical directors' responses revealed considerable variability in the management of primary care in these 112 practices. Group practice characteristics consistently associated with lower nonhospital cost were patient identification of a primary care physician, cost of care profiling, more frequent physician profiling, more patients per hour in the clinic, a higher proportion of primary care physicians in the specialty of family or general practice, and a greater number of physicians in the group practice. CONCLUSIONS: Results of this study demonstrate substantial variation in the management of primary care among group practices participating in a health maintenance organization network. These differences are associated with significant variation in the nonhospital cost of care for enrollees.
- Roblin, D. W. (1996). "Applications of physician profiling in the management of primary care panels." *J Ambul Care Manage* 19(2): 59-74.
A primary care panel is a conceptual and institutional unit, in managed care organizations, that holds an individual primary care physician or a team of primary care physicians accountable for the care of a defined population of enrollees. A persistent question in ambulatory care management is how to align, across a set of primary care panels, the distribution of demand for primary care physician time with the distribution of supply of primary care physician time. Part of the solution to this resource allocation decision requires measurement of primary care panel case mix (a principal influence on demand); and, part of the solution requires measurement of physician practice (an influence on both demand and supply). The goals of this article are several, among them, to outline the reasons research and development in this area are important for the successful functioning of a primary care panel system, and to illustrate how methods of physician profiling may be usefully applied toward evaluating and implementing solutions to the alignment of demand for, and supply of, primary care physician time.
- Roblin, D. W. (1998). "Physician profiling using outpatient pharmacy data as a source for case mix measurement and risk adjustment." *J Ambul Care Manage* 21(4): 68-84.
Pharmacy data are a potential source of information on the health risks of members in a managed care organization (MCO). Reasons why these data can serve as a basis for case mix measurement of MCO populations are reviewed. Several pharmacy-based case mix measurement systems (CDS, KPARx) are summarized. Applications of pharmacy-based case mix systems for profiling of primary physician panels in an MCO are discussed and illustrated with several examples.
- Roos, N. P., K. C. Carriere, et al. (1998). "Factors influencing the frequency of visits by hypertensive patients to primary care physicians in Winnipeg." *Cmaj* 159(7): 777-83.

BACKGROUND: As part of a recent project focused on needs-based planning for generalist physicians, the authors documented the variety of practice styles of primary care physicians for managing patients with hypertension. They investigated the validity of various explanations for these different styles and the relative contributions of physician and patient characteristics to the rates at which hypertensive patients contact physicians. **METHODS:** Retrospective descriptive study using regression analyses to simultaneously adjust for the influence of key patient and physician characteristics. Hypertensive patients in Winnipeg were identified using Manitoba physician claims data for fiscal years 1993/94 and 1994/95. Patients were included if they were 25 years of age or more and had at least one physician contact in both 1993/94 and 1994/95 during which hypertension was diagnosed. In addition, the primary care physician had to be the physician that the patient contacted most frequently in 1993/94 and 1994/95 and with whom she or he had at least 2 visits during this period. Only patients of family practitioners whose practice included at least 50 hypertensive patients were included. **RESULTS:** To control for the effects of large samples and to validate the results, the authors conducted all analyses for half (6282) the sample of hypertensive patients who met the study criteria (12,563). A total of 132 primary care physicians who met the study criteria were identified. The patients made on average 9.3 ambulatory visits to physicians (both general practitioners and specialists) in 1994/95. Those who had more complex medical conditions (i.e., were formally referred to a specialist), those who had 3 or more serious medical problems and those who had been admitted to hospital made more visits to their primary care physician than those without these characteristics. After these and other key patient characteristics were controlled for, a primary care physician's patient recall rate in 1993/94 was strongly related to the number of visits his or her hypertensive patients made to all doctors for any reason in 1994/95. Physicians with high patient recall rates (i.e., who saw their hypertensive patients on average 8 or more times) in 1993/94 also had high recall rates in 1994/95. **INTERPRETATION:** Because patient characteristics most strongly associated with high visit rates were those reflecting patient illness, policy measures aimed at patients (e.g., user fees and deinsurance) do not appear to be the appropriate policy tool for dealing with high visit rates. Given the influence of a physician's patient recall rate on patient visit patterns, physician profiling and feedback may prove more appropriate.

Ruof, J., G. Klein, et al. (2002). "Lipid-lowering medication for secondary prevention of coronary heart disease in a German outpatient population: the gap between treatment guidelines and real life treatment patterns." *Prev Med* 35(1): 48-53.

BACKGROUND: Few published data in particular from the United States indicate that the implementation of guidelines for prevention of coronary heart disease (CHD) is far from optimal. The objective of our study was to identify the type and prevalence of lipid-lowering medications in a German outpatient CHD population and to examine the impact of applied treatment regimens on serum lipid levels. **METHODS:** Retrospective analysis of the washout phase of 2,856 CHD patients requiring lipid-lowering medication. Data are derived from a multicenter, randomized, open-label, parallel group clinical trial comparing the safety and efficacy of atorvastatin versus simvastatin in 591 centers in Germany. Medical history, physical examination, and serum lipid levels were obtained at the beginning of the washout phase (Week -6) and at the end of the washout phase (Week -1, i.e., 5 weeks after the discontinuation of all prior lipid-lowering medications). The data at Week -6 represented the lipid levels under real life conditions. The difference from the data at Week -1 reflected the therapeutic effects achieved by the previous lipid-lowering treatment. **RESULTS:** The mean low-density lipoprotein cholesterol (LDL-C) level at Week -6 was 173.4 +/- 42.5 mg/dl. Only 176 (6.2%) of 2,856 CHD patients were found to meet the target LDL-C level of <115 mg/dl at Week -6, only 76 (2.7%) patients had LDL-C levels <100 mg/dl, and 363 (12.7%) patients had LDL-C levels <130 mg/dl. After discontinuation of all prior lipid-lowering medications, mean LDL-C increased to 187.2 +/- 44.0 mg. This

means that only a marginal 7.4% reduction in LDL-C level was achieved under real life treatment conditions. This limited LDL-C reduction was due mainly to the low prevalence of lipid-lowering treatment (65.5% of patients did not receive any medication at all) and inadequate dosing. With respect to the effect on LDL-C and total cholesterol, statins alone were superior to fibrates. CONCLUSION: The study shows that there is a wide gap between treatment guidelines and real life treatment patterns in Germany. Awareness of the risks of high cholesterol levels has to be increased among both patients and physicians. Available treatment guidelines should be better implemented.

- Salem-Schatz, S., G. Moore, et al. (1994). "The case for case-mix adjustment in practice profiling. When good apples look bad." *Jama* 272(11): 871-4.
OBJECTIVE--To assess the influence of patient characteristics on practice profiling. Using the example of specialty referrals by primary care physicians, we evaluated the impact of adjusting for patient characteristics (age/sex vs case mix) on the estimation of practice variation, the identification of outlier practices, and the evaluation of potential predictors of variation. DESIGN AND SETTING--We applied several measurement strategies to a retrospective cohort of patients (N = 37,830) within 52 physician practices in a large staff-model health maintenance organization during a 1-year period. OUTCOME MEASURES--We calculated unadjusted referral rates and adjusted standardized referral ratios for each physician. Using these, we determined coefficients of variation and statistical "outlier status." RESULTS--Adjustment for patient characteristics decreased the observed variation in referral profiles, with a decrease of more than 50% in the coefficient of variation. Three quarters of the physicians identified as statistical outliers with use of an age/sex-adjusted measure were no longer identified as such with use of an case-mix-adjusted measure. Several key predictors of unadjusted referral rate (including physician age, practice tenure, site of practice, and extent of laboratory test ordering) dropped out of regression models when the outcome variable was adjusted for patient characteristics. CONCLUSION--Failure to adjust for case mix in physician practice profiles may lead to overestimates of variation and misidentification of outliers. To the extent that unadjusted practice profiles are used for decisions about education, sanctions, or employment, physicians may be subject to inequitable decisions and actions. Misinformation about the causes and extent of practice variation may also lead to misdirection of scarce resources for quality improvement efforts.
- Sloan, K. L., A. E. Sales, et al. (2003). "Construction and characteristics of the RxRisk-V: a VA-adapted pharmacy-based case-mix instrument." *Med Care* 41(6): 761-74.
BACKGROUND: Assessment of disease burden is the key to many aspects of health care management. Patient diagnoses are commonly used for case-mix assessment. However, issues pertaining to diagnostic data availability and reliability make pharmacy-based strategies attractive. Our goal was to provide a reliable and valid pharmacy-based case-mix classification system for chronic diseases found in the Veterans Health Administration (VHA) population. OBJECTIVE: To detail the development and category definitions of a VA-adapted version of the RxRisk (formerly the Chronic Disease Score); to describe category prevalence and reliability; to check category criterion validity against ICD-9 diagnoses; and to assess category-specific regression coefficients in concurrent and prospective cost models. RESEARCH DESIGN: Clinical and pharmacological review followed by cohort analysis of diagnostic, pharmacy, and utilization databases. SUBJECTS: 126,075 veteran users of VHA services in Washington, Oregon, Idaho, and Alaska. METHODS: We used Kappa statistics to evaluate RxRisk category reliability and criterion validity, and multivariate regression to estimate concurrent and prospective cost models. RESULTS: The RxRisk-V classified 70.5% of the VHA Northwest Network 1998 users into an average of 2.61 categories. Of the 45 classes, 33 classes had good-excellent 1-year reliability and 25 classes had good-excellent criterion

validity against ICD-9 diagnoses. The RxRisk-V accounts for a distinct proportion of the variance in concurrent ($R^2 = 0.18$) and prospective cost ($R^2 = 0.10$) models. CONCLUSIONS: The RxRisk-V provides a reliable and valid method for administrators to describe and understand better chronic disease burden of their treated populations. Tailoring to the VHA permits assessment of disease burden specific to this population.

Stoddard, J. J., M. Reed, et al. (2003). "Financial incentives and physicians' perceptions of conflict of interest and ability to arrange medically necessary services." *J Ambul Care Manage* 26(1): 39-50.

Many believe that physician payment mechanisms that include incentives to restrain utilization create conflicts of interest for physicians and result in the withholding of needed services. Pooled data from two rounds of the Community Tracking Study physician survey, a nationally representative telephone survey of physicians, are the basis of this analysis. We examine the association between explicit financial incentives linked to physician profiling and perceived conflict of interest, and the reported ability to obtain specific, medically necessary secondary services (referrals, hospitalizations and diagnostic imaging). Logistic regression models were employed to control for potentially confounding influences. After controlling for other factors, physicians subject to profiling linked with financial incentives were much less likely than physicians not affected by profiling to strongly agree that they can make clinical decisions in the best interests of their patients without the possibility of reducing their income. They were also less likely to report that they could always obtain selected medically necessary secondary services for their patients. Physicians subject to explicit financial incentives based on profiling are more likely than other physicians to perceive a conflict of interest. Physicians with financial incentives tied to profiling also experience greater difficulty obtaining medically necessary secondary services for their patients.

Tompkins, C. P., S. Bhalotra, et al. (1996). "Physician profiling in group practices." *J Ambul Care Manage* 19(4): 28-39.

Profiling is a technique that large, multispecialty group practices, like many insurers, can use to monitor and improve quality and efficiency. Groups can examine physician performance by calculating ratios of medical inputs to patient or population outputs. Physician control can help to achieve balance between clinical benefits and economic considerations. Profiles need to reflect a group's multiple missions, such as clinical care, research, and education; philosophy of care; and organizational ethos regarding physician compensation systems. Groups may need to customize standard approaches because of their emphasis on early utilization of specialists and the atypical case mixes often found in referral practices.

Tucker, A. M., J. P. Weiner, et al. (1996). "Profiling primary care physician resource use: examining the application of case mix adjustment." *J Ambul Care Manage* 19(1): 60-80.

Growing emphasis on managed care has led to increased interest in physician practice profiling. Standardized techniques for conducting profiling are not yet well established. One particularly challenging methodologic issue, case mix adjustment, is explored here using actual cost profiles derived from primary care physicians at two independent practice association (IPA)-model health maintenance organizations (HMOs). Specifically, this article examines how the ambulatory care group case mix methodology can be applied to profiling and illustrates that it provides more depth of information with which to assess performance than does standard demographic adjustment alone. This analysis suggests both the potential and methodologic limitations of profiling at the individual physician level.

- Tucker, J. L., 3rd (2000). "The theory and methodology of provider profiling." *Int J Health Care Qual Assur Inc Leadersh Health Serv* 13(6-7): 316-21.
Physician profiling is the comparison of physician practice patterns to determine the existence and effects of significant differences in outcomes. To optimize care quality, these comparisons can be used to influence provider behavior through awareness when outcomes are poor or do not warrant provider expenditures. To maximize the value of such a technique, the underlying bases for comparison and the uses of derived data must be understood. Several factors must be considered or controlled for when determining comparable providers. Additionally, the worth or significance of findings must also be understood. This paper seeks to provide information in these two areas for the benefit of those responsible for managing care at all levels.
- Tufano, J. T., D. A. Conrad, et al. (1999). "Addressing physician compensation and practice productivity." *J Ambul Care Manage* 22(3): 47-57.
Medical groups are challenged to adopt a systematic, evidence-based approach to selecting a physician compensation method that supports the group's overall financial and organizational strategies, including managed care contracting strategies; is consistent with the philosophies, beliefs, and attitudes of the group's membership as they pertain to individual productivity; and can be supported by the organization's information technology, decision support, and management infrastructures. This article explains how research in physician profiling, benchmarking, general compensation theory, and physician productivity provides evidence that can serve as the foundation for a pragmatic approach to evaluating physician compensation method alternatives. It also presents a unique production-based compensation model for illustrative purposes.
- Veugelers, P. J., A. M. Yip, et al. (2004). "Inequalities in health and health services delivery: a multilevel study of primary care and hypertension control." *Chronic Dis Can* 25(3-4): 101-7.
Delivery of health services is an important determinant of health. Restricted availability and access may result in health inequalities. To determine the extent of geographic variation in the delivery of health services and its effect on the health of community residents in terms of under-diagnosis and under-treatment of hypertension, we carried out a multilevel study of participants in the 1995 Nova Scotia Heart Health Survey (n = 3,094). We used individual level survey data and health status measurements linked to geographical level information to examine the importance of adequate delivery of health services to the diagnosis and treatment of hypertension in the universal health care setting of the province of Nova Scotia. The delivery of primary care services across Nova Scotia varied moderately with physician visit rates ranging from 3.3 to 5.5 visits per resident per year. There were neither substantial nor statistically significant differences in the diagnosis and treatment of hypertension among residents of communities varying in the delivery of health services. We concluded that a geographic variation in the delivery of primary care services is a public health concern that is not consistent with the objectives of universal coverage of health services; however, it was not confirmed to result in health inequalities.
- Welch, H. G., M. E. Miller, et al. (1994). "Physician profiling. An analysis of inpatient practice patterns in Florida and Oregon." *N Engl J Med* 330(9): 607-12.
BACKGROUND. Physician profiling is a method of cost control that focuses on patterns of care instead of on specific clinical decisions. It is one cost-control method that takes into account physicians' desire to curb the intrusion of administrative mechanisms into the clinical encounter. To provide a concrete example of profiling, we analyzed the inpatient practice patterns of physicians in Florida and Oregon.
METHODS. Data for 1991 from Medicare's National Claims History File were used to profile 12,720 attending physicians in Florida and 2589 in Oregon. For each attending

physician, we determined the total relative value of all physicians' services delivered during each patient's hospital stay. Relative value was measured in relative-value units (RVUs), according to the resource-based relative-value scale used by Medicare in determining payments to physicians. The mean number of RVUs per admission was then adjusted for the physician's case mix according to the patients' assigned diagnosis-related groups. The influence of the physician's specialty and of selected types of services (such as imaging and endoscopy) was also examined. RESULTS. Florida physicians used markedly more resources, on average, than their colleagues in Oregon (46 vs. 30 case-mix-adjusted RVUs per admission). The difference was apparent for all specialties and all types of service. To illustrate the profiling data potentially available to the medical staffs of individual hospitals, we examined specific data on individual attending physicians and for various types of service for three hospitals' staffs. Despite similar overall profiles that fell below the national mean, each staff had a different practice pattern and would require different efforts to improve efficiency. CONCLUSIONS. In an effort to encourage further debate, we have described one method of physician profiling. Profiling data help identify and characterize differences in practice style to which individual physicians or hospital staffs can respond. Because profiling is not based on rigid rules, it is a cost-containment strategy that can easily accommodate legitimate exceptions; it is therefore preferable to methods in which the appropriateness of each clinical decision is judged separately.