

# Diskussion

---

Die untersuchten quantitativen Methoden unterscheiden sich in ihren Ergebnissen deutlich. Die quantitativen Methoden werden zunächst untereinander verglichen und die Unterschiede, die sich in Prävalenzraten, Verteilung der Deformationen über die Wirbelsäule und Kappa Schätzwerten zeigen, erläutert. Anschließend werden Gründe für die Unterschiede herausgearbeitet. Neben den Indices der Methoden unterscheiden sich die Methoden durch ihre Schwellenwerte. Der Einfluss der Schwellenwerte auf die Ergebnisse wird diskutiert und dabei auf das Problem von Schwellenwerten aus Referenzpopulationen eingegangen. Über den Vergleich der quantitativen Methoden hinaus, erfolgt eine Bewertung der Methoden. Dazu werden die Ergebnisse der quantitativen Methoden mit denen der qualitativen Auswertung verglichen. Dabei zeigt sich die in der Einleitung bereits erwähnte fehlende Spezifität von Höhenminderungen für osteoporotische Wirbelkörperfrakturen. Auf dieses Problem wird im Abschnitt zur differentialdiagnostischen Einteilung der quantitativ erkannten Deformationen ausführlich eingegangen. Basierend auf den Ergebnissen der Arbeit erfolgt dann im letzten Abschnitt eine zusammenfassende Bewertung der untersuchten Methoden.

## Vergleich der quantitativen Methoden untereinander

### Anzahl von Deformationen

Durch die quantitativen Methoden werden signifikant unterschiedliche Prävalenzraten ermittelt. Die Prävalenzraten liegen zwischen 12.92% bei der Methode von Felsenberg 0.75 und 34.10% bei der Methode von Minne (Tabelle 6 auf Seite 25). Die Prävalenzraten von Felsenberg 0.75 (12.92%) und von Melton (13.09%) liegen dicht beieinander, wie auch die Prävalenzraten der Methoden von Felsenberg 0.80 (20.00%) und von Eastell (21.11%). Die Prävalenzrate von McCloskey liegt mit 17.28% etwa zwischen diesen Werten und die von Minne mit 34.10% deutlich über allen anderen Werten.

Unterschiede bei Prävalenzraten von bis zu 400% sind bereits in anderen Arbeiten beschrieben worden (6, 13, 14, 22, 23, 28). In der EVOS Studie wurden Prävalenzraten für eine europäische Querschnittspopulation von 8421 postmenopausalen Frauen von 20.2% mit der Methode von Eastell und 12.0% mit der Methode von McCloskey ermittelt (28). Es ist wichtig, die Methoden an einer Population zu untersuchen. So zeigt sich für die Methode von McCloskey bei den Querschnittspopulationen von OPUS und EVOS ein Unterschied von etwa 5%. In der Publikation zur EVOS Studie ist darauf hingewiesen worden, dass sich die Prävalenzraten zwischen den einzelnen Ländern in Europa deutlich unterscheiden (28). Mit der Methode von Eastell variierten die Prävalenzraten der einzelnen Ländern zwischen 14.9% und 26.6% und mit der Methode von McCloskey zwischen 6.7% und 20.3%.

Die Unterschiede zwischen den Methoden können genauer herausgearbeitet werden, wenn man nicht Prävalenzraten, dh die relative Anzahl von Wirbelsäulen mit einer oder mehreren Wirbelkörperdeformationen untersucht, sondern die einzelnen Wirbelkörper betrachtet. Tabelle 7 auf Seite 25 zeigt die Anzahl der Wirbelkörperdeformationen. Die Anzahl ist wie schon bei den Prävalenzraten bei Minne deutlich höher als bei allen anderen Methoden. Die hohe Anzahl an posterioren Deformationen ist auffällig, während sich die Anzahl der medialen Deformationen wenig von der der anderen Methoden unterscheidet. Bei der Methode von Felsenberg 0.80 ist die Anzahl der anterioren Deformationen verhältnismäßig hoch, obwohl sich die Anzahl der medialen und posterioren Deformationen wenig von den Methoden von Felsenberg 0.75, Eastell, Melton und McCloskey unterscheidet. Auffällig ist auch, dass der Anteil der anterioren Deformationen bei der Methode von Felsenberg mit der Änderung des Schwellenwertes von 0.75 auf 0.80 von 47.04% auf 68.60% ansteigt. Diese Ergebnisse lassen sich nicht mit anderen Studien vergleichen, da keine Publikation entsprechende Aufgliederungen der Deformationen enthalten.

### **Verteilung der Deformationen über die Wirbelsäule**

Die Verteilung der Wirbelkörperdeformationen über die Wirbelsäule wird sowohl bei der Analyse der Schwellenwerte wie auch bei der differentialdiagnostischen Einteilung der Deformationen eine Rolle spielen. Die Verteilung der Deformationen ist bei den einzelnen Methoden ähnlich mit einigen deutlichen Ausnahmen (Abbildung 9 auf Seite 26). Alle Methoden haben einen Gipfel in der mittleren Brustwirbelsäule bei T7 bzw T8 und einen zweiten Gipfel am thorakolumbalen Übergang bei T11 bis L1. Während die Anzahl der Deformationen im thorakolumbalen Übergang zwischen den einzelnen Methoden relativ geringe Unterschiede aufweist, mit Ausnahme der Methode von Minne, unterscheiden sich die Methoden in der mittleren Brustwirbelsäule deutlicher. Insbesondere bei der Methode von Felsenberg 0.80 ist die Anzahl von Deformationen in der mittleren Brustwirbelsäule höher als bei anderen Methoden. Auffällig unterschiedlich von allen anderen Methoden ist die Anzahl der Deformationen in der unteren Lendenwirbelsäule bei der Methode von Minne.

Die zweigipflige Verteilung der Wirbelkörperdeformationen wird in Publikationen von Eastell (10), Melton (24), McCloskey (22) beschrieben. Dort ist zu sehen, dass die Häufigkeit von Deformationen in der mittleren Brustwirbelsäule nicht größer als im thorakolumbalen Übergang ist. Diese Verteilung ist auch in der kanadischen Querschnittsstudie CaMos mit 9424 Probanden, gezeigt worden (17). Vergleiche der Verteilungen über die Wirbelsäule von unterschiedlichen quantitativen Methoden an einer Population sind bisher nicht publiziert worden.

### **Diskordante Mengen von Deformationen**

In den vorangegangenen Abschnitten wurde gezeigt, dass die quantitativen Methoden eine unterschiedliche Anzahl von Deformationen ermitteln und dass sich die Verteilung über die Wirbelsäule unterscheidet. Es stellt sich aber darüber hinaus die Frage, ob bei gleicher Anzahl der Deformationen auch tatsächlich die gleichen Deformationen erkannt werden.

Die Methoden von Felsenberg 0.75 und Melton ermitteln fast identische Prävalenzraten mit 12.92% bzw 13.09%. Betrachtet man die Kontingenztafel der beiden Methoden (Tabelle 8 auf Seite 27), so fällt eine recht große Anzahl an unterschiedlich erkannten Deformationen auf. Die Methode von Felsenberg 0.75 erkennt 35 Wirbelsäulen als deformiert, die die Methode von Melton als nicht deformiert ansieht. Umgekehrt erkennt die Methode von Melton 39 Wirbelsäulen als deformiert, die die Methode von Felsenberg 0.75 als nicht deformiert ansieht. Dh etwa 12% der ermittelten Wirbelsäulen mit Deformationen unterscheiden sich. Das ist deutlich mehr, als von den Prävalenzraten her zu erwarten ist. Mit anderen Worten, zwischen den quantitativen Methoden ist nicht nur die Anzahl der Deformationen unterschiedlich, sondern die Methoden erkennen selbst bei sehr ähnlicher Prävalenzrate unterschiedliche Wirbelsäulen als deformiert an. Noch größere Unterschiede ergeben sich beim Vergleich der Methoden von Felsenberg 0.80 und Eastell. Die Prävalenzraten der Methoden unterscheiden sich nur um etwa 1%, aber etwa 25% der erkannten Deformationen sind diskordant.

Auch McCloskey fand in einer Untersuchung an einer Querschnittspopulation von 946 Frauen Diskordanzen (22). Die Methoden von Eastell und McCloskey differierten dort in 41% der ermittelten Deformationen. McCloskeys Ergebnisse sind weniger aussagekräftig, da sich die Prävalenzraten der verglichenen Methoden um 6% unterscheiden und nur eine kleine Anzahl von Deformationen vorlag. In der OPUS Population differieren die Methoden von Eastell und McCloskey in 32% der ermittelten Deformationen bei einem Unterschied in den Prävalenzraten von 4%.

### **Kappa Schätzwerte der quantitativen Methoden**

Ein Maß, um die Übereinstimmung der quantitativen Methoden zu bemessen, sind Kappa Schätzwerte. In sie fließen sowohl die unterschiedliche Anzahl wie auch die diskordanten Mengen an ermittelten Deformationen ein. Es wurden die Kappa Schätz-

werte für Wirbelsäulen mit Wirbelkörperdeformationen und für Wirbelkörper mit Deformationen ermittelt (Tabellen 10 und 11). Die Kappa Schätzwerte der quantitativen Methoden, abgesehen von der Methode von Minne, liegen zwischen 0.68 und 0.90. Bei der üblichen Einteilung von Kappa Schätzwerten kann damit von einer guten bis sehr guten Übereinstimmung der quantitativen Methoden untereinander gesprochen werden. Die Methode von Minne dagegen weist mit Kappa Schätzwerten von 0.35 bis 0.47 eine schlechte Übereinstimmung mit allen anderen Methoden auf.

Die Kappa Schätzwerte sind aufgrund der verhältnismäßig großen Anzahl nicht deformierter Wirbelsäulen bzw Wirbelkörper allerdings nur bedingt aussagekräftig. Das Verhältnis der diskordant erkannten Wirbelkörperdeformationen zu den konkordant als nicht deformiert erkannten Wirbelkörper ist so klein, dass die Kappa Schätzwerte hoch sind. Selbst wenn Methoden in 25% der erkannten Deformationen differieren, ergibt sich ein hoher Kappa Schätzwert von 0.69.

In verschiedenen Studien sind Kappa Schätzwerte für quantitative Methoden ermittelt worden (1, 6, 12, 22, 33, 34). Die Kappa Schätzwerte waren denen in dieser Arbeit ähnlich. Sie lagen zwischen 0.4 und 0.9. Die Methode von Minne erzielte in vergleichenden Studien einen deutlich niedrigeren Kappa Schätzwert als alle anderen Methoden.

### **Beurteilung der differierenden Ergebnisse**

Die ermittelten Unterschiede in Prävalenzraten, Verteilung der Deformationen über die Wirbelsäule und Kappa Schätzwerten sind beträchtlich. Die hier dargestellten Ergebnisse bestätigen die Ergebnisse aus einer Vielzahl von Studien. Die Vergleichbarkeit von Prävalenzraten der einzelnen Methoden ist bei Unterschieden von bis zu 200% nicht gegeben. Zusätzlich ist problematisch, dass selbst bei gleicher Anzahl von ermittelten Deformationen die Methoden in mehr als 12% der erkannten Deformationen differieren. Die deutlichen Unterschiede zwischen den Methoden führten zu einer ständigen Neuentwicklung von quantitativen Auswertungsverfahren. Mit dem Resultat, dass eine Vielzahl von quantitativen Methoden im Gebrauch ist, deren Ergebnisse nicht vergleichbar sind. Alle Neuentwicklungen brachten keine grundlegenden Verbesserungen. Wie sich später in der Diskussion zeigen wird, sind aufgrund der fehlenden Spezifität von Höhenminderungen für die Erkennung von osteoporotischen Wirbelkörperfrakturen auch keine grundlegenden Verbesserungen von diesen Methoden zu erwarten.

Hier sollen zunächst die Gründe für die Unterschiede herausgearbeitet werden. Die quantitativen Methoden werden anschließend anhand der qualitativen Auswertung und differentialdiagnostischen Einteilung bewertet. In dieser Bewertung wird dann auf die fehlende Spezifität von Höhenminderungen näher eingegangen.

## Gründe für die unterschiedlichen Ergebnisse der quantitativen Methoden

Es fällt nicht schwer, einen Grund für die unterschiedlichen Ergebnisse der quantitativen Methoden zu finden. Vergleicht man die Schwellenwerte der Methoden von Eastell und Melton und die Anzahl der ermittelten Deformationen, so wird der Einfluss der Schwellenwerte deutlich. Die Schwellenwerte der Methode von Eastell sind höher als die von Melton und so auch die Anzahl der ermittelten Deformationen. Da sich die Methoden lediglich in ihren Schwellenwerten unterscheiden, kommt der Einfluss der Schwellenwerte isoliert zum Tragen. Neben den Schwellenwerten unterscheiden sich die quantitativen Methoden durch ihre Indices. Die Schwellenwerte und die Indices sollen in diesem Abschnitt näher untersucht und ihr Einfluss auf die ermittelten Ergebnisse diskutiert werden.

### Schwellenwerte der quantitativen Methoden

Die Schwellenwerte der quantitativen Methoden unterscheiden sich in zweierlei Hinsicht. Die Art wie die Schwellenwerte ermittelt wurden und ob die Schwellenwerte unterschiedliche Werte für einzelne Wirbelkörper und Indices haben oder nicht. Nur bei Felsenberg wird ein fester Schwellenwert verwendet. Bei allen anderen Methoden unterscheiden sich die Schwellenwerte für Wirbelkörper und Indices. Dieser Unterschied kommt durch die Art zustande, wie die Schwellenwerte ermittelt wurden. Es können folgende Arten unterschieden werden:

- Aus Erfahrung gewählte Schwellenwerte (Felsenberg) ohne Berücksichtigung von Referenzpopulationen.
- Schwellenwerte abgeleitet aus Referenzpopulationen ohne Deformationen. Die Referenzpopulationen wurden erstellt aus Fällen ohne erkennbare Veränderungen auf Röntgenbildern (Eastell und Minne) oder ohne Anhalt auf pathologische Veränderungen in der Anamnese (McCloskey). Die Anzahl der Probanden bei Eastell, McCloskey und Minne ist klein: 76 bei Eastell, 100 bei McCloskey und 72 bei Minne.
- Schwellenwerte abgeleitet aus Querschnittspopulation mit Trimmung der Indices. Indices werden getrimmt, um Indices von deformierten Wirbelkörpern vor der Schwellenwertberechnung heraus zu filtern. Verschiedene Trimmungsverfahren werden verwendet. Bei Melton wurden Schwellenwerte aus einer getrimmten Querschnittspopulation mit 762 Probanden ermittelt.

Anhand der Methoden von Felsenberg, Eastell und Melton soll der Einfluss der Schwellenwerte auf die Ergebnisse wie Prävalenzraten und Verteilung der Deformationen erläutert werden. Anhand der Schwellenwerte von Eastell und Melton zeigen sich

Unterschiede, die durch die zugrunde liegenden Referenzpopulationen zustande kommen. Im anschließenden Abschnitt wird der Einfluss der Referenzpopulationen genauer untersucht.

### **Schwellenwerte bei den Methoden von Felsenberg, Eastell und Melton**

Die Wirbelsäule bildet eine physiologische Kyphose in der mittleren Brustwirbelsäule (Abbildung 1 auf Seite 5). In diesem Bereich ist daher das Verhältnis der anterioren zu den posterioren Höhen niedriger als in anderen Bereichen der Wirbelsäule (Abbildung 10 auf Seite 29). Schwellenwerte, die aus Referenzpopulationen abgeleitet sind, haben daher im Bereich der mittleren Brustwirbelsäule bei den a/p Indices ein Minimum (Abbildung 11). Die Schwellenwerte der Methode von Felsenberg bilden diese physiologischen Variationen der a/p Indices nicht ab. Der Schwellenwert von 0.80 bei Felsenberg ist in der mittleren Brustwirbelsäule deutlich höher als bei den anderen Methoden. Dies bedingt die verhältnismäßig hohe Anzahl von anterioren Deformationen in diesem Bereich (Abbildung 12). Der hohe Schwellenwert ist auch der Grund für den bereits erwähnten überproportional großen Anstieg der anterioren Deformationen von einem Schwellenwert von 0.75 zu 0.80.

Bei den a/p Indices wird der Einfluss der eingesetzten Referenzpopulation zur Ermittlung der Schwellenwerte sichtbar. Die Schwellenwerte bei Eastell wurden aus einer Population ohne radiologische, dh auch ohne degenerative Veränderungen errechnet. Wirbelkörper können bei degenerativen Veränderungen keilförmig verändert sein mit dementsprechend niedrigeren a/p Indices. Bei Melton wurde zur Errechnung der Schwellenwerte eine Querschnittspopulation von Frauen mit einem Alter über 50 Jahren verwendet. In einer solchen Population liegen degenerative Veränderungen der Wirbelkörper vor. Erwartungsgemäß liegen die a/p Indices trotz Trimmung unter denen von Eastell und damit auch die Anzahl der erkannten anterioren Deformationen.

Die mediale Wirbelkörperhöhe zeigt eine weitgehend konstante Zunahme von den Wirbelkörpern T4 zu L4 (Abbildung 10). Die m/p Indices zeigen daher deutlich weniger Variationen. Die Schwellenwerte von Felsenberg liegen relativ nah bei den Schwellenwerten von Eastell und Melton (Abbildung 13). Daraus ergibt sich eine recht ähnliche Anzahl von medialen Deformationen (Abbildung 14).

Die posterioren Höhen nehmen bei den Wirbelkörpern T10 bis L1 im Vergleich mit dem darüberliegenden Wirbelkörper überproportional zu (Abbildung 10). Dadurch bedingt sind die p/pu Indices dieser Wirbelkörper höher und die p/pl Indices niedriger als die der anderen Wirbelkörper (Abbildungen 17 und 18). Auch diese anatomische Gegebenheit bilden die Schwellenwerte von Felsenberg nicht ab. Die Auswirkung der posterioren Schwellenwerte auf die Anzahl der Deformationen lässt sich allerdings aufgrund der geringen Anzahl von posterioren Deformationen in der OPUS Population nicht ermessen.

Die Schwellenwerte haben also einen direkten, signifikanten und deutlichen Einfluss auf die Anzahl der ermittelten Deformationen. Sie erklären sowohl die Unterschiede in der Anzahl wie auch der Verteilung der Deformationen. Je nach eingesetzter Referenz-

population kommt es zu deutlich unterschiedlichen Schwellenwerten und einer deutlich unterschiedlichen Anzahl von Deformationen. Feste Schwellenwerte bilden anatomische Gegebenheiten zumindest bei den a/p Indices in der Brustwirbelsäule nicht ab. Der Schwellenwert von 0.80 bei der Methode von Felsenberg führt dadurch zu einer im Vergleich mit anderen Methoden sehr hohen Anzahl von anterioren Deformationen in der mittleren Brustwirbelsäule.

### **Publizierte Schwellenwerte und Schwellenwerte aus der OPUS Population**

Der Einfluss der Referenzpopulation auf Schwellenwerte und damit die Anzahl der ermittelten Deformationen soll in diesem Abschnitt verdeutlicht werden. Dieser Einfluss zeigte sich bereits an den Schwellenwerten der a/p Indices bei Eastell und Melton. Um den Einfluss auch auf die Methoden von McCloskey und Minne darzustellen, wurden Schwellenwerte aus der OPUS Population ermittelt.

Die OPUS Population stellt eine Querschnittspopulation dar. Die Probandenanzahl ist größer als die bei den Referenzpopulationen von Eastell, Melton, McCloskey und Minne. Da es sich um eine Querschnittspopulation handelt, in der Wirbelkörperdeformationen zu erwarten sind, wurden die Indices vor der Schwellenwertberechnung getrimmt. In einer Studie von Black (4) wurden die beiden am häufigsten verwendeten Verfahren zur Trimmung von Querschnittspopulationen untersucht. Nach Angaben der Autoren unterscheiden sich die von Black (3) und Melton (24) beschriebenen Methoden kaum. Für die Trimmung in dieser Arbeit wurde das Verfahren von Melton verwendet. Die aus der OPUS Population ermittelten Schwellenwerte sind in den Tabellen 12, 13 und 14 aufgelistet. Zur Verdeutlichung sind einige der publizierten zusammen mit den aus der OPUS Population berechneten Schwellenwerte grafisch dargestellt (Abbildungen 15, 16, 17, 18).

Der Einfluss der Trimmung an den Schwellenwerten lässt sich an den Indices der Methode von Melton erkennen. Die Trimmung filtert Ausreißer, dh Indices von deformierten Wirbelkörpern, heraus. Man kann an allen Schwellenwerten sehen, dass sich die Werte durch die Trimmung nach oben verschieben. Diese Verschiebung ist in den Regionen der Wirbelsäule am deutlichsten ausgeprägt, in denen die meisten Deformationen liegen. Beim Schwellenwert des m/p Index ist der Effekt der Trimmung im thorakolumbalen Übergang am stärksten. Die Trimmung ist bei den posterioren Schwellenwerten nur schwach ausgeprägt, da es nur wenige posteriore Deformationen gibt. Vergleichbare Effekte der Trimmung sind in einer Publikation zu Schwellenwerten aus der EVOS Studie zu finden (29).

Die für die Methode von Eastell und Melton ermittelten Schwellenwerte liegen durchweg über publizierten Werten. Diese Differenz ist am stärksten in den posterioren Schwellenwerten ausgebildet, dagegen ist sie bei den medialen Schwellenwerten fast nicht vorhanden. Da die Schwellenwerte aus der OPUS Population über denen der publizierten von Eastell und Melton liegen, ist eine höhere Anzahl von Deformationen zu erwarten. Die für die Methode von McCloskey ermittelten Schwellenwerte sind den publizierten ähnlich nur etwas höher. Es ist also zu erwarten, dass die Prävalenzrate

etwas höher ausfällt. Vergleicht man die errechneten und publizierten Schwellenwerten der Methode von Minne zeigen sich geringe Unterschiede in der Brustwirbelsäule und deutlich grössere Unterschiede in der Lendenwirbelsäule. In der Brustwirbelsäule sind die errechneten Schwellenwerte etwas niedriger und in der Lendenwirbelsäule deutlich niedriger. Die Unterschiede nehmen von der Brustwirbelsäule zur Lendenwirbelsäule kontinuierlich zu und sind beim posterioren Index am deutlichsten ausgeprägt. Es ist also eine deutlich niedrigere Prävalenzrate zu erwarten. Vor allem sollte die Anzahl der Deformationen in der Lendenwirbelsäule niedriger ausfallen.

Die oben prognostizierten Änderungen in den Prävalenzraten und der Verteilung der Deformationen über die Wirbelsäule zeigen sich tatsächlich. Mit den OPUS Schwellenwerten liegt die Prävalenzrate bei Melton deutlich höher (23.04% statt 13.09%), bei McCloskey leicht höher (21.85% statt 17.28%) und bei Minne deutlich niedriger (13.76% statt 34.10%) (Tabelle 15 auf Seite 36). In der Verteilung der Deformationen über die Wirbelsäule sind die hohen Zahlen von Deformationen in der Lendenwirbelsäule bei der Methode von Minne verschwunden (Abbildung 19 auf Seite 37). Angesichts der anderen quantitativen Methoden und der qualitativen Auswertung waren die hohen Zahlen auch nicht nachvollziehbar. Gemessen am Kappa Schätzwert zeigt die Methode von Minne mit den Schwellenwerten aus der OPUS Populationen (0.60 statt 0.39) eine deutlich bessere Übereinstimmung mit den anderen Methoden. Berücksichtigt man zusätzlich die nicht nachvollziehbarere Verteilung der Deformationen über die Wirbelsäule, muss die Validität der von Minne publizierten Schwellenwerte, zumindest in der Lendenwirbelsäule, bezweifelt werden.

Eine Vielzahl von Studien hat die Unterschiede von Schwellenwerten aus Referenzpopulationen untersucht. Sie alle kommen zu dem Ergebnis, dass sich die Schwellenwerte signifikant und deutlich unterscheiden, wenn verschiedene Referenzpopulationen verwendet werden. Die Unterschiede ergeben sich nicht nur, wenn ethnisch verschiedene Referenzpopulationen berücksichtigt werden (32), sondern auch wenn Referenzpopulationen einer ethnischen Gruppe verwendet werden (29, 32, 35, 36). Beim Vergleich von einzelnen Zentren der EVOS Population ergaben sich beispielsweise Unterschiede von bis zu 10% bei den Schwellenwerten (29).

### **Beurteilung der Schwellenwerte**

Schwellenwerte aus unterschiedlichen Referenzpopulationen haben also einen deutlichen Einfluss auf Prävalenzraten und Verteilung der Deformationen. Es zeigen sich zwei Probleme. Erstens ergeben einige Referenzpopulationen nicht valide Werte wie die von Minne. Zweitens kommt es selbst bei nachvollziehbaren Schwellenwerten, die aus großen Querschnittspopulationen stammen und mit identischen Verfahren getrimmt wurden, zu deutlichen Unterschieden. Das zeigt sich bei der Methode von Melton mit den publizierten Schwellenwerten und denen aus der OPUS Population. Der Unterschied der ermittelten Prävalenzraten beträgt 175%. Daraus lässt sich folgern, dass Ergebnisse wie Prävalenzraten mit Schwellenwerten aus unterschiedlichen Referenzpopulationen nicht miteinander vergleichbar sind. Bei Studien zur Osteopo-

rose ist es aber zwingend erforderlich, die Anzahl von osteoporotischen Wirbelkörperfrakturen in unterschiedliche Gruppen zu vergleichen. Die Frage ist dann, welche Schwellenwerte bzw Referenzpopulationen verwendet werden sollen.

Es wird häufig vorgeschlagen, Schwellenwerte für die jeweilig untersuchte Population zu ermitteln und zu verwenden. Selbst wenn dies zu vergleichbaren Daten führen würde, ist ein solches Vorgehen meist nicht durchführbar. Anwendbar wäre es nur bei großen, randomisierten Querschnittspopulationen. Eine Beurteilung von Deformationen könnte nur erfolgen, wenn alle Röntgenbilder ausgewertet wären. Bedingungen, denen viele Studien nicht genügen können, zumal die Beurteilungen von Deformationen während der Studie meist schon aus klinischen Gründen erforderlich ist. Risikopopulationen, zB aus Probanden mit mindestens einer osteoporotischen Wirbelkörperfraktur, ließen sich grundsätzlich nicht erstellen.

Auch andere Autoren, wie Black, haben auf die Probleme von aus Populationen ermittelten Schwellenwerten hingewiesen (5). Black stellte fest, dass angesichts der großen Unterschiede bei den Schwellenwerten und der Schwierigkeit repräsentative Schwellenwerte zu ermitteln, es eher zu empfehlen ist, auf feste Werte wie bei der Methode von Felsenberg zurückzugreifen. Dem kann angesichts der Ergebnisse dieser Arbeit nur zugestimmt werden. Eine endgültige Beurteilung der Schwellenwerte wird erst durch die Einbeziehung der qualitativen Auswertung und der differentialdiagnostischen Einteilung quantitativ erkannter Deformationen möglich sein.

### Indices der quantitativen Methoden

Bei der Untersuchung der Schwellenwerte war es möglich den Einfluss der Schwellenwerte auf die Prävalenzraten und andere Ergebnisse zu isolieren. Dazu wurden unterschiedliche Schwellenwerte bei ein und dem selben Index untersucht. Es ist dagegen nicht möglich bei gleichem Schwellenwert unterschiedliche Indices zu betrachten, da jeder Index seine ihm eigenen Schwellenwerte bedingt. Analysiert man die Indices aber für sich und nicht anhand ihren Auswirkungen auf Prävalenzraten, so zeigen sich Probleme bei einzelnen Methoden, auf die hier eingegangen werden soll.

#### Indices der Methode von Minne

Bei der Methode von Minne (20, 26) werden die Höhen aller Wirbelkörper auf die Höhe des Brustwirbelkörpers T4 bezogen. Für den Wirbelkörper T4 ergibt sich daher immer ein Index von 1.0, egal ob der Wirbelkörper deformiert ist oder nicht. Dieser Wirbelkörper kann also mit der Methode von Minne grundsätzlich nicht beurteilt werden. Kann der Wirbelkörper T4 nicht gemessen werden oder ist er deformiert, können sämtliche Wirbelkörper dieser Wirbelsäule nicht mehr ausgewertet werden, da sich keine oder falsch hohe Indices für alle Wirbelkörper ergeben.

Der Bezug der Höhen auf den Wirbelkörper T4 erscheint angesichts der Daten dieser Arbeit problematisch. Der Wirbelkörper T4 ist der mit 12.24% am häufigsten nicht beurteilbare Wirbelkörper (Tabelle 17 auf Seite 38). Andere Studien weisen ebenfalls

daraufhin hin, dass der Wirbelkörper T4 der am häufigsten nicht zu beurteilende Wirbelkörper ist (17, 28). Zwei Gründe sind hauptsächlich dafür verantwortlich. Der Wirbelkörper T4 liegt bei der Aufnahme des Röntgenbildes entfernt vom Zentralstrahl, wird häufiger als andere Wirbelkörper verkippt dargestellt. Außerdem liegt der Wirbelkörper T4 so weit oben in der Brustwirbelsäule, dass es häufig zu Überlagerungen mit der oberen Extremität kommt. Gerade diesen Wirbelkörper als Bezugspunkt auszuwählen erscheint fragwürdig. Es wäre besser, andere Wirbelkörper wie zB T12 als Bezugspunkt zu verwenden. Allerdings gibt es noch ein weiteres Problem, dass auch bei jedem anderen Wirbelkörper zum Tragen kommen würde.

Der Wirbelkörper T4 kann nicht zur Berechnung der Indices herangezogen werden, sofern er deformiert ist, da sich sonst falsch hohe Indices für alle anderen Wirbelkörper ergeben. Die Feststellung einer Deformation des Wirbelkörpers T4 muss allerdings qualitativ erfolgen, da die Methode von Minne kein Verfahren zur quantitativen Beurteilung von T4 enthält. Dieses Problem betrifft jeden Wirbelkörper, der als Bezugspunkt gewählt wird. In der qualitativen Auswertung wurde bei 37 von 2137 auswertbaren Wirbelkörpern T4 eine osteoporotische Wirbelkörperfraktur festgestellt. Die quantitative Methode von Felsenberg ermittelt 37 deformierte Wirbelkörper T4, die von Melton 33 und die von McCloskey 38.

Insgesamt lassen sich also mit der Methode von Minne in der OPUS Population 13.76% der Wirbelsäulen nicht beurteilen. Das ist weit mehr als bei allen anderen Methoden mit denen sich jeweils weniger als 1% der Wirbelkörper quantitativ nicht beurteilen lassen (Abschnitt "Quantitativ nicht beurteilbare Wirbelkörper" auf Seite 39). Darüber hinaus hängt die Feststellung von Deformationen aller Wirbelkörper letztlich von der qualitativen Einschätzung des Wirbelkörpers T4 ab. Dies erscheint widersinnig, da die quantitativen Methoden entwickelt wurden, um eine Auswertung zu gewährleisten, die objektiver als die qualitative ist.

### **Indices der Methoden von Felsenberg, Eastell, Melton**

Die Methoden von Felsenberg, Eastell, Melton und McCloskey umgehen das für die Methode von Minne beschriebene Problem, indem die Höhen eines Wirbelkörpers auf die Höhen des gleichen oder einiger angrenzender Wirbelkörpers bezogen werden. Aus einer Vielzahl von Querschnittsstudien weiß man, dass die posteriore Höhe eines Wirbelkörpers nur selten verändert ist (28). In der OPUS Studie wurden nur 2 kompressionsförmige osteoporotische Frakturen festgestellt. Aus diesem Grund werden die anteriore und mediale Höhe auf die posteriore Höhe des gleichen Wirbelkörpers bezogen und es ergeben sich die Indices  $a/p$  und  $m/p$ .

Ein Problem der Indices  $a/p$  und  $m/p$  besteht darin, dass bei kompressionsförmigen Deformationen alle Höhen, einschließlich der posterioren, vermindert sein können. Damit können die Indices  $a/p$  und  $m/p$  normal erscheinen, obwohl alle Höhen erniedrigt sind. Die Indices  $a/p$  und  $m/p$  reichen aus, um keil-, konkav-, und bikonkavförmige Deformationen zu erkennen. Kompressionsförmige Deformationen können nicht erkannt werden, wenn die  $a/p$  und  $m/p$  Indices nicht vermindert sind. Um auch solche

Deformationen erkennen zu können, führen die Methoden unterschiedliche Indices ein. Die Methoden von Felsenberg, Eastell bzw Melton setzten die posteriore Höhe eines Wirbelkörpers mit denen der beiden angrenzenden in Beziehung und errechnen dazu die Indices  $p/pu$  und  $p/pl$ . Anhand der OPUS Population wurde ermittelt, wie viele der posterioren Deformationen auch über die Indices  $a/p$  und  $m/p$  erkannt werden. Je nach Methode sind es zwischen 85% bis 100% (Tabelle 19 auf Seite 40). Angesichts dieser Daten haben die posterioren Indices  $p/pu$  und  $p/pl$  nur einen kleinen Stellenwert in der Erkennung von kompressionsförmigen Deformationen. Bei zwei kompressionsförmigen Deformationen drängt sich die Frage auf, ob die Erkennung von kompressionsförmigen Deformationen überhaupt notwendig ist. Da aber solche Frakturen in der Regel schwere Frakturen sind, die mit einem deutlich erhöhten Risiko an weiteren osteoporotische Frakturen einhergehen (9, 15, 21), sollten sie in jedem Fall berücksichtigt werden.

### Indices der Methode von McCloskey

McCloskey beschreibt in seiner Publikation die Möglichkeit von mehreren aneinander grenzenden posterioren Deformationen (22, 23). Bei mehr als zwei aneinander grenzend liegenden Deformationen würden die Methoden von Felsenberg, Eastell und Melton bei der Erkennung von mindestens einer posterioren Deformation versagen. In der OPUS Studie liegen keine aneinander angrenzende posterioren Deformationen vor, insbesondere nicht mehrere aneinander angrenzende. Bedenkt man zusätzlich, dass nahezu alle posterioren Deformationen auch mit den Indices  $a/p$  und  $m/p$  erkannt werden konnten, so kann davon ausgegangen werden, dass die von McCloskey erwähnte Problematik bei der Erkennung von osteoporotischen Wirbelkörperfrakturen keine Rolle spielt.

Um das Problem der aneinander grenzenden posterioren Deformationen zu umgehen, führt McCloskey die predicted posterior height ein. Auf diese zu erwartende posteriore Höhe werden bei McCloskey alle anderen Höhen bezogen. Nur errechnet sich diese zu erwartende Höhe aus den posterioren Höhen der angrenzenden Wirbelkörper und ist damit auch vom Problem der aneinander grenzenden posterioren Deformationen betroffen. Darüber hinaus ist die Ermittlung der predicted posterior heights abhängig von Referenzwerten für posteriore Höhen. Diese Werte müssen aus Referenzpopulationen errechnet werden und wie im vorangegangenen Abschnitt gezeigt wurde, sind aus Referenzpopulationen abgeleitete Werte zu vermeiden.

### Beurteilung der Indices

Die Indices der Methode von Minne sind zu vermeiden aufgrund ihres Bezugs auf einen einzigen Wirbelkörper. Insbesondere ist die Wahl des Wirbelkörpers T4 als Bezugspunkt nicht sinnvoll, da er der am häufigsten nicht auswertbare Wirbelkörper ist. Dass bei einer quantitativen Auswertung die Auswertung der gesamten Wirbelsäule letztlich von der qualitativen Beurteilung eines Wirbelkörpers abhängt, ist widersinnig. Das Problem der Methode von Minne wird von den Methoden von Felsenberg, Eastell bzw Melton umgangen. Die Indices der Methoden berücksichtigen das Problem

der posterioren Deformationen hinreichend. Zumindest für die Berechnung der Indices sind die Methoden von Felsenberg, Eastell bzw Melton unabhängig von Referenzpopulationen. Mehrere aneinander grenzende posteriore Wirbelkörperdeformationen liegen zumindest in der OPUS Studie nicht vor. Abgesehen davon, wäre auch die Methode von McCloskey von diesem Problem betroffen. Sie besitzt daher keine Vorteile gegenüber den Methoden von Felsenberg, Eastell bzw Melton. Sie erscheint stattdessen problematisch, da zur Berechnung der Indices Werte aus einer Referenzpopulation notwendig sind.

## Vergleich der quantitativen Methoden mit der qualitativen Auswertung

Ziel der Arbeit ist es nicht nur, quantitative Methoden miteinander zu vergleichen. Die Methoden sollen auch danach bewertet werden, wie gut sie osteoporotische Wirbelkörperfrakturen erkennen. Als Bezugspunkt für die Beurteilung der Methoden ist in dieser Arbeit eine qualitative Auswertung gewählt worden. In den Publikationen zur Etablierung der einzelnen Methoden wird darauf hingewiesen, dass ein Goldstandard zur Erkennung von osteoporotischen Wirbelkörperfrakturen fehlt (6, 8, 13, 22). Keine der quantitativen Methoden ist als Referenzmethode anerkannt. Auch die qualitative Auswertung ist aufgrund ihrer Subjektivität nicht als Goldstandard anerkannt. Meistens werden indirekte Zeichen von Wirbelkörperfrakturen als Bezugspunkt einer Bewertung verwendet. Sowohl klinische Zeichen, wie Rückenschmerzen oder Körpergrößenminderung, als auch andere Messverfahren zur Osteoporosediagnostik, wie Knochendichte, sind verwendet worden (6, 8, 10). Es wäre zwar zu beweisen, erscheint aber offensichtlich, dass die Validierung von quantitativ erkannten Deformationen am Röntgenbild eher angemessen ist, als die Erfragung von Körpergrößenminderung oder Rückenschmerzen. Hinsichtlich der Validierung anhand der Knochendichte, muss angemerkt werden, dass in den letzten Jahren der direkte Zusammenhang von Knochendichte und Wirbelkörperfrakturen zunehmend in Zweifel gezogen wird. In jedem Fall erscheint die direkte Beurteilung von einzelnen Wirbelkörpern am Röntgenbild die angemessenere Methode zur Validierung quantitativ erkannter Deformationen. Leider gibt es keine Studien die quantitative Methoden anhand von qualitativen Auswertungen oder differentialdiagnostischen Einteilungen der ermittelten Deformationen bewerten. Ein Vergleich mit anderen Publikationen ist daher nicht möglich.

In der Einleitung ist bereits erwähnt worden, dass die Verminderung der Wirbelkörperhöhe kein pathognomisches Zeichen für osteoporotische Wirbelkörperfrakturen ist. Höhenminderungen treten bei einer Vielzahl von Wirbelkörperveränderungen auf, wie zB bei traumatischen und degenerativen Prozessen oder einer persistierenden Chorda dorsalis (31). Es ist also zu erwarten, dass die quantitativen Methoden Deformationen erkennen, die keine osteoporotischen Wirbelkörperfrakturen sind. Daher soll geklärt

werden, ob die Methoden alle osteoporotischen Wirbelkörperfrakturen erkennen oder ob sie welche übersehen. In der weiteren Diskussion ist es wichtig, sich zu vergegenwärtigen, dass mit der qualitativen Auswertung nur osteoporotische Wirbelkörperfrakturen, mit den quantitativen Methoden dagegen Deformationen jeglicher Genese gezählt wurden.

### **Anzahl und Verteilung von osteoporotischen Frakturen**

Mit der qualitativen Auswertung ergibt sich eine Prävalenzrate von 14.36% für osteoporotische Wirbelkörperfrakturen. Die Rate liegt im unteren Bereich der quantitativen Methoden. Sie ist den Prävalenzraten der Methoden von Felsenberg 0.75 (12.92%) und Melton (13.09%) am ähnlichsten (Tabelle 6 auf Seite 25). Die Verteilung der Deformationen über die Wirbelsäule gleicht der der quantitativen Methoden. Die Verteilung ist zweigipflig, mit einem Gipfel in der mittleren Brustwirbelsäule und einem im thorakolumbalen Übergang. Die Methode von Felsenberg mit einem Schwellenwert von 0.80 ermittelt allerdings etwa doppelt so viele Deformationen in der mittleren Brustwirbelsäule. Die Methode von Minne ermittelt im gesamten Bereich deutlich mehr Deformationen. Dies betrifft vor allem die Lendenwirbelsäule.

### **Kappa Schätzwerte**

Um die Unterschiede der quantitativen Methoden bewerten zu können, sind Kontingenztabellen und Kappa Schätzwerte bestimmt worden. Wenn im Folgenden Deformationen als falsch positiv bezeichnet werden, so sind dies quantitativ ermittelte Deformationen, die in der qualitativen Auswertung nicht als osteoporotische Wirbelkörperfrakturen angesehen wurden. Umgekehrt werden osteoporotische Wirbelkörperfrakturen, die mit den quantitativen Methoden nicht erkannt wurden, als falsch negativ bezeichnet.

Gemessen an den Kappa Schätzwerten zeigen alle Methoden, mit Ausnahme der Methode von Minne, eine gute bis sehr gute Übereinstimmung mit der qualitativen Auswertung (Tabelle 20 auf Seite 41). Die beste Übereinstimmung der publizierten Methoden zeigt die von Melton mit einem Kappa Schätzwert von 0.86, die schlechteste Übereinstimmung die Methode von Minne mit einem Wert von 0.43. In der Tabelle finden sich zwei Schwellenwerte mit denen die Methode von Felsenberg höhere Kappa Schätzwerte erzielt als die Methode von Melton. Diese Schwellenwerte sind außerhalb dieser Arbeit bisher nicht eingesetzt worden. Sie werden in diese Arbeit lediglich verwendet, um die hohe Anzahl von Deformationen in der mittleren Brustwirbelsäule bei Felsenberg 0.80 näher untersuchen zu können. Auf diese Schwellenwerte wird weiter unten näher eingegangen.

Die Spezifität erscheint durchweg sehr hoch. Bei der Spezifität ist allerdings die gleiche Anmerkung zu machen wie beim Vergleich der quantitativen Methoden zu den Kappa Schätzwerten. Das Verhältnis deformierter Wirbelkörper zur Gesamtzahl von Wirbelkörpern ist sehr klein. Bei der Methode von Melton beispielsweise 582 zu 30575 oder

1.90%. Der Anteil der falsch negativen Deformationen ist noch kleiner mit 52 zu 29925 oder 0.17%. Selbst wenn sich die Anzahl der falsch negativen Deformationen verdoppelt, kommt es nur zu einer sehr geringen Abnahme der Spezifität von 0.9982 auf 0.9965. Die Spezifität ist daher bei allen Methoden hoch, obwohl es zwischen den Methoden verhältnismäßig große Differenzen in der absoluten Zahl an unterschiedlich erkannten Deformationen gibt.

### **Falsch negativ und falsch positiv erkannte Deformationen**

Ein genaueres Bild ergibt sich daher, wenn man die absolute Anzahl der falsch positiven und falsch negativen Deformationen betrachtet. Das beste Ergebnis wäre eine Methode, die weder falsch positive noch falsch negative Deformationen aufweist und alle osteoporotischen Wirbelkörperfrakturen erkennt. Die Methode von Felsenberg mit einem Schwellenwert von 0.80 ermittelt nur eine einzige falsch negative Deformation, dh sie hat bis auf eine osteoporotische Wirbelkörperdeformation alle erkannt. Allerdings hat sie mit 214 falsch negativen Deformationen relativ viele nicht osteoporotische Deformationen ermittelt. Die Methode von Felsenberg mit einem Schwellenwert von 0.75 und die Methode von Melton ermitteln nur 52 falsch negative Deformationen. Allerdings werden mit beiden Methoden eine beträchtliche Anzahl osteoporotischer Wirbelkörperfrakturen nicht erkannt. Bei Felsenberg 0.75 sind es 12.62% der osteoporotische Wirbelkörperfrakturen und bei Melton 18.46%.

Der Zusammenhang zwischen falsch positiv und falsch negativ erkannten Deformationen und ihre Abhängigkeit von Schwellenwerten zeigt sich an der Methode von Felsenberg deutlich. Zur Veranschaulichung des Zusammenhangs wurden neben den Schwellenwerten 0.75 und 0.80 noch die Schwellenwerte 0.775 und 0.825 untersucht. Ist der Schwellenwert niedrig gewählt, so ist die Anzahl falsch positiver Deformationen gering und die der falsch negativen Deformationen hoch. Bei einem Schwellenwert von 0.75 werden 52 Deformationen falsch positiv und 145 falsch negativ erkannt. Ist der Schwellenwert hoch gewählt, so ist die Anzahl falsch positiver Deformationen gering und die der falsch negativen Deformationen hoch. Bei einem Schwellenwert von 0.825 werden 1353 Deformationen falsch positiv und keine falsch negativ erkannt.

Ein Ziel wäre es die Anzahl falsch erkannter Deformationen so niedrig wie möglich zu halten und für das Verhältnis von falsch positiven und falsch negativen Deformationen ein möglichst optimales Verhältnis zu finden. Die Kappa Schätzwerte sind hierfür ein quantitatives Maß. Nach ihnen ist, gemessen an der qualitativen Auswertung, von den publizierten Methoden die von Melton die beste zur Erkennung von prävalenten osteoporotischen Wirbelkörperfrakturen. Allerdings werden mit der Methode knapp 20% der osteoporotische Wirbelkörperfrakturen übersehen und das kann nicht Ziel einer quantitativen Methode sein.

### **Fehlende Spezifität von Höhenminderungen**

Angesichts der Daten scheint eine optimale Methode nicht zu existieren. Eine Veränderung der Schwellenwerte wird aus einem einfachen Grund auch keine optimale Methode ergeben. Die Höhenminderung ist kein pathognomisches Zeichen für eine osteoporotische Wirbelkörperfraktur. Wäre sie es und käme sie nur bei osteoporotischen Wirbelkörperfrakturen vor, so könnte man durch die richtige Wahl des Schwellenwertes eine optimale Methode erhalten. Ein Zugewinn an Spezifität und damit diagnostischer Güte ist nur zu erwarten, wenn andere Faktoren wie Strahlentransparenz, Trabekelzeichnung, Dicke der Kortikalis und weitere radiologisch differentialdiagnostische Kriterien berücksichtigt werden.

Die beste Methode lässt sich daher weniger durch die Wahl der Indices oder des Schwellenwertes nach dem besten Kappa Schätzwert erzielen, als vielmehr durch die Wahl eines Schwellenwertes mit dem möglichst alle osteoporotischen Deformationen erkannt werden und einer anschließenden differentialdiagnostischen Beurteilung aller ermittelten Deformation. Diese anschließende Beurteilung würde die oben erwähnten differentialdiagnostische Kriterien berücksichtigen. Damit könnten alle nicht osteoporotischen Wirbelkörperdeformationen herausgefiltert werden. Einen vergleichbaren Zugewinn an Spezifität für osteoporotische Wirbelkörperdeformationen zu erzielen, ist aufgrund der fehlenden Spezifität der Höhenminderung von keiner quantitativen Methode zu erwarten.

Anhand der differentialdiagnostischen Einteilung der quantitativ Deformationen soll auf die fehlende Spezifität der Höhenminderung näher eingegangen werden und geklärt werden, um welche Arten von Deformationen es sich bei den falsch positiven Deformationen handelt.

## **Differentialdiagnose quantitativ ermittelter Deformationen**

### **Anteil osteoporotischer und nicht osteoporotischer Deformationen**

Die Ergebnisse der differentialdiagnostischen Einteilung sind in Tabelle 21 auf Seite 42 dargestellt. Betrachtet man die Ergebnisse für die Methode von Melton zeigt sich, dass es sich bei 91.07% aller ermittelten Deformationen tatsächlich um osteoporotische Wirbelkörperfrakturen handelt. Ein ähnlich gutes Ergebnis erzielt auch die Methode von Felsenberg 0.75. Bei der Methode von Felsenberg mit einem Schwellenwert von 0.80 zeigt sich, dass es sich nur bei 75.20% der ermittelten Deformationen um osteoporotische Wirbelkörperfrakturen handelt. 21.32% der ermittelten Deformationen sind degenerativ bedingt.

Wie im vorangegangenen Abschnitt soll der Einfluss der Schwellenwerte anhand der Methode von Felsenberg dargestellt werden. Erhöht man den Schwellenwert von 0.75 auf 0.80 nimmt die Anzahl der erkannten osteoporotischen Frakturen von 505 auf 649 zu. Gleichzeitig nimmt aber auch die Anzahl der erkannten degenerativen Deformationen überproportional von 33 auf 184 zu. Erkennt die Methode also mehr osteoporotische Frakturen, erkennt sie auch deutlich mehr degenerative Deformationen. Anders als die Anzahl der degenerativen Deformationen ändert sich die Zahl der ermittelten traumatischen Frakturen bei Änderung des Schwellenwertes nicht. Die Anzahl der ermittelten traumatischen Frakturen ist bei allen Methoden und allen Schwellenwerten identisch.

Es ist noch eine kurze Anmerkung zu den nicht zu bestätigenden Deformationen zu machen, dh zu den Deformationen in der Kategorie "keine" in Tabelle 21. Die von den Methoden von Felsenberg und Melton ermittelten Deformationen konnten in der differentialdiagnostischen Auswertung bis auf wenige Ausnahmen als Deformationen bestätigt werden. Die Methoden von Eastell, McCloskey und vor allem Minne ermitteln dagegen eine erhebliche Anzahl nicht nachvollziehbarer Deformationen. Wie aus der Abbildung 20 zu entnehmen ist, sind es bei Eastell vor allem Deformationen der Wirbelkörper T4 und L3 und bei McCloskey Deformationen der Wirbelkörper T6 und L4. Die Schwellenwerte dieser Wirbelkörper sind jeweils deutlich höher als die aller anderen Wirbelkörper. Bei der Methode von Minne sind in der gesamten Wirbelsäule vor allem aber in der Lendenwirbelsäule eine große Anzahl von nicht nachvollziehbaren Deformationen zu finden (Abbildung 20). Angesichts der Ergebnisse der differentialdiagnostischen Einteilung muss davon ausgegangen werden, dass mit den eingesetzten Populationen keine validen Schwellenwerte erzielt wurden. Für die Methoden von Eastell und McCloskey betrifft das einzelne Wirbelkörper, für die Methode von Minne allerdings die gesamte Wirbelsäule.

### **Indices osteoporotisch und degenerativ bedingter Deformationen**

Die fehlende Spezifität der Höhenminderung wird deutlich, wenn man die Indices deformierter Wirbelkörper unterschiedlicher Genese miteinander vergleicht. Für die OPUS Population sind dazu die a/p und m/p Indices der Wirbelkörper getrennt nach den Differentialdiagnosen osteoporotisch, degenerativ und traumatisch zusammengestellt worden. Die Abbildungen 21 und 22 stellen die Ergebnisse dar. Traumatische Frakturen weisen so große Höhenminderungen auf, dass sich Indices ergeben, die weit unter allen eingesetzten Schwellenwerten liegen. Degenerative Deformationen weisen nur leichte Höhenminderungen auf und es ergeben sich Indices die um 0.80 liegen. Osteoporotische Frakturen zeigen anders als degenerative Deformationen und traumatische Frakturen Höhenminderungen und damit Indices unterschiedlichster Art. Sie reichen von leichten Höhenminderungen, die im Bereich der degenerativen Deformationen liegen, bis zu großen Höhenminderungen ähnlich denen der traumatischen Frakturen. Eine Differenzierung zwischen degenerativen Deformationen, tra-

matischen Frakturen und osteoporotischen Frakturen ist daher aufgrund von Höhenminderungen nicht möglich. Setzt man beispielsweise den eingesetzten Schwellenwert niedrig, um möglichst wenige degenerative Deformationen zu erkennen, übersieht man dabei gleichzeitig die leichteren osteoporotischen Frakturen, die ähnlich geringe Höhenminderungen aufweisen. Wenn man die Auswertung allein auf eine Höhenminderung stützt, muss man entweder eine beträchtliche Anzahl von nicht osteoporotischen Deformationen oder man eine beträchtliche Anzahl übersehener osteoporotische Frakturen akzeptieren.

Die Anzahl von osteoporotischen Wirbelkörperdeformationen mit einem a/p oder m/p Index über 0.8 ist in den Abbildungen 21 und 22 so groß, dass das Ergebnis von einer falsch negativ erkannter Deformation mit der Methode Felsenberg 0.80 falsch erscheint. Diese scheinbare Diskrepanz ist folgendermaßen zu erklären. Bei konkavförmigen, osteoporotischen Wirbelkörperfrakturen kann der a/p Index über 0.8 liegen, wenn die Fraktur anhand des m/p Index erkannt wird. Genauso kann der m/p Index einer keilförmigen, osteoporotischen Wirbelkörperfraktur über 0.8 liegen, wenn die Fraktur anhand des a/p Index erkannt wird.

### Deformationen der mittleren Brustwirbelsäule

Das Problem ähnlicher Indices bei degenerativen und osteoporotischen Deformationen zeigt sich bei der Methode von Felsenberg mit unterschiedlichen Indices besonders deutlich. Die Änderung des Schwellenwertes von 0.75 zu 0.80 führt dazu, dass bis auf eine osteoporotische Fraktur alle erkannt werden. Gleichzeitig nimmt die Anzahl der falsch positiv erkannten Deformationen von 52 auf 214 zu. Betrachtet man die Verteilung der Deformationen, ist die größte Zunahme an erkannten Deformationen in der mittleren Brustwirbelsäule sichtbar. Es wurde bereits gezeigt, dass diese Zunahme vor allem durch die hohe Anzahl der anterioren Deformationen in diesem Bereich zustande kommt. Grund ist der verhältnismäßig hohe Schwellenwert von 0.80 in diesem Bereich. Aus der differentialdiagnostischen Einteilung der quantitativ erkannten Deformationen ergibt sich, dass die Methode mit einem Schwellenwert von 0.80 im Bereich der mittleren Brustwirbelsäule eine große Anzahl von degenerativen Deformationen erkennt (Abbildung 20 auf Seite 43).

Eine Möglichkeit die Methode von Felsenberg 0.80 zu optimieren, wäre es isoliert nur die Schwellenwerte der a/p Indices in der mittleren Brustwirbelsäule, dh bei T6, T7 und T8, zu vermindern. So könnte die Erkennung der vielen falsch positiven, degenerativen Deformationen in diesem Bereich umgangen werden. Um diesen Ansatz zu untersuchen sind für die Methode von Felsenberg zwei weitere Schwellenwerte untersucht worden, die hier mit Felsenberg 0.80 (0.75) und Felsenberg 0.80 (0.775) bezeichnet werden. Für die a/p Indices im Bereich der Wirbelkörper T6, T7 und T8 sind dabei Schwellenwerte von 0.75 bzw 0.775 gewählt worden, während alle anderen Schwellenwerte bei 0.80 belassen wurden. Aufschlussreich ist, dass mit diesen Schwellenwerten keine wirklichen Vorteile erreicht werden. Zwar erzielt die Methode von Felsenberg mit den beiden Schwellenwerte die besten Kappa Schätzwerte von allen untersuchten

Methoden, aber sie übersieht 39 (6%) bzw 26 (4%) osteoporotische Wirbelkörperfrakturen (Tabelle 20 auf Seite 41). Offensichtlich liegen im Bereich der mittleren Brustwirbelsäule osteoporotische Wirbelkörperfrakturen vor, die nur geringe Höhenminderungen aufweisen und deren Indices nur wenig unter dem Schwellenwert von 0.80 liegen. Eine Senkung der Schwellenwertes in diesem Bereich auf 0.75 bzw 0.775 führt neben einer Abnahme der erkannten degenerativen Deformationen gleichzeitig zu einer Abnahme der erkannten osteoporotischen Deformationen. Die hohe Anzahl von Deformationen im Bereich der mittleren Brustwirbelsäule, die der hauptsächliche Kritikpunkt an dem Schwellenwert von 0.80 war, muss also in Kauf genommen werden, um möglichst alle osteoporotischen Wirbelkörperfrakturen zu erkennen.

### **Kombination quantitativer und qualitativer Methoden**

Es wäre falsch, die quantitativen Methoden abzulehnen, nur weil sie nicht in der Lage sind, osteoporotisch, degenerativ und traumatisch bedingte Deformationen zu unterscheiden. Die quantitativen Methoden bieten den Vorteil eines objektiven Kriteriums, ab wann eine Deformation als pathologisch anzusehen ist. Gerade bei osteoporotischen Wirbelkörperfrakturen, die ein großes Kontinuum an Höhenminderungen aufweisen, ist dies ein entscheidender Punkt zur Objektivierung der Diagnose von osteoporotischen Wirbelkörperfrakturen. Das Kriterium kann völlig zu Recht die Höhenminderung sein. Denn wie sich auch in dieser Arbeit zeigte, weist jede osteoporotische Wirbelkörperfraktur eine Höhenminderung auf. Nur ist nicht jede Höhenminderung auch eine osteoporotische Fraktur. Welcher Schwellenwert eine pathologische Höhenminderung am besten erfasst, lässt sich ebenfalls an dieser Arbeit ableiten. Es ist ein Schwellenwert von 0.80 bei der Methode von Felsenberg. Denn mit diesem Schwellenwert wird nur eine einzige osteoporotische Wirbelkörperfraktur übersehen. Der Schwellenwert von 0.825, mit dem alle osteoporotischen Wirbelkörperfrakturen erkannt wurden, stellt sich nicht als praktikabel dar, da eine unverhältnismäßig hohe Anzahl von falsch positiven Deformationen erkannt wird.

Durch die Kombination einer quantitativen Auswertung und einer anschließenden differentialdiagnostischen Einteilung der ermittelten Deformationen können die Vorzüge einer quantitativen Methode mit denen einer qualitativen Auswertung kombiniert und die Nachteile beider Methoden minimiert werden. Vorteil der quantitativen Methode ist ihre Objektivität in der Entscheidung, ob eine Formveränderung bereits pathologisch ist oder nicht, und Nachteil ihre fehlende Spezifität, da Höhenminderungen kein pathognomisches Zeichen für eine osteoporotische Wirbelkörperfraktur sind. Vorteil der qualitativen Auswertung ist ihre hohe Spezifität und Nachteil ihre fehlende Objektivität in der Entscheidung, ob eine Formveränderung als pathologisch anzusehen ist oder nicht.

## Bewertung der quantitativen Methoden

Bei der Bewertung der quantitativen Methoden anhand der qualitativen Auswertung und differentialdiagnostischen Einteilung ist eine Einschränkung notwendig. Diese Auswertungen sind nicht als Goldstandard zur Erkennung von osteoporotischen Wirbelkörperfrakturen anerkannt und gelten als subjektiv. Die Frage ist, ob die Ergebnisse dieser Arbeit generalisierbar sind oder ob sich bei der qualitativen Auswertung durch einen anderen Auswerter ganz andere Bewertungen ergeben würden. Zunächst einmal sind die meisten Ergebnisse dieser Arbeit unabhängig von der qualitativen Auswertung. Die Feststellung gravierender Differenzen zwischen den quantitativen Methoden, die Problematik von Schwellenwerten aus Referenzpopulationen, die Kritik an den Indices der Methoden von Minne und McCloskey ergaben sich bereits aus dem Vergleich der quantitativen Methoden untereinander. Die fehlende Spezifität von Höherminderungen für osteoporotische Wirbelkörperfrakturen ist eine anerkannte radiologische Tatsache. Die differentialdiagnostische Einteilung belegte diese Tatsache und verdeutlichte die daraus resultierende Probleme. Die Schlüsse die aus der differentialdiagnostischen Einteilung gemacht wurden, sind nicht abhängig von einer bestimmten Anzahl von Differentialdiagnosen in der Auswertung. Einzig und allein der genaue Schwellenwert, mit dem alle osteoporotischen Wirbelkörperfrakturen erkannt werden können, ist abhängig von einer bestimmten Menge von osteoporotischen Wirbelkörperfrakturen in der qualitativen Auswertung. Bei diesem Schwellenwert ist eine Abweichung je nach Auswerter möglich. Der Wert von 0.80 der in dieser Arbeit als der optimalste ermittelt wurde, wäre daher anhand von qualitativen Auswertungen anderer Untersucher zu validieren. Solch eine Validierung sollte durch Personen erfolgen, die in der Diagnose von osteoporotischen Wirbelkörperfrakturen möglichst erfahren sind. Dies erfordert die Zusammenarbeit von radiologischen Auswertungszentren, die sich auf die Diagnose von osteoporotischen Wirbelkörperfrakturen spezialisiert haben und geht daher über diese Arbeit hinaus. Es lässt sich aber feststellen, dass alle Ergebnisse bis auf den exakten Schwellenwert von 0.80 nicht von einer möglichen Subjektivität der qualitativen Auswertung abhängig sind.

Beim Vergleich der quantitativen Methoden untereinander zeigte sich der maßgebliche Einfluss, den Schwellenwerte auf die mit den quantitativen Methoden ermittelten Ergebnisse haben. In dieser Arbeit erwiesen sich Schwellenwerte aus Referenzpopulationen als problematisch. Bis auf eine Methode erschienen alle publizierten Schwellenwerte aus Referenzpopulationen zumindest bei einzelnen Wirbelkörpern als nicht valide. Aber selbst bei nachvollziehbaren Schwellenwerten, die aus großen Querschnittspopulationen stammen und mit identischen Verfahren getrimmt wurden, ergeben sich derart differente Ergebnisse, dass diese nicht mehr vergleichbar sind. Nach dieser Arbeit sollte auf die Verwendung von Schwellenwerten aus Referenzpopulationen verzichtet werden, zumal die Ergebnisse mit festen Schwellenwerten wie 0.80 den Ergebnissen mit den untersuchten Schwellenwerten aus Referenzpopulationen überlegen waren.

Die Methode von Minne ist nicht empfehlenswert. Die Prävalenzrate ist zu hoch, insbesondere in der Lendenwirbelsäule werden eine nicht nachvollziehbare hohe Anzahl von Wirbelkörperdeformationen ermittelt. Werden nicht die von Minne publizierten Schwellenwerte verwendet, sondern die aus der OPUS Population errechneten Schwellenwerte, sind die Ergebnisse wie Prävalenzrate und Kappa Schätzwert besser. Allerdings bleiben die Ergebnisse weiterhin deutlich schlechter als die aller anderen Methoden. Die Normierung der Wirbelkörperhöhen mit den Höhen des Wirbelkörpers T4 erscheint problematisch. Die quantitative Auswertung aller Wirbelkörper beruht auf der qualitativen Beurteilung eines einzigen Wirbelkörpers, wobei der Wirbelkörper T4 der am häufigsten nicht auswertbare Wirbelkörper ist.

Die mit der Methode von McCloskey eingeführten predicted posterior heights haben nach dieser Arbeit keine Vorteile gegenüber den posterioren Indices der Methoden von Felsenberg, Eastell bzw Melton. Mehrere aneinander angrenzende kompressionsförmige Deformationen, wie sie von McCloskey als Grund für die Einführung der predicted posterior heights erwähnt wurden, kommen zumindest in der OPUS Studie nicht vor. Darüber hinaus beruht die Berechnung der predicted posterior heights auf Referenzwerten, die aus einer Referenzpopulation errechnet werden müssen. Die Schwellenwerte von McCloskey für die Wirbelkörper T6 und L4 erscheinen nicht repräsentativ. Gemessen am Kappa Schätzwert im Vergleich zur qualitativen Auswertung ist die Methode anderen wie der von Melton und Felsenberg 0.80 unterlegen. Die Verwendung von Schwellenwerten aus der OPUS Population ergab keine Verbesserung der Ergebnisse im Vergleich zur qualitativen Auswertung.

Die Methoden von Felsenberg, Eastell und Melton unterscheiden sich lediglich in ihren Schwellenwerten. Die Schwellenwerte von Melton erzielen den besten Kappa Schätzwert aller publizierten Methoden im Vergleich mit der qualitativen Auswertung. Allerdings werden 18.46% der osteoporotischen Frakturen übersehen, die in der qualitativen Auswertung erkannt wurden. Mit den Schwellenwerten von Eastell werden 11% der osteoporotischen Frakturen übersehen und die Anzahl von erkannten, nicht osteoporotischen Deformationen ist um ein fünffaches höher. Der Kappa Schätzwert ist daher im Vergleich zu anderen quantitativen Methoden eher niedrig. Die Schwellenwerte für die Wirbelkörper T4 und L3 bei Eastell erscheinen angesichts der differentialdiagnostischen Einteilung nicht repräsentativ. Die Verwendung von Schwellenwerten aus der OPUS Population ergab keine Verbesserung der Ergebnisse im Vergleich zur qualitativen Auswertung. Der Kappa Schätzwert der Methode von Felsenberg mit einem Schwellenwert von 0.80 liegt nur wenig unter dem der Methode von Melton (0.8544 statt 0.8575). Zudem bietet die Felsenberg 0.80 die Vorteile bis auf eine osteoporotische Fraktur alle erkannt zu haben und nicht auf Schwellenwerten aus Referenzpopulationen zu beruhen.

Anhand der qualitativen Auswertung und der differentialdiagnostischen Einteilung zeigte sich die fehlende Spezifität von Höhenminderungen zur Diagnose von osteoporotischen Wirbelkörperfrakturen. Diese Problem weisen alle Methoden gleichermaßen auf. Durch die Wahl des Schwellenwertes lässt sich eine Differenzierung von Wirbelkörper-

perdeformationen osteoporotischer und nicht osteoporotischer Genese nicht erzielen. Da die Methoden möglichst alle osteoporotischen Wirbelkörperfrakturen erkennen sollten, muss damit eine beträchtliche Zahl nicht osteoporotischer Deformationen in Kauf genommen werden. Einen Ausweg bietet lediglich die Kombination einer quantitativen Methode mit einer anschließenden differentialdiagnostischen Einteilung aller quantitativ ermittelten Deformationen. Die quantitative Methode sollte bei einer solchen Kombination möglichst sensitiv sein. Die Spezifität der quantitativen Methode spielt eine untergeordnete Rolle, da sie durch die differentialdiagnostische Einteilung erzielt werden kann.

Zur Erkennung von osteoporotischen prävalenten vertebrealen Deformationen ist die Kombination einer hoch sensitiven quantitativen Methode ohne Schwellenwerte aus einer Referenzpopulation mit einer anschließenden differentialdiagnostischen Beurteilung der von der quantitativen Methode erkannten Deformationen zu empfehlen. Als quantitative Methode bietet sich, angesichts der qualitativen Auswertung dieser Arbeit, die von Felsenberg mit einem Schwellenwert von 0.80 an.