

R. Rau  
S. Wassenberg

für die Kommission „Bildgebende Verfahren“ der Deutschen Gesellschaft für Rheumatologie (Sprecher: Rolf Rau, Ratingen)

## Bildgebende Verfahren in der Rheumatologie: Scoring-Methoden bei der rheumatoiden Arthritis

### Imaging techniques in rheumatology: scoring methods in rheumatoid arthritis

■ **Zusammenfassung** Es wird eine kritische Übersicht über die verschiedenen Scoring-Methoden gegeben, ihre Grundlagen, ihr Messsystem, ihre Vor- und Nachteile.

Das Röntgenbild ist noch immer – da überall verfügbar, interpretierbar und kostengünstig – das beste Mittel der Verlaufsbeurteilung der RA mit bildgebenden Verfahren. Scoring-Methoden dienen der semiquantitativen Messung der im Röntgenbild erkennbaren Veränderungen, besonders

der erosiven Gelenkdestruktion, teilweise auch des Knorpelverlusts. Sie sind gut validiert, gut reproduzierbar, geben bei klinischen Prüfungen annähernd vergleichbare Resultate. Die meisten Methoden überbewerten frühe Veränderungen und haben einen deutlichen Deckeneffekt. Bei (zeitlich limitierten) klinischen Prüfungen ist die langsame Progression der Erkrankung ein genereller Nachteil der Destruktionsmessung und bedingt ihre geringe Änderungssensitivität. Zudem werden kleine, nicht randständige oder überlagerte Erosionen nicht erfasst. Nachteil der Larsen-Methode ist die Berücksichtigung der im Röntgenbild schlecht erkennbaren und potenziell rasch reversiblen Weichteilschwellung im Stadium I. Nachteil der Gelenkspaltnessung bei der Sharp-Methode und ihren Modifikationen ist, dass sie eher Fehlstellungen oder Fehlprojektionen als Knorpeldestruktion misst. Für alle klinischen Prüfungen muss die Messgenauigkeit der Methode bei wiederholter Messung am jeweiligen Röntgenbild-Satz und damit die kleinste sicher messbare Veränderung (MDC) bestimmt werden.

Röntgenbild und Scoring-Methoden sind bisher zur Messung der Destruktionsbeeinflussung einer Substanz unverzichtbar. Eine

zukünftige Aufgabe ist die Bewertung reparativer Veränderungen, die sich bisher mit den Scoring-Methoden nicht beschreiben lassen.

■ **Summary** This is a critical overview of existing scoring methods including their basis, their measurement systems, their advantages and disadvantages.

Conventional radiography is still – since available everywhere, interpretable and cost effective – the best imaging method to evaluate the course of RA. Scoring methods are designed to semiquantitatively measure radiographically visible changes, especially erosive destruction and – in part – cartilage loss. These methods are well validated, reproducible, and yield similar results in clinical trials. Most methods overestimate early changes and have a clear ceiling effect. Within the time frame of clinical trials radiographic evaluation is not very sensitive to change since the progression of destruction in RA is relatively slow. Moreover, small erosions cannot be detected if they are not at the margin of the bone or if they are superimposed by other bones. Within the Larsen method the definition of grade 1 by soft tissue swelling is disadvantageous: soft tissue swelling is difficult to identify on x-rays, it is

Eingegangen: 17. März 2003  
Akzeptiert: 14. April 2003

Prof. Dr. med. Rolf Rau (✉)  
Dr. med. Siegfried Wassenberg  
Rheumaklinik  
Evangelisches Fachkrankenhaus  
Rosenstr. 2  
40882 Ratingen, Germany  
Tel.: 021 02/20 62 46  
Fax: 021 02/70 93 29  
E-Mail: rrau@uni-duesseldorf.de

a measure of disease activity and not of destruction and is quickly reversible. Joint space narrowing, measured with Sharp's method and its modifications, may be caused by misprojection due to soft tissue swelling with flexion or subluxation of the joint rather than cartilage loss. Since the measurement error of a scoring method is very much dependent on the severity of the disease of the

patient population and on the quality of the radiographs, the measurement error and thereby the minimal detectable change (MDC) should be stated for every single clinical trial.

Conventional radiographs and scoring methods are still indispensable to measure the influence of treatment on radiographic progression in RA. A future task will be to include in scoring methods

the rating of reparative changes which cannot be scored so far.

■ **Schlüsselwörter** Rheumatoide Arthritis – Radiologie – radiologische Progression – Scoring-Methoden

■ **Key words** Rheumatoid arthritis – radiology – radiologic progression – scoring methods

## Grundlagen und Prinzip der Scoring-Methoden

Die Beeinflussung der radiologischen Progression gilt als wichtigstes Kriterium der Wirksamkeit eines Basistherapeutikums. Verschiedene Scoring-Methoden wurden entwickelt, um die im Röntgenbild erkennbaren Veränderungen quantifizieren zu können. Tabelle 1 nach Resnick (1) zeigt, welche pathologisch-anatomischen Veränderungen der rheumatoiden Arthritis sich im Röntgenbild darstellen. Weichteilschwellung und gelenknahe Osteoporose hängen in ihrer Beurteilbarkeit sehr stark von der Aufnahmequalität, d. h. insbesondere deren Härte ab, sind damit unzuverlässig und zur Verlaufsbeurteilung ungeeignet. Sie spiegeln zudem die Aktivität der Erkrankung wider und sind rasch reversibel. Das entscheidende Maß für das Fortschreiten der Erkrankung aber ist die (kaum rückbildungsfähige) Gelenkzerstörung. Hauptaufgabe des Scorens ist demnach die Quantifizierung der Destruktion. Versuche mittels Computertechnik die Destruktionen zu messen waren bisher wenig erfolgreich (2). Die semi-quantitativen Scoring-Methoden erfassen insbesondere die erosiven Veränderungen. Sie zählen die Erosionen pro Gelenk (Sharp), bewerten teilweise die Größe der Erosionen mit (v. d. Heijde), schätzen den prozentualen Anteil der destruierten Gelenkoberfläche (Ratingen-Score) oder bewerten die Gelenkzerstörung global, teilweise unter Zuhilfenahme von Standardreferenzfilmen oder Skizzen (Larsen). – Neben den erosiven Veränderungen wird bei einigen Methoden auch die Gelenkspaltweite als Indikator für die Knorpeldestruktion herangezogen (Sharp, v.d.Heijde); die Gelenkspaltverschmälerung kommt aber größtenteils nicht durch Knorpeldestruktion, sondern durch Kapsel- und Bandüberdehnung zustande (siehe Tabelle 1); die darauf beruhenden Subluxationen und Luxationen werden von Kaye und Nance (3) gesondert bewertet, von van der Heijde (4) bewusst in den Gelenkspaltverschmälerungs-Score aufgenommen, von Sharp (5) jedoch nicht berücksichtigt.

**Tab. 1** Pathologisch anatomischer Befund und Röntgenbefund im Vergleich (nach Resnick) (1)

Pathologische Anatomie	Röntgenbefund
Entzündung der Synovialis und Flüssigkeitsansammlung Hyperämie und Knochenmineralverlust	Weichteilschwellung und Gelenkspalterweiterung Gelenknahe Osteoporose
Zerstörung des Knorpels durch Pannus	Gelenkspaltverschmälerung
Zerstörung des „ungeschützten“ Knochens am Gelenkrand durch Pannus	Marginale Erosionen
Zerstörung des subchondralen Knochens durch Pannus	Knochenerosionen und subchondrale Zysten
Fibröse und knöcherne Ankylosen	Ankylose
Kapsel- und Bandinstabilität	Deformationen und Subluxationsfehlstellungen

Grundübereinstimmung bei der Entwicklung der Scoring-Methoden war die Annahme, dass es eine Besserung destruktiver Veränderungen nicht gibt. Demgemäß kann nur die Progression, allenfalls deren Stillstand, nicht aber eine Reparation beschrieben werden.

Die Scoring-Methoden wurden zur Anwendung an den kleinen Gelenken der Hände und Vorfüße entwickelt, die am frühesten und häufigsten betroffen sind und sich besser beurteilen lassen als die großen Gelenke. Veränderungen an Händen und Vorfüßen korrelieren mit denen der großen Gelenke, sind also repräsentativ für die Gesamterkrankung.

## Technische Voraussetzungen und Durchführung der Untersuchung

Die Zuverlässigkeit der Score-Werte ist entscheidend abhängig von der Beurteilbarkeit, d. h. der Qualität der Röntgenaufnahme. Für eine technisch optimale Aufnahme müssen folgende Voraussetzungen erfüllt sein (siehe auch Kapitel „Konventionelles Röntgen“):

- Verwendung einer hochauflösenden, feinstzeichnenden Film-Rückfolien-Kombination.
- Optimale Positionierung (streng dorso-plantar oder dorso-palmar):
  - Hände flach auf Kassette liegend
  - Vorfüße im Sitzen auf die auf dem Boden liegende Kassette gesetzt
- Empfohlenes Format 18×24 cm für je eine Hand oder für beide Vorfüße
- Zentralstrahl auf das MCP-3-Köpfchen oder zwischen die Großzehengrundgelenke
- Alternativ: Beide Hände auf eine Kassette 24×30 cm, aber einzeln belichten
- Einblenden soweit wie möglich, aber vollständige Darstellung der ganzen Hand inclusive distale Radius- und Ulnaenden
- Konstante Belichtung

Halbschräge Aufnahmen sind unnötig, da die gleiche Projektion im Verlauf nicht erreicht wird. Weichteil-aufnahmen bringen gegenüber guten Standardaufnahmen keinen Vorteil.

Die Auswertung der Bilder kann einzeln erfolgen („single“), d.h. Bilder unterschiedlicher Patienten werden gemischt. Diese Methode ist kaum verbreitet. Im Allgemeinen werden alle Bilder eines Patienten (oft 3–6 Zeitpunkte) zusammen ausgewertet („setweise“). Dies kann bei bekannter zeitlicher Reihenfolge („chronological order“) erfolgen oder aber bei unbekannter Reihenfolge („random order“). Die unbekannte zeitliche Reihenfolge wird heute bevorzugt, um das Vorurteil einer ständigen Progression zu umgehen. Sie führt aber auch zu einer konservativen Beurteilung durch den Untersucher (um Fehler zu vermeiden) und damit zu geringerer Änderungssensitivität.

Bei multizentrischen Studien werden neuerdings die Röntgenbilder digitalisiert und dann auf dem Bildschirm befundet.

## Darstellung der Scoring-Methoden

Unterschiedliche Methoden, anwendbar für Querschnitts- oder Folgeuntersuchungen, wurden beschrieben:

1. Methoden zur globalen Bewertung des ganzen Patienten (z. B. Steinbrocker)
2. Globale Beurteilung einzelner Gelenke (z. B. Larsen)
3. Separate Messung von Erosionen und Gelenkspaltverschmälerung (z. B. Sharp).

## 1. Methoden mit globaler Bewertung des gesamten Patienten

*Steinbrocker 1949 (6):*

In die „Therapeutischen Kriterien“ der American Rheumatism Association (6) wurden auch röntgenologische Parameter aufgenommen mit einer Einteilung in vier Stadien bzw. Schweregrade:

Grad 1 = keine Veränderung, allenfalls gelenknahe Osteoporose  
Grad 2 = Erosion, ggf. Osteoporose/Gelenkspaltverschmälerung  
Grad 3 = Erosionen + Subluxation/Fehlstellung  
Grad 4 = Ankylose

Bewertet wird das am stärksten betroffene Gelenk, das allerdings für die Gesamterkrankung nicht unbedingt repräsentativ ist.

*Kellgren 1963 (7):*

Atlas von Standardreferenzfilmen mit typischen Vergleichsbildern in fünf Schweregraden.

Grad 0 = normal  
Grad 1 = fragliche Veränderung  
Grad 2 = geringe Veränderung  
Grad 3 = mäßige Veränderung  
Grad 4 = schwerwiegende Veränderung

Auch bei dieser Methode bestimmt das am stärksten betroffene Gelenk die Zuordnung zum jeweiligen Schweregrad.

## 2. Methoden mit globaler Beurteilung pro Gelenk

*Larsen 1977 (8):*

Die Einteilung in 6 Schweregrade (0–5) berücksichtigt im Wesentlichen die erosive Destruktion. Die Grade wurden ursprünglich durch Standardreferenzfilme illustriert, da diese aber nicht befriedigten, wurden sie durch Strichzeichnungen (Abb. 1) ersetzt.

Die Schweregrade werden folgendermaßen definiert:

Grad 0 = normal  
Grad 1 = geringe Veränderungen: Weichteilschwellung, gelenknahe Osteoporose oder geringe Gelenkspaltverschmälerung  
Grad 2 = definitive Veränderungen: eine oder mehrere kleine Erosionen vorhanden  
Grad 3 = deutliche Veränderungen: ausgeprägte Erosionen sind vorhanden  
Grad 4 = schwere Veränderungen: große Erosionen vorhanden, nur Teile der ursprünglichen Gelenkfläche noch erhalten  
Grad 5 = mutilierende Veränderungen: die ursprüngliche Gelenkfläche ist verschwunden, schwere Deformität möglich

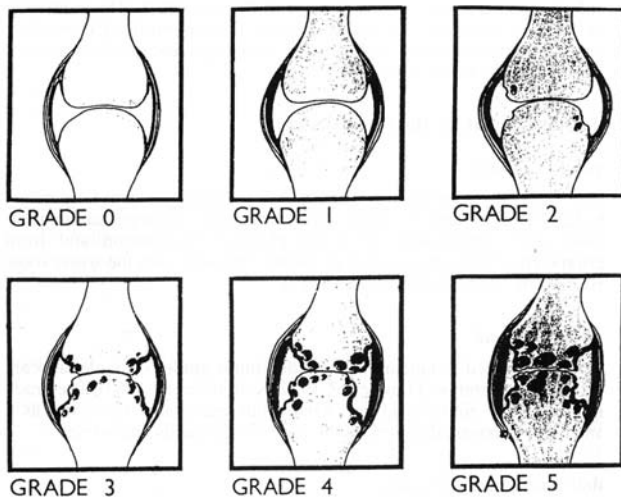


Abb. 1 Röntgenologische Grade oder Stadien der RA nach Larsen (8) aus Larsen und Thoen (9)

Diese Definitionen beziehen sich auf die Veränderungen an Händen und Füßen, für die großen Gelenke sind sie teilweise leicht modifiziert.

Folgende Gelenke werden bewertet: PIP II–V, IP I (Daumen), MCP I–V, Handgelenk, MTP-Gelenke I–V. – Bei insgesamt 32 Gelenken beträgt der Gesamt-Score zwischen 0 und 160.

Diese Methode wurde durch Larsen selbst und andere Autoren mehrfach modifiziert, die zu untersuchenden Gelenke wurden nicht geändert.

#### Modifikation von Larsen und Thoen 1987 (9)

Um dem Handgelenk mehr Gewicht zu geben wird der Handgelenks-Score mit 5 multipliziert. Der Gesamt-Score beträgt dann 0–200. *Diese Multiplikation ist aber ungerechtfertigt, wenn (häufig) nur eine Stelle des Handgelenks erosiv ist, die damit 5-mal höher bewertet wird als die gleiche Erosion an einem Fingergelenk.*

#### Modifikation von Rau und Herborn 1994 (10):

In dieser Modifikation werden die Stadien 2–5 durch den prozentualen Anteil der zerstörten an der gesamten Gelenkoberfläche definiert und dadurch Verständlichkeit und Lernbarkeit deutlich verbessert.

Grad 0 = normal
Grad 1 = Weichteilschwellung und/oder Gelenkspaltverschmälerung/subchondrale Osteoporose
Grad 2 = erosive Destruktion der Gelenkoberfläche von < 25%
Grad 3 = Gelenkoberflächendestruktion 26–50%
Grad 4 = Gelenkoberflächendestruktion 51–75%
Grad 5 = Gelenkoberflächendestruktion > 75%

#### Modifikation von Scott 1995 (11)

Die Zahl der Schweregrade wird auf 0–4 reduziert, der maximale Gesamt-Score beträgt 128.

Grad 0 = normal
Grad 1 a = periartikuläre Osteoporose/Gelenkschwellung (falls wesentlich)
Grad 1 b = Erosionen/Cysten an zwei Stellen < 1 mm vorhanden
Grad 2 = eine oder mehrere Erosionen mit Unterbrechung der Cortex um > 1 mm
Grad 3 = signifikante Erosionen an beiden Gelenkflächen, Teile der Gelenkfläche noch erhalten
Grad 4 = Subluxation

In den Grad 1 werden kleine Erosionen oder Zysten aufgenommen.

Die völlige Destruktion der Gelenkflächen ist weggelassen, damit eine Differenzierung im Bereich schwerer Veränderungen (Grade 3–5) kaum möglich. *Die Aufnahme einer Subluxation – noch dazu mit dem höchsten Score-Wert – widerspricht dem „Geist“ des Larsen-Scores, der nur destruktive Veränderungen bewertet.*

#### Modifikation von Edmonds 1999 (12)

Es werden wieder fünf Stadien eingeführt.

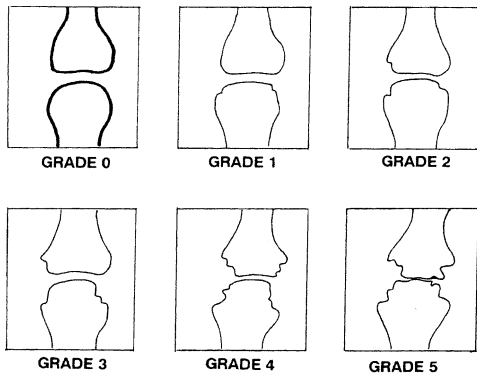
Grad 0 = normal oder nur degenerative Veränderungen
Grad 1 = leichte Gelenkspaltverschmälerung und/oder ausgeprägte periartikuläre Osteoporose und/oder deutliche Weichteilschwellung und/oder Verdacht auf Erosionen/Cysten an 2 Stellen < 1 mm
Grad 2 = eine oder mehr Erosionen mit > 2 mm Cortex-Unterbrechung
Grad 3 = ausgeprägte Erosionen an beiden Gelenkflächen, Gelenkoberfläche weitgehend erhalten
Grad 4 = große Erosionen beidseits, Gelenkfläche nur noch teilweise erhalten und/oder Subluxation
Grad 5 = Schwere Deformität mit völliger Destruktion der Gelenkflächen und/oder Status nach Operation

*Die Aufnahme einer Subluxation als Grad 4 widerspricht dem „Geist“ des Larsen-Scores, der nur destruktive Veränderungen bewertet.*

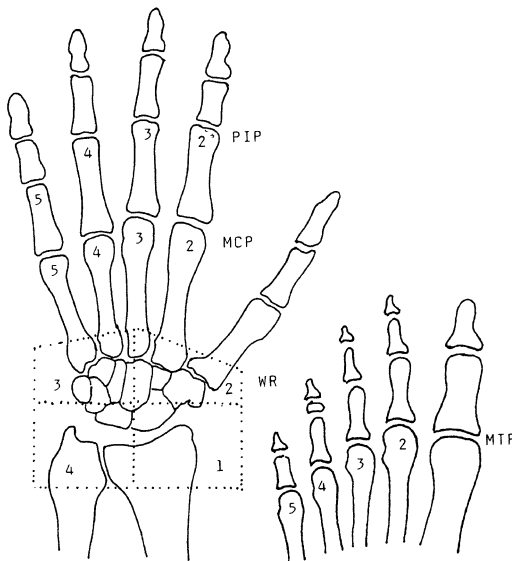
Ein operiertes Gelenk als im Stadium 5 befindlich zu bezeichnen, ist nicht gerechtfertigt, da häufig Fusionen (z.B. Handgelenk) oder Resektionen (z.B. Zehengrundgelenke) an wenig veränderten Gelenken erfolgen.

#### Modifikation für Langzeitsstudien von Larsen 1995 (13)

Hierbei reagierte Larsen auf die Schwierigkeit Weichteilschwellung und gelenknahe Osteoporose bei wechselnder Qualität der Röntgenaufnahmen zu bewerten und definierte deshalb das Stadium 1 neu: eine oder mehrere Erosionen mit einem Durchmesser von < 1 mm oder Gelenkspaltverschmälerung [sehr kleine Erosionen sind allerdings bei schlechter Aufnahmequalität (unscharf, zu hart, zu weich) ebenfalls schlecht erkennbar].



**Abb. 2** Schematische Darstellung der Graduierung bzw. der Stadieneinteilung in der Larsen-Modifikation (13)



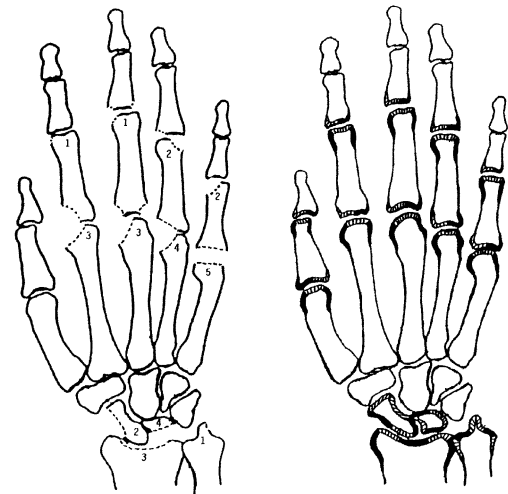
**Abb. 3** Gelenke oder Regionen, die in der Modifikation von Larsen (13) gescore werden

Stadium 2: eine oder mehrere kleine Erosionen mit einem Durchmesser von  $>1$  mm.

Die Definition der anderen Stadien bleibt unverändert (Abb. 2). Die zu scorenden Gelenke und Regionen gehen aus Abbildung 3 hervor. Bei insgesamt 32 Gelenken beträgt der maximale Gesamt-Score 160.

#### Ratingen Score 1998 (14)

Im Unterschied zur Larsen-Methode und zu ihrer Modifikation durch Rau und Herborn (10), die noch schnell rückbildungsfähige Elemente der Krankheitsaktivität (Schwellung/Osteoporose) enthält, werden nur destruktive Veränderungen am Knochen gewertet. Eine Bewertung der Gelenkspaltverschmälerung erfolgt nicht. Das Handgelenk wird, wie bei der Sharp-Methode, in mehrere Regionen unterteilt.



**Abb. 4** Ratingen-Score: Schematische Darstellung der verschiedenen Grade sowie der Einteilung der Gelenkoberfläche in jeweils 20%-Abschnitte (14)

Stadium 0 = normales Gelenk
Stadium 1 = eine oder mehrere Erosionen, $<20\%$ der Gelenkoberfläche sind zerstört
Stadium 2 = $21\% - 40\%$ der Gelenkoberfläche zerstört
Stadium 3 = $41\% - 60\%$ der Gelenkoberfläche zerstört
Stadium 4 = $61\% - 80\%$ der Gelenkoberfläche zerstört
Stadium 5 = $>80\%$ der Gelenkoberfläche zerstört

Folgende Gelenke werden bewertet: PIP II–V, IP I (Daumen), MCP I–V, 4 Handgelenksregionen (Naviculare, Lunatum, distaler Radius, distale Ulna), MTP II–V, IP I (Großzehenendgelenk).

Zur Erleichterung der Abschätzung der destruierten Gelenkfläche kann die proximale Gelenkfläche gedanklich in drei, die distale Gelenkfläche in zwei Abschnitte unterteilt werden, die jeweils etwa  $20\%$  der Gelenkoberfläche entsprechen (Abb. 4). Abbildung 5 zeigt beispielhaft Röntgenbilder von PIP-Gelenken in den verschiedenen Graden oder Stadien.

### 3. Methoden mit getrennter Bewertung von Erosionen und Gelenkspaltverschmälerung

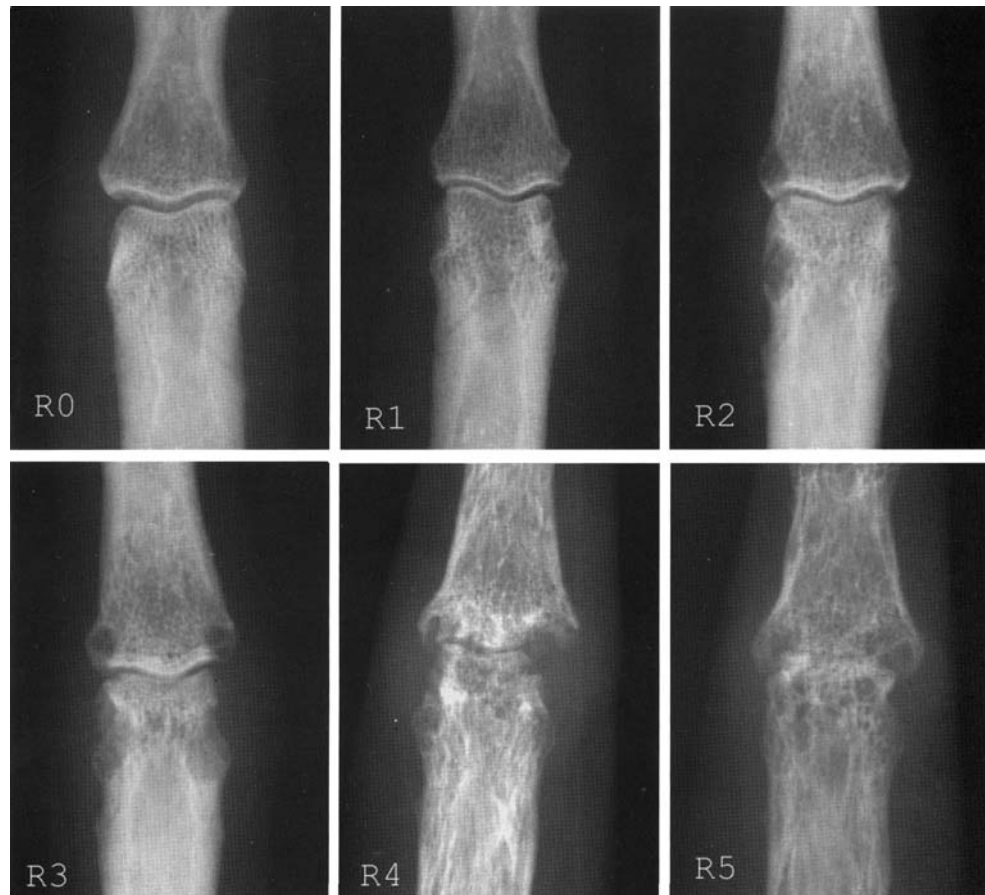
#### Sharp 1971 (5)

Ursprünglich schlug Sharp vor 27 Regionen der Finger- und Handgelenke getrennt auf Erosionen und Gelenkspaltverschmälerung zu scoren. Angewendet wird heute aber nur noch seine Modifikation von 1985 (15).

#### Sharp 1985 (15)

Bei dieser Modifikation wurde die Zahl der zu bewertenden Gelenke auf 17 (Erosionen) bzw. 18 (Ge-

**Abb. 5** Beispiele von Röntgenbildern der PIP-Gelenke in den verschiedenen Stadien des Ratingen-Scores



lenkspaltverschmälerung) reduziert. Die Erosionen werden unabhängig von ihrer Größe gezählt. Bei schwer abgrenzbaren Erosionen (z.B. ineinander übergehende Usuren) soll auch der Anteil der Gelenkoberfläche (ohne genaue Anweisung) berücksichtigt werden. Der Höchstwert von 5 ist bei 5 Erosionen erreicht, oder wenn die Hälfte einer der beiden Gelenkflächen destruiert ist.

Erosive Veränderungen	Gelenkspaltverschmälerung
Grad 0 = keine Erosionen	0 = normal
Grad 1 = 1 Erosion	1 = fokale Gelenkspaltverschmälerung
Grad 2 = 2 Erosionen	2 = diffuse Gelenkspaltverschmälerung um weniger als 50%
Grad 3 = 3 Erosionen	3 = diffuse Gelenkspaltverschmälerung um mehr als 50%
Grad 4 = 4 Erosionen	4 = Ankylose
Grad 5 = >4 Erosionen oder >50% der Gelenkoberfläche auf einer Seite des Gelenks zerstört	

Bei 34 bewerteten Gelenken (siehe Abb. 6) beträgt der maximale Erosions-Score 170 (34×5). Der maxi-

male Gelenkspaltverschmälerungs-Score beträgt 144 (36 Gelenke×4). Subluxationen oder Luxationen werden nicht gewertet. Der maximale Gesamt-Score beträgt 314.

Neuerdings bewertet Sharp auch 12 Gelenke an den Füßen (10 MTP-, 2 IP-Gelenke der Großzehen). Der Erosionsscore beträgt 0–60, der JSN-Score 0–48.

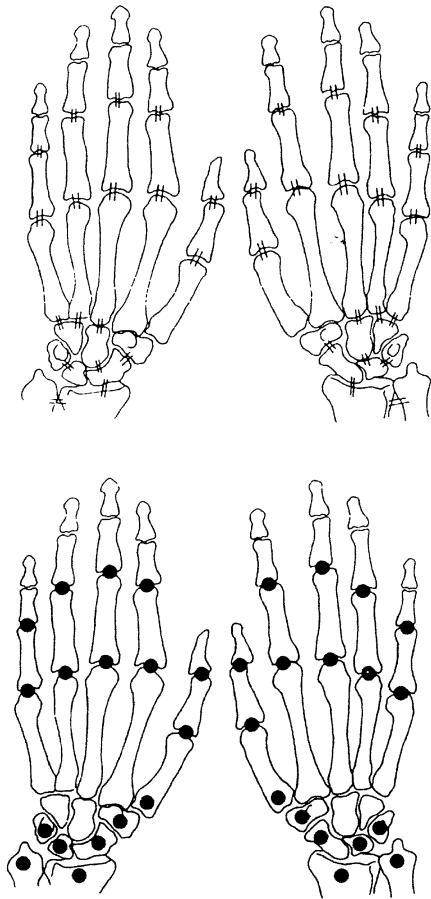
Gesamt-Score für Hände und Füße beträgt somit jetzt 0–422.

#### *Van der Heijde-Modifikation der Sharp-Methode 1989 (16)*

Frau van der Heijde führte als wesentlichen Unterschied zur Sharp-Methode (15) die Mitbewertung der Füße ein.

Der Erosions-Score wird gegenüber der Sharp-Methode derart modifiziert, dass eine kleine Erosion als 1 bezeichnet wird, eine längere Erosion als 2, eine die Mittellinie überschreitende Erosion als 3 (siehe Abb. 7). Die Summe dieser Score-Werte darf an einem Gelenk aber niemals 5 überschreiten.

An den Füßen beträgt der maximale Gesamt-Score pro Gelenk 10, begründet damit, dass hier häufiger langstreckige Usuren vorkommen.



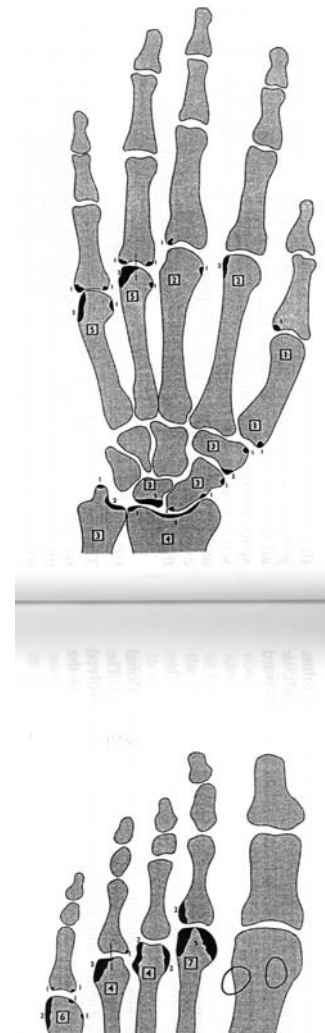
**Abb. 6** Gelenke, die bei der Sharp-Methode (15) bewertet werden. Die mit einem Punkt markierten Gelenke werden auf Erosionen, die mit Parallelstrichen markierten Gelenke auf Gelenkspaltverschmälerung untersucht

Im Gelenkspaltverschmälerungs-Score wird zusätzlich zur jeweiligen Sharp-Definition für Grad 3 die Subluxation und Grad 4 die Luxation aufgenommen.

Erosions-Score*	Gelenkspaltverschmälerungs-Score**
Grad 0 = normal	0 = normal
Grad 1 = 1 kleine Erosion	1 = fokale Gelenkspaltverschmälerung
Grad 2 = 1 längere oder 2 kleine Erosionen	2 = diffuse Gelenkspaltverschmälerung um weniger als 50%
Grad 3 = die Mittellinie überschreitende oder 3 kleine Erosionen oder 2 + 1	3 = diffuse Gelenkspaltverschmälerung um mehr als 50% oder Subluxation
Grad 4 = 4 × 1/2 + 2 oder 3 + 1	4 = Ankylose oder Luxation
Grad 5 = Summe der Erosionen 5 oder mehr	

\* 22 Gelenke einer Körperseite werden bewertet: PIP II–V, IP I (Daumen), MCP I–V, Basis des Metacarpale I, Multangulum, Naviculare, Lunatum, distaler Radius, Ulna, MTP I–V, IP I (Großzehe).

\*\* Folgende 21 Gelenke einer Körperseite werden bewertet: PIP II–V, MCP I–V, Carpometacarpalgelenke I – III, Trapezionaviculargelenk, Gelenk zwischen Naviculare/Lunatum und Capitatium, Radiocarpalgelenk, MTP-Gelenke I–V, IP I (Großzehe)



**Abb. 7** Sharp/van der Heijde-Methode in der letzten Modifikation (16a). Schematische Darstellung der Gelenke, an denen Gelenkspaltverschmälerung (a) bzw. Erosionen (b) gescoret werden und Schema für das Scoring je nach Größe der Erosion (c)

Die Bewertung maximaler erosiver Destruktion mit 10 an den (kleineren) Zehengelenken gegenüber 5 an den (größeren) Fingergelenken gibt den Vorfüßen (unberechtigterweise) ein größeres Gewicht. Der maximale Erosions-Score beträgt 280 (32 × 5 an den Händen, 12 × 10 an den Füßen), der maximale Gelenkspaltverschmälerungs-Score beträgt 168 (42 Gelenke × 4), der maximale Gesamt-Score beträgt 448.

Genant 1983 (17)

Bei dieser Modifikation der Sharp-Methode werden ebenfalls Erosionen und Gelenkspaltverschmälerung beurteilt, die Definition der Grade ist allerdings allgemeiner gefasst:

Erosions-Score*	Gelenkspaltverschmälerungs-Score**
Grad 0 = normal	0 = normal
Grad 0+ = diskrete Veränderung	0+ = fraglich oder diskret
Grad 1 = leicht	1 = leicht
Grad 1+ = leicht, schlechter	1+ = leicht, schlechter
Grad 2 = mäßig	2 = mäßig
Grad 2+ = mäßig, schlechter	2+ = mäßig, schlechter
Grad 3 = schwer	3 = schwer
Grad 3+ = schwer, schlechter	3+ = schwer, schlechter
	4 = Ankylose oder Dislokation

+ = 0,5

\* 14 Gelenke pro Hand werden untersucht: IP I, PIP II–V, MCP I–V, CMC I, Scaphoid, distaler Radius, distale Ulna. Maximaler Score = 98

\*\* 13 Gelenke pro Hand werden untersucht; IP I, PIP II–V, MCP I–V, Kombination der CMC III–V, Kombination von Capitulum, Scaphoid, Lunatum und Radiocarpalgelenk. Maximaler Score = 104

Die Methode lässt sich entsprechend auch für die Zehengrundgelenke und das Großzehendgelenk anwenden.

### Kaye und Nance 1991 (3)

Bei dieser Methode werden Erosionen und Gelenkspaltverschmälerung etwas anders definiert als bei Sharp (15) und Genant (17); dazu kommt als Besonderheit der Methode ein Fehlstellungs-Score:

Erosionsscore	JSN-Score*	Fehlstellungs-Score (malalignment)
Grad 0 = normal	Grad 0 = normal	Grad 0 = normal
Grad 2 = leichte Veränderung	Grad 2 = leichte Veränderung	Grad 2 = Subluxation
Grad 3 = mäßige Veränderung	Grad 3 = mäßige Veränderung	–
Grad 4 = schwere Veränderung	Grad 4 = schwere Veränderung	Grad 4 = Dislokation
–	Grad 5 = Ankylose	

\*JSN = Joint Space Narrowing = Gelenkspaltverschmälerung

Folgende Gelenke werden beurteilt:

Erosionen und Gelenkspaltverschmälerung: PIP, MCP, CMC I, III, IV, V, 7 Stellen der Handwurzel.

Fehlstellung: PIP, MCP, 5 Stellen der Handwurzel.

## ■ Vor- und Nachteile der Methoden

### a. Vorteile

Das Röntgenbild ist kostengünstig weltweit verfügbar, archivierbar und damit auch später noch zu bewerten. Mit seiner Interpretation bestehen jahrzehntelange Erfahrungen. Es bildet die Summe der Veränderungen ab, die zwischen zwei Aufnahmezeitpunkten am Gelenk eingetreten sind, reflektiert somit die kumulative Konsequenz der Synovitis.

Röntgenveränderungen sind sehr spezifisch und somit auch diagnostisch verwertbar. Die Scoring-Methoden reflektieren die Progression und den aktuellen Krankheitsstatus objektiver als das wechselhafte klinische Bild und die sich im Krankheitsverlauf ändernde Selbsteinschätzung des Patienten. Die Präzision der Methoden bei Befundung durch unterschiedliche Untersucher ist mit Intraclass-Korrelations-Koeffizienten von  $>0,9$  gut (siehe Validierung der Befunde).

### b. Nachteile

Die Scoring-Methoden erfassen im eindimensionalen Röntgenbild nur randständige Defekte, an der Volar- oder Dorsalseite gelegene gar nicht oder allenfalls als „Zysten“. Durch Überlagerung, z. B. am Handgelenk oder bei Fehlstellungen, werden auch randständige Usuren verdeckt. Bei tomographischer Technik, wie beim MRT, werden mehr Erosionen sichtbar.

Das nur langsame Auftreten destruktiver Veränderungen und die schlechte Erfassbarkeit kleinster Defekte bedingen eine geringe Änderungssensitivität. Das Röntgenbild spiegelt die Krankheitsaktivität erst mit einer Zeitverzögerung von 6–12 Monaten. Eine Destruktionshemmung unter Behandlung kann meist erst nach 6 Monaten festgestellt werden. In den bisherigen Studien zeigt die Mehrzahl der Patienten (70 –  $>90\%$ ) keine Score-Änderung. Die erforderliche Patientenzahl ließe sich reduzieren und die Diskriminierung zwischen unterschiedlichen Behandlungsmethoden verbessern, wenn nur Patienten mit großem Progressionspotenzial (hohe Krankheitsaktivität, Seropositivität, Früherosivität) in Studien aufgenommen würden. Änderungen sind bei Frühfällen sicherer erfassbar als bei bereits stark destuierten Gelenken.

Ungleiche Schritte auf der numerischen Skala führen zur Überbewertung früher Veränderungen und zu einem frühen Deckeneffekt (methodisch bedingte Abflachung der Progressionskurve). Im Larsen-Score repräsentiert eine kleine Erosion (Grad 2) bereits 40% des maximal möglichen Scores. Im Sharp-System entsprechen 4 kleine Erosionen (Grad 4) bereits 80% des maximal möglichen Scores; eine 50%ige Destruktion einer der beiden Gelenkflächen ergibt bereits den Höchstscore von 5, d. h. eine weitere Destruktion kann nicht beschrieben werden. Die oft beschriebene stärkere Progression in der Frühphase der Erkrankung kann somit rein methodisch bedingt sein. Im Ratingen-Score sind die Skalenschritte mit jeweils 20%iger Destruktion der Gelenkoberfläche gleich. Das Stadium 1 im Larsen-Score (Gelenkschwellung, gelenknahe Osteoporose) ist im Röntgenbild schwer zu erfassen und erhöht damit die Intra- und Inter-Untersucher-Varianz. Die Besserung der Schwellung im



Rahmen abnehmender Krankheitsaktivität und die Zunahme der Destruktion können sich gegenseitig teilweise aufheben und die Änderungssensitivität des Larsen-Scores beeinträchtigen.

Da sich Knorpel im Röntgenbild nicht darstellt, wird die Gelenkspaltverschmälerung als Indikator einer Knorpeldestruktion betrachtet. Eine Gelenkspaltverschmälerung kann aber durch unkorrekte Lagerung in Folge Gelenkschwellung, Kontrakturen, Überdehnung der Gelenkkapsel, Subluxationen oder Luxationen vorgetäuscht werden. Subluxation und Luxation werden von Sharp nicht berücksichtigt, erhalten aber bei van der Heijde die höchsten Grade (3 und 4) im Gelenkspaltverschmälerungs-Score. Damit werden Knorpeldestruktion und Gelenkfehlstellung (Kapsel- und Bandlaxizität) in den gleichen Topf geworfen. Der Gelenkspaltverschmälerungs-Score diskriminiert oft schlechter zwischen zwei Therapien als der Erosions-Score (18–20).

Das bis vor kurzem gültige Dogma, dass Erosionen prinzipiell nicht heilen können, führte beim Scoring dazu, dass eine einmal (auch fälschlich) gesehene Erosion weiter gezählt wurde, auch wenn sie nicht mehr sichtbar war („once an erosion, always an erosion“). Eine Änderung war also immer nur im Sinne einer Verschlechterung feststellbar mit folgender Überschätzung der Progression. Beim Lesen in unbekannter Reihenfolge wurden aber Score-Reduktionen beschrieben (18, 19, 21). Heilungen am einzelnen Gelenk sind beim Scoring nicht beschreibbar, eine „aktive“ Usur erhält den gleichen Score-Wert wie eine inaktivierte abgeheilte Usur.

## Validierung der Methoden

Zur Messung der Gelenkdestruktion gibt es keinen (externen) Goldstandard, der mit den semiquantitativen Scoring-Methoden verglichen werden kann. Daher erfolgt die Beurteilung ihrer Zuverlässigkeit (Reliabilität) und Gültigkeit (Validität) im Wesentlichen durch Vergleich wiederholter oder von verschiedenen Untersuchern erhobener Werte. Für alle Methoden ergab sich dabei eine gute Assoziation mit Korrelationskoeffizienten von  $>0,9$ . Der Korrelationskoeffizient hängt sehr stark von Extremwerten ab (22) und ist nach Bland und Altman (23) nicht wirklich ein Maß der Übereinstimmung. Besser geeignet ist die Varianzanalyse (ANOVA), die den Grad der Veränderung (Progression) zur Fehlerbreite der Methode in Beziehung setzt: die Standardabweichung (SD) der Veränderung wird durch die SD des Untersuchers bei wiederholter Befundung (Wiederholbarkeit) oder die SD mehrerer Untersucher (Übereinstimmung) geteilt. Je größer der Quotient  $>1$  ist,

desto eher ist die gemessene Änderung real und nicht durch Messfehler vorgetäuscht.

Durch Multiplikation der SD mit 2,8 lässt sich die kleinste messbare Veränderung (minimal detectable change=MDC) ermitteln. Sie ist ein Maß für die Messempfindlichkeit (sensitivity to change) und damit ein wichtiges Qualitätsmerkmal.

Für den Larsen-Score wurde anhand von im Abstand eines Jahres aufgenommenen Röntgenbildern der MDC für den gleichen Untersucher mit 3,8% des maximalen Scores, für mehrere Untersucher mit 5,2% des maximalen Scores bestimmt (24). Für den Sharp-Score ergab sich bei im Abstand von 1½ Jahren aufgenommenen Röntgenbildern ein MDC von 10,1% des maximalen Scores (25).

In Ratingen wurde von 6 erfahrenen Untersuchern – J. Sharp, A. Larsen, D. v.d. Heijde, M. Wijnands, G. Herborn, S. Wassenberg – ein Vergleich der Methoden von Sharp und Larsen und ein Jahr später ein Vergleich dieser Methoden mit dem Ratingen-Score vorgenommen. Der Vergleich erfolgte anhand

**Tab. 2** Wiederholbarkeit, Präzision und Zeitbedarf verschiedener Scoring-Methoden

Methode	Wiederholbarkeit	Präzision (MDC*)	Zeitbedarf
<i>Van der Heijde-Score</i>			
Alle Untersucher	2,8	2,9%	22,22
Bester/schlechtester Untersucher	3,8/1,8	2,1/3,7%	11,13/42,55
Ratinger-Untersucher	2,8	3,3	14,74
<i>Larsen-Score</i>			
Alle Untersucher	2,7	3,5%	9,88
Bester/schlechtester Untersucher	3,3/2,3	2,4/4,6%	5,54/15,75
Ratinger Untersucher	2,6	4,2%	7,06
<i>Ratingen-Score</i>			
Alle Untersucher	2,6	3,3%	9,01
Bester/schlechtester Untersucher	3,2/2,2	2,6/3,8%	6,90/12,93
Erfahrene Untersucher	2,6	3,3%	8,99
Unerfahrene Untersucher	2,6	3,3%	9,04

\* MDC=Minimal Detectable Change Wiederholbarkeit=Quotient Standardabweichung der Befunde/Standardabweichung der Untersucher bei wiederholter Messung; Präzision=Standardabweichung der Untersucher bei wiederholter Messung $\times$ 2,8; Ergebnisse standardisiert auf den Prozentsatz des maximalen Score-Werts; Zeitbedarf=durchschnittlicher Zeitbedarf für die Befundung einer Serie von 7 aufeinander folgenden Bildern von Händen und Füßen in Minuten

*van der Heijde- & Larsen-Score:*  
alle Untersucher=JS, AL, DvdH, MW, GH, SW  
Ratinger Untersucher=GH, SW

*Ratingen-Score:*  
erfahrene Untersucher=GH, SW  
unerfahrene Untersucher=VFK, TL, CS  
alle Untersucher=erfahrene und unerfahrene Untersucher

von Röntgenbildern der Hände und Füße von 30 Patienten (leichtere und schwere Verläufe) mit 7 Messzeitpunkten über 5 bzw. 10 Jahre.

Die Ergebnisse dieser beiden Untersuchungen sind in Tabelle 2 dargestellt.

## Indikationen und Kontraindikationen

Die Anwendung von Scoring-Methoden zur Auswertung von Röntgenbildern ist indiziert bei allen Basistherapiestudien über einen Zeitraum von einem Jahr oder länger. Eine einfach anzuwendende Methode wie der Ratingen-Score kann auch im klinischen Alltag angewandt werden. Die Dokumentation mehrerer Zeitpunkte auf einem Dokumentationsbogen veranschaulicht Befund und Verlauf besser als die üblichen Beschreibungen und spart Zeit beim Befunden wie auch beim Veranschaulichen und Beurteilen der Befunde.

## Dokumentation der Befunde

Die Dokumentation der Score-Werte für jedes einzelne Gelenk und jeden Zeitpunkt erfolgt auf entsprechenden Formblättern. Die Auswertung kann erfolgen für Patientengruppen, einzelne Patienten, Gelenkgruppen und einzelne Gelenke.

**Tab. 3** Mitzuteilende Daten bei Publikationen

- Zahl der Patienten mit Erosionen bei Therapiebeginn
- Zahl der Patienten mit Röntgenbildern im Verlauf
- Mittel- und Medianwerte mit SD oder Koifidenzintervallen im Verlauf
- Zahl der Patienten mit Progression > MDC\*
- Zahl der Patienten mit Rückgang des Scores
- Publikation in Score-Einheiten oder (besser) % des max. möglichen Scores

\* MDC=minimal detectable change – synonym: SDD=smallest detectable difference

Bei Publikationen sollen die in Tabelle 3 aufgeführten Daten mitgeteilt werden.

Durch Angabe der Score-Werte in % des max. möglichen Scores können Daten, die mit unterschiedlichen Methoden oder Modifikationen erhoben wurden, verglichen werden. Alle Röntgenbilder (auch die von Abbrechern) sollten zu den vorgeschriebenen Zeitpunkten angefertigt und ausgewertet werden – last value carried forward würde Abbrecher wegen (noch) geringer Progression begünstigen.

## Kosten und Zeitbedarf

Die wesentlichen Kosten entstehen durch das Anfertigen der Röntgenaufnahmen. Für das Scoring incl. Aufhängen der Bilder und Dokumentation muss für einen Patienten (ein Set) im Durchschnitt eine halbe Stunde für Untersucher und Hilfskraft angesetzt werden. Die Verlaufsbeurteilung ist bei Spätfällen zeitintensiver als bei Frühfällen. Der Zeitbedarf ist für den Sharp-Score und seine Modifikationen doppelt so groß wie für den Larsen- oder Ratingen-Score. Zusätzliche Kosten entstehen durch Versand der Aufnahmen, Randomisierung, Verblindung, ggf. Digitalisierung und Apparatur zum Lesen digitalisierter Aufnahmen.

Für den in der Beurteilung von Skelett-Röntgenaufnahmen von Rheumatikern Erfahrenen ist das Erlernen der jeweiligen Methode relativ einfach. Trainingssitzungen sind unabdingbar, wenn Aufnahmen einer Studie von mehreren Untersuchern beurteilt werden sollen. Bei jeder Studie muss durch wiederholtes Lesen eines Teils der Röntgenbilder die Fehlerbreite (Präzision) und damit die kleinste messbare Änderung (minimal detectable change) bestimmt werden, da diese sehr stark abhängig ist von der Qualität der Röntgenbilder, der Schwere der vorliegenden Destruktion (Früh- oder Spätfälle), dem Ausmaß der Destruktion und der Erfahrung (der Einsatzbereitschaft) des oder der Untersucher.

## Literatur

1. Resnick D, Niwayama G (1998) Rheumatoid arthritis and the seronegative Spondylarthropathies: Radiographic and Pathologic concepts in Resnick and Niwayama: Diagnosis of Bone and Joint Disorders, Vol. 2. WB Saunders Company, Philadelphia, PA
2. Sharp JT, Gardner JC, Bennett EM (2000) Computer-based methods for measuring joint space and estimating erosion volume in the finger and wrist joints of patients with rheumatoid arthritis. *Arthritis Rheum* 43: 1378–86
3. Kaye JJ (1991) Radiographic methods of assessment (scoring) of rheumatic disease. *Rheum Dis Clin North America* 17(3):457–469
4. van der Heijde DM (1999) How to read radiographs according to the Sharp/van der Heijde method. *J Rheum* 26:743–745
5. Sharp JT, Lidsky MD, Collins LC, Moreland J (1971) Methods of scoring the progression of radiologic changes in rheumatoid arthritis. *Arthritis Rheum* 14:706–720
6. Steinbrocker O, Traeger CH, Batterman RT (1949) Therapeutic criteria in rheumatoid arthritis. *JAMA* 140: 659–662
7. Kellgren HJ (1956) Radiological signs of rheumatoid arthritis. A study of observer differences in the reading of hand films. *Ann Rheum Dis* 15:55–60
8. Larsen A, Dale K, Eek M (1977) Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. *Acta Radiol Diagn* 18:481–491
9. Larsen A, Thoen J (1987) Hand radiography of 200 patients with rheumatoid arthritis repeated after an interval of one year. *Scand J Rheum* 16:395–401
10. Rau R, Herborn G (1995) A modified version of Larsen's scoring method to assess radiologic changes in rheumatoid arthritis. *J Rheumatol* 22:1976–1982
11. Scott D, Houssien D, Laasonen L (1995) Proposed modification to Larsen's scoring method for hand and wrist radiographs. *Br J Rheumatol* 34:56
12. Edmonds J, Saudan A, Lassere M, Scott D (1999) Introduction to reading radiographs by the Scott modification of the Larsen method. *J Rheumatol* 26:740–742
13. Larsen A (1995) How to apply Larsen score in evaluating radiographs of rheumatoid arthritis in longterm studies? *J Rheumatol* 22:1974–1975
14. Rau R, Wassenberg S, Herborn G, Stucki G, Gebler A (1998) A new method of scoring radiographic change in rheumatoid arthritis. *J Rheumatol* 25:2094–2107
15. Sharp JT, Young DY, Bluhm GB, Brook A, Brower AC et al (1985) How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis? *Arthritis Rheum* 28:1326–1335
16. van der Heijde DMFM, van Riel PL, Nuvér-Zwart HH et al (1989) Effects of hydroxychloroquine and sulfasalazine on progression of joint damage in rheumatoid arthritis. *Lancet* i:1036–1038
- 16a. van der Heijde DMFM (1996) Plain X-rays in rheumatoid arthritis: overview of scoring methods, their reliability and applicability. *Bailliere's Clinical Rheumatology* 10:435–453
17. Genant HK (1983) Methods of assessing radiographic change in rheumatoid arthritis. *Am J Med* 75:35–47
18. Boers M, Verhoeven AC, Markusse HM, van de Laar MAFJ, Westhovens R et al (1997) Randomised comparison of combined step-down prednisolone, methotrexate and sulfasalazine with sulfasalazine in early rheumatoid arthritis. *Lancet* 350:309–318
19. Lipsky PE, van der Heijde DMFM, St Clair EW et al (2000) Infliximab and methotrexate in the treatment of rheumatoid arthritis. *N Engl J Med* 343:1594–1602
20. Bathon JM, Martin RW, Fleischmann RM et al (2000) A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. *N Engl J Med* 343:1586–1593
21. Rau R, Wassenberg S, Zeidler H (2000) Low dose prednisolone therapy (LDPT) retards radiographically detectable destruction in early rheumatoid arthritis – Preliminary results of a multicenter, randomized, parallel, double-blind study. *Z Rheumatol* 59:Suppl II/90–96
22. Rückmann A, Ehle B, Trampisch HJ (1995) How to evaluate measuring methods in the case of non-defined external validity. *J Rheumatol* 22: 1998–2000
23. Bland JM, Altman DD (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, pp 307–310
24. O'Sullivan MM, Lewis PA, Newcombe RG et al (1990) Precision of Larsen grading of radiographs in assessing progression of rheumatoid arthritis in individual patients. *Ann Rheum Dis* 49:286–289
25. Salaffi F, Carotti M (1997) Observer variation in quantitative analysis of hand radiographs in rheumatoid arthritis: Comparison of 3 different reading procedures. *J Rheum* 24: 2055–2056