

Scoringmethoden bei der rheumatoiden Arthritis

R. RAU, S. WASSENBERG

■ Grundlagen und Prinzip der Scoringmethoden

Die Beeinflussung der radiologischen Progression gilt als wichtigstes Kriterium der Wirksamkeit eines Basistherapeutikums. Verschiedene Scoringmethoden wurden entwickelt, um die im Röntgenbild erkennbaren Veränderungen quantifizieren zu können. Tabelle 2.1 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 semiquantitativen Scoringmethoden erfassen insbesondere die erosiven Veränderungen. Sie zählen

Tabelle 2.1. Pathologisch anatomischer Befund und Röntgenbefund im Vergleich (nach Resnick) [1]

Pathologische Anatomie	Röntgenbefund
■ Entzündung der Synovialis und Flüssigkeitsansammlung	Weichteilschwellung und Gelenkspalterweiterung
■ Hyperämie und Knochenmineralverlust	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

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 2.1); 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.

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

Die Scoringmethoden 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 Scorewerte 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 dorsoplantar oder dorsopalmar):
 - 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 inklusive distale Radius- und Ulnaenden,
- konstante Belichtung

Halbschräge Aufnahmen sind unnötig, da die gleiche Projektion im Verlauf nicht erreicht wird. Weichteilaufnahmen 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 („setwise“). 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 konservativeren 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. Schon digital erhobene Röntgenbilder werden auf CD versandt und am Computerbildschirm befundet.

■ Darstellung der Scoringmethoden

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 4 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 5 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 [8], da diese aber nicht befriedigten, wurden sie durch Strichzeichnungen (Abb. 2.1) ersetzt [9].

Die Schweregrade werden folgendermaßen definiert [9]:

Grad 0 = normal
 Grad 1 = geringe Veränderungen: Weichteilschwellung, gelenknahe Osteoporose oder geringe Gelenkspaltverschmälerung
 Grad 2 = definitive Veränderungen: eine oder mehrere kleine Erosionen vorhanden, Gelenkspaltverschmälerung nicht obligat
 Grad 3 = deutliche Veränderungen: ausgeprägte Erosionen und Gelenkverschmälerung 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

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 Gesamtscore 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 Handgelenksscore mit 5 multipliziert. Der Gesamtscore 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.*

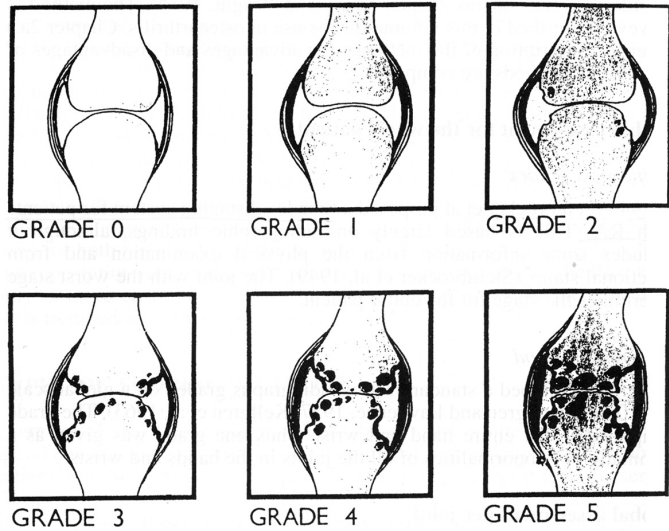


Abb. 2.1. Röntgenologische Grade oder Stadien der RA nach Larsen [8] aus Larsen und Thoen [9]

■ **Modifikation von Rau und Herborn 1994** [10]. In dieser schon 1986 [10 a] angewandten 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 Gesamtscore beträgt 128.

Grad 0 = normal

Grad 1a = periartikuläre Osteoporose/Gelenkschwellung (falls wesentlich)

Grad 1b = Erosionen/Zysten an zwei Stellen < 1 mm vorhanden

Grad 2 = eine oder mehrere Erosionen mit Unterbrechung der Kortex 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 Scorewert – widerspricht dem „Geist“ des Larsen-Scores, der nur destruktive Veränderungen bewertet.*

■ **Modifikation von Edmonds 1999** [12]. Es werden wieder 5 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/Zysten an 2 Stellen < 1 mm
Grad 2 = eine oder mehr Erosionen mit > 2 mm Kortexunterbrechung
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).

Das Stadium 2 wird jetzt durch eine oder mehrere kleine Erosionen mit einem Durchmesser von > 1 mm definiert.

Die Definition der anderen Stadien bleibt unverändert (Abb. 2.2). Die zu scorenden Gelenke und Regionen gehen aus Abbildung 2.3 hervor. IP I, MCP I und MTP I werden nicht gescored, am Handgelenk werden 4 Quadranten getrennt bewertet. Bei insgesamt 32 Gelenken beträgt der maximale Gesamtscore 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) als Stadium 1 enthält, werden nur destruktive Veränderungen am Knochen ge-

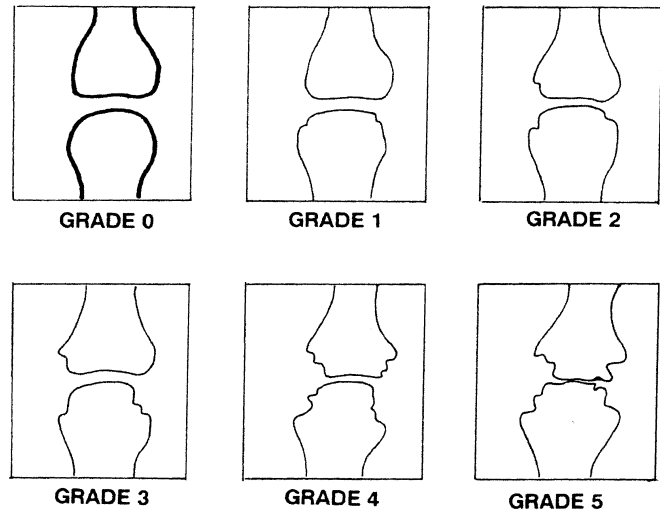


Abb. 2.2. Schematische Darstellung der Graduierung bzw. der Stadieneinteilung in der Larsen-Modifikation [13]

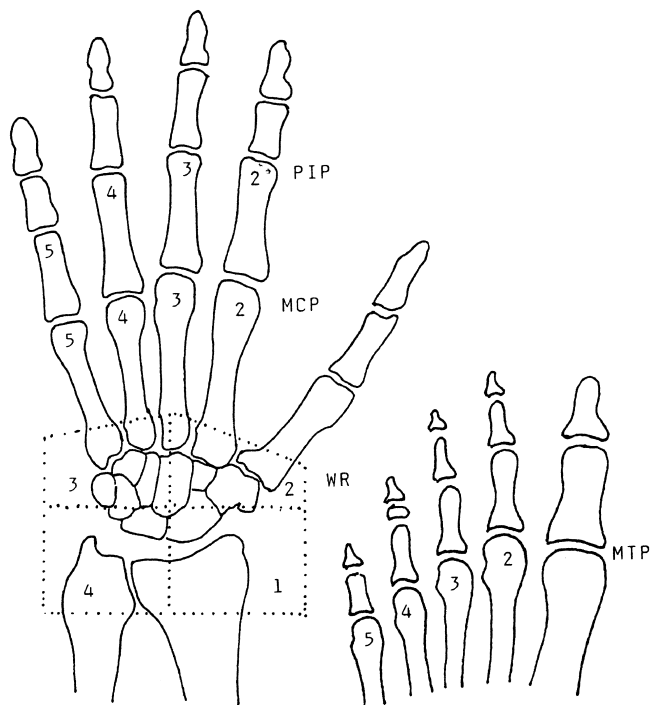


Abb. 2.3. Gelenke oder Regionen, die in der Modifikation von Larsen [13] gescored werden

wertet. Eine Bewertung der Gelenkspaltverschmälerung erfolgt nicht. Am Handgelenk werden, wie bei der Sharp-Methode, nur einzelne Skelettabschnitte (die änderungssensitivsten und am häufigsten betroffenen) berücksichtigt.

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 (Großzehengelenk).

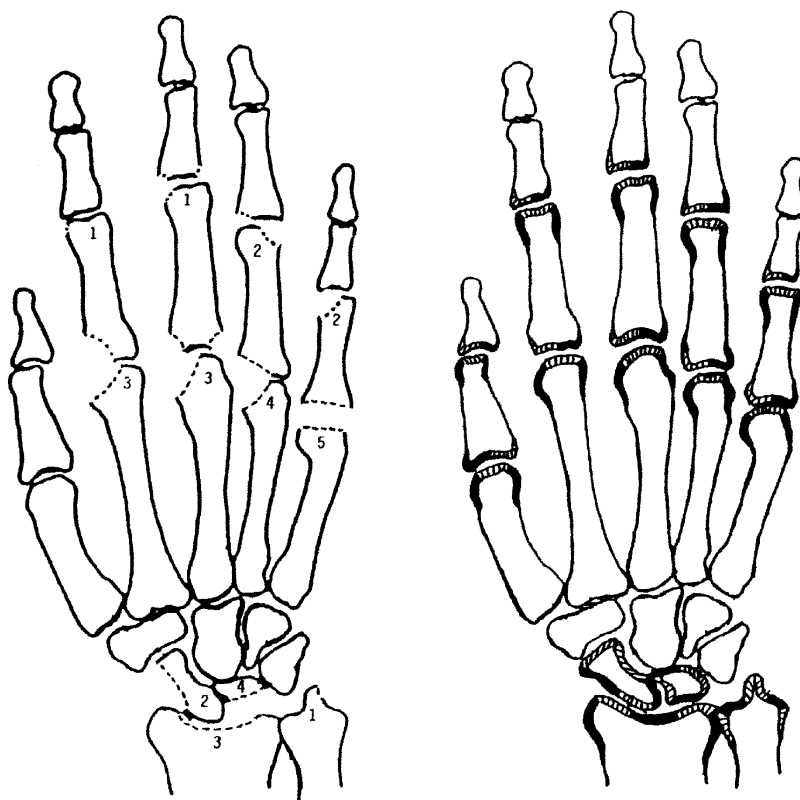


Abb. 2.4. Ratingen-Score: Schematische Darstellung der verschiedenen Grade sowie der Einteilung der Gelenkoberfläche in jeweils 20%-Abschnitte [14]

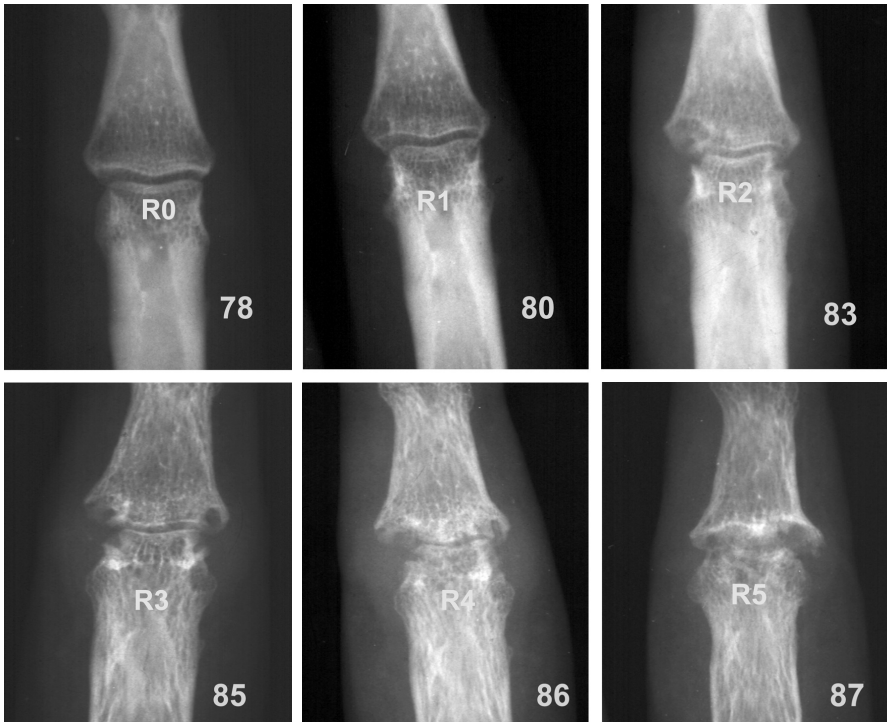


Abb. 2.5. Beispiele von Röntgenbildern des PIP-Gelenks III der gleichen Patientin (E, I) in den verschiedenen Stadien des Ratingen-Scores

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. 2.4). Abbildung 2.5 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 (Gelenkspaltverschmälerung) 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. 2.6) beträgt der maximale Erosionsscore 170 (34×5). Der maximale Gelenkspaltverschmälerungsscore beträgt 144 (36 Gelenke×4). Subluxationen oder Luxationen werden nicht gewertet. Der maximale Gesamtscore beträgt 314.

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

Gesamtscore 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.

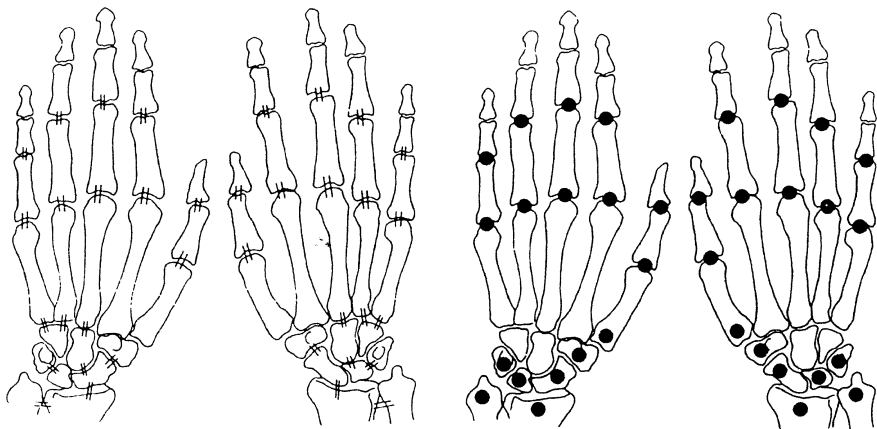


Abb. 2.6. Gelenke, die bei der Sharp-Methode (15) bewertet werden. Die mit einem Punkt markierten Gelenke werden auf Erosionen, die mit Parallelstrichen markierten Gelenke auf Gelenkspaltverschmälerung untersucht

Der Erosionsscore wird seit 1996 [16a] 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. 2.7). Die Summe dieser Scorewerte darf an einem Gelenk der Hand aber niemals 5 überschreiten.

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

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

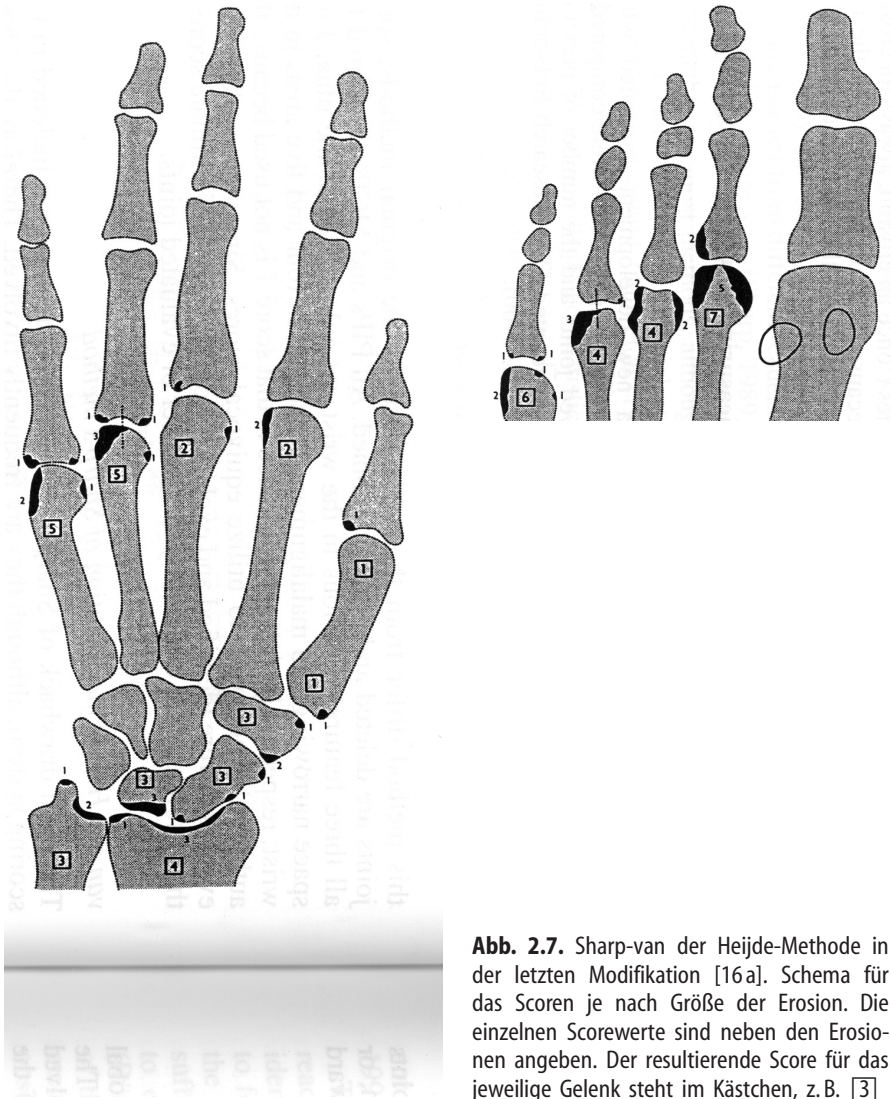


Abb. 2.7. Sharp-van der Heijde-Methode in der letzten Modifikation [16a]. Schema für das Scoring je nach Größe der Erosion. Die einzelnen Scorewerte sind neben den Erosionen angegeben. Der resultierende Score für das jeweilige Gelenk steht im Kästchen, z. B. [3]

Erosionsscore*	Gelenkspaltverschmälerungsscore**
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 oder 2×2 oder 1×3+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 Metakarpale 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, Karpometakarpalgelenke I–III, Trapezionavikulargelenk, Gelenk zwischen Naviculare/Lunatum und Capitatum, Radiokarpalgelenk, MTP-Gelenke I–V, IP I (Großzehe)

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 Erosionsscore beträgt 280 (32×5 an den Händen, 12×10 an den Füßen), der maximale Gelenkspaltverschmälerungsscore beträgt 168 (42 Gelenke×4), der maximale Gesamtscore 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:

Erosionsscore*	Gelenkspaltverschmälerungsscore**
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

- + wird mit 0,5 Punkten bewertet
- * 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 Capitatum, Scaphoid, Lunatum und Radiokarpalgelenk. Maximaler Score =104, Gesamtscore 0–202

Die Methode lässt sich entsprechend auch für die Zehengrundgelenke und das Großzehenendgelenk 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 Fehlstellungsscore:

Erosionsscore	JSN-Score*	Fehlstellungsscore (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.

■ **Simplifizierte Scoringmethode von van der Heijde 1977** [18]. Bei diesem Simple Erosion Narrowing Score (SENS) werden die gleichen Gelenke bewertet wie in der Sharp-van der Heijde-Methode. Ein Punkt wird vergeben, falls Erosionen an diesem Gelenk vorhanden sind, unabhängig von Zahl und Größe; dazu kommt ein weiterer Punkt, falls eine Gelenkspaltverschmälerung vorhanden ist, unabhängig von deren Schwere. Die Gelenke können somit von 0–2 graduiert werden. Eigentlich wird lediglich die Zahl der befallenen Gelenke ohne Differenzierung des Schweregrades festgestellt. Da die Zahl der nicht betroffenen Gelenke im Verlauf schnell abnimmt, wird die Methode im Laufe der Zeit immer unempfindlicher Änderungen registrieren.

Vor- und Nachteile der Methoden

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 Scoringmethoden reflektieren die Progression und den aktuellen Krankheits-

status 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 Intraclasskorrelationskoeffizienten von $>0,9$ gut (siehe Validierung der Befunde).

Nachteile

Die Scoringmethoden 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 Änderungsensitivität. 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 Scoreänderung. 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 Interuntersuchervarianz. 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 Gelenkspaltverschmälerungsscore. Damit werden Knorpeldestruktion und Gelenkfehlstellung (Kapsel- und Bandlaxizität) in den gleichen Topf geworfen. Der Gelenkspaltverschmälerungsscore diskriminiert oft schlechter zwischen zwei Therapien als der Erosionsscore [19–21].

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 Scorereduktionen beschrieben [19, 20, 22]. Heilungen am einzelnen Gelenk ohne Verkleinerung des Defektes („recortication“) 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 Scoringmethoden 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 [23] und ist nach Bland und Altman [24] 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 [25]. Für den Sharp-Score ergab sich bei im Abstand von 1½ Jahren aufgenommenen Röntgenbildern ein MDC von 10,1% des maximalen Scores [26].

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

Tabelle 2.2. Wiederholbarkeit, Präzision und Zeitbedarf verschiedener Scoringmethoden

Methoden	Wiederholbarkeit	Präzision (MDC)	Zeitbedarf
Van der Heijde-Score			
■ Alle Untersucher	2,8	2,9%	22,22
■ Bester/schlechtester Untersucher	3,8/1,8	2,1/3,7%	11,13/42,55
■ Rater Untersucher	2,8	3,3%	14,74
Larsen-Score			
■ Alle Untersucher	2,7	3,5%	9,88
■ Bester/schlechtester Untersucher	3,3/2,3	2,4/4,6%	5,54/15,75
■ Rater Untersucher	2,6	4,2%	7,06
Ratingen-Score			
■ 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×2,8; Ergebnisse standardisiert auf den Prozentsatz des maximalen Scorewerts; 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
- Rater Untersucher: GH, SW

Ratingen-Score:

- erfahrene Untersucher: GH, SW
- unerfahrene Untersucher: VFK, TL, CS
- alle Untersucher: erfahrene und unerfahrene Untersucher

Methoden mit dem Ratingen-Score vorgenommen. Der Vergleich erfolgte anhand 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 2.2 dargestellt.

■ **Indikationen und Kontraindikationen**

Die Anwendung von Scoringmethoden 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

Tabelle 2.3. Mitzuteilende Daten bei Publikationen

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

MDC „minimal detectable change“ – synonym: SDD „smallest detectable difference“

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 Scorewerte für jedes einzelne Gelenk und jeden Zeitpunkt erfolgt auf entsprechenden Formblättern. Die Auswertung kann erfolgen für Patientengruppen, einzelne Patienten, Gelenkgruppen und einzelne Gelenke.

Bei Publikationen sollen die in Tabelle 2.3 aufgeführten Daten mitgeteilt werden.

Durch Angabe der Scorewerte 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 Scoren 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 zeitaufwändiger 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 Skelettröntgenaufnahmen 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.

■ Nachtrag

John T. Sharp hat in den letzten beiden Jahren eine bedeutsame Änderung bzw. Ergänzung seiner Methode eingeführt. In mehreren Publikationen [27, 28, 29] hat er einen „Modified Sharp Score“ angegeben. Darin werden die Grade seines Erosionsscores folgendermaßen beschrieben:

Grad 1: eine Erosion oder Destruktion der Gelenkoberfläche bis 20%

Grad 2: zwei Erosionen oder Destruktion der Gelenkoberfläche 21–40%

Grad 3: drei Erosionen oder Destruktion der Gelenkoberfläche 41–60%

Grad 4: vier Erosionen oder Destruktion der Gelenkoberfläche 61–80%

Grad 5: fünf Erosionen oder Destruktion der Gelenkoberfläche > 80%.

Sharp empfiehlt bei gut abgrenzbaren Erosionen seine ursprüngliche Version zur Grundlage des Scores zu nehmen, bei weniger gut abgrenzbaren Erosionen (was häufig der Fall ist) die betroffene Fläche als Grundlage für den Score heranzuziehen.

Sharp hat jetzt also den Ratingen-Score exakt übernommen, allerdings ohne die Autoren des Ratingen-Scores zu zitieren. Dieses Vorgehen spricht für die Vorzüge des Ratingen-Scores und wird seine Verbreitung sicher fördern. Andererseits ist der Sharp-Score nicht mit dem Ratingen-Score kompatibel. Beispielsweise können 2 oder 3 kleine abgrenzbare Usuren, die im Sharp-Score mit 2 oder 3 bewertet werden, nur eine Oberflächendestruktion von <20% ausmachen und würden damit mit 1 bewertet. Ein Konfluieren kleiner Erosionen könnte also zu einer Scorereduktion Anlass geben, obwohl die Destruktionsfläche und damit die Größe der Destruktion zugenommen hat.

■ 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* 43:1378–186

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
- 10a. Rau R, Herborn G, Karger T, Menninger H, Elhard D, Schmitt J (1991) A double blind randomized parallel trial of intramuscular methotrexate and gold sodium thiomalate in early erosive rheumatoid arthritis. *J Rheumatol* 18:328–333
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, Nuyt-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. van der Heijde D, Dankert T, Nieman F, Rau R, Boers M (1999) Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis. *Rheumatology, Oxford* 38:941–947
19. Boers M, Verhoeven AC, Markkuse 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
20. 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
21. 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
22. 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
23. 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
 24. Bland JM, Altman DD (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, pp 307–310
 25. 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
 26. 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
 27. Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hoa Y et al (2004) Radiographic, clinical and functional outcomes of treatment with adalimumab in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum* 50: 1400–1411
 28. Breedveld FC, Weisman HM, Kavanaugh AF, Cohen FB, Pawelka K et al (2006) The PREMIER study. *Arthritis Rheum* 54:26–37
 29. SONORA Studie – noch nicht publiziert