



UNIVERSITAT AUTÒNOMA DE BARCELONA
FACULTAT DE MEDICINA
DEPARTAMENT DE CIRURGIA

TESI DOCTORAL

**ANÀLISI METODOLÒGIC DELS ESTUDIS DE
CONCORDANÇA INTEROBSERVADOR I
REPRODUCTIBILITAT INTRAOBSERVADOR EN LA
CARACTERITZACIÓ DE FRACTURES DE L'EXTREM
PROXIMAL DE L'HÚMER.**

Autora

Elisa Cassart i Masnou

Directors

Enric Cáceres i Palou
Carlos Torrens i Cánovas

Autora

Elisa Cassart i Masnou

Directors

Enric Cáceres i Palou

Carlos Torrens i Cánovas

ANÀLISI METODOLÒGIC DELS ESTUDIS DE CONCORDANÇA
INTEROBSERVADOR I REPRODUCTIBILITAT INTRAOBSERVADOR EN LA
CARACTERITZACIÓ DE FRACTURES DE L'EXTREM PROXIMAL DE
L'HÚMER

Departament de cirurgia

Facultat de Medicina

Universitat Autònoma de Barcelona

2014



UNIVERSITAT AUTÒNOMA DE BARCELONA
FACULTAT DE MEDICINA
DEPARTAMENT DE CIRURGIA

Tesi doctoral:

**ANÀLISI METODOLÒGIC DELS ESTUDIS DE CONCORDANÇA
INTEROBSERVADOR I REPRODUCTIBILITAT
INTRAOBSERVADOR EN LA CARACTERITZACIÓ DE FRACTURES
DE L'EXTREM PROXIMAL DE L'HÚMER**

Tesi realitzada per Elisa Cassart i Masnou, sota la direcció dels professors Dr. Enric Cáceres i Palou i Dr. Carlos Torrens i Cánovas, per optar al grau de doctor en cirurgia per la Universitat Autònoma de Barcelona

Barcelona, juny 2014

Doctorand:

Directors:

Dr. Enric Cáceres i Palou

Catedràtic de Cirurgia Ortopèdica i Traumatologia

Facultat de Medicina

Universitat Autònoma de Barcelona

Dr Carlos Torrens i Cánovas

Professor associat de Cirurgia
Ortopèdica i Traumatologia

Facultat de Medicina

Universitat Autònoma de Barcelona

I ens ha costat Déu i ajuda
arribar fins aquí.....

Mai deixarem d'explorar
i el final de l'exploració serà
arribar al punt de partida
i conèixer el lloc per primera vegada.

T.S. ELIOT

Als meus pares, Josep i M^a Àngels, perquè sense ells no hauria arribat mai fins aquí. Per l'esforç, l'educació, el sacrifici, la humilitat, la bondat i la honestat. Al meu pare per la xerrada anual a Palamós. A la meua mare perquè també és la preferida.

Per en Joan, per creure sempre en mi i seguir el meu camí. Per donar-me cada dia l'equilibri just. Per no perdre mai la perspectiva. Per estar sempre content. Perquè t'estimo com mai a ningú.

Pel meu germà Francesc, el meu pol oposat, per ser tot allò que jo no he pogut ser mai.

A l'Aleix per existir

AGRAÏMENTS

Al Dr Cáceres, catedràtic de Cirurgia Ortopèdica i Traumatologia de la Universitat Autònoma de Barcelona, per acceptar dirigir aquesta tesi.

A en Carles Torrens, el nostre guia espiritual. Per la humilitat, la honestedat, la simplicitat i el pragmatisme, en l'espatlla i en la vida. Per qüestionar allò inqüestionable. Per la confiança i l'ensenyament. Per convertir l'espatlla en la meva passió.

A Fernando Santana, per compartir camí, amicitat i passió per l'espatlla. Perquè no perdem mai el nord en aquest món de bojós.

A Josep M^a Mora, Rocío Goberna, Antonio Pablo Rosales, Ferran Gàmez, Mónica Corrales i Aina Ruiz per totes les hores dedicades a la lectura i selecció d'imatges d'aquest treball. Una col·laboració que no té preu.

Per Joan Miquel, Albert Alier, Xavier Piqueres, Toni Quinjoan, i la resta de components del grup d'estudi de l'espatlla, perquè poguem seguir gaudint d'aquells caps de setmana lúdico-científics tan profitosos.

A Sergi Mojal, per la incalculable ajuda en l'anàlisi estadístic.

Per Cohen...el del kappa.

Pels meus companys i amics del servei de traumatologia de l'hospital Germans Trias i Pujol que m'han recolzat en aquest camí des del primer dia. A en Leo pels ànims, consells i correccions.

Als qui al llarg de la meva vida, volent o sense voler m'han portat fins aquí.

Als que m'he deixat

SUMARI GENERAL

1.JUSTIFICACIÓ I MOTIVACIONS DEL TREBALL	17
2.INTRODUCCIÓ	19
2.1 Epidemiologia de les fractures de l'extrem proximal de l'húmer.....	19
2.2 Aspectes anatòmics i biomecànics de les fractures de l'extrem proximal de l'húmer.....	20
2.3 Avaluació clínic-radiològica.....	23
2.4 Sistemes de classificació de les fractures.....	25
2.4.1 Classificació en les fractures de l'húmer proximal.....	26
2.4.2 Nous sistemes de caracterització.....	31
2.5 Concordança i reproductibilitat en la classificació de Neer.....	35
2.5.1 Concordança interobservador.....	35
2.5.2 Reproductibilitat intraobservador.....	38
2.5.3 El coeficient Kappa.....	39
2.5.3.1 Les paradoxes del Kappa.....	42

3.HIPÒTESI I OBJECTIUS DEL TREBALL 51

3.1 Hipòtesi del treball..... 51

3.2 Objectius secundaris..... 51

4. MATERIAL I MÈTODES 53

1ª PART

4.1 Estudi de la concordança interobservador i la reproductibilitat intraobservador en la caracterització de fractures de l'húmer proximal. Influència de la seqüència de la TAC (axial i completa) i de l'experiència de l'observador (experts i no experts)... 53

4.1.1 Selecció de la mostra..... 54

4.1.2 Variables a estudi..... 54

4.1.3 Selecció dels observadors..... 65

4.1.4 Observacions..... 65

4.1.5 Anàlisi estadística..... 66

2ª PART

4.2 Influència del biaix de prevalença en la concordança interobservador i reproductibilitat intraobservador de la classificació de Neer..... 67

4.2.1 Mostra consecutiva.....	67
4.2.1.1 Selecció de la mostra	67
4.2.1.2 Variables a estudi.....	68
4.2.1.3 Selecció dels observadors.....	68
4.2.1.4 Observacions.....	68
4.2.1.5 Anàlisi estadística.....	68
4.2.2 Mostra seleccionada.....	69
4.2.2.1 Selecció de la mostra.....	69
4.2.2.2 Variables a estudi.....	69
4.2.2.3 Selecció dels observadors.....	70
4.2.2.4 Observacions.....	70
4.2.2.5 Anàlisi estadística.....	70
4.2.3 Comparació entre mostra consecutiva i mostra seleccionada.....	71

5. RESULTATS

73

1ª PART

5.1 Estudi de la concordança interobservador i reproductibilitat intraobservador en la caracterització de fractures de l'húmer proximal. Influència de la seqüència de la TAC (axial i completa) i de l'experiència de l'observador (experts i no experts)..... 73

5.1.1 Concordança interobservador en TAC axial i TAC completa amb tots els observadors..... 73

5.1.2 Reproductibilitat intraobservador en TAC axial i TAC completa amb tots els observadors..... 77

5.1.3 Comparació de la concordança interobservador entre TAC axial i TAC completa segons grau d'experiència..... 79

5.1.4. Comparació de la concordança interobservador entre experts i no experts segons l'estudi radiològic..... 84

2ª PART

5.2 Influència del biaix de prevalença en la concordança i reproductibilitat de la classificació de Neer..... 88

5.2.1 Concordança interobservador en la classificació de Neer en la mostra consecutiva..... 88

5.2.2 Reproductibilitat intraobservador en la classificació de Neer en la mostra consecutiva..... 92

5.2.3. Concordança interobservador en la classificació de Neer en la mostra seleccionada..... 92

5.2.4	Reproductibilitat intraobservador en la classificació de Neer en la mostra seleccionada.....	96
5.2.5	Diferències estadístiques en la concordança interobservador i la reproductibilitat intraobservador de la classificació de Neer segons tipus de mostra.....	97
6.	DISCUSSIÓ	99
6.1	De la influència de la prevalença.....	100
6.2	De l'estudi radiològic.....	103
6.3	Dels observadors i l'experiència.....	108
7.	APLICABILITAT CLÍNICA	113
8.	CONCLUSIONS	117
8.1	Conclusions de la hipòtesi de treball.....	117
8.2	Conclusions dels objectius secundaris.....	117
9.	REFERÈNCIES BIBLIOGRÀFIQUES	119

SUMARI DE FIGURES

Fig. 1 Segments anatòmics de l'extrem proximal de l'húmer.....	20
Fig. 2 Desplaçaments dels fragments segons tracció muscular.....	21
Fig. 3 Irrigació i innervació de l'espatlla.....	21
Fig. 4 Projeccions radiològiques.....	24
Fig. 5 Combinacions possibles de fractura segons Codman.....	27
Fig. 6 Classificació de Neer.....	28
Fig. 7 Fractura en 4 parts en valg de Jakob.....	29
Fig. 8 Classificació AO/ASIF.....	30
Fig. 9 Sistema binari de Hertel.....	32
Fig. 10 Criteris addicionals de Hertel.....	33
Fig. 11 Dependència del coeficient Kappa de la prevalença del diagnòstic.....	45
Fig. 12 Contacte entre cap-diàfisi menor del 33%.....	55
Fig. 13 A. Angle cefalodiafisari B. Desplaçament lateromedial.....	56
Fig. 14 A. Fractura metafisària associada B. Fulcre intern no preservat.....	57
Fig. 15 A i B: Impactació diàfisi-cap.....	57
Fig. 16 Desplaçament diàfisi-tuberositat gran.....	58

Fig. 17 Desplaçament de la tuberositat gran respecte el cap major de 0,5 cm.....	59
Fig. 18 A.Orientació cefaloglenoidea suficient. B Orientació cefaloglenoidea insuficient.....	60
Fig. 19 A.Estoc ossi del cap suficient. B. Estoc ossi del cap insuficient.....	60
Fig. 20 Luxació posterior del cap humeral.....	61
Fig. 21 Fractura en 1 part de Neer.....	63
Fig. 22 Fractura en 2 parts de Neer.....	63
Fig. 23 Fractura en 3 parts de Neer.....	63
Fig.24 Fractura en 4 parts de Neer.....	63

SUMARI DE TAULES

Taula 1. Protocol de lectura radiològica de Mora i cols.....	34
Taula 2.Valors Kappa segons nombre de categories en la classificació de Neer en el treball de Sidor.....	36
Taula 3. Valors de Kappa intraobservador en sistemes de classificació de Neer i AO.....	39
Taula 4. Taula de contingència dos observadors valorant una variable dicotòmica.....	40
Taula 5. Criteris de Landis i Koch.....	41
Taula 6. Mostra A Concordança interobservador amb un acord observat de 0,8 i una prevalença del 50%.....	43
Taula 7 Mostra B Concordança interobservador amb un acord observat de 0,8 i una prevalença del 12%.....	44
Taula 8. Disbalanç simètric perfecte.....	47
Taula 9. Disbalanç imperfecte.....	47
Taula 10. Publicacions que valoren acord en classificació de fractures d'húmer proximal. Nombre de casos i selecció de la mostra.....	49
Taula 11. Resum de les diferents característiques radiològiques valorades.....	64
Taula 12. Resultats de concordança interobservador de tots els observadors amb intervals de confiança per cada una de les variables estudiades.....	74

Taula 13. Resultats de reproductibilitat intraobservador de tots els observadors amb intervals de confiança per cada una de les variables estudiades.....	77
Taula 14. Concordança interobservador amb TAC axial i TAC completa en els observadors experts.....	81
Taula 15. Concordança interobservador amb TAC axial i TAC completa en els observadors no experts.....	82
Taula 16-30. Taules de contingència per parelles d'observadors en la classificació de Neer simplificada en quatre categories en la mostra consecutiva.....	89-91
Taula 31. Reproductibilitat en la mostra consecutiva de cada observador.....	92
Taula 32-46. Taules de contingència per parelles d'observadors en la classificació de Neer simplificada en 4 categories en la mostra seleccionada.....	93-95
Taula 47. Reproductibilitat en la mostra seleccionada de cada observador.....	96
Taula 48. Prevalença observada en cada mostra dels diferents tipus de fractura.....	97
Taula 49. Resultats de concordança interobservador i reproductibilitat intraobservador en la classificació de Neer simplificada en la mostra consecutiva i la mostra seleccionada.....	98

SUMARI DE GRÀFICS

Gràfic 1: Comparació concordança interobservador en TAC axial entre experts i no experts..... 84

Gràfic 2: Comparació concordança interobservador en TAC completa entre experts i no experts..... 86

ABREVIACIONS

AO/ASIF: Arbeitsgemeinschaft für Osteosynthesefragen/ Association of the Study of Internal Fixation

TAC: Tomografia axial computaritzada

1. JUSTIFICACIÓ I MOTIVACIONS DEL TREBALL

L'ús dels sistemes de classificació és bàsic en la nostra especialitat, ja que ens ajuda a descriure el tipus de lesió, la gravetat d'aquesta, ens orienta en la presa de decisions terapèutiques i ens serveix com a factor pronòstic. A llarg termini ens serveix per valorar l'eficàcia d'un tractament i per comparar els nostres resultats amb altres membres de la comunitat científica que utilitzen diferents procediments terapèutics.

Els principals sistemes de classificació descrits per les fractures de l'extrem proximal de l'húmer, el sistema de Neer i el sistema de la AO/ASIF, han demostrat una pobra concordança interobservador i moderada reproductibilitat intraobservador. Diferents treballs han analitzat la influència de l'estudi radiològic utilitzat en aplicar un sistema de classificació i la influència dels observadors, pensant que podrien ser aquests factors els causants de la baixa concordança de la classificació, sense concloure quin és l'estudi radiològic més eficient i quina és la influència dels observadors en la fiabilitat de la classificació.¹⁻¹⁴

La motivació del treball, donat que la concordança i la reproductibilitat dels sistemes de classificació de fractures d'húmer proximal segueix sent molt baixa, és valorar en un sistema de caracterització de fractures de l'húmer proximal i en la classificació de Neer, l'estudi radiològic més eficient per obtenir la millor concordança quan definim característiques radiològiques, valorar la influència de l'experiència de l'observador i la influència de la selecció de la mostra en els resultats.

2.INTRODUCCIÓ

2.1 Epidemiologia de les fractures de l'extrem proximal de l'húmer

Les fractures de l'extrem proximal de l'húmer representen aproximadament el 4% de totes les fractures i al voltant del 45% de totes les fractures de l'húmer.¹⁵

En pacients joves es produeixen per un mecanisme d'alta energia mentre que en el pacient d'edat avançada la fractura es produeix per caigudes de baixa energia, seguint un patró de distribució bimodal. En pacients majors de 40 anys, el 75% de les fractures d'húmer tenen lloc en aquesta localització, sent més freqüents en dones en una proporció de 3:1. En dones majors de 60 anys es troben directament relacionades amb la presència d'osteoporosi i la seva incidència va augmentant conforma ho fa l'esperança de vida de la població.¹⁶⁻²⁰

Es calcula que el nombre de fractures en l'ancià és triplicarà en les pròximes tres dècades. Pel seu caràcter osteoporòtic són fractures que es comporten epidemiològicament de forma similar a les fractures de maluc i vertebrals i en moltes ocasions són la primera fractura a presentar-se.²¹

Igual que aquestes, i pel fet de donar-se en dones d'edat avançada, representen també un augment de la mortalitat respecte a la població general durant el primer any de la fractura.^{22,23}

2.2 Aspectes anatòmics i biomecànics de les fractures de l'extrem proximal de l'húmer

La porció més proximal de la diàfisi humeral es troba connectada amb el cap de l'húmer pel coll quirúrgic que es troba just per sota les tuberositats (gran o troquíter i petita o troquíin) Per sobre d'aquestes trobem el coll anatòmic que les separa del fragment articular del cap. *Figura 1* La superfície articular és un segment d'esfera amb un diàmetre de curvatura que varia entre 37 i 57mm i el cap es troba inclinat uns 130° respecte a la diàfisi. La retroversió que presenta el cap respecte a la diàfisi és difícil de determinar amb variacions entre 18 a 30° descrites en diferents estudis anatòmics així com variacions en la mateixa persona en estudis realitzats en espatlles sanes. La superfície glenoidea, on s'articula el cap humeral, és una estructura poc profunda que mesura aproximadament un terç del cap humeral, fet que confereix a l'articulació de l'espatlla un gran arc de moviment i alhora una major inestabilitat en comparació a altres articulacions. El labrum i la càpsula contribueixen a donar-li estabilitat.



Figura 1. Diferents segments anatòmics de l'extrem proximal de l'húmer

En la tuberositat gran trobem tres caretes per la inserció dels tendons dels muscles supraespinós , infraespinós i rodó menor i en la tuberositat petita trobem la inserció del subescapular. El solc bicipital, conté el tendó de la porció llarga del bíceps que es troba cobert pel lligament transvers humeral i separa ambdues tuberositats.

Els muscles i tendons de la còfia rotadora realitzen un paper crucial en el patró de desplaçament de cada fractura. *Figura 2* La tracció del supraespinós i l'infraespinós desplacen el troquíter a superior i a posterior respectivament, mentre el troquín es desplaçarà en direcció medial per la tracció del subescapular. El pectoral major desplaçarà la diàfisi medialment. En funció del traç de fractura i nombre de fragments es produiran els diferents tipus de desplaçaments.

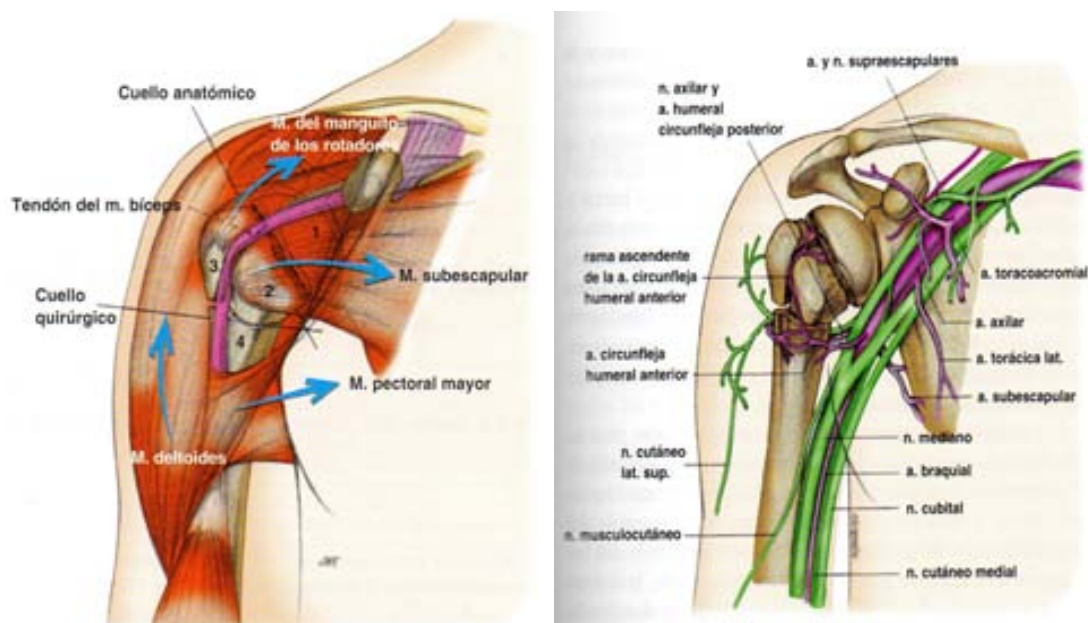


Figura 2. Esquerra. La tracció dels muscles que s'inserten en cada segment anatòmic produirà el desplaçament dels fragments i ens donarà els diferents patrons de fractura.

Figura 3. Dreta. Irrigació i innervació de l'espatlla

Rockwood & Matsen *The Shoulder*. 3ª ed. New York: Wirth&Lippitt.; 2006. p. 355-411.

L'aport arterial principal del cap humeral prové de l'artèria circumflexe humeral anterior, branca de l'artèria axil·lar la qual pot lesionar-se fàcilment en alguns tipus de fractura per la seva proximitat al coll quirúrgic i la tracció de la diàfisi a medial que provoca el pectoral major. *Figura 3*

Estudis recents però, han demostrat el paper de l'artèria circumflexe humeral posterior en la irrigació del cap, que segons estudis de cadàver podria contribuir en fins un 64% de l'aport sanguini i explicaria la taxa relativament baixa de necrosi avascular del cap en les fractures desplaçades multifragmentàries on sol estar conservada.²⁴

La innervació de l'espatlla depèn del plexe braquial. El nervi supraescapular (C5, C6) s'origina directament del plexe i innerva el múscle supraespinós i infraespinós, del cordó posterior (C5-C8) origina el nervi subescapular que innerva el múscle subescapular i aquest cordó acaba en el nervi axil·lar, que un cop ha travessat l'espai quadrilàter gira al voltant de l'húmer proximal en la superfície profunda del múscle deltoide i es ramifica en tres branques majors que aporten innervació motora al rodó menor i al deltoide. *Figura 3*

La majoria de fractures de l'húmer proximal es produeixen a través d'os osteoporòtic en pacients ancians per un mecanisme indirecte en caigudes de baixa energia sobre l'extremitat. Traumatismes d'alta energia poden produir lesions en aquesta localització a qualsevol edat. El contacte de l'acròmion i el marge de la glena, combinat amb la tracció dels tendons del manegot dels rotadors inicia els plans de fractura. Una vegada els fragments s'han separat, les forces musculars contribueixen a mantenir el desplaçament. La diàfisi es desplaça a anterior i medial per acció del pectoral major. El troquíter és traccionat a superior pel supraespinós i a posterior per l'infraespinós. El subescapular desplaça a medial un fragment de troquíen o bé provoca rotació interna del cap si no s'ha produït fractura d'aquest.^{15,25}

Sembla que la reducció anatòmica de les tuberositats és necessària per una completa restitució de la funció de l'espatlla quan s'ha produït una fractura, principalment de la tuberositat gran.²⁶

La força que ha de realitzar el deltoide en la flexoabducció del braç es veuria alterada quan hi ha una mala reducció de la tuberositat així com els vectors de força de supraespinós i infraespinós.

Diferents treballs han valorat la importància d'una correcta reducció de les tuberositats en les fractures tant en la cirurgia protèsica d'hemiartroplàstia com amb diferents muntatges de sutures que fixin les tuberositats.²⁷⁻²⁹ La corredera bicipital s'estableix com a referència per la reducció i el càlcul de l'alçada en la restitució anatòmica en cirurgia protèsica.³⁰

2.3 Avaluació clínic-radiològica

Davant una fractura de l'extrem proximal de l'húmer, és necessària una adequada valoració radiològica amb projeccions anteroposterior (AP) en el pla de l'escàpula, perfil en el pla de l'escàpula (Y d'escàpula) i axil·lar. La projecció anteroposterior permet valorar principalment el desplaçament del cap respecte la diàfisi i el desplaçament de la tuberositat gran, així com lesions associades en la glenoide. La projecció AP en rotació externa és lleument superior al resta de projeccions per detectar desplaçaments de la tuberositat gran.³¹ En la projecció en perfil d'escàpula valorem la presència de luxació i el desplaçament de les tuberositats. La projecció axil·lar, utilitzada per valorar principalment la posició del cap respecte la glenoide i el desplaçament de la tuberositat gran, és dolorosa i difícil d'obtenir en un pacient amb una fractura i a vegades pot canviar la posició de la fractura. Aquesta es pot substituir per una projecció axil·lar de

Velpreau. *Figura 4* D'aquesta manera es poden obtenir 3 projeccions ortogonals entre si , sense moure l'extremitat del cabestrell.

Si la fractura és relativament estable o porta diverses setmanes d'evolució, una projecció AP en rotació externa proporcionarà millor visualització del troquíter. Això pot obtenir-se fàcilment retirant el cabestrell i rotant lleument 20° el braç.

Si les radiografies no ens permeten realitzar una bona valoració de la fractura es sol·licita una TAC que a part de definir amb més precisió els traços i la localització dels fragments permetrà descartar fractures de glena i avaluar luxacions. En les fractures tributàries de tractament quirúrgic, aquest serà imprescindible per fer una bona planificació.

?

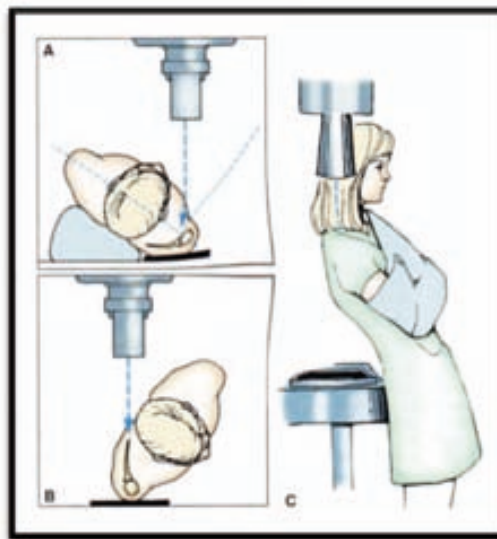


Figura 4. Projeccions radiològiques en l'estudi de fractures de l'húmer proximal. A. Projecció anteroposterior en el pla de l'escàpula per valorar el desplaçament del cap respecte la diàfisi i de la tuberositat gran. B. Projecció en Y d'escàpula, ens permet valorar luxacions i el desplaçament de la tuberositat gran. C. Projecció de Velpreau, es realitza quan hi ha dificultats en realitzar la projecció en Y.

Rockwood & Matsen The Shoulder. 3ª ed. New York: Wirth&Lippitt.; 2006. p. 355-411.

2.4 Sistemes de classificació de les fractures

Els sistemes de classificació són utilitzats en la nostra especialitat per tenir una forma estandarditzada i comú per descriure els diferents patrons d'una fractura. Els utilitzem per la decisió terapèutica i per avaluar, a curt o a llarg termini, els resultats d'aquest i poder-los comparar. Assumim que apliquem un determinat tractament en funció de la gravetat de la lesió i aquesta ens la determina el mateix sistema de classificació. Per tant aquest sistema ha de ser fiable, reproducible i concordant entre diferents observadors, ja que ens determinarà, junt amb altres factors particulars de cada fractura i pacient, el tractament a realitzar. Un bon sistema ha de suggerir un mètode de tractament i també una estimació del resultat d'aquest tractament.³²

Considerem que un sistema és útil quan un mateix observador obté el mateix resultat en diferents observacions, reproductibilitat intraobservador, i quan diferents observadors estan d'acord en les observacions, concordança interobservador. Alhora el sistema de classificació ha de ser obtingut per sèries radiològiques estandarditzades, reproduïbles i fàcils de realitzar.⁶

Les diferències entre observadors en un sistema de classificació no sempre tindran les mateixes implicacions. Si la diferència és petita, probablement no influeixi en el tractament a realitzar, però si les diferències són grans, el tractament a realitzar en una determinada fractura pot diferir molt.

Un dels principals problemes quan avaluem els sistemes de classificació és que utilitzem fórmules que ens valoren grau d'acord entre observadors, però no la precisió, és a dir la fiabilitat. El sistema fiable és el que s'acosta a la realitat. Com no tenim un patró or , ja que tots els sistemes depenen de la interpretació d'uns observadors, hem de fixar-nos en l'acord entre aquests esperant una distribució randomitzada dels errors i que hi haurà teòricament acord només quan

els observadors diguin el correcte. Allò que pot tenir un acord excel·lent pot ser molt poc precís perquè no s'acosti a la realitat.

Així els sistemes de classificació han de ser fiables i reproduïbles. Els resultats dels estudis realitzats per valorar sistemes de classificació en la nostra especialitat han estat decebedors amb pobre concordança interobservador i moderada reproductibilitat intraobservador.³³⁻³⁶

2.4.1 Classificació en les fractures d'húmer proximal

El primer en classificar les fractures d'húmer proximal fou Kocher³⁷, dividint-les en supratubercular, pertubercular, infratubercular i subtubercular.

El 1934 Codman³⁸ introdueix el concepte de 4 segments, observant que les fractures de l'húmer proximal tendeixen a produir quatre fragments majors en diverses combinacions: cap, tuberositat major, tuberositat menor i diàfisi. Codman introdueix el concepte de fragment i descriu les diferents combinacions possibles de fractura a través de les línies fissàries. *Figura 5*

La classificació de Watson-Jones, va intentar classificar les fractures segons el mecanisme de producció: impactada en adducció, impactada en abducció i mínimament desplaçada "*contusion-crack fractures*" Però les fractures en abducció o en adducció poden ser fàcilment confoses realitzant mínimes rotacions alhora de realitzar la projecció radiològica.³⁹

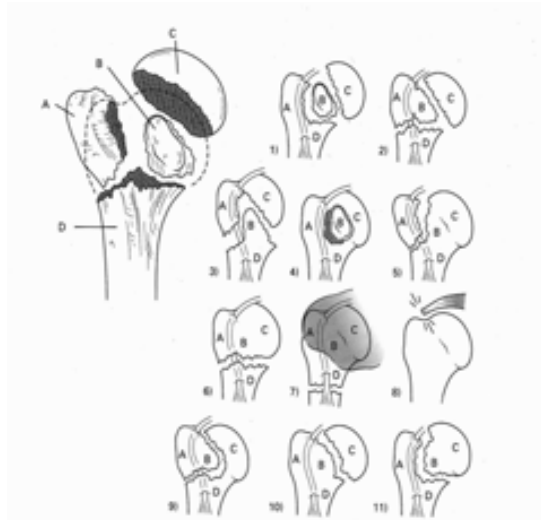


Figura 5. Combinacions possibles de fractura segons Codman. Rockwood and Green's Fractures in Adults. 5ª ed. Philadelphia: Lippincott Williams & Wilkins; 2003. p. 997-1040.

El 1953, Neer recolza l'observació de Codman i senyala que el patró en 4 fragments varia només en el grau de desplaçament de cada un dels fragments individuals i crida l'atenció sobre la possibilitat de necrosi avascular del fragment cefàlic en les fractures-luxacions, els mals resultats de les quals el van portar a dissenyar la seva pròtesi de substitució del cap humeral.⁴⁰

El 1970 Neer, utilitzant una sèrie de 300 fractures agafades a l'atzar, descriu el seu sistema de classificació en 4 fragments: articular, diàfisi, tuberositat major i tuberositat menor, fent èmfasi en els patrons de desplaçament més que en la localització de les línies de fractura. Segons Neer, els fragments es converteixen en parts quan algun dels 4 fragments es troba desplaçat entre si més d'1 cm o angulat més de 45°, de manera que segons aquest sistema, encara vigent actualment, podem trobar 16 tipus diferents de fractures en aquesta localització. *Figura 6* Tota fractura que no presenti aquest desplaçament es

considera com a mínimament desplaçada i es considera en 1 part, independentment del nombre i localització de les línies de fractura.⁴¹

Fracturas desplazadas


















	2 partes	3 partes	4 partes	Superficie articular
Cuello anatómico				
Cuello quirúrgico				
Troquíter				
Troquíin				
Fractura-luxación	Anterior 			
	Posterior 			
Con dehiscencia de la cabeza				

Figura 6. Sistema de classificació de Neer de fractures de l'húmer proximal. Segons l'afectació i desplaçament dels quatre segments de Codman podem trobar 16 combinacions possibles de fractura Rockwood & Matsen The Shoulder. 3ª ed. New York: Wirth&Lippitt.; 2006. p. 355-411.

Així, la fractura més freqüent és la fractura en 2 parts que podem subdividir en : 2 parts del coll quirúrgic, de la tuberositat major, de la tuberositat menor que són rars i les fractures aïllades del coll anatómic que són també molt poc freqüents.

Una fractura amb desplaçament en 3 parts conté diàfisi i, o bé la tuberositat major (amb més freqüència) o la menor (rarament) separades del cap, en el qual segueix unida l'altra tuberositat.

Una fractura en 4 parts es presenta quan el cap es troba separat de tots els altres segments, inclòs quan les tuberositats estan juntes, així pot haver-hi tres peces una de les quals conté les dues tuberositats. En una fractura en quatre parts "clàssica" el cap està luxat respecte a la glena, sense insercions de teixits tous. En la fractura en 4 parts "impactada en valg" el cap es troba girat a superior però manté algunes insercions amb periosti i càpsula.⁴² *Figura 7* Aquestes últimes tenen molt millor pronòstic que la fractura en 4 parts clàssica, ja que el risc de necrosi avascular és menor i el desplaçament en valg dona millors resultats funcionals, tot i així aquesta diferència no es reflecteix en el sistema de classificació.

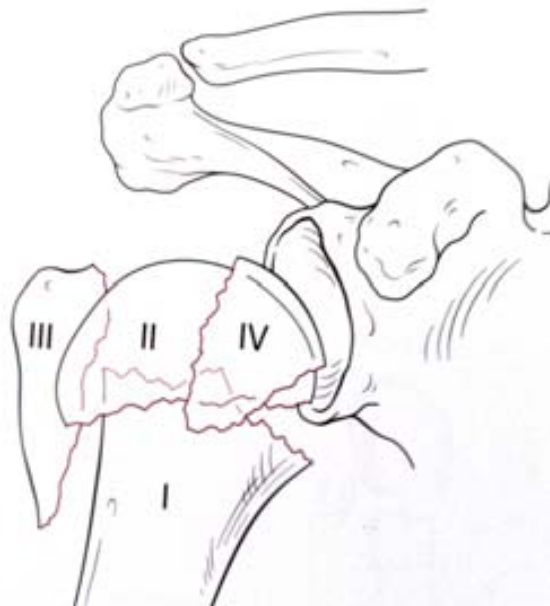


Figura 7. Fractura en 4 parts en valg de Jakob. Fractura en 4 parts amb el cap impactat en valg, al contrari de la fractura en 4 parts clàssica on el cap es troba desplaçat en var. Rockwood and Green's Fractures in Adults. 5ª ed. Philadelphia: Lippincott Williams & Wilkins; 2003. p. 997-1040.

Jakob i cols.⁴² van proposar una classificació basada en l'estudi de 730 fractures que es va incorporar al sistema de classificació del grup AO/ASIF, en el que apareixien 27 subgrups de fractures. El sistema AO posa més èmfasi en l'aport vascular del fragment articular. Si alguna de les tuberositats i el seu manegot associat romanen unides al fragment articular, l'aport vascular es considera adequat. La baixa incidència de necrosi avascular en les fractures en 4 parts impactades en valg es creu deguda a la integritat dels teixits tous de la part medial del coll anatómic. Aquest sistema pot tenir certa utilitat en la investigació clínica, però no ha demostrat ser més fiable que el sistema de Neer i presenta una complexitat més elevada.^{42,43} *Figura 8*

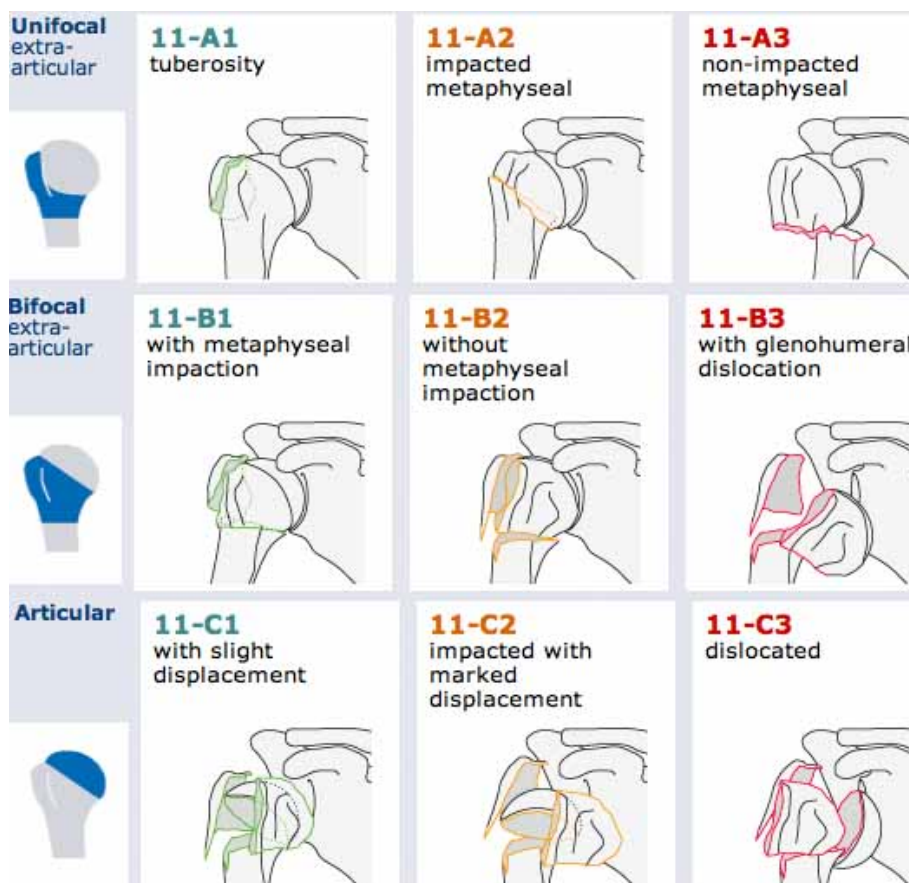


Figura 8. Sistema de classificació AO/ASIF

En estudis comparatius entre la classificació de Neer i la classificació de AO pel que fa a la concordança interobservador i la reproductibilitat intraobservador els resultats han estat decebedors. Tal com van observar Siebenrock i cols.⁷, seguint els criteris de Landis i Koch⁴⁴, el problema principal, per observadors experimentats, era la localització del troquín en la radiologia simple. Els talls de TAC i la valoració intraoperatòria podrien ajudar en la classificació final. Aquests autors obtenien només moderada concordança intra i interobservador.^{2-5,7} Estudis recents comparant ambdues classificacions conclouen que en ambdues manca informació rellevant per la descripció de la fractura.⁴⁵

2.4.2 Nous sistemes de caracterització

La pobre concordança tant en el sistema de Neer com en el sistema AO, ha portat més recentment a descriure sistemes que intenten entendre els diferents patrons de fractura i es tendeix a definir característiques que ens ajudin en la descripció i en la desició terapèutica.⁴⁶⁻⁴⁹

Així, Hertel⁴⁶ el 2001, proposa un sistema de descripció, basat en un qüestionari d'avaluació de la fractura, es què es respon, a unes preguntes simples, de l'existència de línies de fractura en els principals segments, segons les quals la combinació dels 5 plans de fractura bàsics resulten en 12 patrons de fractura. Els plans es localitzen entre tuberositat gran i cap, tuberositat gran i diàfisi, tuberositat petita i cap, tuberositat petita i diàfisi i entre tuberositat gran i tuberositat petita. Això resulta en 6 possibles fractures en 2 fragments, 5 possibles fractures en 3 fragments i 1 fractura en 4 fragments. *Figura 9*

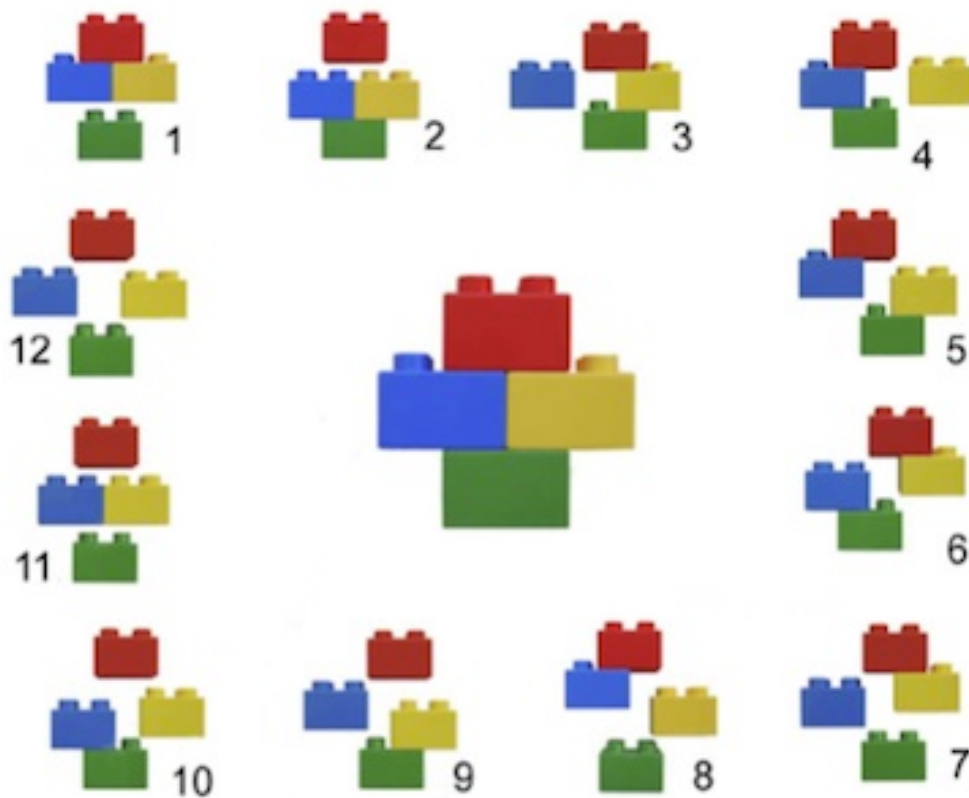


Figura 9. Sistema binari de Hertel. Hertel utilitzà el sistema de LEGO^R per representar les diferents combinacions de fractura. Cada una de les quatre peces, representa un segment de Codman. (Hertel R et al. Morphologic classification of fractures of the proximal humerus. A validated, teachable and practicable alternative. Presented at the 8th International Conference of Shoulder Surgery (ICSS) 2001)

El sistema valora també altres criteris addicionals:

1. La longitud de l'extensió metafisària posteromedial del fragment cefàlic, a més longitud més probabilitat de bona perfusió del cap.
2. La integritat cefalometafisària medial com a predictor d'isquèmia i capacitat de reducció. Es mesura el desplaçament de la diàfisi respecte al cap en l'aspecte més posteromedial
3. Comminució del fragment cefàlic amb un o dos plans intrarticulats.

Figura 10

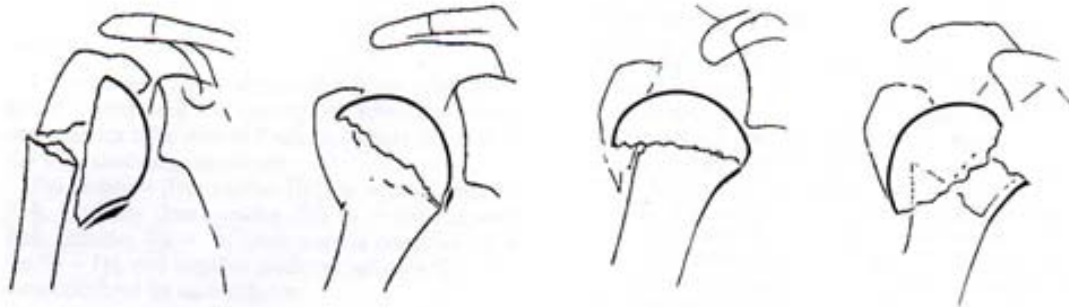


Figura 10. Criteris addicionals de Hertel. L'extensió metafisària posteromedial i la integritat del calcar medial seran factors de bon pronòstic pel risc d'isquèmia del fragment cefàlic.

El mateix autor en un estudi que valorava els factors associats a la mala perfusió del cap segons la seva classificació, determinava que la longitud de l'extensió metafisària dorsomedial (<8mm) , la disrupció del calcar medial i el tipus de fractura eren els factors predictors d'isquèmia més rellevants. Contràriament al que s'havia suggerit en estudis previs, el desplaçament dels fragments era menys important.⁵⁰ Aquesta classificació modificada, ha estat recentment validada obtenint millor concordança interobservador que les classificacions de Neer i AO.⁵¹

De la mateixa manera Mora^{47,52} proposa un protocol de lectura d'imatges en radiologia simple i TAC que utilitzarem per realitzar aquest treball en el qual l'observador ha de valorar diferents variables agrupades en: relació cefalodiafisària, relació tuberodiafisària, relació cefalotuberositària, relació glenohumeral i descripció dels fragments.

Taula 1

Grup	Característica
Relació cefalodiafisària	Impactació (si/no)
	Contacte (No, <20%; 20-50%; >50%)
	Desplaçament lateromedial
	Angulació en var/valg/normal
	Fractura metafisària associada
	Fulcre intern conservat
Relació tuberodiafisària	Desplaçament ≥ 10 mm de la tuberositat gran
	Desplaçament ≥ 10 mm de la tuberositat petita
Relació cefalotuberositària	Desplaçament ≥ 5 mm de la tuberositat gran
	Desplaçament ≥ 10 mm de la tuberositat petita
	Tuberositat gran: gran/petit ≥ 3 cm; ample/estret ≥ 1 cm; Comminut/íntegre
Cap humeral i relació cefaloglenoidea	Estoc ossi del cap (bo/adequat/deficient)
	Orientació del cap humeral (bo: 0-20°; suficient: 20-45°; deficient: >45°)
	Luxació del cap humeral
	Fractura per impactació de la superfície articular >20% o split
Descripció dels fragments	Nombre de fragments
	Fractura articular/extrarticular
	Fractura de tuberositat: no/gran/petita/ambdues

Taula 1. Protocol de lectura radiològica de Mora i cols. Es divideixen les variables segons les relacions entre els fragments, cap, diàfisi i tuberositats, relació amb la glena i descripció dels fragments.

2.5 Concordança i reproductibilitat en la classificació de Neer

Un capítol a part, ja que és objecte de part de l'estudi, mereix una anàlisi de la concordança i la reproductibilitat del sistema de classificació de Neer i de com estadísticament la valorem, ja que malgrat la descripció de nous sistemes de classificació i caracterització de les fractures, cap d'ells ha demostrat encara ser més fiable que el sistema de Neer, i segueix sent encara avui el més utilitzat.

2.5.1 Concordança interobservador

Kristiansen¹ el 1988 fou el primer en valorar la concordança de la classificació de Neer. En un treball de 100 radiografies simples que incloïen fractures del coll quirúrgic de l'húmer, utilitzant dues parelles d'observadors, obtenia uns valors de concordança, de 0,07 a 0,48, observant grau d'acord elevat quan les fractures no estaven desplaçades i en canvi dificultats en l'acord en les fractures desplaçades.

Els estudis de Sidor³ i Siebenrock⁷ de 1993, utilitzen tres projeccions radiològiques en lloc de dues i tenen en compte el grau d'experiència de l'observador. Cal destacar que Sidor³ és l'únic autor que té en compte les característiques de la mostra de fractures per valorar la concordança. Tal com veurem més endavant el coeficient kappa es troba influenciat per la prevalença de la característica estudiada en la mostra estudiada. En el treball de Sidor uns observadors independents valoraven la qualitat de l'estudi radiològic i que les diferents variables a estudiar estiguessin igual de representades en la mostra. Aquests autors, analitzen els resultats de 5 observadors en parelles, obtenint una concordança mitjana de 0,50 i 0,40 respectivament.

Sidor³, pensà que la complexitat de la classificació podia ser una factor influent en el grau d'acord, tot i així en simplificar la classificació de les

16 categories inicials (*figura 5*) a 6 categories: 1 part, 2 parts, 3 parts, 4 parts, fractura-luxació i fractura-articular, no s'obtingueren diferències en els valors d'acord. *Taula 2*

Sidor 1993	Classificació Neer	Classificació de Neer simplificada
Concordança interobservador	0.48	0.42
Reproductibilitat intraobservador	0.66	0.66

Taula 2. Valors kappa segons nombre de categories en la classificació de Neer en el treball de Sidor . No s'observen canvis en la concordança interobservador ni en la reproductibilitat intraobservador quan es simplifica la classificació de Neer de les 16 categories originals a 6 categories, una, dues, tres, quatre parts, fractura-luxació i fractura-articular.

El 1996, Bernstein i cols.⁶ valoren la classificació de Neer afegint la TAC en l'estudi radiològic. Suposadament l'acord entre observadors hauria de millorar si tenen informació en radiologia simple associada a TAC , ja que aquesta ha de facilitar valorar desplaçaments i línies de fractura. La concordança interobservador obtinguda fou de 0,52 quan els observadors disposaven només de la radiologia simple i de 0,50 quan disposaven de radiologia simple i TAC. Amb qualsevol dels dos estudis radiològics els graus d'acord no milloraven respecte els estudis previs de Sidor³ i Siebenrock⁷ amb radiologia simple. Igualment Sallay i cols.⁸

no observaren milloria en la concordança en afegir la TAC a la radiologia simple.

Bernstein⁶, pensà que potser la dificultat de la classificació de Neer es trobava en decidir quan un fragment es trobava el suficientment desplaçat o angulat com per ser considerat part, per això demanà als observadors de definir fragments independentment del desplaçament, la concordança obtinguda fou de 0,56.

Únicament, el grau d'experiència de l'observador en el treball de Bernstein⁶ sembla millorar la concordança, 0,64 en observadors experts.

Dos anys després Sjödén⁴, valorà també el grau d'acord amb radiologia simple associada a la TAC en tres dimensions, obtenint els mateixos resultats que Bernstein⁶. El coeficient Kappa per la classificació de Neer amb 6 categories fou de 0,49 en una mostra de 20 fractures valorada per 7 observadors, i de 0,32 en la classificació AO/ASIF.

Ja en 2009, Brorson i cols.⁵³, valoren el grau de concordança interobservador en fractures desplaçades en 4 parts segons la classificació de Neer, en un metanàlisi de 5 estudis observacionals, els graus d'acord oscil·laren de 0,16 a 0,48, sense diferències en funció de l'estudi radiològic utilitzat (radiologia simple, radiologia simple amb TAC 2 dimensions, radiologia simple amb TAC 3 dimensions) En general el grau d'acord era inferior en les fractures en 4 parts que en la resta de la classificació de Neer. Quan la població era seleccionada per un observador independent en lloc de ser una mostra consecutiva la concordança era millor. El mateix autor demostrà la millora en la concordança després de l'entrenament en la classificació de Neer, de 0,27 a 0,62, independentment de l'experiència dels observadors. En separar aquests, els no especialistes milloraven de 0,26 a 0,51 i els especialistes de 0,30 a 0,79.^{10,54}

L'estudi de Brunner¹² demostrava una milloria en el coeficient kappa de moderada a bona, segons criteris de Landis i Koch⁴⁴, tant en la classificació de Neer com en la classificació AO , utilitzant un sistema d'estereo-visualització d'imatges en tres dimensions traduïdes de dades de TAC en dues dimensions. Tot i així estudis més recents comparant la TAC-2D amb la TAC-3D no observaven superioritat en el grau d'acord de la TAC-3D, obtenint un valor kappa de 0,18 en 2D i de 0,17 en 3D per la classificació AO. Els observadors estaven més d'acord amb el desplaçament de la tuberositat gran amb la TAC en dues dimensions que en tres.⁵⁵ En un estudi recent comparant rx, TAC i TAC-3D s'observava un millor acord amb la TAC però només entre els observadors amb més experiència.⁵⁶

Ramappa⁵⁷ en un estudi recent conclou que la TAC si pot millorar l'acord en la caracterització de fractures però no en millora ni la classificació ni les recomanacions de tractament.

Mahadeva i cols¹³ al 2011 valorant el sistema de radiologia digital (PACS) obtenien valors de concordança interobservador bons, 0,6-0,8 segons els criteris de Landis i Koch⁴⁴, entre parelles d'observadors amb diferent grau d'experiència.

2.5.2 Reproductibilitat intraobservador

Considerem reproductibilitat quan el mateix observador dóna el mateix resultat en repetir la mateixa observació, és a dir està d'acord amb si mateix. Es considera que cal deixar un temps entre les diferents observacions i durant aquest temps els observadors no poden tenir accés a les dades. Idealment els observadors tampoc han de saber que seran revalorats una altre vegada. Els casos es presenten en ordre invers. Un sistema té fiabilitat si, a part de la concordança entre

observadors, té una bona reproductibilitat intraobservador, ja que vol dir que arribem al mateix resultat repetint les observacions.

La reproductibilitat intraobservador en estudis previs els sistemes de Neer i de AO per la classificació de fractures de l'extrem proximal de l'húmer es representa en la taula 3.

Kappa intraobservador	Sistema Neer	Sistema AO
Sidor	0,66	-
Siebenrock	0,60	0,58
Bernstein	0,72	-
Sjödén	0,56	0,52
Brunner	0,58	0,74
Marhadeva	0,83	-

Taula 3. Valors de Kappa intraobservador en sistemes de classificació de Neer i AO. S'expressen treballs que valoren les dues classificacions i altres que valoren només la classificació de Neer.

2.5.3 El coeficient Kappa

El coeficient Kappa és una mesura quantitativa del grau d'acord que hi ha entre uns observadors. La variació interobservador pot realitzar-se en qualsevol situació en que dos observadors estiguin avaluant la mateixa cosa.

El càlcul es basa en la diferència entre quant acord hi ha entre aquests observadors, el que anomenem acord observat (P_o), comparat amb quant acord hauríem esperat per atzar, el que anomenem acord esperat (P_e).⁵⁸ *Taula 4*

Observador B	Observador A		Total
	Si	No	
Si	a	b	M1
No	c	d	Mo
Total	N1	No	n

Taula 4. Taula de contingència dos observadors valorant una variable dicotòmica.

Acord observat (P_o)= $a+d/n$

Acord esperat (P_e)= $(N1/n) * (M1/n) + (N0/n)*(M0/n)$

Kappa (K)= $(p_o-p_e) / (1-p_e)$

Els valors de Kappa oscil·len de -1 a 1. El valor 1 representa un acord perfecte, 0 és l'acord esperat per atzar i els valors negatius indiquen menys acord de l'esperat per atzar.

L'estadístic més utilitzat és el Kappa de Cohen⁵⁹ limitat a dos observadors, tot i que hi ha altres fórmules com el Kappa de Fleiss⁶⁰, aplicable amb múltiples observadors. Fleiss⁶⁰ i Cohen⁵⁹ no tracten de la mateixa manera el factor atzar. En una determinada observació, segons la discordança en les respostes, amb la fórmula de Cohen podem obtenir un acord millor que l'atzar, mentre que amb Fleiss el valor pot ser inferior a zero, és a dir un acord pitjor que el que s'obtidria per atzar.

Tot i així, en l'estudi de Komagata⁶¹, es demostra que el Kappa de Cohen és perfectament aplicable a múltiples observadors. Segons les discordances entre els observadors, les dues fórmules poden donar-nos resultats diferents en la mateixa mostra, inferiors si utilitzen el Kappa de Fleiss.

Alguns estudis que valoren concordança en classificació de Neer utilitzen coeficients diferents al Kappa de Cohen, com Siebenrock⁷ que utilitza el Kappa de Fleiss i Bernstein⁶ el Kappa de Dunn⁶². Els resultats utilitzant diferents fórmules no poden ser comparables.

La interpretació del Kappa, es categoritza generalment segons els criteris de Landis i Koch⁴⁴, tot i que existeixen en la literatura altres tipus de categoritzacions. *Taula 5*

Siebenrock i cols⁷. categoritzen el valor de Kappa segons Svanholm i cols.⁶³, segons els quals un valor major de 0,75 es considera bo, de 0,5 a 0,75 baix i menor de 0,5 pobre.

Kappa	Acord
<0	Menys acord que per atzar
0,01-0,20	Lleu
0,21-0,40	Pobre
0,41-0,60	Moderat
0,61-0,80	Substancial
0,81-1	Excel·lent o quasi perfecte

Taula 5. Criteris de categorització del Kappa de Landis i Koch

El coeficient no té en compte el grau de desacord entre els observadors i tracta qualsevol desacord com a desacord total. És a dir, donarà la mateixa importància a desacords petits o grans. Així, per exemple, quan tenim una variable ordenada per categories, on el valor menor indica menys gravetat i el major gravetat màxima, és preferible utilitzar el Kappa ponderat per categories de manera que diferents nivells de desacord contribueixen diferent en el valor de Kappa.^{58,59}

2.5.3.1 Les paradoxes del Kappa

L'estadístic kappa es troba afectat per dues paradoxes:

El kappa es troba afectat per la prevalença de la variable a estudiar a la mostra. Així si per exemple estem observant la presència o absència d'una fractura-luxació d'espatlla, en una mostra de pacients on aquest tipus de lesió és molt poc prevalent, podem tenir un acord d'absència d'aquesta lesió entre els observadors del 90% i en canvi el valor del Kappa resultar proper a 0. Per variables molt o poc prevalents en una determinada mostra, valors de Kappa molt baixos no necessàriament reflecteixen baix percentatge d'acord.

És important conèixer aquesta limitació a l'hora d'analitzar treballs on s'expressen graus d'acord entre observadors i no seria apropiat comparar valors de kappa entre estudis amb diferents poblacions.⁵⁸

Exemple:

Imaginem dues mostres diferents, A i B, de 100 pacients amb diferents prevalences de la variable a estudi. En A les dues opcions de la variable estan distribuïdes a parts iguals, mentre que en B, la prevalença d'una de les dues opcions és molt elevada mentre que la de l'altre és molt baixa. Amb els mateix percentatge d'acord observat entre els observadors, els valors del coeficient kappa seran molt diferents.⁶⁴ *Taules 6 i 7*

Observador B	Observador A		Total
	Si	No	
Si	40	10	50
No	10	40	50
Total	50	50	100

Taula 6. Mostra A: Resultats de dos observadors valorant una variable dicotòmica. L'acord observat és 0,80, els dos observadors estant d'acord en el 80% de les observacions (Po). Amb una prevalença d'un diagnòstic positiu de 0,5, és a dir 50% de si i 50% de no, el resultat de Kappa en aplicar la fórmula seria de 0,60

Observador B	Observador A		Total
	Si	No	
Si	2	10	12
No	10	78	88
Total	12	88	100

Taula 7. Mostra B: Resultats de dos observadors valorant una variable dicotòmica. L'acord observat és 0,80, estant d'acord en el 80% de les observacions (Po). Amb una prevalença d'un diagnòstic positiu de 0,12, és a dir 12% de si i per tant 88% de no, el resultat de Kappa en aplicar la fórmula seria de 0,05

El grau d'acord entre els dos observadors observat en ambdues mostres és de 0,80, el que anomenem acord observat (Po).

A: $0,40 + 0,40 = 0,80$; B: $0,02 + 0,78 = 0,80$

En la mostra A, la prevalença d'un diagnòstic positiu en aquesta població és de 0,5 ($(40 + [10 + 10] / 2) / 100$) i el valor de Kappa quan apliquem la fórmula serà de 0,60

En la mostra B, la prevalença d'un diagnòstic positiu és molt inferior ($(2 + [10 + 10] / 2) / 100 = 0,12$) i el valor de Kappa quan apliquem la fórmula serà de 0,05

Els observadors han tingut el mateix percentatge d'acord i en canvi els valors de kappa són molt diferents. Si només valorem el valor del Kappa sense tenir en compte el biaix de prevalença podem pensar que la concordança interobservador en la mostra B és molt dolenta.

Això succeirà tant si la variable és molt prevalent com si ho és molt poc. Així doncs amb el mateix percentatge d'acord tindrem valors de Kappa molt diferents en funció de la prevalença. *Figura 11*

?

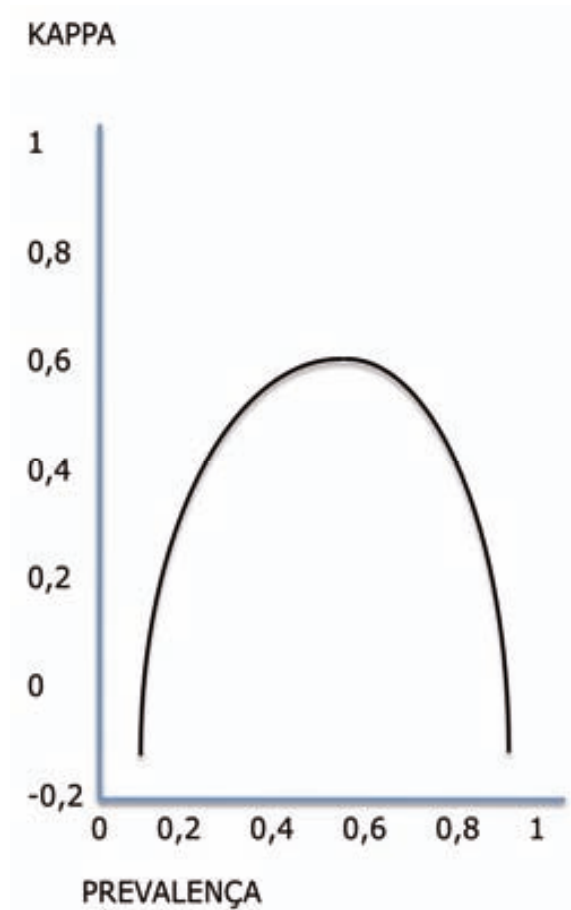


Figura 11. Dependència del coeficient Kappa de la prevalença del diagnòstic. La corba mostra els valors de Kappa amb un acord observat de 0,8, amb diferents valors de prevalença. Quan tenim el 50% de prevalença del diagnòstic és quan obtenim el major valor de kappa.

Aquest efecte, no depèn del nombre de casos de la mostra. Si les mostres haguessin sigut de 10.000 pacients en lloc de 100, amb la mateixa distribució en les taules de contingència, els resultats hagueren estat els mateixos.

?

on

La dependència del Kappa de la prevalença no ha estat prèviament estudiada, comparant valors obtinguts de mostres amb prevalències del diagnòstic diferents.

Alhora, la fiabilitat d'un test s'expressa pels valors predictius i aquests depenen també de la prevalença del diagnòstic. La dependència del kappa de la prevalença implica que no podem basar únicament la reproductibilitat d'un diagnòstic només amb aquest valor. Els resultats d'un estudi interobservador no s'haurien de presentar únicament amb els valors de Kappa sinó amb els resultats originals en les taules de contingència perquè el lector es faci una idea de la influència de la prevalença.

Feinstein i Cicchetti^{65,66} i Byrt i cols.⁶⁷ han publicat treballs on s'analitzen les paradoxes del kappa i es proposen fórmules per corregir-les.

El valor de kappa, a part d'estar influït per una molt baixa o molt alta prevalença de la variable a estudi, que donarà un desbalanç en els totals a la taula de contingència, també dependrà de com aquest desbalanç es distribueixi. Si el desbalanç és perfectament simètric ens donarà un valor de Kappa més baix que si el desbalanç és més imperfecte. *Taules 8 i 9*

Observador B	Observador A		Total
	Si	No	
Si	85	5	90
No	5	5	10
Total	90	10	100

Taula 8. Disbalanç simètric perfecte. En aquest cas, la prevalença dels diagnòstics positius és molt superior als negatius però el disbalanç és perfectament simètric. L'acord observat (Po) és de 0,90, 90% de coincidència en les respostes, i en valor Kappa (K) 0,44

Observador B	Observador A		Total
	Si	No	
Si	70	10	80
No	0	20	20
Total	70	30	100

Taula 9. Disbalanç imperfecte. Si les respostes, no estan balancejades, igual que en la Taula 8, però a més el disbalanç és imperfecte, amb el mateix acord observat (Po 0,90) el valor de Kappa serà superior (K 0,74)

Un altre factor a tenir en compte és el biaix de l'observador. Assumim que a priori cada observador té una probabilitat relativament fixa de donar una resposta positiva o negativa. Teòricament respondran el que observen en cada cas, però això sol estar influenciat per l'expectativa que tenen aquests d'observar, si coneixen quin tipus de població estudien, amb una distribució normal o anormal, o si la decisió que prenen té unes implicacions o unes altres. L'instrument de mesura no és un procediment tecnològic inanimat sinó un humà amb respostes subjectives que a vegades pot estar d'acord només per atzar. La situació ideal és que els observadors siguin cecs per la població a estudi i que no facin deduccions a mesura que van fent l'observació.

L'expressió del Kappa com a coeficient simple fa la seva interpretació i comparació difícil. S'ha de descompondre el resultat de manera que reflecteixi, el grau d'acord real observat, les influències de la prevalença i el biaix.

Una manera per corregir aquestes paradoxes i evitar la influència en el kappa del desbalanç de la mostra seria obtenir una població a estudi ben balancejada tant per les respostes positives com per les negatives.

Així en les fractures de l'extrem proximal de l'húmer els estudis epidemiològics indiquen una elevada freqüència de fractures en una part, quasi el 50%, menor proporció de fractures en 2 i 3 parts i un incidència de fractures en 4 parts menor al 5%.^{18,68} Els estudis de concordança publicats fins al moment contenen un petit nombre de casos, en mostres recollides de forma consecutiva de manera que alguns autors atribueixen al biaix de prevalença els baixos valors del coeficient Kappa. Únicament dos autors tenen en compte aquest factor en la selecció de la mostra, però sense especificar com s'ha fet exactament la selecció dels casos.^{3,55} *Taula 10*

Autors	Nombre de casos	Selecció de la mostra
Kristiansen¹	100	"Quasi" consecutiva
Sidor³	50	Consecutiva
Siebenrock⁷	95	Consecutiva
Brien⁹	28	Segons qualitat de les imatges
Bernstein⁶	20	Segons qualitat imatges de 46 casos consecutius
Sjöden⁴	26	Consecutius
Sjödén⁵	24	Consecutius
Sanchez-Sotelo⁷⁰	113	Consecutius
Sallay⁸	12	Consecutius
Brorson⁵⁴	42	Consecutius
Sukthankar⁵¹	47	Consecutius
Brunner¹²	40	Consecutius
Marhadeva¹³	50	Consecutius randomitzat dies alterns
Bruinsma⁵⁵	15	Segons patró de fractura

Taula 10. Publicacions que valoren concordança interobservador en classificació de fractures de l'extrem proximal de l'húmer. S'expressen nombre de casos i com s'ha seleccionat la mostra a estudi. Únicament Sidor, tot i tenir 50 casos consecutius, valora que els diferents tipus de fractura tinguin la mateixa representació i Bruinsma selecciona 15 casos segons el patró de fractura per tenir representació de tots els tipus.

3.HIPÒTESI I OBJECTIUS

3.1 HIPÒTESI DE TREBALL

Hipòtesi de treball: El biaix de prevalença del coeficient Kappa de Cohen, influeix en la baixa concordança interobservador i reproductibilitat intraobservador observada en la classificació de Neer de fractures de l'extrem proximal de l'húmer.

Hipòtesi nul·la: El biaix de prevalença del coeficient Kappa de Cohen, no influeix en la baixa concordança interobservador i reproductibilitat intraobservador observada en la classificació de Neer de fractures de l'extrem proximal de l'húmer.

3.2 OBJECTIUS SECUNDARIS

1. Analitzar la concordança interobservador i reproductibilitat intraobservador en valorar diferents característiques radiològiques de les fractures de l'húmer proximal.
2. Avaluar la influència de la informació aportada per la TAC en la caracterització i classificació de fractures de l'húmer proximal.
3. Avaluar la influència de l'experiència de l'observador en la caracterització i classificació de fractures de l'húmer proximal.

4.MATERIAL I MÈTODE

Per tal de demostrar la hipòtesi de treball i els objectius secundaris, el treball s'estructura en dues parts. En la primera part es selecciona una mostra de casos consecutius i es valoren els acords interobservador i intraobservador de diferents característiques radiològiques i de la classificació de Neer. Aquesta part ens serveix per valorar els objectius secundaris i l'acord de la classificació de Neer en una mostra consecutiva de la població. En la segona part, es valora l'acord interobservador i intraobservador de la classificació de Neer en una mostra seleccionada de fractures balancejada pels quatre tipus principals de fractura i per tant ens permetrà valorar, comparant amb la mostra consecutiva, els efectes del biaix de prevalença.

1ª PART

4.1 Estudi de la concordança interobservador i la reproductibilitat intraobservador en la caracterització de fractures de l'húmer proximal. Influència de la seqüència de la TAC (axial i completa) i de l'experiència de l'observador (experts i no experts).

En la primera part de l'estudi es valorarà quines variables es poden caracteritzar amb radiologia simple i TAC, i quina informació aporta el tall de TAC axial en comparació amb la TAC completa. Es vol valorar si la informació aportada per totes les seqüències de la TAC, ajuda als observadors a millorar el seu grau d'acord o si al contrari tenir més

informació els genera més desacord. Es valora també com influeix l'experiència de l'observador en els graus d'acord.

4.1.1 Selecció de la mostra

Hem seleccionat de forma prospectiva 153 fractures de l'extrem proximal de l'húmer en tractament ortopèdic recollides de forma consecutiva en un període de sis mesos, en quatre centres hospitalaris. Tots els casos disposaven d'estudi radiològic simple en dues projeccions, anteroposterior i perfil en "Y" d'escàpula, realitzat en el moment de la fractura. En tractar-se d'un estudi prospectiu multicèntric els responsables de cada centre s'encarregaren d'obtenir les projeccions correctes i estandarditzades. Durant les tres setmanes posteriors a la fractura es realitzava una TAC en dues dimensions amb talls axial, sagital i coronal. Les imatges de cada un dels casos en guardaven en carpetes amb l'estudi radiològic simple i cada un dels talls de TAC.

4.1.2 Variables a estudi

En cada un d'aquests casos s'estudien les següents característiques radiològiques segons el protocol de Mora i cols. *Taula 1* S'han introduït algunes variacions en les variables respecte l'estudi original. *Taula 11*

Dividim les variables en funció de les relacions entre el cap humeral, les tuberositats, la diàfisi i la glena: relació cefalodiafisària, relació tuberodiafisària, relació cefalotuberositària i relació cefaloglenoidea. Aquest últim grup inclou variables de caracterització del cap humeral. Es valora també el nombre de fragments i la classificació de Neer simplificada en quatre categories.

Cada una de les característiques radiològiques es detallen a continuació.

RELACIÓ CEFALODIAFISÀRIA:

1. Contacte: Es valora i quantifica el contacte entre el cap humeral i la diàfisi, independentment de si existeixen fractures en les tuberositats. Cal determinar si no hi ha contacte, hi ha contacte però menor al 33% de la circumferència de la diàfisi, contacte entre el 33-66% o més del 66% de contacte. Els percentatges foren modificats respecte a l'estudi original. Taules 1 i 11 . *Figura 12*

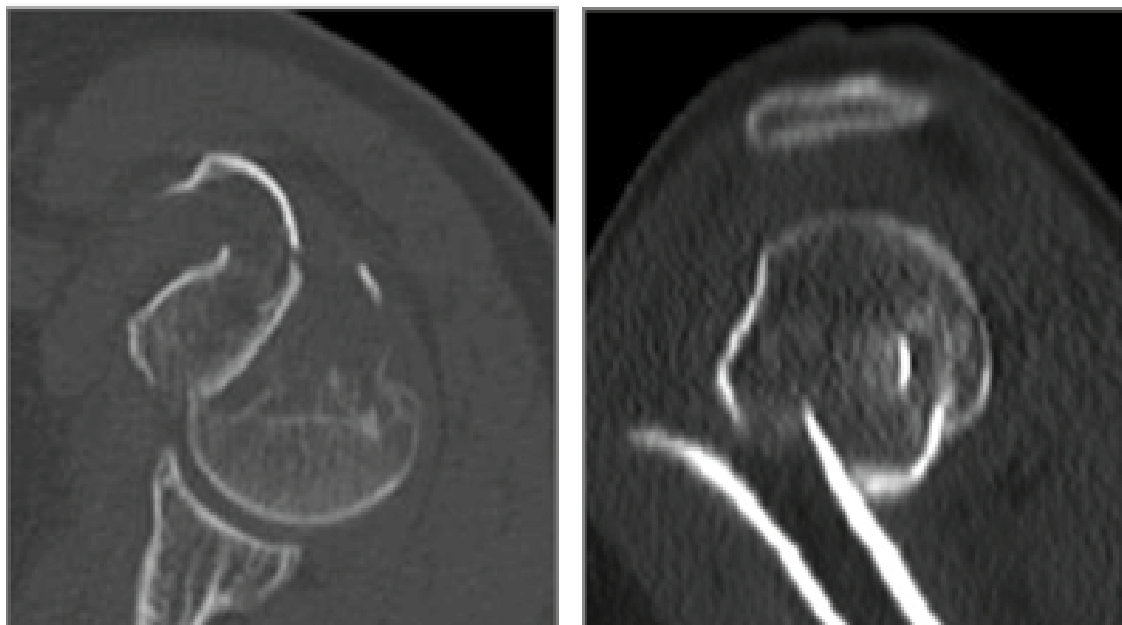


Figura 12. Contacte entre cap diàfisi menor del 33%

2. Angle cefalodiafisari: Es calcula l'angle entre el cap i la diàfisi traçant una línia en el coll anatòmic i calculant l'angle entre la perpendicular a aquesta línia i una línia paral·lela a la diàfisi. Es considera que hi ha un

desplaçament angular quan hi ha $>30^\circ$ en var o en valg comparat amb l'angle cefalodiafisari mig normal de 130° . Es defineix angle normal (entre 100° i 145°), valg (major de 145°) i var (menor de 100°).

Figura 13A

?

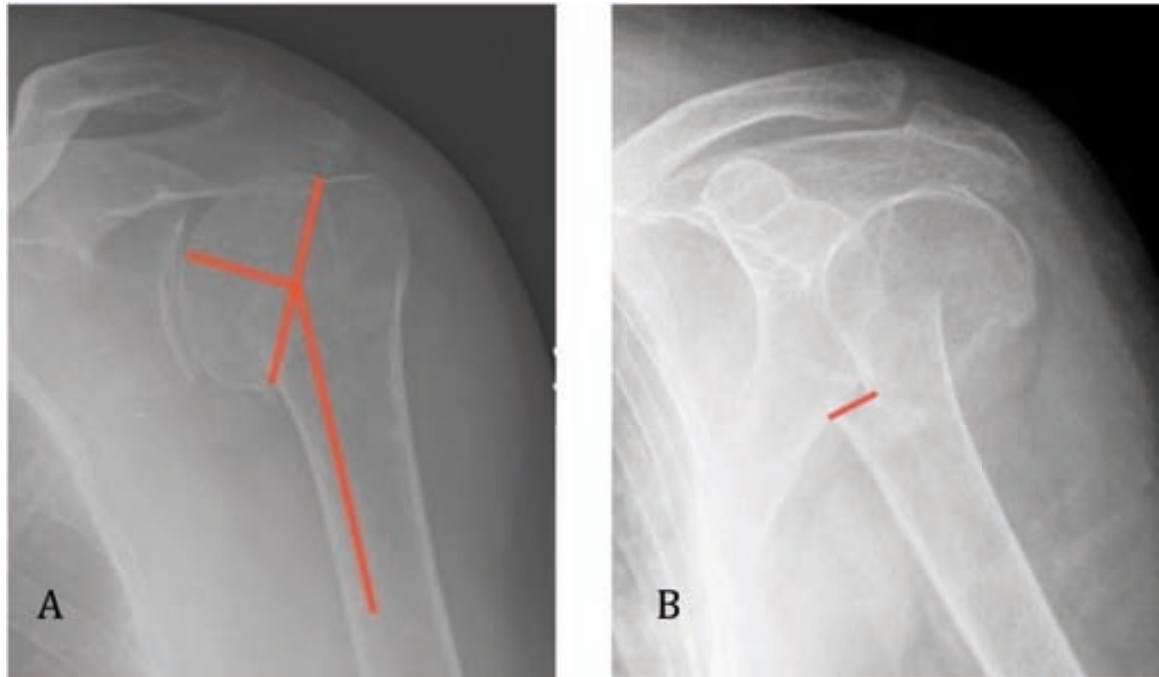


Figura 13. A. Angle cefalodiafisari B. Desplaçament lateromedial

3. Desplaçament lateromedial de la diàfisi respecte al cap $>1\text{cm}$. Es defineix entre si o no. *Figura 13B*

4. Desplaçament anteroposterior de la diàfisi respecte al cap $>1\text{cm}$. Es defineix si o no.

5. Fractura metafisària associada: Presència de fractura en la metàfisi en un fragment o comminuta $>1\text{cm}$. Es defineix si o no. *Figura 14A*

6. Fulcre intern conservat: Es considera el fulcre intern preservat quan existeix continuïtat entre el còrtex intern del cap humeral i la diàfisi. Es defineix si o no. *Figura 14B*

?

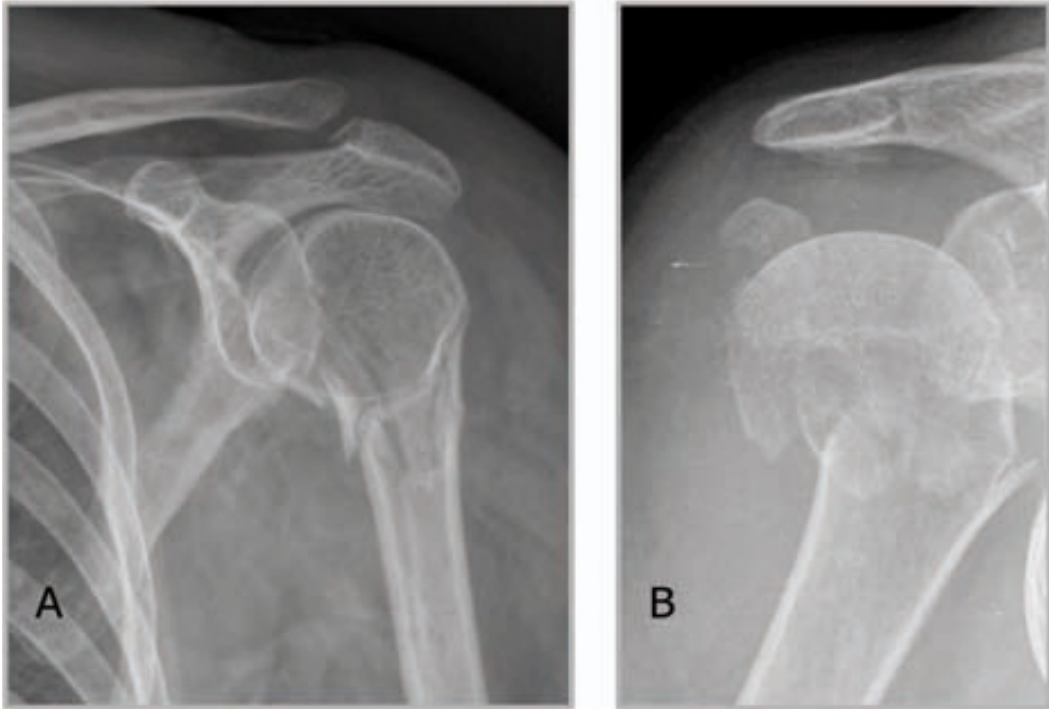


Figura 14. A. Fractura metafisària associada. B. Fulcre intern no preservat

7. Fractura impactació: Es considera fractura impactada quan el 50% o més de la diàfisi humeral es manté en contacte amb el cap i ha penetrat en l'os esponjós d'aquest. Es defineix si o no. *Figura 15 A i B*

?

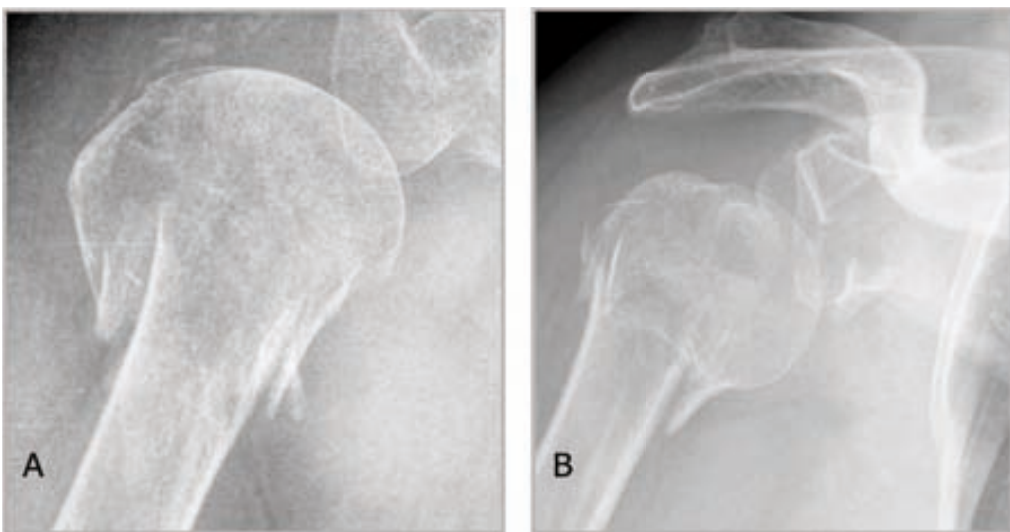


Figura 15. A i B. Impactació diàfisi-cap

?

RELACIÓ TUBERODIAFISÀRIA:

1. Desplaçament $\geq 1\text{cm}$ de la tuberositat gran com un fragment independent respecte la diàfisi. Es defineix si o no. *Figura 16*
2. Desplaçament $\geq 1\text{cm}$ de la tuberositat petita com un fragment independent respecte a diàfisi. Es defineix si o no



Figura 16. Desplaçament diàfisi-tuberositat gran

RELACIÓ CEFALOTUBEROSITÀRIA:

1. Desplaçament \geq de 0,5 cm entre la tuberositat gran i el cap. Es defineix si o no. *Figura 17*
2. Desplaçament $\geq 1\text{cm}$ entre la tuberositat petita i el cap. Es defineix si o no
3. Integritat o cominució de la tuberositat gran. Es considera cominuta quan hi ha tres o més fragments

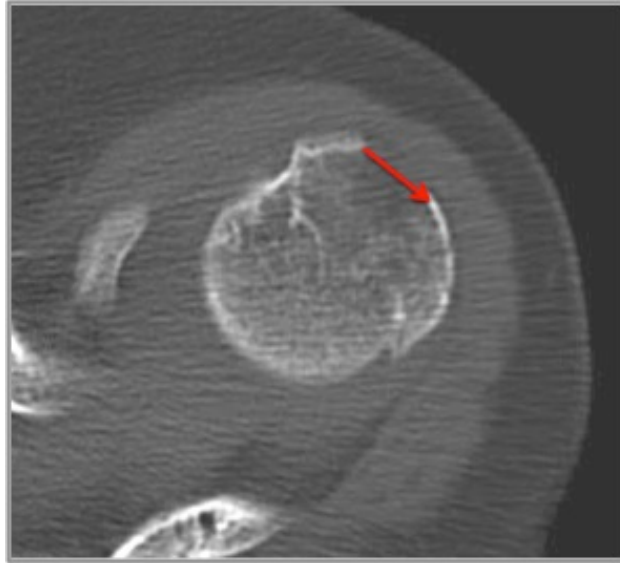


Figura 17. Desplaçament de la tuberositat gran respecte el cap major de 0,5 cm

RELACIÓ CEFALOGLENOIDEA:

1. Orientació de la superfície articular: Es defineix com la relació entre la superfície articular del cap humeral i la glena sense luxació. Es calcula l'angle cefaloglenoideu, en el tall axial de TAC, traçant una línia en la superfície articular i una línia en la superfície glenoidea i calculant l'angle entre les perpendiculars a aquestes dues. Es considera suficient orientació cefaloglenoidea, quan l'àpex del cap està orientat a la glena o si no ho està l'angle és $\leq 45^\circ$. Es considera una orientació insuficient quan l'àpex del cap no té contacte amb la glena i l'angle és $>45^\circ$. Es considera aquesta relació independentment de la relació del cap amb la tuberositat gran. *Figura 18 A i B*

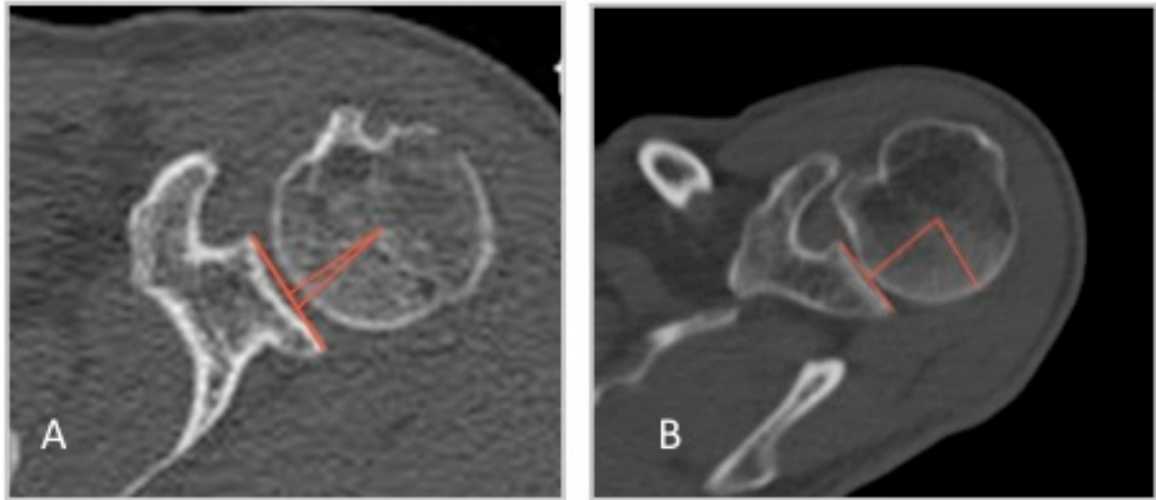


Figura 18. A.Orientació cefaloglenoidea suficient. B.Orientació cefaloglenoidea insuficient

2. Estoc ossi del cap: Es defineix suficient quan l'os esponjós està present en la meitat o més del semicercle del cap. Es considera insuficient quan és inferior al semicercle del cap. *Figura 19 A i B*

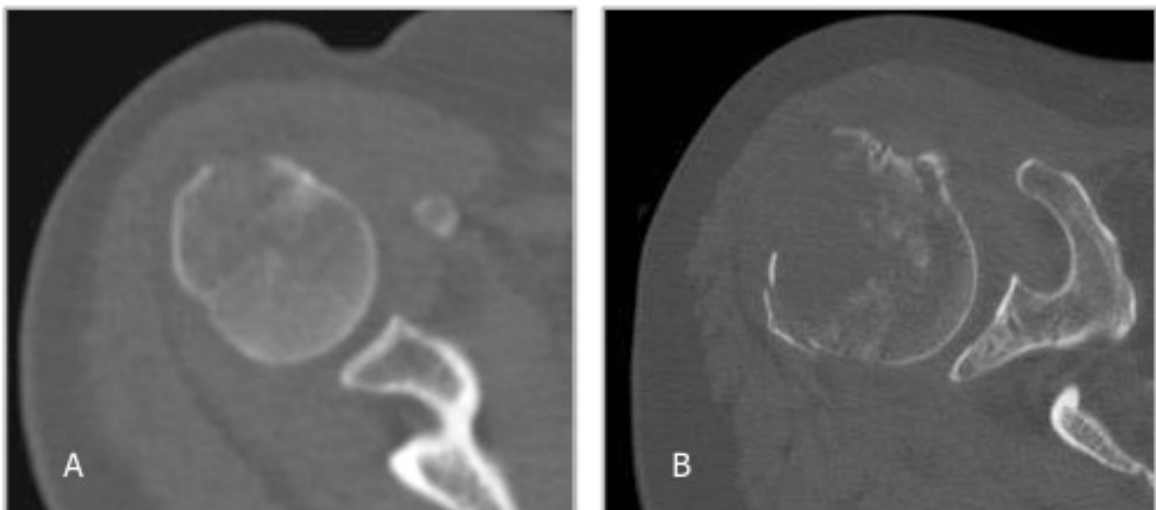


Figura 19A. Estoc ossi del cap suficient. B. Estoc ossi del cap insuficient

3. Luxació del cap: Pèrdua de contacte entre el cap humeral i la glena en el pla frontal o interposició de la diàfisi entre el cap i la glena. Es defineix si o no. En cas de si, es defineix la direcció: anterior, posterior o lateral.*Figura 20*

-Anterior: Quan el cap humeral és anterior al pla frontal de la glena

-Posterior: Quan el cap humeral és posterior al pla frontal de la glena

-Lateral: Quan l'àpex del cap es troba oposat a la glena. El cap ha girat 180°, i la diàfisi es troba interposada entre el cap i la glena.

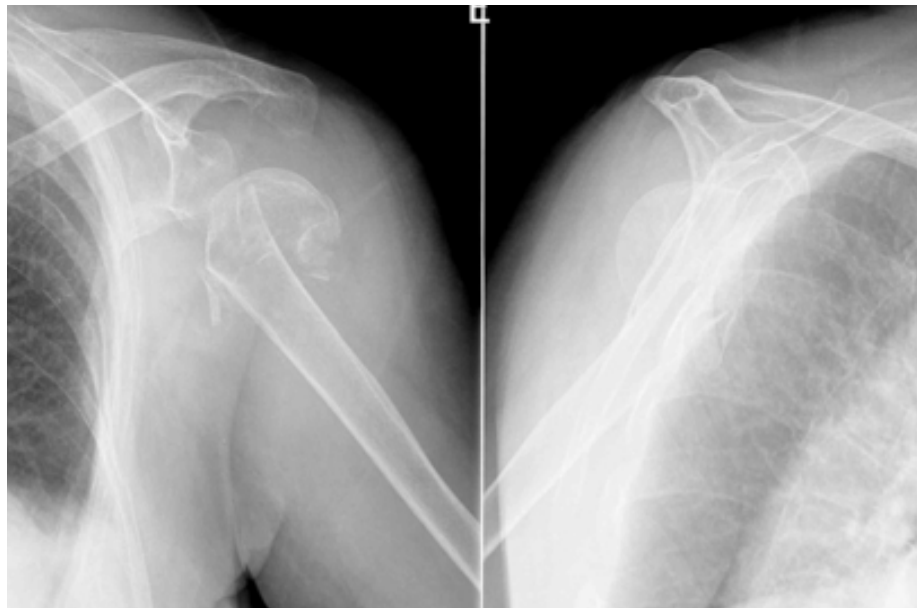


Figura 20. Luxació posterior del cap humeral

4. Fractura impactació de la superfície articular: Fractura descrita per Hill-Sachs secundària a luxació glenohumeral. Es defineix presència si o no.

5. Fractura tipus "split": Quan es produeix fractura de la superfície articular en fragments separats afectant un mínim del 20% de la superfície articular. Es defineix presència o absència d'aquest tipus de fractura.

NOMBRE DE FRAGMENTS:

Es defineix el nombre de fragments segons les quatre parts anatòmiques descrites per Codman, independentment del desplaçament. Es defineix si la fractura és intraarticular o extrarticular. Es considera intrarticular quan afecta al coll anatòmic.

CLASSIFICACIÓ DE NEER:

S'aplica la classificació de Neer simplificada en una, dues, tres o quatre parts, per minimitzar el nombre d'opcions de resposta. No es diferencia entre 4 parts en valg o en var. Si que es consideren les fractures en 4 parts luxació i en 4 parts impactació, ja que són variables que ja s'han valorat prèviament. Segons els criteris de Neer, es considera part quan un fragment està desplaçat més 1 cm o angulat més de 45°. *Figures 21-24*



Figura 21. Fractura en 1 part



Figura 22. Fractura en 2 parts del coll quirúrgic



Figura 23. Fractura en 3 parts de la tuberositat gran.



Figura 24. Fractura en 4 parts

A la taula 11 es resumeixen les diferents característiques radiològiques separades per grups. Veure les variacions amb la classificació original de Mora i cols. *Taula 1*

Grup	Característica
Relació cefalodiafisària	Contacte (No, <33%; 33-66%; >66%)
	Angulació en var/valg/normal
	Desplaçament lateromedial >1cm (si/no)
	Desplaçament anteroposterior >1cm (si/no)
	Fractura metafisària associada (si/no)
	Fulcre intern conservat (si/no)
	Impactació (si/ no)
Relació tuberodiafisària	Desplaçament \geq 10mm de la tuberositat gran (si/no)
	Desplaçament \geq 10mm de la tuberositat petita (si/no)
Relació cefalotuberositària	Desplaçament \geq 5mm de la tuberositat gran (si/no)
	Desplaçament \geq 10mm de la tuberositat petita (si/no)
	Comminució de la tuberositat gran (si/no)
Cap humeral i relació cefaloglenoidea	Orientació de la superfície articular (suficient/insuficient)
	Estoc ossi del cap (suficient/insuficient)
	Luxació del cap humeral (anterior/posterior/lateral)
	Fractura per impactació de la superfície articular (si/no)
	Fractura tipus split (si/no)
Descripció dels fragments	Nombre de fragments
	Fractura articular/extrarticular
Classificació de Neer	1,2,3,4 parts

Taula 11. Resum de les diferents característiques radiològiques

4.1.3 Selecció dels observadors

Es seleccionaren 6 observadors especialistes en cirurgia ortopèdica, subespecialitzats en cirurgia d'espatlla i per tant habituats a la lectura radiològica de fractures d'húmer proximal. Entre aquests, dos eren especialistes amb més de deu anys d'experiència en el tractament de fractures d'espatlla i quatre eren especialistes amb menys de deu anys d'experiència. Els observadors només coneixien que participaven en un estudi de lectura radiològica de fractures d'húmer proximal i en una reunió conjunta se'ls explicaren cada una de les variables a valorar. No coneixien els objectius del treball. Els sis observadors pertanyien a sis centres diferents, per tant no tenien contacte entre ells.

4.1.4 Observacions

Cada observador va rebre per correu electrònic una taula de Microsoft Excel^R on realitzar la recollida de dades, amb el número codificat de cada cas i cada una de les variables amb les opcions de resposta. En un servei d'allotjament d'arxius (dropbox^R) on els sis observadors tenien accés, es col·locaven els casos en carpetes amb les imatges necessàries per cada observació. Al final d'aquesta es retiraven i es tornaven a col·locar les carpetes quan s'iniciava la següent observació.

En una primera lectura els observadors disposaren de la taula de Microsoft Excel^R, les imatges de radiologia simple en dues projeccions i la seqüència axial de TAC de cada cas. Disposaven d'un mes per llegir, emplenar i enviar la taula. Un cop enviades les respostes les imatges es retiraven del centre d'emmagatzematge.

Després de dos mesos de descans, els observadors van rebre una nova taula de Microsoft Excel^R corresponent a la segona lectura, que contenia els mateixos casos en el mateix ordre i les imatges amb la radiologia simple i la TAC en seqüència completa, talls axial, sagital i

coronal. Disposaren d'un mes per llegir, emplenar i enviar la taula de Microsoft Excel^R.

En una tercera lectura, després de dos mesos més de descans, els observadors van rebre una nova taula , amb els mateixos casos però en ordre invers i les carpetes amb les imatges amb radiologia simple i seqüència de TAC axial. Disposaren també d'un mes per enviar les respostes.

Finalment i després de dos mesos, es realitzà una quarta lectura. Els observadors van rebre la taula de Microsoft Excel^R amb els casos en ordre invers i en el centre d'imatges les carpetes amb la radiologia simple i la seqüència completa de la TAC de cada cas. Disposaren també d'un mes per realitzar la lectura i enviar els resultats.

4.1.5 Anàlisi estadística

L'anàlisi estadístic es realitzà per SPSS versió 18.0 i programa R 3.0 (The R project for statistical computing) a l'Institut Municipal d'Investigació Mèdica de Barcelona (IMIM)

Es calcula el coeficient Kappa de Cohen no ponderat per parelles d'observadors i es realitza una mitjana de valors Kappa per cada variable. S'expressen intervals de confiança per cada resultat.

Es calcula la concordança interobservador i la reproductibilitat intraobservador utilitzant la segona lectura amb l'ordre invers dels casos.

Es separen els resultats entre els dos observadors sènior, considerats experts i els quatre observadors junior, considerats no experts

Es calculen les diferències entre experts i no experts i entre observació amb la TAC axial i la TAC completa.

Per calcular les diferències de concordança es realitza l'estadístic T Student per dades no aparellades. Es consideren diferències significatives valors de $p < 0,05$

2ª PART

4.2 Influència del biaix de prevalença en la concordança interobservador i reproductibilitat intraobservador de la classificació de Neer

Valorem en aquesta part la influència de la prevalença en els resultats de concordança interobservador i reproductibilitat intraobservador comparant els resultats en la classificació de Neer entre una mostra consecutiva de fractures i una mostra seleccionada, per tal de valorar si una mostra amb una distribució més balancejada de la variable a estudi corregeix el biaix de prevalença i millora el grau d'acord en els mateixos observadors.

4.2.1 Mostra consecutiva

4.2.1.1 Selecció de la mostra

S'utilitza en aquesta part la mostra de la primera part de l'estudi que consistia en 153 fractures recollides de forma consecutiva durant un període de 6 mesos en quatre centres hospitalaris. Tots els casos tenien estudi amb radiologia simple en dues projeccions,

anteroposterior i perfil en "Y" d'escàpula i TAC realitzada dins les primeres tres setmanes de la fractura.

4.2.1.2 Variables a estudi

La variable estudi fou la classificació de Neer simplificada en 4 categories , una dues, tres o quatre parts. Segons Neer es defineix part quan un fragment de fractura està desplaçat més 1 cm o angulat més de 45°

4.2.1.3 Selecció dels observadors

Els observadors foren els mateixos que en la primera part de l'estudi.

4.2.1.4 Observacions

Les observacions corresponen a la primera part de l'estudi, tal com s'ha explicat en l'apartat 4.1.4. Per aquesta part s'utilitzen els resultats en les observacions amb la seqüència de TAC completa.

4.2.1.5 Anàlisi estadística

L'anàlisi estadístic es realitza amb SPSS versió 18.0 i programa R 3.0 (The R project for statistical computing) a l'Institut Municipal d'Investigació Mèdica de Barcelona (IMIM)

Es calcula la concordança interobservador i la reproductibilitat intraobservador per les variables amb el coeficient Kappa de Cohen no ponderat per parelles de dos i expressant la mitjana i els intervals de confiança.

4.2.2 Mostra seleccionada

Es pretén valorar en aquesta part de l'estudi una mostra de fractures on cada una de les quatre opcions de resposta de la classificació de Neer hi estigui representada en la mateixa proporció. En una mostra consecutiva, tal com observem les fractures d'húmer proximal habitualment, cal esperar una elevada proporció de fractures en una i dues parts, menys de tres parts i un baix nombre de fractures en 4 parts.

4.2.2.1 Selecció de la mostra

Per tal de seleccionar una mostra amb aquestes característiques, tres investigadors independents, no observadors en l'estudi, seleccionaren dins una base de dades de fractures d'húmer proximal, els quatre tipus de fractures i crearen una mostra de 92 casos amb una representació equitativa de cada un dels 4 tipus. Els tres investigadors, especialistes també en la interpretació d'imatges de fractures d'húmer proximal, havien d'estar d'acord en cada un dels tipus de fractura. Els casos disposaven del mateix estudi radiològic que en la mostra consecutiva, amb radiologia simple en dues projeccions i TAC en seqüència completa.

4.2.2.2 Variables a estudi

La variable estudi fou, igual que en la mostra consecutiva, la classificació de Neer simplificada en 4 categories, una dues, tres o quatre parts. Segons Neer es defineix part quan un fragment de fractura està desplaçat més 1 cm o angulat més de 45°

4.2.2.3 Selecció dels observadors

Els observadors foren els mateixos que en la primera part de l'estudi, amb la mostra consecutiva, i desconeixien que estaven en aquest cas valorant una mostra seleccionada de fractures. Tal com s'ha comentat en la primera part, només coneixien que participaven en un estudi de lectura d'imatges radiològiques de fractures de l'extrem proximal de l'húmer.

4.2.2.4 Observacions

Les observacions es realitzaren, de la mateixa manera que en la primera part de l'estudi. En aquest cas només es realitzaren dues lectures amb radiologia simple i la seqüència completa de TAC. La segona, després de dos mesos de descans, amb l'ordre invers dels casos per valorar la reproductibilitat intraobservador.

4.2.2.5 Anàlisi estadística

L'anàlisi estadística es realitza amb SPSS versió 18.0 i programa R 3.0 (The R project for statistical computing) a l'Institut Municipal d'Investigació Mèdica de Barcelona (IMIM)

Es calcula la concordança interobservador i la reproductibilitat intraobservador amb el coeficient Kappa de Cohen no ponderat per parelles de dos i expressant la mitjana i els intervals de confiança.

4.2.3 Comparació entre mostra consecutiva i mostra seleccionada

Es comparen les diferències en coeficient Kappa entre la mostra consecutiva i la mostra seleccionada. S'aplica l'estadístic T Student per dades no aparellades. Es consideren diferències significatives valors de p inferiors a 0,05

Tots els pacients firmaren un consentiment informat segons participaven d'un estudi observador de fractures d'húmer proximal i consentien la realització de la TAC i l'ús anònim de les imatges.

5.RESULTATS

1ª PART

5.1 Estudi de la concordança interobservador i reproductibilitat intraobservador en la caracterització de fractures de l'húmer proximal. Influència de la seqüència de la TAC (axial i completa) i de l'experiència de l'observador (experts i no experts)

Es comenten els resultats per grups de variables: variables de relació cefalodiafisària, variables relacionades amb les tuberositats, que inclou variables de relació tuberodiafisària i cefalotuberositària, i variables de relació cefaloglenoidea i de característiques del cap humeral. Com a variables a part, nombre de fragments i classificació de Neer.

5.1.1 Concordança interobservador en TAC axial i TAC completa amb tots els observadors

A la taula 12 s'expressen els valors del coeficient Kappa no ponderat per cada una de les variables, entre tots els observadors, obtinguts utilitzant radiologia simple amb tall axial de TAC i utilitzant radiologia simple amb la seqüència completa de la TAC. El coeficient Kappa es realitza per parelles i després es fa una mitja de tots els resultats. S'expressen els intervals de confiança.

KAPPA INTEROBSERVADOR (INTERVAL DE CONFIANÇA)	AXIAL	COMPLET	P
Contacte	0,39 (0,10-0,67)	0,34 (0,07-0,60)	0,329
Desplaçament lateromedial >1cm	0,53 (0,31-0,75)	0,35 (0,07-0,63)	0,001
Desplaçament anteropost. >1cm	0,23 (-0,01-0,47)	0,32 (0,04-0,60)	0,075
Angulació cefalodiafisària	0,37 (0,20-0,53)	0,36 (0,15-0,57)	0,776
Fractura metafisària associada	0,28 (-0,05-0,63)	0,29(-0,09-0,68)	0,886
Fulcre intern conservat	0,46 (0,16-0,76)	0,41 (0,15-0,66)	0,340
Impactació	0,31 (-0,02-0,65)	0,33(-0,09-0,76)	0,784
Desplaçament diàfisi-tub. gran	0,34 (0,10-0,57)	0,39 (0,19-0,59)	0,224
Desplaçament diàfisi tub.petita	0,09 (-0,09-0,28)	0,10(-0,06-0,27)	0,774
Desplaçament cap-tub.gran	0,33 (0,01-0,64)	0,41(-0,02-0,85)	0,27
Desplaçament cap- tub. petita	0,20 (-0,008-0,41)	NA	NC
Comminució tuberositat gran	0,15 (-0,01-0,33)	0,14 (-0,03-0,32)	0,768
Estoc ossi del cap	0,16 (-0,11-0,43)	NA	NC
Split cap	0,13 (-0,12-0,40)	NA	NC
Fractura impactació articular	0,06 (-0,22-0,35)	0,16 (-0,2-0,59)	0,156
Orientació superfície articular	0,24 (-0,13-0,62)	0,17(-0,25-0,59)	0,356
Luxació superfície articular	0,15 (-0,17-0,48)	0,19 (-0,3-0,74)	0,641
Nombre de fragments	0,36 (0,22-0,50)	0,34 (0,07-0,60)	0,613
Neer	0,43 (0,17-0,68)	0,43 (0,15-0,71)	1

Taula 12. Resultats coeficient Kappa interobservador entre tots els observadors amb intervals de confiança del 95% (entre parentèsi). Rx+ TAC axial / rx + TAC completa. NA: no aplicable; kappa no ponderat. Valor de p: diferències entre TAC axial i TAC completa. NC no calculable.

En general el grau d'acord fou pobre en els dos estudis radiològics segons els criteris de Landis i Koch. Únicament en la variable desplaçament lateromedial s'observen diferències estadísticament significatives, amb millor acord entre els observadors amb la TAC axial.

En la TAC axial el menor valor obtingut fou de 0,06 en fractura impactació de la superfície articular i el valor major fou de 0,53 en desplaçament lateromedial major d'1 cm, corresponent a un grau d'acord moderat segons els criteris de Landis i Koch⁴⁴. Per grups de variables, les corresponents a la relació cefalodiafisària obtenien valors lleugerament superiors a la resta de grups, oscil·lant de 0,23 en el desplaçament anteroposterior a 0,53 en el desplaçament lateromedial en la TAC axial. En la TAC completa els graus d'acord foren similars excepte per al desplaçament lateromedial on s'observa un empitjorament en el valor de Kappa, passant d'un acord moderat en axial (0,53) a pobre en TAC completa (0,35) amb diferències estadísticament significatives ($p=0,001$). No s'observen diferències estadísticament significatives en la resta de variables d'aquest grup ($p>0,05$).

Les variables relacionades amb les tuberositats obtenien graus d'acord pobres, principalment les relacionades amb la tuberositat petita. En el desplaçament diàfisi-tuberositat petita el grau d'acord fou proper a zero, en els dos estudis radiològics. No s'observen diferències estadísticament significatives entre els dos estudis radiològics ($p>0,05$).

Les variables de relació cefaloglenoidea i de caracterització del cap humeral, obtenien també valors d'acord lleus, tant en la TAC axial com en l'estudi amb TAC completa. No s'observen diferències estadísticament significatives entre els dos estudis radiològics ($p>0,05$).

En nombre de fragments s'obtenia un valor de kappa de 0,36 en TAC axial i de 0,34 en la TAC completa sense diferències estadísticament significatives ($p=0,61$). En aquesta variable, s'observà un error d'interpretació d'un dels observadors en analitzar les dades, pel que aquest fou retirat de l'anàlisi.

En la classificació de Neer el grau d'acord fou de 0,43 en els dos estudis radiològics, sense diferències estadísticament significatives. (p=1)

Si comparem el grau d'acord entre nombre de fragments (0,34) i parts (0,43) en TAC completa, tal com han fet altres autors en la literatura, no observem diferències significatives entre aquestes dues variables (p= 0,317).

Algunes variables en TAC completa el valor fou no valorable (NV) per excessiva homogeneïtat en les respostes dels observadors.

5.1.2 Reproductibilitat intraobservador en TAC axial i TAC completa amb tots els observadors

KAPPA INTRA OBSERVADOR (INTERVAL DE CONFIANÇA)	AXIAL	COMPLET	P
Contacte	0,54 (0,3-0,78)	0,57 (0,47-0,67)	0,664
Desplaçament lateromedial >1cm	0,63 (0,53-0,73)	0,58 (0,34-0,82)	0,255
Desplaçament anteropost. >1cm	0,45 (0,29-0,61)	0,65 (0,45-0,85)	0,000
Angulació cefalodiafisària	0,56 (0,48-0,64)	0,63 (0,53-0,73)	0,002
Fractura metafisària associada	0,60 (0,56-0,76)	0,62 (0,48-0,76)	0,568
Fulcre intern conservat	0,58 (0,5-0,66)	0,68 (0,54-0,82)	0,001
Impactació	0,58 (0,08-1,8)	0,66 (0,36-0,96)	0,409
Desplaçament diàfisi-tub. gran	0,55 (0,17-0,93)	0,61 (0,53-0,69)	0,363
Desplaçament diàfisi tub.petita	0,33 (0,2-0,51)	0,36 (0,14-0,58)	0,824
Desplaçament cap-tub.gran	0,57 (0,51-0,63)	0,68 (0,54-0,82)	0,000
Desplaçament cap- tub. petita	0,30 (0,0-0,6)	0,32 (0,18-0,82)	0,835
Comminució tuberositat gran	0,32 (0,22-0,42)	0,43 (0,19-0,67)	0,019
Estoc ossi del cap	0,31 (0,11-0,41)	0,55 (0,51-0,59)	0,000
Split cap	0,36 (0,16-0,46)	0,35 (0,0-0,73)	0,895
Fractura impactació articular	0,31 (0,01-0,6)	0,47 (0,27-0,78)	0,017
Orientació superfície articular	0,50 (0,44-0,56)	0,59 (0,31-0,87)	0,075
Luxació superfície articular	0,63 (0,17-1,43)	0,56 (0,06-1,18)	0,681
Nombre de fragments	0,50 (0,54-0,58)	0,56 (0,48-0,64)	0,000
Neer	0,74 (0,4-1,08)	0,72 (0,56-0,88)	0,748

Taula 13. Resultats de reproductibilitat intraobservador en TAC axial i en TAC completa. S'observa una millora en l'acord en la TAC completa respecte a la TAC axial en totes les variables, excepte en el desplaçament lateromedial, split i luxació. Els valors foren moderats en la majoria de variables. En desplaçament anteroposterior, angulació, conservació del fulcre, desplaçament cap-tub.gran, comminució tuberositat gran, estoc, fractura impactació articular i nombre de fragments, les diferències foren estadísticament significatives.

Els valors de Kappa en la reproductibilitat intraobservador foren moderats en la majoria de variables.

En les variables de relació cefalodiafisària s'obtingueren valors de 0,45 a 0,63 en axial i de 0,57 a 0,68 en la TAC completa. S'observa una millora en l'acord en la TAC completa respecte a l'estudi axial en totes les variables, excepte en el desplaçament lateromedial. En les variables desplaçament anteroposterior ($p=0,000$), angulació cefalodiafisària ($p=0,002$) i conservació del fulcre intern ($p=0,001$) s'observen diferències estadísticament significatives.

En les variables relacionades amb les tuberositats, els pitjors resultats s'observaren en les variables relacionades amb la tuberositat petita, on els graus d'acord intraobservador foren pobres en els dos estudis radiològics. S'observa millor concordança intraobservador en TAC completa en el desplaçament de la tuberositat major respecte al cap amb diferències estadísticament significatives ($p = 0,000$)

En les variables de relació cefaloglenoidea i característiques del cap humeral, s'observa millor acord intraobservador en la TAC completa en l'estoc ossi del cap humeral ($p=0,000$) i en fractura impactació ($p=0,017$) amb diferències estadísticament significatives. El grau d'acord fou pobre en presència de fractura tipus split tant en TAC axial com en la TAC completa. La resta de variables milloren amb l'estudi amb TAC complet excepte luxació de la superfície articular on s'observa menor acord intraobservador, sense diferències estadísticament significatives ($p>0,05$).

En nombre de fragments el grau d'acord fou moderat en els dos estudis radiològics amb diferències estadísticament significatives a favor de l'estudi amb la TAC completa ($p=0,000$).

La concordança intraobservador fou substancial en la variable classificació de Neer, 0,74 en la TAC axial i 0,72 en la TAC completa, sense diferències estadísticament significatives ($p=0,74$).

5.1.3 Comparació de la concordança interobservador entre TAC axial i TAC completa segons grau d'experiència

En les taules 14 i 15 s'expressen els resultats del coeficient Kappa interobservador en els dos estudis radiològics quan dividim els observadors segons el seu grau d'experiència. Dos observadors amb més de deu anys d'experiència en patologia d'espatlla, considerats experts, i quatre observadors amb un temps inferior a deu anys, considerats no experts.

Entre els dos observadors experts, s'observa, empitjorament en el grau d'acord en la TAC completa respecte a la TAC axial, en algunes variables de relació cefalodiafisària, especialment contacte i desplaçament lateromedial amb diferències estadísticament significatives en aquestes dues variables (contacte $p=0,004$; desplaçament lateromedial ($p=0,002$).

Obtenen graus d'acord molt lleus i en alguns casos no valorables, en les variables relacionades amb la tuberositat petita i en les variables de relació cefaloglenoidea. En les variables de desplaçament de la tuberositat gran l'acord fou superior en la TAC completa però sense diferències estadísticament significatives respecte a la TAC axial (desplaçament tuberositat gran-diàfisi $p= 0,18$; desplaçament tuberositat gran-cap $p=0,06$).

L'acord en la fractura impactació de la superfície articular fou significativament millor amb la TAC completa ($p=0,04$).

La variable nombre de fragments no es va poder calcular en aquest grup, ja que un dels observadors s'havia retirat de l'anàlisi per un error de lectura.

En la classificació de Neer, els dos experts obtenien menys acord en l'estudi amb la TAC completa ($k=0,43$), respecte a la TAC axial ($k=0,56$) sense diferències estadísticament significatives ($p= 0,06$).

KAPPA INTEROBSERVADOR EXPERTS (INTERVAL DE CONFIANÇA)	AXIAL	COMPLET	P
Contacte	0,60 (0,47-0,72)	0,37 (0,23-0,51)	0,004
Desplaçament lateromedial >1cm	0,67 (0,53-0,81)	0,37 (0,19-0,55)	0,002
Desplaçament anteropost. >1cm	0,24 (0,10-0,37)	0,21 (0,05-0,38)	0,19
Angulació cefalodiafisària	0,45 (0,33-0,57)	0,43 (0,31-0,56)	0,20
Fractura metafisària associada	0,46 (0,25-0,67)	0,28 (0,06-0,5)	0,06
Fulcre intern conservat	0,61 (0,49-0,73)	0,50 (0,38-0,62)	0,05
Impactació	0,44 (0,29-0,58)	0,50 (0,36-0,64)	0,14
Desplaçament diàfisi-tub. gran	0,28 (0,14-0,41)	0,31 (0,18-0,43)	0,18
Desplaçament diàfisi tub.petita	-0,007(-0,04-0,02)	NV	NC
Desplaçament cap-tub.gran	0,48 (0,34-0,63)	0,59 (0,45-0,73)	0,06
Desplaçament cap- tub. petita	0,13(-0,11-0,37)	NV	NC
Comminució tuberositat gran	0,12(0,011-0,23)	0,07 (-0,02-0,17)	0,12
Estoc ossi del cap	0,008(-0,06-0,08)	NV	NC
Split cap	-0,017(-0,030-0,002)	NV	NC
Fractura impactació articular	-0,01 (-0,02-0,006)	0,33 (-0,16-0,81)	0,04
Orientació superfície articular	NV	NV	NC
Luxació superfície articular	NV	NV	NC
Nombre de fragments	NC	NC	NC
Neer	0,52 (0,41-0,63)	0,43 (0,32-0,54)	0,06

Taula 14. Concordança interobservador experts en rx+ TAX axial/ rx + Tac completa amb intervals de confiança. NV: no valorable; NC: no calculable; Kappa no ponderat En les variables, contacte i desplaçament lateromedial els experts obtenen significativament més acord en TAC axial que en TAC completa. En la variable fractura impactació l'acord fou millor amb TAC completa. En la resta de variables no s'observen diferències estadísticament significatives

KAPPA INTEROBSERVADOR NO EXPERTS (INTERVAL DE CONFIANÇA)	AXIAL	COMPLET	P
Contacte	0,34 (0,14-0,54)	0,32 (0,13-0,50)	0,729
Desplaçament lateromedial >1cm	0,46 (0,23-0,69)	0,29 (0,07-0,51)	0,030
Desplaçament anteropost. >1cm	0,18 (-0,09-0,45)	0,38 (0,14-0,62)	0,026
Angulació cefalodiafisària	0,41 (0,21-0,60)	0,32 (0,06-0,57)	0,202
Fractura metafisària associada	0,22(-0,06-0,51)	0,25 (-0,15-0,66)	0,781
Fulcre intern conservat	0,40 (0,07-0,73)	0,35 (0,10-0,60)	0,576
Impactació	0,24(-0,05-0,55)	0,23(-0,14-0,61)	0,924
Desplaçament diàfisi-tub. gran	0,29 (0,04-0,54)	0,37 (0,16-0,57)	0,261
Desplaçament diàfisi tub.petita	0,11 (-0,09-0,35)	0,14 (0,04-0,24)	0,6
Desplaçament cap-tub.gran	0,26 (-0,02-0,54)	0,31 (-0,10-0,73)	0,647
Desplaçament cap- tub. petita	0,24(-0,07-0,49)	0,08 (-0,22-0,40)	0,091
Comminució tuberositat gran	0,24 (0,15-0,33)	0,16 (0,01-0,31)	0,059
Estoc ossi del cap	0,18(-0,07-0,44)	0,18 (-0,25-0,63)	1
Split cap	0,12(-0,18-0,44)	0,02 (-0,12-0,17)	0,216
Fractura impactació articular	0,03(-0,11-0,17)	0,04 (-0,16-0,26)	0,859
Orientació superfície articular	0,31 (0,16-0,45)	0,19(-0,25-0,64)	0,276
Luxació superfície articular	0,25 (0,02-0,48)	0,23(-0,38-0,85)	0,889
Nombre de fragments	0,37 (0,26-0,48)	0,34(0,038-0,65)	0,677
Neer	0,36 (0,07-0,65)	0,38(0,020,74)	0,840

Taula 15. Concordança interobservador en observadors no experts en rx+ TAC axial/ rx + Tac complet amb intervals de confiança. Kappa no ponderat. Només en les variables desplaçament cefalodiafisari lateromedial i desplaçament cefalodiafisari anteroposterior s'observen diferències estadísticament significatives, a favor de TAC axial en la primera, i de la TAC completa en la segona. En la resta de variables no s'observen diferències en el grau d'acord entre TAC axial i TAC complet.

Els observadors considerats no experts, obtenien graus d'acord pobres en les variables de relació cefalodiafisària en els dos estudis radiològics. S'observen diferències estadísticament en les variables desplaçament lateromedial ($p=0,03$), amb millor resultat en TAC axial, i en desplaçament anteroposterior ($p=0,026$), amb millor resultat amb la TAC completa.

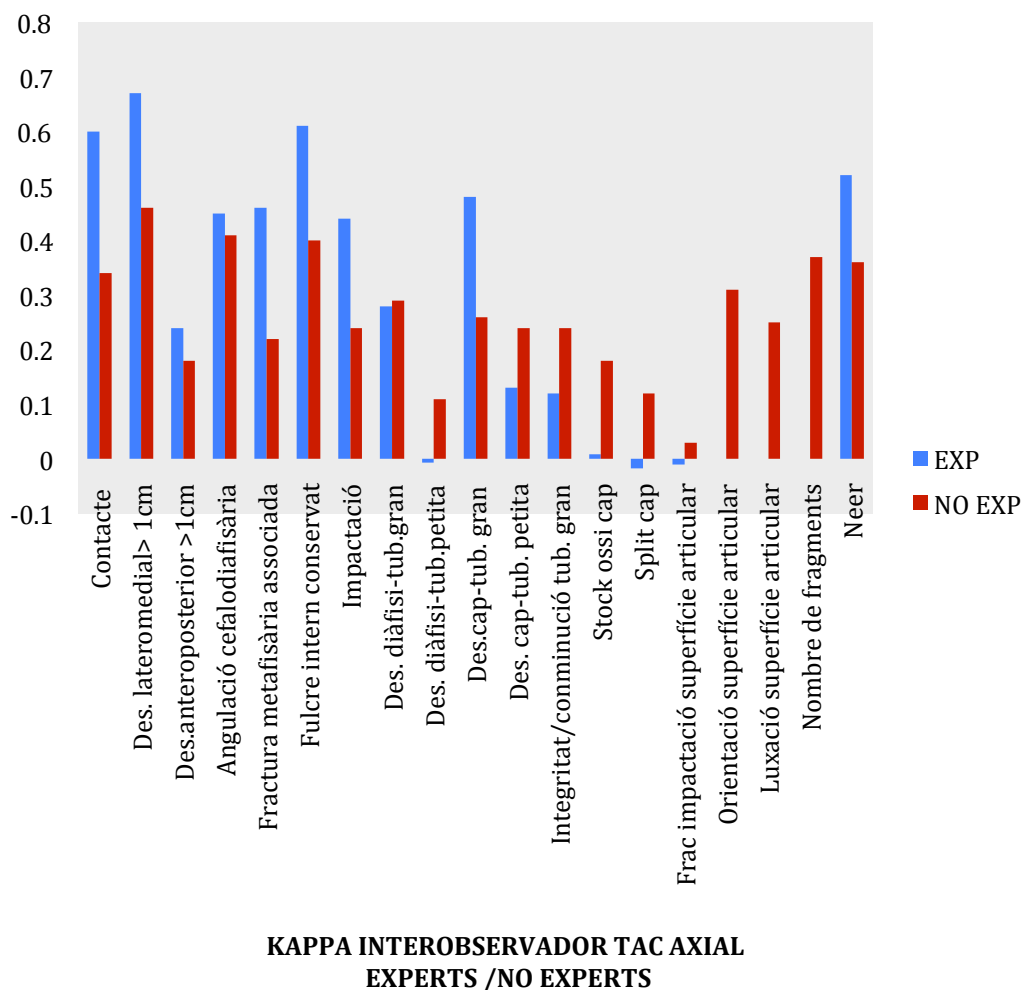
En les variables relacionades amb la tuberositat gran, els valors d'acord foren pobres en els dos estudis radiològics sense diferències estadísticament significatives ($p>0,05$).

Igual que en l'anàlisi de tots els observadors , els valors foren lleus, pròxims a zero , en les variables relacionades amb la tuberositat petita i en estoc, split i fractura impactació de la superfície articular en el grup de relació cefaloglenoidea.

En nombre de fragments i classificació de Neer, l'acord fou pobre en els dos estudis radiològics. No s'observaren diferències estadísticament significatives (nombre de fragments $p=0,67$, Neer $p= 0,84$).

5.1.4 Comparació de la concordança interobservador entre experts i no experts segons l'estudi radiològic

En els gràfics 1 i 2 es comparen els resultats entre experts i no experts, de cada un dels estudis radiològics, TAC axial i TAC completa.



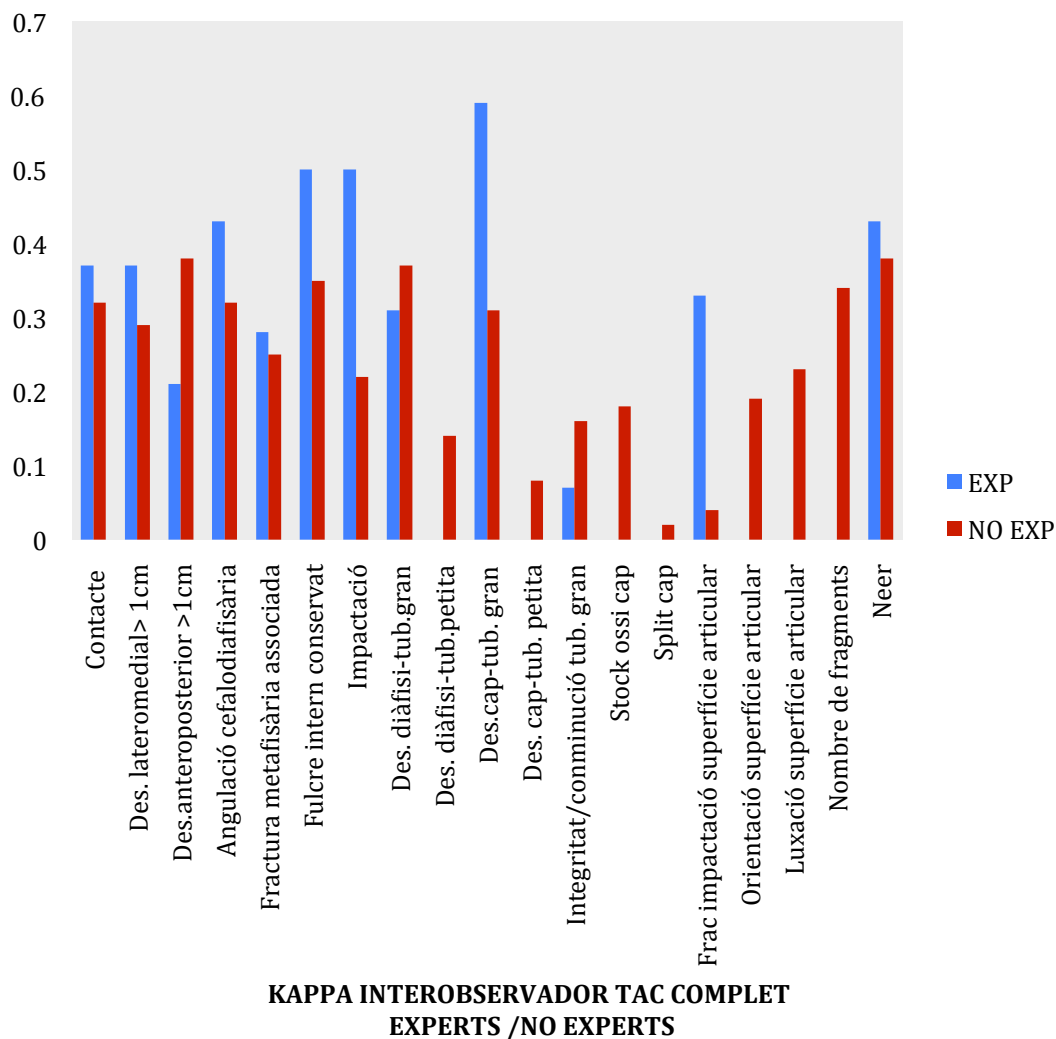
Gràfic 1. Comparació concordança interobservador en tac axial entre experts i no experts. S'observen diferències estadísticament significatives en les variables contacte, desplaçament lateromedial, fractura metafisària, fulcre intern, impactació, desplaçament cap-tuberositat gran i en la classificació de Neer, amb millor acord entre els observadors amb més experiència. Algunes variables foren no valorables en experts per excessiva homogeneïtat en les respostes.

En TAC axial, els observadors experts obtenien millor acord que els no experts en totes les variables de relació cefalodiafisària, amb acord substancial en variables com contacte i desplaçament lateromedial major de 1 cm. Les diferències foren estadísticament significatives, amb millor acord entre els observadors experts, en les variables contacte ($p=0,002$), desplaçament lateromedial ($p=0,007$), fractura metafisària associada ($p= 0,011$), fulcre intern conservat ($p= 0,028$) i impactació ($p=0,027$)

En les variables relacionades amb les tuberositats, l'acord fou millor per als experts en el desplaçament cap-tuberositat gran, amb diferències estadísticament significatives ($p=0,013$). En la resta de variables del grup, els experts obtenen pitjor acord, sense diferències estadísticament significatives ($p>0,05$).

En les variables de relació cefaloglenoidea i característiques del cap humeral, l'acord fou similar amb valors propers a zero i en orientació de la superfície articular i luxació en experts el kappa fou no valorable per excessiva homogeneïtat en les respostes.

En la classificació de Neer en la TAC axial, els experts obtenien un acord moderat (0,52), mentre que el valor fou pobre entre els observadors no experts (0,36), amb diferències estadísticament significatives ($p= 0,04$)



Gràfic 2. Comparació concordança interobservador en l'estudi amb la TAC completa entre experts i no experts. S'observa significativament millor acord en experts en les variables fulcre intern, impactació i desplaçament cap-tuberositat gran. No s'observen diferències estadísticament significatives en la classificació de Neer. Algunes variables foren no valorables en experts per excessiva homogeneïtat en les respostes.

En TAC completa, els experts obtenien millor acord, estadísticament significatiu en les variables conservació del fulcre intern ($p=0,035$) i impactació ($p=0,017$). En les variables desplaçament lateromedial i angulació cefalodiafisària l'acord era lleument superior en experts sense diferències estadísticament significatives. En el desplaçament anteroposterior els no experts obtenien significativament millor resultat ($p=0,019$)

En les variables relacionades amb les tuberositats, els experts obtenien un acord moderat en el desplaçament cap-tuberositat gran mentre que en els no experts l'acord era pobre, amb diferències estadísticament significatives ($p=0,024$). En canvi en la integritat de la tuberositat gran el resultat era significativament millor en els no experts ($p=0,035$).

En les variables de relació cefaloglenoidea i característiques del cap humeral l'acord fou no valorable en totes les variables en experts, excepte en fractura impactació de la superfície articular on tenien un acord pobre, lleu en els no experts, amb diferències estadísticament significatives ($p=0,001$)

En la classificació de Neer, els experts obtenien un valor moderat (0,43), mentre que l'acord en no experts fou pobre (0,38), però sense diferències estadísticament significatives. ($p=0,534$)

2ª PART

5.2 Influència del biaix de prevalença en la concordança i reproductibilitat de la classificació de Neer

5.2.1 Concordança interobservador en la classificació de Neer en la mostra consecutiva

Tal com s'ha expressat en l'apartat 5.1.1 per la classificació de Neer, simplificada en una, dues, tres i quatre parts el grau d'acord interobservador en la mostra consecutiva amb la TAC completa per tots els observadors fou de 0,43 amb un interval de confiança de 0,15-0,71

Expressem a continuació les taules de contingència per parelles de dos observadors per aquesta variable. Anomenem els observadors com: Observador (O) i la numeració del u al sis, O1,O2,O3,O4,O5,O6. Els observadors O2 i O6 són els observadors experts i els observadors O1,O3,O4,O5 són els observadors no experts. Taules 16-30

Podem observar a les taules, les concordances i discrepàncies en les respostes de cada parella. Els observadors experts no són la parella que obté millor resultat.

NEER	O1				
O2	1	2	3	4	TOTAL
1	34	1	0	0	35
2	16	64	4	0	84
3	4	17	13	0	34
4	0	0	0	0	0
TOTAL	54	82	17	0	153
KAPPA	0,543 Taula 16				

NEER	O1				
O3	1	2	3	4	TOTAL
1	19	0	0	0	19
2	31	54	3	0	88
3	3	27	11	0	41
4	0	0	3	0	3
TOTAL	53	81	17	0	151
KAPPA	0,276 Taula 17				

NEER	O1				
O4	1	2	3	4	TOTAL
1	42	7	0	0	49
2	10	66	8	0	84
3	0	7	9	0	16
4	0	0	0	0	0
TOTAL	52	80	17	0	149
KAPPA	0,623 Taula 18				

NEER	O1				
O5	1	2	3	4	TOTAL
1	52	31	5	0	88
2	2	46	6	0	54
3	0	5	6	0	11
4	0	0	0	0	0
TOTAL	54	82	17	0	153
KAPPA	0,466 Taula 19				

NEER	O1				
O6	1	2	3	4	TOTAL
1	49	14	1	0	64
2	5	53	3	0	61
3	0	14	8	0	22
4	0	1	5	0	6
TOTAL	54	82	17	0	153
KAPPA	0,549 Taula 20				

NEER	O2				
O3	1	2	3	4	TOTAL
1	19	0	0	0	19
2	15	66	7	0	88
3	1	15	25	0	41
4	0	1	2	0	3
TOTAL	35	82	34	0	151
KAPPA	0,542 Taula 21				

NEER	O2				
O4	1	2	3	4	TOTAL
1	30	15	4	0	49
2	3	64	17	0	84
3	0	4	12	0	16
4	0	0	0	0	0
TOTAL	33	83	33	0	149
KAPPA	0,510 Taula 22				

NEER	O2				
O5	1	2	3	4	TOTAL
1	34	45	9	0	88
2	1	36	17	0	54
3	0	3	8	0	11
4	0	0	0	0	0
TOTAL	35	84	34	0	153
KAPPA	0,256 Taula 23				

NEER	O2				
O6	1	2	3	4	TOTAL
1	33	25	6	0	64
2	2	50	9	0	61
3	0	9	13	0	22
4	0	0	6	0	6
TOTAL	35	84	34	0	153
KAPPA	0,43 Taula 24				

NEER	O3				
O4	1	2	3	4	TOTAL
1	19	26	3	0	48
2	0	59	24	0	84
3	0	1	13	0	16
4	0	0	0	0	0
TOTAL	19	86	40	3	148
KAPPA	0,357 Taula 25				

NEER	O3				
O5	1	2	3	4	TOTAL
1	19	58	9	0	86
2	0	29	25	0	54
3	0	1	7	3	11
4	0	0	0	0	0
TOTAL	19	88	41	3	151
KAPPA	0,092 Taula 26				

NEER	O3				
O6	1	2	3	4	TOTAL
1	19	37	6	0	62
2	0	47	14	0	61
3	0	4	17	1	22
4	0	0	4	2	6
TOTAL	19	88	41	3	151
KAPPA	0,350 Taula 27				

NEER	O4				
O5	1	2	3	4	TOTAL
1	48	37	0	0	85
2	1	45	7	0	53
3	0	2	9	0	11
4	0	0	0	0	0
TOTAL	49	84	16	0	149
KAPPA	0,478 Taula 28				

NEER	O4				
O6	1	2	3	4	TOTAL
1	42	20	0	0	62
2	6	50	3	0	59
3	1	10	11	0	22
4	0	4	2	0	6
TOTAL	49	84	16	0	149
KAPPA	0,505 Taula 29				

NEER	O5				
O6	1	2	3	4	TOTAL
1	63	1	0	0	64
2	20	40	1	0	61
3	4	11	7	0	22
4	1	2	3	0	6
TOTAL	88	54	11	0	153
KAPPA	0,538 Taula 30				

Taules 16-30: Taules de contingència per parelles d'observadors en la classificació de Neer simplificada en 4 categories en la mostra consecutiva. Observem el valor de Kappa de cada parella i la dispersió en les respostes entre ells. La parella amb millor acord, O1 i O4, obtenen un valor de 0,62 i la parella amb pitjor acord, O3 i O5 obtenen 0,092. En la taula 26, s'aprecia la gran discordança en les respostes en aquesta parella.

5.2.2 Reproductibilitat intraobservador en la classificació de Neer en la mostra consecutiva

La reproductibilitat per la classificació de Neer en la mostra consecutiva amb radiologia simple i TAC completa fou de 0,72

A la taula 31 s'expressa la reproductibilitat obtinguda per cada un dels observadors

CONSECUTIVA	KAPPA
O1	0,587
O2	0,600
O3	0,931
O4	0,706
O5	0,787
O6	0,767
MITJA	0,72

Taula 31. Reproductibilitat en la mostra consecutiva de cada observador

5.2.3 Concordança interobservador en la classificació de Neer en la mostra seleccionada

El grau d'acord kappa obtingut en seleccionar una mostra balancejada pels quatre tipus de fractura de la classificació de Neer amb radiologia simple i TAC completa, dels mateixos sis observadors fou de 0,45 amb un interval de confiança de 0,21-0,69

A continuació s'expressen les taules de contingència i el valor Kappa de cada parella d'observadors. Taules 32-46

NEER	O1				
O2	1	2	3	4	TOTAL
1	17	4	0	1	22
2	5	19	2	2	28
3	2	5	16	5	28
4	0	0	1	9	10
TOTAL	24	28	19	17	88
KAPPA	0,585 Taula 32				

NEER	O1				
O3	1	2	3	4	TOTAL
1	7	1	0	0	8
2	13	18	4	2	37
3	4	10	14	6	34
4	1	0	1	9	11
TOTAL	25	29	19	17	90
KAPPA	0,369 Taula 33				

NEER	O1				
O4	1	2	3	4	TOTAL
1	21	2	0	0	23
2	4	23	5	4	36
3	0	4	14	6	24
4	0	0	0	7	7
TOTAL	25	29	19	17	90
KAPPA	0,619 Taula 34				

NEER	O1				
O5	1	2	3	4	TOTAL
1	23	9	1	2	35
2	2	20	10	3	35
3	0	0	8	7	15
4	0	0	0	5	5
TOTAL	25	29	19	17	90
KAPPA	0,476 Taula 35				

NEER	O1				
O6	1	2	3	4	TOTAL
1	20	9	2	0	31
2	4	17	5	3	29
3	1	3	12	5	21
4	0	0	0	9	9
TOTAL	25	29	19	17	90
KAPPA	0,514 Taula 36				

NEER	O2				
O3	1	2	3	4	TOTAL
1	8	0	0	0	8
2	13	20	3	0	36
3	1	7	23	2	33
4	0	1	2	8	11
TOTAL	22	28	28	10	88
KAPPA	0,538 Taula 37				

NEER	O2				
O4	1	2	3	4	TOTAL
1	16	5	1	0	22
2	5	22	7	2	36
3	1	1	18	3	23
4	0	0	2	5	7
TOTAL	22	28	28	10	88
KAPPA	0,571 Taula 38				

NEER	O2				
O5	1	2	3	4	TOTAL
1	21	12	2	0	35
2	1	16	15	1	33
3	0	0	11	4	15
4	0	0	0	5	5
TOTAL	22	28	28	10	88
KAPPA	0,448 Taula 39				

NEER	O2				
O6	1	2	3	4	TOTAL
1	20	7	3	0	30
2	1	17	10	0	28
3	1	4	13	3	21
4	0	0	2	7	9
TOTAL	22	28	28	9	88
KAPPA	0,515 Taula 40				

NEER	O3				
O4	1	2	3	4	TOTAL
1	8	12	3	1	24
2	1	21	13	1	36
3	0	4	17	3	24
4	0	0	1	6	7
TOTAL	9	37	34	11	91
KAPPA	0,392 Taula 41				

NEER	O3				
O5	1	2	3	4	TOTAL
1	9	23	3	1	36
2	0	13	22	0	35
3	0	1	9	5	15
4	0	0	0	5	5
TOTAL	9	37	34	11	91
KAPPA	0,179 Taula 42				

NEER	O3				
O6	1	2	3	4	TOTAL
1	8	17	6	1	32
2	1	15	13	0	29
3	0	5	12	4	21
4	0	0	3	6	9
TOTAL	9	37	34	11	91
KAPPA	0,255 Taula 43				

NEER	O4				
O5	1	2	3	4	TOTAL
1	23	12	1	0	36
2	1	23	11	0	35
3	0	1	11	3	15
4	0	0	1	4	5
TOTAL	24	36	24	7	91
KAPPA	0,526 Taula 44				

NEER	O4				
O6	1	2	3	4	TOTAL
1	19	12	1	0	32
2	4	16	9	0	29
3	1	6	12	2	21
4	0	2	2	5	9
TOTAL	24	36	24	7	91
KAPPA	0,399 Taula 45				

NEER	O5				
O6	1	2	3	4	TOTAL
1	24	8	0	0	32
2	10	16	3	0	29
3	2	10	9	0	21
4	0	1	3	5	9
TOTAL	36	35	15	5	91
KAPPA	0,415 Taula 46				

Taules 32-46. Taules de contingència per parelles d'observadors en la classificació de Neer simplificada en 4 categories en la mostra seleccionada. Observem el valor de Kappa de cada parella i la dispersió en les respostes entre ells. La parella amb millor acord, O1 i O4, obtenen un valor de 0,61 i la parella amb pitjor acord, O3 i O5 obtenen 0,179. Els resultats entre les parelles es mantenen en comparació a la mostra consecutiva, sent les mateixes parelles les que obtenen millor i pitjor resultat.

5.2.4 Reproductibilitat intraobservador en la classificació de Neer en la mostra seleccionada

La reproductibilitat intraobservador en la classificació de Neer en aquesta mostra fou de 0,746 amb un interval de confiança de 0,68-0,8

A la taula 47 s'expressa la reproductibilitat obtinguda per cada un dels observadors.

SELECCIONADA	KAPPA
O1	0,773
O2	0,731
O3	0,729
O4	0,774
O5	0,849
O6	0,626
MITJA	0,746

Taula 47. Reproductibilitat en la mostra seleccionada de cada observador

5.2.5 Diferències estadístiques en la concordança interobservador i reproductibilitat intraobservador de la classificació de Neer segons tipus de mostra.

A la taula 48 s'expressa la prevalença observada per cada tipus de fractura en les dues mostres.

Tipus de fractura	Mostra consecutiva	Mostra seleccionada
Una part	34%	26%
Dues parts	47%	34%
Tres parts	15%	24%
Quatre parts	1,1%	10%

Taula 48. Prevalença observada en cada mostra dels diferents tipus de fractura. Si comparem les prevalences de cada mostra, observem un predomini de les fractures d'una i dues parts en la mostra consecutiva, amb només un 1,1% de fractures en quatre parts i en canvi en la mostra seleccionada, la variable presenta una millor distribució amb més percentatge de fractures en tres i quatre parts. En seleccionar la mostra es va aconseguir una millor distribució de la variable a estudi.

A la taula 49 s'expressen els resultats del coeficient Kappa interobservador i intraobservador per la classificació de Neer en les dues mostres

NEER	MOSTRA CONSECUTIVA	MOSTRA SELECCIONADA	p
KAPPA INTEROBSERVADOR	0,43 (0,15-0,71)	0,45 (0,21-0,69)	0,68
KAPPA INTRAOBSERVADOR	0,72 (0,64-0,8)	0,74 (0,68-0,8)	0,14

Taula 49. Resultats de concordança interobservador i reproductibilitat intraobservador en la classificació de Neer simplificada en la mostra consecutiva i la mostra seleccionada. No s'observen diferències estadísticament significatives.

No s'observen diferències estadísticament significatives en la concordança interobservador ni en la reproductibilitat intraobservador en la classificació de Neer simplificada entre una mostra consecutiva de la població i una mostra seleccionada per tipus de fractura (p=0,68).

6. DISCUSSIÓ

Ha estat i segueix sent en l'actualitat un gran repte pels cirurgians d'espatlla ser capaços de definir, caracteritzar i classificar les fractures de l'extrem proximal de l'húmer. Sent aquestes, en alguns casos, patologies de difícil tractament, pel fet de donar-se en pacients d'edat avançada, amb una qualitat òssia deficient i menors capacitats de recuperació funcional.

La classificació de Neer, encara avui el patró or en la classificació d'aquestes fractures, ha generat molta controvèrsia per la baixa concordança i reproductibilitat obtinguda en els estudis realitzats fins al moment. Diferents autors han analitzat les causes d'aquests resultats, focalitzant-se en l'estudi radiològic i en l'experiència del cirurgià, mentre altres han intentat definir altres sistemes de classificació més reproduïbles, basant-se en el sistema dels quatre segments de Codman. Fins al moment, no ha pogut definir-se un sistema millor, però sí que s'han definit altres factors que no té en compte la classificació de Neer, com la disrupció del calcar medial i l'extensió metafisària descrits per Hertel.⁵⁰

A part de l'estudi radiològic i de l'experiència de l'observador, que no han demostrat encara avui ser el principal factor de la baixa fiabilitat de la classificació, alguns autors han atribuït al biaix de prevalença del coeficient kappa els mals resultats dels estudis de concordança en aquesta classificació. Hem analitzat els efectes d'aquest biaix en un estudi clínic així com la metodologia utilitzada per diferents investigadors en realitzar aquest tipus d'estudis.

6.1 De la influència de la prevalença

Audigé i cols.⁶⁰ en una revisió sistemàtica de 44 estudis que valoren concordança en classificacions de fractures, recomanen seleccionar la mostra tal com es presentaria en una situació real, independentment de la qualitat de les imatges, patrons de fractura o tractament utilitzat, quan valorem la validesa d'un sistema de classificació.

Brorson⁵³ en una revisió sistemàtica d'estudis de concordança en classificació de Neer en fractures en 4 parts, observà millor grau d'acord en estudis que tenien poblacions seleccionades per fractures desplaçades, que no en estudis amb poblacions no seleccionades, però cap estudi previ ha valorat l'efecte de la mostra en els resultats de concordança.

Trobem una gran dispersió en la literatura de com els autors seleccionen la mostra per valorar la validesa de la classificació de fractures d'húmer proximal. De tots els estudis publicats, únicament Bruinsma i cols.⁵⁵ seleccionen els casos segons el patró de fractura, la resta d'estudis seleccionen els casos de forma consecutiva amb mostres de desde 12 casos en l'estudi de Sallay⁸ a 113 casos de l'estudi de Sanchez-Sotelo.⁷⁰

Hem valorat amb els mateixos observadors quin efecte té el canvi de la mostra, en els resultats de concordança interobservador, basant-nos en que el biaix de prevalença fa tendir a zero el resultat del Kappa quan estudiem una categoria poc o molt prevalent. Ja que disposem d'una mostra consecutiva de 153 fractures en tractament ortopèdic, tindrem, segona la prevalença descrita en aquest tipus de fractura, una molt elevada proporció de fractures en una part i probablement molt poques o cap fractura en 4 parts, per tant hi ha categories de la variable a estudi que poden no existir en aquesta mostra. Hipotèticament, si tenim una mostra balancejada per les quatre categories de la classificació, això ens hauria de millorar els resultats

dels mateixos observadors, i per tant el factor prevalença explicaria els mals resultats de concordança obtinguts en estudis previs amb mostres consecutives de la població o no balancejades pels tipus de fractura.

Valorant la classificació de Neer, únicament Sidor³, malgrat tenir una mostra de casos consecutius, comprova amb dos cirurgians ortopèdics no observadors en l'estudi, que la majoria de patrons de la classificació estan representats en els 50 casos del seu estudi. No s'especifica com es fa aquesta valoració i si es produeix algun canvi en la mostra. Per la prevalença descrita dels diferents tipus de fractura de la classificació de Neer, podem deduir que en cinquanta casos consecutius no tinguem probablement una igual representació dels diferents tipus de fractura. Bruinsma⁵⁵, valorant classificació de la AO, selecciona 15 casos de manera que representin els diferents patrons de fractura, no s'especifica de quina manera.

Prèviament, alguns autors com Feinstein i Cicchetti⁶⁶, buscaren fórmules per corregir el biaix de prevalença del coefficient Kappa sense èxit, sent conscients que seleccionar una mostra balancejada dificultaria els estudis de concordança interobservador, ja que la població a estudi ha de ser primer avaluada per veure si té una distribució adequada de la variable.

Alguns autors atribueixen al biaix de prevalença els baixos valors de Kappa i la pobra concordança en la classificació de Neer.⁵³ En el nostre treball hem obtingut els mateixos valors de concordança quan hem demanat als mateixos observadors, valorar la classificació de Neer en una mostra on la variable a estudi tenia teòricament les diferents opcions igualment representades, per tant demostrem que el biaix de prevalença no és la causa dels baixos valors de concordança en aquesta classificació. Tot i així el problema es troba en com seleccionem una mostra balancejada, ja que els seleccionadors d'aquesta no deixen en certa manera de ser també observadors, tot i

que els tres seleccionadors eren especialistes en cirurgia d'espatlla que es posaven d'acord entre ells sobre el tipus de fractura que estaven seleccionant. La mostra seleccionada ideal seria potser la seleccionada de casos quirúrgics en què realitzant cirurgia oberta s'ha pogut observar com és el patró de fractura, però això només seria útil per fractures de tractament quirúrgic obert i per poder valorar tota la classificació precisem també de fractures en tractament conservador o fractures en què el tractament quirúrgic seria tancat.

Si comparem la prevalença observada de cada tipus de fractura en les dues mostres, veiem que la mostra seleccionada té una distribució molt més heterogènia pels quatre tipus de fractura que la mostra consecutiva on la prevalença de fractures en 2 parts era quasi del 50% mentre que les fractures en 4 parts representaven l'1,1%. Malgrat que es depèn del criteri dels seleccionadors de la mostra, es va aconseguir una mostra amb una distribució més heterogènia de la variable a estudi.

Malgrat les recomanacions en estudis de validesa de classificacions són utilitzar mostres representatives de la població i no mostres seleccionades, hem volgut demostrar en aquest estudi la influència del biaix de prevalença en el resultat del coeficient Kappa, ja que molts autors han suggerit aquest factor com a causa dels mals resultats obtinguts en la valoració de la classificació de Neer. Queda demostrat doncs que el biaix de prevalença inherent en el coeficient Kappa no influeix quan valorem la classificació de Neer simplificada en 4 categories. Això no exclou que no hàgim d'interpretar amb cautela els resultats d'aquest i expressar-lo sempre detallant els intervals de confiança i els resultats per cada parella d'observadors i categoria.

Potser ens caldria una prova estadística més adequada o diferents criteris d'interpretació. Podem utilitzar el coeficient per valorar factors que ens influeixen en la validesa de la classificació però no utilitzar les

guies d'interpretació per concloure que una classificació és o no és vàlida. A més, no hi ha un acord entre les diferents guies descrites, sobre què seria un resultat de Kappa acceptable. Per Landis i Koch⁴⁴ un valor $>0,8$ és un acord quasi perfecte, mentre que per altres autors com Emam⁷¹ un valor $>0,75$ és excel·lent.

Finalment, les discrepàncies en la classificació, no impliquen necessàriament discrepàncies en el tractament. Malgrat que no s'ha valorat en aquest estudi, altres autors han observat desacords entre els especialistes pel que fa al tractament a aplicar, però millor grau d'acord respecte a la classificació.^{72,73}

6.2 De l'estudi radiològic

Fins al moment, la major raó per explicar la pobra concordança de la classificació de Neer, ha estat la dificultat en transformar informació observada en una imatge en grups determinats segons un concepte anatomopatològic. No és el mateix definir el que observem que decidir-nos per un grup segons què pensem que ha succeït anatomopatològicament.

En el protocol de lectura radiològica de Mora i cols.⁴⁷ es pretén definir unes característiques radiològiques que poden ser rellevants en definir la fractura i definir un protocol de lectura amb l'objectiu de definir millor la fractura i d'obtenir millors resultats en el tractament. En els estudis d'aquest autor es definien les característiques comparant radiologia simple amb radiologia amb TAC tal com es va fer en altres estudis de concordança en classificació de Neer i de AO. Aquests autors obtenien millors valors de concordança amb l'ús de la TAC en les variables nombre de fragments, desplaçament de la tuberositat petita, i fractura extra articular o articular.

En aquest treball es revaloren les característiques radiològiques d'aquest protocol, quin acord tenen diferents especialistes d'espalla en la lectura radiològica i si la valoració de les diferents projeccions de la TAC contribueix a millorar el grau d'acord. Les variables utilitzades en aquest estudi corresponen a una actualització posterior d'aquest autor.⁵² En l'estudi original, l'únic on es presenten resultats de concordança, no s'utilitzen totes les variables o no s'utilitzen de la mateixa manera, per tant aquestes no es poden comparar.

L'estudi original incloïa 30 fractures desplaçades, en el nostre cas la sèrie inclou 153 fractures recollides de forma consecutiva. Mora i cols.⁴⁷ eliminaven fractures no desplaçades i fractures en dos parts ben definides amb l'objectiu que els observadors tinguessin més probabilitats de diferir. En tenir una mostra consecutiva, malgrat l'elevat nombre de casos, cal esperar un elevat nombre de fractures no desplaçades en aquesta mostra, ja que així es dona en una situació real, s'esperava que aquest fet hauria d'augmentar el grau d'acord dels observadors, ja que teòricament tindríem menys probabilitats de diferir entre ells, per contra si la distribució d'una variable fos massa homogènia, el resultat del Kappa es veuria teòricament afectat pel biaix de prevalença.

En les variables de relació cefalodifiària, els resultats foren molt inferiors als obtinguts en l'estudi original, només valorades amb la radiologia simple, amb valors de kappa pobres en pràcticament totes les característiques, mentre que en l'estudi original els valors oscil·laven de 0,50 a 0,79, acord moderat segons els criteris de Landis i Koch. S'entén que, en aquest grup de variables, en l'observació amb radiologia simple i TAC axial, els observadors es guiaven per la radiologia simple, ja que són variables poc valorables en TAC axial però sí en TAC completa. Quan es van valorar amb aquest no s'obtenien diferències en el grau d'acord. Contràriament l'acord empitjorava significativament en l'ús de la TAC completa en la variable de

desplaçament lateromedial. Podem concloure que en aquest grup de variables l'ús de la TAC no és útil per definir aquestes característiques.

En les variables relacionades amb les tuberositats, els valors de kappa foren també inferiors als de l'autor original, però tampoc l'ús de la TAC completa contribuïa a millorar l'acord entre els observadors en cap de les variables. Cal remarcar que en l'estudi original no es definia respecte a on es produïa el desplaçament de la tuberositat, cap o diàfisi.

En les variables de relació cefalogenoidea i estoc ossi del cap s'observen els mateixos resultats, sense diferències entre la TAC axial i la TAC completa. En l'estudi original es valorava fractura tipus split i/o fractures impactació de la superfície articular com una única variable.

Una de les possibles explicacions a graus d'acord baixos és el biaix de prevalença del Kappa, ja que en ser una mostra de casos consecutius, és possible que algunes d'aquestes variables tinguin molt poca o molta presència en la mostra i això ens acosti a zero el grau d'acord, tot i així els valors són baixos en totes les variables, així doncs no pot ser aquesta l'única causa. Sí que observem, que les variables amb més probabilitat d'heterogeneïtat en la mostra com angulació, contacte, conservació del fulcre o les relacionades amb la tuberositat gran, tenen valors més elevats que variables com les relacionades amb la tuberositat menor, split del cap o luxació on els valors són encara més propers a zero o no calculables en alguns casos, sent variables probablement amb una distribució molt més homogènia.

Crida l'atenció també, la gran amplitud en els intervals de confiança que s'obté en totes les variables, fet que confereix poca precisió al resultat. Tant en l'estudi previ d'aquest sistema de caracterització com en la resta d'estudis de fiabilitat de sistemes de classificació de fractures d'húmer proximal, exceptuant els estudis de Brorson^{10,11,53,72} i Bruinsma⁵⁵, els intervals de confiança no hi són reflectits, per tant no

podem saber quant precisos són els resultats expressats. Un resultat amb un interval de confiança estret li confereix molta més precisió que no si aquest és molt ampli.

No podem comparar aquests resultats amb els obtinguts en l'estudi original, ja que en el cas de les variables de relació cefalodiafisària no es van valorar amb la TAC, la mostra estava seleccionada per fractures desplaçades, i no coneixem si es va calcular el coeficient kappa ponderat o no ponderat, que caldria especificar en l'anàlisi estadística d'aquest tipus d'estudis.

En relació a la variable nombre de fragments, ja es va demostrar en l'estudi de Bernstein i cols.⁶ que el fet de valorar presència de línies de fractura, independentment del grau de desplaçament o angulació, no contribuïa a millorar els valors de concordança, obtenint en el seu treball, valors de 0,50 quan valorava parts i de 0,56 quan valorava fragments. En l'estudi original de Mora i cols.⁴⁷ els valors de Kappa foren molt més elevats quan es valoraven fragments, 0,72, en comparació amb 0,44 en parts de Neer. En la nostra sèrie els valors foren inferiors en valorar fragments (0,34) sense diferències en funció de la seqüència de la TAC, i sense diferències quan es comparaven amb parts, concordant amb l'estudi de Bernstein⁶.

Un altre factor on s'observa una gran dispersió en els estudis de concordança en la classificació de Neer és en la simplificació de la classificació. Cal tenir en compte que no tots els estudis miren la classificació de Neer de la mateixa manera, alguns valoren la classificació original en 16 categories^{2-6,11} altres la classificació simplificada en 6(8) i altres només nombre de parts^{7,47,51}. Alguns autors valoren la classificació original i després la simplifiquen en 6 categories^{3,4,6}.

Els graus d'acord en la classificació de Neer simplificada en quatre categories, foren similars als obtinguts en la sèrie de Mora⁴⁷, també

amb la mateixa simplificació. Tot i que els resultats són concordants amb la resta de sèries publicades, és discutible si aquests serien comparables. Com més opcions de resposta disposa l'observador en una variable, més probabilitats de dispersió en les respostes, fet que ens afecta el percentatge d'acord entre els observadors i per tant els resultats del coeficient kappa. Estudis previs han conclòs que simplificar la classificació no millora l'acord, però estadísticament no està clar si podem comparar dues variables en la que una té 16 opcions de resposta i l'altre 6 o 4. En la literatura s'accepta utilitzar el Kappa de Cohen, independentment de les opcions de resposta, però aquest coeficient fou descrit per dos observadors avaluant una variable dicotòmica.

Els graus d'acord foren superiors en la reproductibilitat intraobservador tant en la TAC axial com en completa en totes les variables. En algunes com desplaçament anteroposterior, angulació, conservació del fulcre, desplaçament cap-tuberositat gran, integritat de la tuberositat gran, fractura impactació i nombre de fragments, la TAC completa contribueix de forma significativa a millorar la reproductibilitat. Destaca l'elevada reproductibilitat en la classificació de Neer en els dos estudis radiològics, concordant amb la resta d'estudis publicats. Així doncs, l'ús de la TAC completa no ens millora l'acord amb altres observadors però si ens dóna més seguretat en les nostres pròpies observacions.

La introducció de la radiologia digital tampoc ha contribuït a millorar l'acord entre observadors. Tot i que pugui semblar d'entrada un sistema molt més precís que la radiologia simple convencional que podria ajudar-nos a millorar l'acord, ja que té eines per calcular distàncies, angulacions, mm de desplaçament i permet fer canvis de contrast i magnificar la imatge, és igualment dependent de l'observador, per exemple en identificar un punt d'inici i final respecte a una angulació o un desplaçament, és a dir dos observadors poden diferir igualment en la seva calibració si no interpreten el mateix.

Tot i que l'ús de la TAC no millora el grau d'acord entre observadors en la classificació de fractures de l'húmer proximal, en la nostra pràctica clínica habitual seguim utilitzant la TAC quan ens trobem davant d'una fractura complexa, per tal de definir millor línies de fractura i desplaçament dels fragments i ajudar-nos en la decisió terapèutica i planificació quirúrgica. L'ús de la seqüència completa de la TAC no seria necessària ni per definir característiques radiològiques de la fractura ni per aplicar la classificació de Neer, però això no implica que no ens sigui útil per la planificació quirúrgica. Aportar més informació no ajuda a l'observador a definir millor la fractura. Amb la seqüència de la radiologia simple i la TAC axial, seria suficient per definir el desplaçament de les tuberositats i la posició del cap respecte a la diàfisi, factors més importants a l'hora de prendre decisions terapèutiques.

6.3 Dels observadors i l'experiència

La major part d'estudis de concordança interobservador, utilitzen observadors amb diferents graus d'experiència i amb diferents subespecialitzacions, i per tant amb diferents nivells de coneixement de la classificació a estudi.

No hi ha a la literatura treballs que ens indiquin com han de ser els observadors en un estudi de concordança ni el nombre que necessitem. Alguns autors intenten reproduir una situació real utilitzant observadors de diferents graus d'experiència, altres en canvi utilitzen especialistes en la classificació assumint que en la vida real seran aquests els que aplicaran la classificació, els altres, de menys experiència, normalment els consultaran en cas de dubte i a l'hora de prendre decisions.

Respecte al nombre d'observadors no necessàriament més observadors ens donaran millor resultat, ja que dependrà de com siguin aquests observadors, però estadísticament tenim més probabilitats de donar un resultat més precís com més observadors tinguem, que no si agafem només dos observadors a l'atzar. Si tenim un nombre elevat d'observadors, els resultats de les parelles que tinguin un molt bon o molt mal resultat quedaran compensats amb les parelles amb un resultat mitjà, si escollim únicament dos observadors, podem obtenir per atzar uns resultats molt bons o molt dolents.

Un altre potencial problema dels estudis interobservador és el nombre de casos. No existeix una fórmula per calcular el nombre de casos necessari en aquest tipus d'estudi tot i que en estudis previs es recomana un mínim de 40 pacients.⁶⁴ Tal com s'ha expressat a la taula 10, no tots els estudis contenen aquest nombre de casos.

En la literatura els observadors són molt diferents. Kristiansen¹ utilitzava quatre observadors especialistes en cirurgia ortopèdica amb diferents anys d'experiència, Sidor³ utilitzà cinc observadors, un especialista en espatlla, un especialista en cirurgia traumàtica, un radiòleg especialista en patologia musculoesquelètica, un resident de cinquè any i un resident de primer any. Siebenrock⁷ en canvi utilitzava cinc especialistes en cirurgia ortopèdica amb "especial interès" en la patologia d'espatlla. Bernstein⁶ utilitza quatre observadors, dos residents de cinquè any i dos especialistes d'espatlla. Sjöden⁵ set observadors, 4 especialistes en ortopèdia i tres especialistes en radiologia. En l'estudi de Mora i cols.⁴⁷ els observadors foren un especialista d'espatlla, un ortopeda general experimentat, un ortopeda general junior i un resident de quart any.

En el nostre estudi, participaven observadors especialistes en cirurgia ortopèdica subespecialitzats en patologia de l'espatlla tant traumàtica com ortopèdica. Els dividim segons els anys d'experiència, ja que dos

d'ells tenien més de deu anys d'experiència i els altres quatre menys de 10 anys en aquesta patologia. El fet de tenir tots els observadors especialistes en espatlla feia esperar globalment millors graus d'acord tant interobservador com intraobservador, però els resultats no han estat així, tant si comparem amb l'estudi de Mora i cols.⁴⁷ en les variables radiològiques tant si comparem amb la resta d'estudis publicats de la classificació de Neer.

Dels autors que valoren els observadors més experts i menys experts per separat, únicament Bernstein⁶ presenta millors graus d'acord entre els experts, sense valorar-ne les diferències. En el nostre estudi, els observadors més experts sí que obtenien millor grau d'acord en algunes variables de forma significativa, tant en TAC axial com en completa. En la classificació de Neer en TAC axial els més experts obtenien millor grau d'acord de forma significativa. Quan valorem l'estudi radiològic segons el grau d'experiència, crida l'atenció que els observadors experts obtenen millor acord en TAC axial que en TAC complet, és a dir en aquests, tenir més informació implica menys acord.

Brorson i cols.⁵⁴ demostraren que l'entrenament dels observadors sí que podia fer-ne augmentar el grau d'acord. En el treball es valorava acord interobservador en la classificació de Neer, en un grup de cirurgians ortopèdics sense realitzar explicacions de la classificació i després de realitzar una sessió de 45 minuts on s'ensenyava la classificació. En observadors entrenats per la classificació de Neer observà un grau d'acord de 0,29 abans de l'entrenament i de 0,62 després, al contrari que el grup control que obtenia un valor de Kappa de 0,28 basal i de 0,33 en la segona valoració sense entrenament. També Rasmussen i cols.⁷⁴ van observar milloria en l'acord quan es donaven instruccions als observadors sobre com valorar la classificació.

En el nostre treball, tots els observadors utilitzen la classificació de Neer de forma habitual i saben aplicar-la, per tant podríem concloure que el fet de tenir observadors subespecialitzats en la classificació no fa obtenir millors valors de Kappa que en altres sèries amb observadors de menys experiència. Tot i així quan els dividim segons grau d'experiència si que obtenim millors resultats en algunes variables, per tant el grau d'experiència si que millora l'acord entre observadors.

Tot i que tal com s'ha comentat prèviament, teòricament més observadors ens donaran un resultat més precís, el percentatge d'acord entre tots els observadors disminuirà conforma augmenti en nombre d'observadors. Sidor³ millorava el grau d'acord quan analitzava les dades amb menys observadors. Amb cinc observadors obtenia un 32% d'acord i amb tres observadors el 88%. Així es recomana, en aquest tipus d'estudis, expressar els percentatges d'acord o bé els resultats de kappa en les taules de contingència per parelles d'observadors, tal com hem realitzat. Com més observadors tinguem l'acord serà menor, però serà més precís.

Finalment cal tenir en compte factors subjectius personals dels observadors. El fet de participar en un estudi pot fer augmentar el grau d'atenció dels observadors, però alhora com que la seva decisió no té implicacions en el tractament i en els resultats clínics d'un determinat pacient, pot fer-los menys rigorosos en l'observació.

Un cop analitzada la literatura i davant els nostres resultats, pot afirmar-se que la gran dispersió entre els estudis publicats en referència al nombre d'observadors, selecció d'aquests, graus d'experiència, nombre de casos, selecció de la mostra, categories de la classificació, proves d'imatge i la prova estadística utilitzada fa no comparables els estudis publicats fins al moment de concordança i reproductibilitat de la classificació de Neer. La comparació del coeficient

Kappa estaria indicada únicament quan la distribució de les categories és similar en dues mostres o quan s'utilitza la mateixa mostra en diferents ocasions. Únicament els estudis de Sjöden^{4,5} en què els autors utilitzaven la mateixa mostra i el mateix mètode canviant només les proves d'imatges, serien comparables.

7. APLICABILITAT CLÍNICA

En la pràctica diària, el clínic té la informació sobre el pacient, edat, sexe, nivell d'activitat prèvia a la fractura, demanda funcional i qualitat òssia que l'ajuden en la presa de decisions sobre el tractament de la fractura i l'ajuden en poder pronosticar un resultat determinat per cada cas en concret. Quan expressem graus d'acord, només basats en unes proves d'imatge no estem expressant acord en el diagnòstic ni en les decisions terapèutiques, que poden ser molt diferents en cada cas. És evident, que les decisions terapèutiques no han de prendre's només segons el resultat d'una classificació. Tot i així les classificacions de les fractures són necessàries per comparar els resultats amb altres autors i per valorar els nostres propis resultats evolutius així com a eina de comunicació amb altres professionals.

Queda evidenciat amb aquest estudi, que l'observació de característiques radiològiques en fractures d'húmer proximal entre observadors especialistes en cirurgia d'espatlla independentment de la seqüència de TAC, dona molt poca concordança entre els observadors, és a dir especialistes en observar fractures d'espatlla, no observen el mateix o no interpreten la característica radiològica de la mateixa manera. Alhora s'han valorat en el treball característiques amb radiologia simple i TAC però no coneixem en quin dels dos estudis es van basar els observadors per donar les seves respostes.

Més que definir una nova classificació de fractures de l'extrem proximal de l'húmer, potser caldria definir una nova sistemàtica de lectura radiològica on es defineixin millor les característiques i en quin estudi radiològic poden ser valorades, eliminant característiques que no poden ser valorades o interpretades i poden generar confusió. Ja s'ha demostrat en estudis previs en la literatura que l'ús de la TAC associada a la radiologia simple no millora l'acord en la classificació de

les fractures. En el nostre cas, la informació aportada per la TAC en seqüència completa ha creat més confusió entre els observadors, especialment entre els més experts, disminuint en alguns casos l'acord entre ells.

Així amb la radiologia simple en tindrem suficient per valorar les variables de relació cefalodiafisària i el desplaçament de la tuberositat gran respecte la diàfisi. Caldria definir millor els punts de mesura del desplaçament anterior i lateral, i com es mesura el contacte i el desplaçament de la tuberositat gran respecte la diàfisi. Cal tenir en compte també que per mesurar l'angle cefalodiafisari es necessita una projecció anteroposterior en rotació neutra, ja que les rotacions de l'húmer poden donar variacions d'aquest angle. No és valorable ni quantificable el desplaçament de la tuberositat petita respecte a la diàfisi.

Amb la TAC axial podem valorar el desplaçament de la tuberositat gran respecte al cap, la orientació de la superfície articular, i l'estoc ossi del cap tal com es valoren en el protocol. És també més difícil valorar el desplaçament de la tuberositat petita.

Probablement amb aquesta informació de la fractura, seria suficient per prendre una decisió terapèutica i pronosticar-ne el resultat.

Respecte a la classificació de Neer, seguirà sent el patró or en la classificació de fractures de l'extrem proximal de l'húmer, ja que fins al moment no tenim una classificació ni un protocol de lectura radiològica amb més fiabilitat. Els resultats obtinguts han estat concordants amb la resta d'estudis en la literatura independentment de la seqüència de TAC. El biaix de prevalença del Kappa, al qual alguns autors han atribuït els mals resultats, influeix menys del que es pensava quan valorem aquesta classificació i no és la causa dels baixos nivells de concordança.

Malgrat no haver trobat una causa del baix acord en aquesta classificació, el treball ha evidenciat la gran dispersió en la literatura de la metodologia utilitzada per valorar la fiabilitat de les classificacions. Molts treballs publicats utilitzen diferents proves estadístiques, diferents guies d'interpretació del kappa, comparen variables diferents, manquen els intervals de confiança i les taules de contingència, i seleccionen mostra i observadors de manera diferent. És doncs evident que manca en la literatura una guia per realitzar aquests tipus d'estudis, ja que ara com ara cap dels treballs publicats és comparable ni aplicable a la població general.

8.CONCLUSIONS

8.1 Conclusions de la hipòtesi de treball

El biaix de prevalença del coeficient Kappa no contribueix de forma significativa en els baixos resultats de concordança interobservador i reproductibilitat intraobservador observats en valorar la classificació de Neer de fractures de l'extrem proximal de l'húmer.

8.2 Conclusions dels objectius secundaris

1. Els graus d'acord entre els observadors de l'estudi han estat pobres en la caracterització de fractures de l'húmer proximal amb radiologia simple i TAC.

2. Aportar més informació radiològica, no ajuda a millorar el grau d'acord entre observadors quan valorem característiques radiològiques ni en aplicar la classificació de Neer, però si en millora la reproductibilitat.

3. El grau d'experiència augmenta l'acord entre observadors de forma significativa en valorar algunes característiques radiològiques i en la classificació de Neer utilitzant radiologia simple i el tall axial de la TAC.

9.REFERÈNCIES BIBLIOGRÀFIQUES

1. Kristiansen B, Andersen UL, Olsen CA, Varmarken JE. The Neer classification of fractures of the proximal humerus. An assessment of interobserver variation. *Skeletal Radiol.* 1988;17(6):420-2.
2. Sidor ML, Zuckerman JD, Lyon T, Koval K, Schoenberg N. Classification of proximal humerus fractures: The contribution of the scapular lateral and axillary radiographs. *J Shoulder Elb Surg Am Shoulder Elb Surg Al.* 1994 Jan;3(1):24-7.
3. Sidor ML, Zuckerman JD, Lyon T, Koval K, Cuomo F, Schoenberg N. The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility. *J Bone Joint Surg Am.* 1993 Dec;75(12):1745-50.
4. Sjødén GO, Movin T, Aspelin P, Güntner P, Shalabi A. 3D-radiographic analysis does not improve the Neer and AO classifications of proximal humeral fractures. *Acta Orthop Scand.* 1999 Aug;70(4):325-8.
5. Sjødén GO, Movin T, Güntner P, Aspelin P, Ahrengart L, Ersmark H, et al. Poor reproducibility of classification of proximal humeral fractures. Additional CT of minor value. *Acta Orthop Scand.* 1997 Jun;68(3):239-42.
6. Bernstein J, Adler LM, Blank JE, Dalsey RM, Williams GR, Iannotti JP. Evaluation of the Neer system of classification of proximal humeral fractures with computerized tomographic scans and plain radiographs. *J Bone Joint Surg Am.* 1996 Sep;78(9):1371-5.

7. Siebenrock KA, Gerber C. The reproducibility of classification of fractures of the proximal end of the humerus. *J Bone Joint Surg Am*. 1993 Dec;75(12):1751–5.
8. Sallay PI, Pedowitz RA, Mallon WJ, Vandemark RM, Dalton JD, Speer KP. Reliability and reproducibility of radiographic interpretation of proximal humeral fracture pathoanatomy. *J Shoulder Elb Surg Am Shoulder Elb Surg Al*. 1997 Feb;6(1):60–9.
9. Brien H, Noftall F, MacMaster S, Cummings T, Landells C, Rockwood P. Neer’s classification system: a critical appraisal. *J Trauma*. 1995 Feb;38(2):257–60.
10. Brorson S, Hróbjartsson A. Training improves agreement among doctors using the Neer system for proximal humeral fractures in a systematic review. *J Clin Epidemiol*. 2008 Jan;61(1):7–16.
11. Brorson S, Bagger J, Sylvest A, Hróbjartsson A. Low agreement among 24 doctors using the Neer-classification; only moderate agreement on displacement, even between specialists. *Int Orthop*. 2002;26(5):271–3.
12. Brunner A, Honigmann P, Treumann T, Babst R. The impact of stereo-visualisation of three-dimensional CT datasets on the inter- and intraobserver reliability of the AO/OTA and Neer classifications in the assessment of fractures of the proximal humerus. *J Bone Joint Surg Br*. 2009 Jun;91(6):766–71.
13. Mahadeva D, Dias RG, Deshpande SV, Datta A, Dhillon SS, Simons AW. The reliability and reproducibility of the Neer classification system--digital radiography (PACS) improves agreement. *Injury*. 2011 Apr;42(4):339–42.
14. Majed A, Macleod I, Bull AMJ, Zyto K, Resch H, Hertel R, et al. Proximal humeral fracture classification systems revisited. *J*

- Shoulder Elb Surg Am Shoulder Elb Surg Al. 2011 Oct;20(7):1125–32.
15. Levine W BT, Bigliani LU. Fracturas proximales de húmero. Rockwood & Matsen The Shoulder. 3^a ed. New York: Wirth&Lippitt.; 2006. p. 355–411.
 16. Lanting B, MacDermid J, Drosdowech D, Faber KJ. Proximal humeral fractures: a systematic review of treatment modalities. J Shoulder Elb Surg Am Shoulder Elb Surg Al. 2008 Feb;17(1):42–54.
 17. Bengnér U, Johnell O, Redlund-Johnell I. Changes in the incidence of fracture of the upper end of the humerus during a 30-year period. A study of 2125 fractures. Clin Orthop. 1988 Jun;(231):179–82.
 18. Court-Brown CM, Garg A, McQueen MM. The epidemiology of proximal humeral fractures. Acta Orthop Scand. 2001 Aug;72(4):365–71.
 19. Kannus P, Palvanen M, Niemi S, Parkkari J, Järvinen M, Vuori I. Osteoporotic fractures of the proximal humerus in elderly Finnish persons: sharp increase in 1970-1998 and alarming projections for the new millennium. Acta Orthop Scand. 2000 Oct;71(5):465–70.
 20. Nordqvist A, Petersson CJ. Incidence and causes of shoulder girdle injuries in an urban population. J Shoulder Elb Surg Am Shoulder Elb Surg Al. 1995 Apr;4(2):107–12.
 21. Palvanen M, Kannus P, Niemi S, Parkkari J. Update in the epidemiology of proximal humeral fractures. Clin Orthop. 2006 Jan;442:87–92.

22. Johnell O, Kanis JA, Odén A, Sernbo I, Redlund-Johnell I, Petterson C, et al. Mortality after osteoporotic fractures. *Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA*. 2004 Jan;15(1):38–42.
23. Olsson C, Petersson CJ. Clinical importance of comorbidity in patients with a proximal humerus fracture. *Clin Orthop*. 2006 Jan;442:93–9.
24. Hettrich CM, Boraiah S, Dyke JP, Neviasser A, Helfet DL, Lorich DG. Quantitative assessment of the vascularity of the proximal part of the humerus. *J Bone Joint Surg Am*. 2010 Apr;92(4):943–8.
25. Flatow E. Fracturas de húmero proximal. *Rockwood and Green's Fractures in Adults*. 5^a ed. Philadelphia: Lippincott Williams & Wilkins; 2003. p. 997–1040.
26. Bono CM, Renard R, Levine RG, Levy AS. Effect of displacement of fractures of the greater tuberosity on the mechanics of the shoulder. *J Bone Joint Surg Br*. 2001 Sep;83(7):1056–62.
27. Frankle MA, Greenwald DP, Markee BA, Ondrovic LE, Lee WE 3rd. Biomechanical effects of malposition of tuberosity fragments on the humeral prosthetic reconstruction for four-part proximal humerus fractures. *J Shoulder Elb Surg Am Shoulder Elb Surg Al*. 2001 Aug;10(4):321–6.
28. Abu-Rajab RB, Stansfield BW, Nunn T, Nicol AC, Kelly IG. Re-attachment of the tuberosities of the humerus following hemiarthroplasty for four-part fracture. *J Bone Joint Surg Br*. 2006 Nov;88(11):1539–44.
29. Boileau P, Krishnan SG, Tinsi L, Walch G, Coste JS, Molé D. Tuberosity malposition and migration: reasons for poor outcomes after hemiarthroplasty for displaced fractures of the proximal

- humerus. *Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons* . [et al]. 2002;11:401–12.
30. Torrens C, Corrales M, Melendo E, Solano A, Rodríguez-Baeza A, Cáceres E. The pectoralis major tendon as a reference for restoring humeral length and retroversion with hemiarthroplasty for fracture. *J Shoulder Elb Surg Am Shoulder Elb Surg Al*. 2008 Dec;17(6):947–50.
 31. Parsons BO, Klepps SJ, Miller S, Bird J, Gladstone J, Flatow E. Reliability and reproducibility of radiographs of greater tuberosity displacement. A cadaveric study. *J Bone Joint Surg Am*. 2005 Jan;87(1):58–65.
 32. Burstein AH. Fracture classification systems: do they work and are they useful? *J Bone Joint Surg Am*. 1993 Dec;75(12):1743–4.
 33. Nielsen JO, Dons-Jensen H, Sørensen HT. Lauge-Hansen classification of malleolar fractures. An assessment of the reproducibility in 118 cases. *Acta Orthop Scand*. 1990 Oct;61(5):385–7.
 34. Thomsen NO, Overgaard S, Olsen LH, Hansen H, Nielsen ST. Observer variation in the radiographic classification of ankle fractures. *J Bone Joint Surg Br*. 1991 Jul;73(4):676–8.
 35. Andersen E, Jørgensen LG, Heddam LT. Evans' classification of trochanteric fractures: an assessment of the interobserver and intraobserver reliability. *Injury*. 1990 Nov;21(6):377–8.
 36. Frandsen PA, Andersen E, Madsen F, Skjødt T. Garden's classification of femoral neck fractures. An assessment of inter-observer variation. *J Bone Joint Surg Br*. 1988 Aug;70(4):588–90.

37. Kocher T. Beiträge zur Kenntnis einiger Praktisch Wichtiger Frakturformen. Basel: Carl Sallmann; 1896.
38. Codman EA CE. Rupture of the supraespinatus tendon and other lesions in or about the subacromial bursas. The Shoulder. G Miller& Co Medical Publishers; 1934. p. 262–331.
39. Watson-Jones R. Fracture of the neck of the humerus. Fractures and Other Bone and Joint Injuries. Baltimore: Williams and Wilkens; 1940. p. 289–97.
40. Neer, CS CS, Brown TH Jr, McLaughlin HL. Fracture of the neck of the humerus with dislocation of the head fragment. Am J Surg. 1953 Mar;85(3):252–8.
41. Neer CS 2nd. Displaced proximal humeral fractures. II. Treatment of three-part and four-part displacement. J Bone Joint Surg Am. 1970 Sep;52(6):1090–103.
42. Jakob RP, Miniaci A, Anson PS, Jaberg H, Osterwalder A, Ganz R. Four-part valgus impacted fractures of the proximal humerus. J Bone Joint Surg Br. 1991 Mar;73(2):295–8.
43. Muller ME. The Comprehensive Classification of Fractures of Long Bones. Springer-Verlag; 1990.
44. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159–74.
45. Brorson S, Eckardt H, Audigé L, Rolauffs B, Bahrs C. Translation between the Neer- and the AO/OTA-classification for proximal humeral fractures: do we need to be bilingual to interpret the scientific literature? BMC Res Notes. 2013;6:69.

46. Hertel R. Morpho-logic classification of fractures of the proximal humerus. A validated, teachable and practicable alternative. the 8th International Conference on Shoulder Surgery (ICSS); 2001.
47. Mora Guix JM, Gonzalez AS, Brugalla JV, Carril EC, Baños FG. Proposed protocol for reading images of humeral head fractures. *Clin Orthop*. 2006 Jul;448:225–33.
48. Cuny C, Baumann C, Mayer J, Guignand D, Irrazi M, Berrichi A, et al. AST classification of proximal humeral fractures: introduction and interobserver reliability assessment. *Eur J Orthop Surg Traumatol Orthopédie Traumatol*. 2013 Jan;23(1):35–40.
49. Carrerra E da F, Wajnsztein A, Lenza M, Netto NA. Reproducibility of three classifications of proximal humeral fractures. *Einstein São Paulo Braz*. 2012 Dec;10(4):473–9.
50. Hertel R, Hempfing A, Stiehler M, Leunig M. Predictors of humeral head ischemia after intracapsular fracture of the proximal humerus. *J Shoulder Elb Surg Am Shoulder Elb Surg Al*. 2004 Aug;13(4):427–33.
51. Sukthankar AV, Leonello DT, Hertel RW, Ding GS, Sandow MJ. A comprehensive classification of proximal humeral fractures: HGLS system. *J Shoulder Elb Surg Am Shoulder Elb Surg Al*. 2013 Jul;22(7):e1–6.
52. Mora Guix JM, Pedrós JS, Serrano AC. Updated classification system for proximal humeral fractures. *Clin Med Res*. 2009 Jun;7(1-2):32–44.
53. Brorson S, Bagger J, Sylvest A, Hrobjartsson A. Diagnosing displaced four-part fractures of the proximal humerus: a review of observer studies. *Int Orthop*. 2009 Apr;33(2):323–7.

54. Brorson S, Bagger J, Sylvest A, Hrøbjartsson A, Høbjartsson A. Improved interobserver variation after training of doctors in the Neer system. A randomised trial. *J Bone Joint Surg Br.* 2002 Sep;84(7):950-4.
55. Bruinsma WE, Guitton TG, Warner JJP, Ring D, Science of Variation Group. Interobserver reliability of classification and characterization of proximal humeral fractures: a comparison of two and three-dimensional CT. *J Bone Joint Surg Am.* 2013 Sep 4;95(17):1600-4.
56. Foroohar A, Tosti R, Richmond JM, Gaughan JP, Ilyas AM. Classification and treatment of proximal humerus fractures: inter-observer reliability and agreement across imaging modalities and experience. *J Orthop Surg.* 2011;6:38.
57. Ramappa AJ, Patel V, Goswami K, Zurakowski D, Yablon C, Rodriguez EK, et al. Using computed tomography to assess proximal humerus fractures. *Am J Orthop Belle Mead NJ.* 2014 Mar;43(3):E43-47.
58. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005 May;37(5):360-3.
59. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull.* 1968 Oct;70(4):213-20.
60. Fleiss JL. *Statistical methods for rates and proportions.* Ed 2. New York, John Wiley and Sons; 1981.
61. Nobo Komagata. Chance agreement and significance of the Kappa statistic.

62. Dunn G. Design and analysis of Reliability studies. The statistical evaluation of measurement errors. 1989.
63. Svanholm H, Starklint H, Gundersen HJ, Fabricius J, Barlebo H, Olsen S. Reproducibility of histomorphologic diagnoses with special reference to the kappa statistic. *APMIS Acta Pathol Microbiol Immunol Scand*. 1989 Aug;97(8):689–98.
64. Gjørup T. Reliability of diagnostic tests. *Acta Obstet Gynecol Scand Suppl*. 1997;166:9–14.
65. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol*. 1990;43(6):551–8.
66. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*. 1990;43(6):543–9.
67. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993 May;46(5):423–9.
68. Bhandari M, Matthys G, McKee MD, Evidence-Based Orthopaedic Trauma Working Group. Four part fractures of the proximal humerus. *J Orthop Trauma*. 2004 Feb;18(2):126–7.
69. Audigé L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. *Acta Orthop Scand*. 2004 Apr;75(2):184–94.
70. Shrader MW, Sanchez-Sotelo J, Sperling JW, Rowland CM, Cofield RH. Understanding proximal humerus fractures: image analysis, classification, and treatment. *J Shoulder Elb Surg Am Shoulder Elb Surg AI*. 2005 Oct;14(5):497–505.
71. Emam K. Benchmarking Kappa: Interrater agreement in software process assessment.

72. Brorson S, Olsen BS, Frich LH, Jensen SL, Sørensen AK, Krogsgaard M, et al. Surgeons agree more on treatment recommendations than on classification of proximal humeral fractures. *BMC Musculoskelet Disord*. 2012;13:114.
73. Petit CJ, Millett PJ, Endres NK, Diller D, Harris MB, Warner JJP. Management of proximal humeral fractures: surgeons don't agree. *J Shoulder Elb Surg Am Shoulder Elb Surg Al*. 2010 Apr;19(3):446–51.
74. Rasmussen S, Madsen PV, Bennicke K. Observer variation in the Lauge-Hansen classification of ankle fractures. Precision improved by instruction. *Acta Orthop Scand*. 1993 Dec;64(6):693–4.

