Nonsensus in the treatment of proximal humerus fractures: uncontrolled, blinded, comparative behavioural analysis between Homo chirurgicus accidentus and Macaca sylvanus

Sam Razaeian,1 Birgitt Wiese,2 Dafang Zhang,3 Affi Harb,1 Christian Krettek,1 Naël Hawi1

ABSTRACT

OBJECTIVE
To investigate the inter-rater reliability of Barbary macaques compared with an expert group of surgeons for the choice of treatment and predicted outcome of proximal humerus fractures.

DESIGN
Uncontrolled, blinded, comparative behavioural analysis.

SETTING
Germany and United States.

PARTICIPANTS
10 blinded experts in the field of orthopaedic trauma surgery (Homo chirurgicus accidentus), with special focus on upper extremity surgery from Germany and the US, and five Barbary macaques (Macaca sylvanus) from a semi-free range enclosure.

MAIN OUTCOME MEASURES
The reliability of agreement between raters assessed with Fleiss’ k.

RESULTS
Barbary macaques seem to have inferior inter-rater reliability in comparison with experts for choice of treatment (non-surgical v surgical), but for the geriatric age group most frequently affected by proximal humeral fractures, they performed similarly to the experts in their choices of treatment and choice of surgical procedure. Agreement about predicted outcome was poor among the macaques and slight among the experts. All experts almost always predicted the outcome incorrectly and tended to underestimate it. While only 4 (4.4%) of 90 experts’ predictions were correct, 13 (28.9%) of 45 macaques’ predictions were correct.

CONCLUSIONS
Consensus on treatment and expected outcomes of proximal humeral fractures is lacking even beyond the human species. Although Barbary macaques tend to predict the clinical outcome more accurately, their reliability to assist surgeons in making a consistent decision is limited. Future high quality research is needed to guide surgeons’ decision making on the optimal treatment of this common injury.

WHAT THIS STUDY ADDS
Consensus on the treatment and expected outcomes of proximal humeral fractures is lacking even beyond the human species
Future high quality research is needed to guide surgeon decision making on the optimal treatment of this common injury

WHAT IS ALREADY KNOWN ON THIS TOPIC
To date, no consensus has been reached on the optimal treatment of proximal humerus fractures
Increasing evidence suggests that non-operative management might have functional outcomes similar to those of operative management but with lower risks of complications and reoperation
Evidence based guidelines are lacking to inform decision making between different interventions, and expert consensus is considered to be poor

Introduction
Proximal humeral fractures are a common injury, representing about 6% of all adult fractures. Around 70% of these fractures occur in patients over the age of 60, with the greatest reported incidence among people aged 80 or older. The incidence of proximal humeral fractures has been increasing over the past few decades owing to an ageing population and the associated increase in osteoporosis and low energy falls from standing height. The incidence of proximal humeral fractures is about 60 per 100,000 people in the United States, but in the population aged 65 or older, the incidence is fourfold higher at 253 per 100,000 people. In Finland, the incidence of proximal humeral fractures tripled between 1970 and 2002 to 105 per 100,000 people aged 60 or above. Therefore, management of proximal humeral fractures will increase in healthcare systems.

Although it is well known that most proximal humeral fractures (nearly 75%) can be treated non-operatively with acceptable functional results, surgery became popular with advancements in the field of osteosynthetic implants, such as locking nails, plates, and prosthetic shoulder joint replacements. The rates of surgically treated patients are higher than 25% in some institutions, leading to substantial variation worldwide in the management of this common injury.

In a review of a large sample of US Medicare data, Bell and coauthors found a significant increase in the number of surgical procedures for proximal humeral fractures without a corresponding increase in their incidence in the study period, and moreover, with significant regional variation in the rates of surgery, ranging from 0% to 68%. This heterogeneity of treatment is in the setting of a lack of scientific consensus on the optimal treatment for these fractures. The latest Cochrane review suggests that non-operative management might have similar functional outcomes to operative management with lower risks of complications and reoperation, but there is insufficient evidence from current randomised controlled trials to inform decision making between different non-surgical, surgical, or rehabilitation interventions for these fractures.

But there is still hope. Deep in the Thuringian basin of Germany, between the mottled sandstone...
hills of Windleite and the shell limestone formations of Hainleite, surrounded by the murmuring sound of Wernröder stream, the Barbary macaques (Macaca sylvanus) live and still roam the vast beech forests of Germany in one of the biggest semi-free range enclosures in Europe. Besides humans, the Barbary macaques are the only free-living primates in Europe, and besides geriatric patients with proximal humerus fractures, one of the most endangered species in the world (fig 1).

As evidence based guidelines are lacking and expert consensus is considered to be poor, this species could be promising for future decision making processes owing to its impartiality and the ability to put itself into the same threatened position as patients with proximal humeral fractures. The aim of this behavioural analysis is to investigate inter-rater reliability of Barbary macaques in comparison with an expert group of surgeons for the choice of treatment and prediction of outcome of proximal humerus fractures, and to determine the extent of consensus on treatment of this common injury.

Methods
This behavioural analysis was carried out in accordance with the ethical standards of the 1964 Declaration of Helsinki as updated in 2004. No animal was forced to participate or was in any way misused, abused, or damaged. Some human beings might feel themselves so, however, after reading this analysis.

We identified independent experts in the field of orthopaedic trauma surgery with special focus on upper extremity surgery from Germany and the US and invited them by e-mail to participate in an anonymous web based survey (SoSci Survey GmbH, Munich, Germany). The survey consisted of nine case reports of acute proximal humerus fractures. The case presentations included radiographs and a reconstructed three dimensional computed tomography (CT) image, patient demographics, information about secondary illnesses, and general health state before the injury given in the form of the three level version of the EuroQol five-dimensional instrument (EQ-5D-3L; supplementary material). All cases were randomly selected from a prospective, observational registry study (Hannover Humerus Registry, NCT03060876). Two independent study nurses evaluated all cases with a clinical and radiological follow-up of 12 months.

The Hannover Humerus Registry is a prospective, CT based, single centre registry study of a supraregional level 1 trauma centre, aiming at investigating the healing process of proximal humerus and humeral shaft fractures. We informed all the experts about the intention of this analysis. They were blinded only to the actual treatment procedure and outcome. Besides details of memberships and professional working experience of the experts, their response to the following questions with corresponding answer options was evaluated:

1. Which treatment regimen would you recommend? Answer: Non-operative or operative.
2. Which procedure would you recommend, if you had to treat surgically? Answer: Locking plate, cement augmented locking plate, intramedullary nail, hemiarthroplasty, reverse shoulder arthroplasty, allograft augmented locking plate, or something else.
3. Which outcome (Constant score adapted for age and sex) would you expect one year after conservative treatment? Answer: less than or equal to 59, 60-69, 70-79, 80-89, or 90-100, out of 100 points.

Similarly, the behaviour of Barbary macaques was evaluated for the same nine cases and questions in one of the biggest semi-free range enclosures in Europe (Affenwald Straußberg, Sonderhausen, Thuringia, Germany) during the winter season in January 2020. The general public were excluded to guarantee the anonymity of participating macaques. The web based case presentations were printed as 29.7×42 cm coloured posters and positioned serially with the related and aforementioned three questions using a customer stopper from a local ice cream vendor in the enclosure (fig 2). With the aid of internationally accepted and validated rating scales, consisting of disposable, cellulose kidney dishes and laminated pictograms, the behaviour of the macaques was observed (fig 2). Each kidney dish represented one of the aforementioned possible responses. Equal doses of Mediterranean sultanas and peanuts (Nutwork GmbH, Hamburg, Germany) and Californian walnuts (Märsch Importhandel GmbH, Ulm, Germany) functioned as

Fig 1 | Conservation status of proximal humeral fractures and Barbary macaques according to the International Union for Conservation of Nature. The figure shows a minimally displaced proximal humerus fracture according to the most commonly used Neer classification. A 70 year old woman was treated with an intramedullary nail at a German trauma centre in 2019, probably owing to its biomechanical superiority over extramedullary implants. After only three months, conversion to reverse shoulder arthroplasty was performed, probably owing to its biomechanical superiority over intramedullary nails.

doi: 10.1136/bmj.m4429 | BMJ 2020;371:m4429 | the bmj
environmental enrichment and were placed in the kidney dishes. The first grasp into a kidney dish was defined as a treatment or outcome selection, and this behaviour was noted. For question number 2, any non-responding among the macaques was defined as the response option “something else.” Apart from that, macaques that did not respond completely to all cases and those with apparently severe conflict of interests were excluded from evaluation (fig 3).

As this behavioural analysis was to be carried out voluntarily by the macaques in their familiar enclosure under uncontrolled conditions, calculation of the number of complete responds was not possible in advance. Therefore, it was necessary to begin with the analysis of the macaques followed by analysis of the experts in order to arrive at approximately equally sized groups. For this reason, the web based survey was closed to the experts after a comparable number of responds were obtained.

**Statistical analysis**

To assess the reliability of agreement between raters, Fleiss’ $\kappa$ was determined. We used the benchmark scale developed by Landis and Koch to interpret the strength of agreement for Fleiss’ $\kappa$ values as indicated in figure 4.16 For the analyses, SPSS 25 (IBM, Armonk, NY) was used.

**Species analysed**

*Macaca sylvanus*

*M sylvanus* (fig 5), also known as Barbary macaque or colloquially called maggot, is the only native species of primate to live in Europe, and the only macaque species found outside Asia.11 The species can live in a wide range of habitats, but prefers high altitude forests, and is also found in coastal scrub and on rocky slopes. As the Barbary macaques’ habitat is threatened by human activity, their habitat availability and population have decreased considerably in recent decades.11 Thus macaques are listed as endangered by the IUCN (International Union for Conservation of Nature) red list of threatened species.11

Barbary macaques are sociable, living in mixed sex groups, which can vary in size from about 13 to 80 individuals. Both sexes have their own hierarchies, while female Barbary macaques form strictly matrilineal hierarchies.17-19 Their diet consists mainly of plant food as leaves, fruits, and seeds, but insects and mushrooms are also preferred.17 19

*Homo chirurgicus accidentus*

*H chirurgicus accidentus* (fig 5), also known as orthopaedic trauma surgeon or colloquially called the ox,20 is a species of surgeon unique for its wide distribution in the world. While it can live in a wide range of habitats, it is frequently encountered at bigger healthcare centres in urban regions.

Sometimes there are adversarial tensions between this species and others, such as anaesthetists, anaesthesia nurses, anaesthesia nurse assistants, and some orthopaedic trauma surgeons recommending non-operative treatments, but these territorial conflicts are usually solved without physical harm and instead on an intellectual level.20 Its sociocultural competences and mating behaviours are unknown owing to its extreme work ethos. Both are currently subjects of intensive research.

<table>
<thead>
<tr>
<th>Fleiss’ $\kappa$</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤0</td>
<td>Poor agreement</td>
</tr>
<tr>
<td>0.01-0.20</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>0.21-0.40</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>0.41-0.60</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>0.61-0.80</td>
<td>Substantial agreement</td>
</tr>
<tr>
<td>0.81-1.00</td>
<td>Almost perfect agreement</td>
</tr>
</tbody>
</table>

Fig 2 | Case report presented on a customer stopper of a local ice cream vendor in the enclosure, and condition of the validated rating scales for questions 2 and 3 regarding treatment procedure and expected outcomes after the analysis. A two piece rating scale in analogous fashion for question 1 and its two response options (non-operative or operative) is not shown as it could not be secured in intact condition out of the macaques’ hands and was lost to follow-up.

Fig 3 | A senior macaque with apparently severe conflicts of interest is biasing one of its inferior subjects.

Fig 4 | Benchmark scale according to Landis and Koch for interpretation of strength of agreement, assessed with Fleiss’ $\kappa$ values.
H chirurgicus accidentus is nocturnal and therefore moody during most of the daytime, forming into groups of equally moody men, who are patriarchal, with their hierarchy determined by direct lineage or personal favour of the lead man. Its diet consists mainly of bone fractures, damaged cartilage, infected soft tissues, and broken prosthetics, but profitable, elective outpatient operations are also preferred.

**Patient and public involvement**

Patients’ clinical and radiographic records were used from an observational registry study (Hannover Humerus Registry, NCT03060876) for the survey, and we thank them for their records. We did not involve patients in the design or analysis of the data.

**Results**

Ten independent experts in the field of orthopaedic trauma surgery with special focus on upper extremity surgery from Germany and the US were available for the survey. The responder rate of the experts in the US was higher than in Germany (5/10 vs 5/20). Only five of 22 macaques provided complete responses to all cases, probably fearing loss of reputation. Reactions of non-responders among the experts ranged from a diplomatic German “funny idea, but I am out, sry” to a warm hearted American “it’s wild what my alma mater and former colleagues are studying nowadays.”

Figure 6 provides details of the experts’ professional qualifications. Professional qualifications or memberships of the macaques remained uncertain, but all were obviously fellowship trained in picking one’s nose and delousing each other.

Among the experts, operative treatment was the preferred treatment (51 of 90 selections), but the macaques more often chose non-operative treatment (25 of 45 selections). Overall inter-rater agreement for this choice was moderate among the experts and poor among the macaques, although there were marked differences between the two different nations (fig 7, fig 8, and fig 9).

In a post hoc subgroup analysis of the cases by patient age, the inter-rater agreement of the experts for choice of treatment and of surgical procedure was as poor as that of the macaques for patients aged over 65, and only slight for patients aged 65 or under (fig 7 and fig 10). Once again there were marked differences between the nations.

While the US experts achieved unanimous agreement with respect to non-operative treatment for patients over the age of 65, German experts reached only poor agreement, with four (26.7%) of 15 tending to choose surgical treatment (fig 7).

Of the nine presented cases, all patients were actually treated non-operatively, with an excellent clinical outcome. Agreement about prediction of outcome was poor among the macaques and slight among the experts (fig 11). All experts almost always predicted the outcome incorrectly and tended to underestimate it. While only 4 (4.4%) of 90 experts’ predictions were correct, 13 (28.9%) of 45 macaques’ predictions were correct (fig 12). Figure 9 gives details of experts’ and macaques’ selections of treatment and preferred surgical procedure.

**Discussion**

**Principal findings**

This study investigates inter-rater reliability of Barbary macaques in comparison with an expert group for management and prediction of clinical outcome of proximal humerus fractures. Barbary macaques seem to have inferior inter-rater reliability compared with the experts for choice of treatment (non-surgical vs surgical), but they performed similarly to the experts...
findings highlight the continuing controversy and lack of expert consensus on the optimal treatment of these fractures even outside the human species.4 6 21

Findings in the light of national trends and evidence
Surgical treatment of proximal humeral fractures has been associated with complication rates as high as 49% and reoperation rates of 14%. Growing evidence from randomised controlled trials and meta-analyses showing similar outcomes between surgical and non-surgical management of proximal humeral fractures has called into question surgical treatment of such fractures for patients older than 65.5 6 21 22 In addition to prospective trials, pooled data of previous studies in a recent Cochrane review showed no clinically important difference in functional outcomes and quality of life between surgical and non-surgical treatment of proximal humerus fractures at one to two year follow-up.8 16

Nonetheless, surgical treatment of this injury has been increasingly used over the past two decades.5 Ironically, it was for the relevant age group of patients over the age of 65 that inter-rater agreement across the two species was equally poor.

The marked differences between the two nations should be considered in the context of published national treatment trends. According to an analysis of the national inpatient database, the percentage of surgically treated proximal humeral fractures increased by 6% between 2004 and 2012 in the US, but nonetheless non-operative treatment remained the most common treatment modality in 59% of patients.23 Conversely, according to a recent trend analysis of German Federal Statistical Office data, surgical procedures increased by 39% in Germany between 2007 and 2016, with about 68.9% of all procedures being performed in elderly patients.24

Locking plate fixation was the most commonly used procedure within all age groups, although it has already been identified as an independent risk factor for inpatient adverse events and mortality in patients.
operative treatment, assessed with Fleiss’ \( \kappa \) Fig 11 | Inter-rater reliability of the analysed species for outcome prediction of non-operative treatment, assessed with Fleiss’ \( \kappa \)

<table>
<thead>
<tr>
<th>Species</th>
<th>All proximal humerus fractures</th>
<th>Age ( \leq 65 )</th>
<th>Age ( &gt; 65 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>All experts</td>
<td>0.09</td>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>United States</td>
<td>0.15</td>
<td>0.19</td>
<td>-0.03</td>
</tr>
<tr>
<td>Germany</td>
<td>0.02</td>
<td>0.08</td>
<td>-0.13</td>
</tr>
<tr>
<td>Barbary macaques</td>
<td>-0.10</td>
<td>-0.16</td>
<td>-0.12</td>
</tr>
</tbody>
</table>

Fig 10 | Inter-rater reliability of the analysed species for recommended surgical procedure, assessed with Fleiss’ \( \kappa \)

older than 65 compared with non-operative inpatient treatment.\(^9\)\(^24\) This risk is even more concerning from a health economic view, as previous epidemiological and cost analyses have shown that fractures of the shoulder contribute substantially to the rising treatment costs for upper limb fractures.\(^3\)

### Study limitations

Our study has some limitations that should be considered. Although it is a promising observation that the macaques chose non-operative treatment more often than the experts, their agreement about optimal treatment was consistently poor. A systematic confounding behaviour was unfortunately seen during the whole study. Some senior primates with apparently severe conflicts of interest biased responders during their selections (fig 3). We believe that this might have adversely affected the results of these responders, and that their agreement and their outcome prediction ability would be much better without this disruptive factor. As this behavioural analysis was to be carried out voluntarily by the macaques in their familiar enclosure under uncontrolled conditions, any attempt to prevent or minimise this occurrence was omitted. We chose winter for this analysis to avoid general public access and to guarantee the anonymity of participating macaques; this choice might have been poor, however, as the authors did not know that conflicts of interest among Barbary macaques are a seasonal affair beginning in November and lasting until March.\(^25\)

Self-reported conflicts of interest are also common in orthopaedic trauma surgeons, and it is known that they can influence reported outcomes.\(^26\) The conspicuous finding that all experts almost always underestimated and predicted the outcome of non-operatively treated proximal humeral fractures incorrectly suggests the possibility of interference among the allegedly independent experts. However, it remains unclear whether, to what extent, and how the experts examined here were affected, as the self-disclosure referred only to details about scientific memberships and professional working experience.

Furthermore, the smaller number of only five macaques compared with 10 experts should be considered as a limitation when interpreting overall inter-rater agreement of the two species.

In addition, in retrospect, the mixture of Mediterranean sultanas, peanuts, and Californian walnuts as environmental enrichment was an unfavourable choice by the authors. Unfortunately, significant differences in the popularity of these treats could be observed in the aforementioned order. This difference led in parts to dependent selections, when the kidney dishes were not refilled equally immediately.

This form of selection bias must be seen as a major methodological weakness. The authors recommend Californian walnuts as single treats for future behavioural analysis.

### Table 1: Actual outcome in points versus predicted outcome in points

<table>
<thead>
<tr>
<th>Case</th>
<th>Age</th>
<th>Actual outcome in points</th>
<th>Predicted outcome in points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>52</td>
<td>87</td>
<td>8</td>
</tr>
<tr>
<td>Case 2</td>
<td>44</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>Case 3</td>
<td>55</td>
<td>93</td>
<td>4</td>
</tr>
<tr>
<td>Case 4</td>
<td>60</td>
<td>100</td>
<td>3</td>
</tr>
<tr>
<td>Case 5</td>
<td>62</td>
<td>83</td>
<td>5</td>
</tr>
<tr>
<td>Case 6</td>
<td>77</td>
<td>100</td>
<td>2</td>
</tr>
<tr>
<td>Case 7</td>
<td>60</td>
<td>85</td>
<td>9</td>
</tr>
<tr>
<td>Case 8</td>
<td>86</td>
<td>94</td>
<td>3</td>
</tr>
<tr>
<td>Case 9</td>
<td>80</td>
<td>74</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig 12 | Number of predictions for Constant score adapted for age and sex and actual outcome after one year of non-operative treatment according to question number 3. *All patients were treated non-operatively. The score is given in points out of 100 possible points. Selections of the experts and macaques are given in blue and red digits, respectively

Blue digits

Red digits
Conclusion

Consensus on treatment and expected outcomes of proximal humeral fractures is lacking even beyond the human species. Although Barbary macaques tend to predict the clinical outcome more accurately, their reliability to assist surgeons in making a consistent decision is limited. Future high quality research is needed to guide surgeon decision making on the optimal treatment of this common injury.

We thank Silvio Dietzel, park ranger of Affenwald Straubling (Sonnderhausen, Thuringia, Germany) and supervisor of this analysis, for his support enabling this study and for his non-commercial funding in the form of Californian walnuts.

Contributors: SR designed, conducted, and wrote this analysis. BW performed the statistical analysis. DZ, NH, and CK identified and contacted experts for the survey. AH assisted with the behavioural analysis and edited the manuscript. SR is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no other meetings the criteria have been omitted.

Funding: No funding has been obtained.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Patient consent: The manuscript includes images or information that might identify a patient. Signed consents have been obtained.

Ethical approval: This analysis was carried out in accordance with the ethical standards of the 1964 Declaration of Helsinki as updated in 2004. No animal was forced to participate or was in any way misused, abused, or damaged. The behavioural analysis of the Barbary macaques in their familiar enclosure under uncontrolled conditions was carried out under supervision of a responsible park ranger. The local ethical committee of Hannover Medical School was asked to deliver an opinion on this behavioural analysis, but it did not assume responsibility for satirical analysis such as this (supplementary material).

Data sharing: Complete dataset is available from the corresponding author at sam.r@hotmail.de.

The manuscript’s guarantor affirms that the manuscript is an honest, accurate, and transparent account of the study being reported, that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Dissemination to participants and related patient and public communities: In order to disseminate the results to a scientific community, it is planned to submit this topic for the Science-Slam Event at the Annual German Congress of Orthopaedics and Trauma Surgery (DKOUL) in October 2021. Furthermore, if the article is reviewed, it is planned to provide a direct link to the study with a synopsis containing background information about the scientific key problem, and our motivation as study conductor, on our institutional website (https://www.mhh-unfallchirurgie.de/) and on the website of the involved enclosure (https://www.affenwald.info/) in order to disseminate the results and message in a transparent, and easily understandable way for patients and non-specialists.

Provenance and peer review: Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

18 Prud’Homme J, Chaisip B. Aggressive interventions and matirnline dominance relations in semifree-ranging barbary macaques (Macaca sylvan). Primates 1993;34:271-83. doi:10.1007/BF02382621

Web appendix: Supplementary material