HAL
open science

Fiabilité de l'interprétation radiographique dans la prise en charge en urgence des traumatismes fermés récents du poignet

Jean-Maxime Alet

To cite this version:

Jean-Maxime Alet. Fiabilité de l'interprétation radiographique dans la prise en charge en urgence des traumatismes fermés récents du poignet. Médecine humaine et pathologie. 2016. dumas-01341766

HAL Id: dumas-01341766

https://dumas.ccsd.cnrs.fr/dumas-01341766

Submitted on 4 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

université de BORDEAUX

U.F.R. DES SCIENCES MÉDICALES

Année 2016 Thèse n°3033

Thèse pour l'obtention du

DIPLOME d'ÉTAT de DOCTEUR EN MÉDECINE

Présentée et soutenue publiquement
le 15 juin 2016

Par Jean-Maxime Alet
né le 13/06/1985 à Toulouse

Fiabilité de l'interprétation radiographique dans la prise en charge en urgence des traumatismes fermés récents du poignet

---

Directeur de thèse
Professeur Louis-Rachid Salmi

Jury

Pr Philippe Pelissier	Président
Pr Vincent Casoli	
Pr Vincent Pinsolle	
Dr Caroline Leclercq	
Pr Louis-Rachid Salmi	
Dr Benjamin Sommier	Rapporteur

A Christelle,

Je ne suis plus qu'à moitié dans le vrai monde depuis que je te connais,

Tout ce que tu apportes dans ma vie ne tient sur aucune ligne.

Ensemble nous irons partout autour du monde,

Et le reste n'a pas d'importance.

REMERCIEMENTS

À notre maître et président du jury,

Monsieur le Professeur Philippe PELISSIER
Professeur des Universités - Praticien Hospitalier
Service de chirurgie plastique - main - brûlés
CHU de Bordeaux

Vous me faites l'honneur de présider ce jury de thèse et je vous en remercie.

Vous avez guidé mes premiers pas d'interne avec la générosité qui est la vôtre, en me transmettant la passion de votre spécialité.

Votre approche de la pédagogie, de la psychologie et de la relation humaine sont pour moi un modèle. J'apprécie le calme que vous savez instaurer autour de vous.

Pour toutes ces raisons et bien d'autres, soyez assuré de ma gratitude et de mon profond respect.

À notre directeur de thèse,

Monsieur le Professeur Louis-Rachid Salmi
Professeur des Universités - Praticien Hospitalier
Service d'information médicale
Directeur de l'Institut de Santé Publique, d'Epidémiologie et de Développement
CHU de Bordeaux

Vous m'avez fait l'honneur de diriger ce travail et je vous en remercie.

Vous m'avez accueilli avec bienveillance dans un milieu qui peut sembler hostile à un chirurgien. J'ai pu bénéficier de la grande qualité de votre encadrement, qui m'a permis d'assimiler toutes ces notions nouvelles.

Votre sens du détail et votre capacité à prendre du recul sont exemplaires. J'espère que notre collaboration pourra perdurer dans le futur.

Soyez assuré de ma gratitude et de mon profond respect.

À nos juges et maîtres,

Monsieur le Professeur Vincent CASOLI
Professeur des Universités - Praticien Hospitalier
Service de chirurgie plastique - main - brûlés
CHU de Bordeaux

Vous me faites l'honneur de juger ce travail et je vous en remercie.

J'admire votre esprit d'équipe et votre capacité de délégation qui vous permet de garder le contrôle sur toutes les activités tout en nous responsabilisant.

Votre accessibilité, votre forte présence et votre encadrement permanent font partie des clés majeures de la qualité de la formation dans le service.

Soyez assuré de ma gratitude et de mon profond respect.

Monsieur le Professeur Vincent PINSOLLE
Professeur des Universités - Praticien Hospitalier
Service de chirurgie plastique - main - brûlés
CHU de Bordeaux

Vous me faites l'honneur de juger ce travail et je vous en remercie.

Votre vision juste et paternelle de l'enseignement de la chirurgie, fondée sur la confiance, force le respect de tous ceux qui travaillent avec vous.

J'admire votre capacité à exprimer vos jugements sans détours, qu'ils soient mauvais ou bons, mais toujours justifiés et toujours pour nous pousser en avant. Cette franchise est la clé du respect que vous portent tous vos élèves.

Soyez assuré de ma gratitude et de mon profond respect.

Madame le Docteur Caroline LECLERCQ
Chirurgien
Institut de la Main
Clinique Jouvenet, Paris

Vous me faites l'honneur de vous déplacer pour juger ce travail et je vous en remercie.

Je suis très heureux d'avoir pu passer ces quelques semaines à votre contact, malgré les difficultés d'organisation.

Cette courte période était suffisante pour susciter mon admiration de vos qualités chirurgicales, votre grand souci du détail, la finesse de vos indications, et votre capacité à pouvoir justifier chacun de vos gestes.

J'espère avoir de nouveau l'occasion de me former à vos côtés.

Soyez assurée de ma gratitude et de mon profond respect.

À notre rapporteur,

Monsieur le Docteur Benjamin SOMMIER
Chirurgien, Praticien Attaché des Hôpitaux
Hôpital Privé Saint Martin, Pessac
Service de chirurgie plastique - main - brûlés
CHU de Bordeaux

Merci d'avoir pris le temps de lire cette thèse en intégralité.

Ta présence dans ce jury de thèse était une évidence. Pas besoin d'en faire trop, tu sais l'estime que j'accorde à tes conseils, professionnels et moins professionnels. Ton influence dans ma formation a été majeure, et il y aura beaucoup de toi dans le chirurgien que je deviendrai.

Je suis fier et heureux que notre entente ait dépassé les frontières de l'hôpital et de pouvoir te compter parmi mes amis.

Sois assuré de ma gratitude, de mon profond respect, et qu'il nous reste encore beaucoup de donuts à partager en pleurant nos cheveux disparus.

A MES MAITRES D'INTERNAT

Chirurgie plastique : Dr Jean-Claude Castède ; Dr Marie-Laure Abi-Chahla (grande prêtresse de la chirurgie de la main bordelaise. C'est en grande partie à toi que je dois mon intérêt pour la chirurgie de la main. C'est un vrai plaisir d'être formé par toi) ; Dr Hussein Choughri ; Dr Cici Isacu ; Dr Romain Weigert ; Dr Eva Gachie

Chirurgie vasculaire : Pr Dominique Midy ; Pr Eric Ducasse ; Dr Xavier Bérard ; Dr Gérard Sassoust ; Dr Vincent Dubuisson (le chirurgien hors plastique qui m'aura le plus marqué de toute ma formation, le monde serait plus drôle et meilleur avec plus de gens comme toi) ; Dr Jeremy Chayet ; Dr Stéphane Mayeux ; Dr Emilie Charmoille

Chirurgie maxillo-faciale : Pr Claire Majoufre ; Pr Francis Siberchicot ; Pr Philippe Caix ; Dr Matthieu Laurentjoye ; Dr Anne-Sophie Ricard ; Dr Matthieu Bondaz ; Dr Olivier Nau

Chirurgie orthopédique : Dr Eric Hammel ; Dr Henri Duraffour ; Dr André Seguette ; Dr Gilles Mouret ; Dr Igor Benezis ; Dr Maxime Laporte

Chirurgiens de l'Institut de la Main :
Dr Alain Gilbert ; Dr Christophe Mathoulin ; Dr Alexandre Kilinc ; Dr Francesco Brunelli ; Dr Eric Lenoble ; Dr Philippe Sauzières ; Dr Eric Roulot ; Dr Mathilde Gras

A MA FAMILLE

Maman, malgré toutes tes maladies graves imaginaires je t'aime et je t'admire d'avoir élevé seule quatre enfants qui ne t'ont pas ménagée, et d'être arrivée à en faire malgré tout des personnes (presque) adaptées.
Bernard, votre maîtrise des subtilités de l'art de la gestion de maman ne peut que susciter l'admiration.

Nadine et Guy, vous m'avez accueilli avec tellement de simplicité et de naturel, c'est un vrai plaisir de faire partie de votre famille.

Laeti, le roc de la famille, la survie de maman à ses trois garçons a reposé sur toi, et on a tous eus beaucoup de chance d'avoir une grande sœur comme toi.
Romane, c'est dans ce livre que se trouve la recette secrète de la soupe aux tomates extra-terrestres. Tonton sera toujours là pour répondre à tes questions sur le sens de la vie.
Annah, comme tu sais pas lire, tu auras une dédicace dans mon prochain roman.
Steph, merci d'avoir apporté un peu de sport aux Alet qui clairement sont déficitaires dans le domaine. Garde ton calme légendaire qui donne l'impression que les choses glissent sur toi, et ta détermination sans failles à aller au bout de tes ambitions. Coach !

Louis, une version de moi en différent. Ils sont bien loin les boudignes à la chenille, mais même lorsqu'on ne s'est pas parlés depuis un moment (oui, ca arrive souvent, vu qu'on a tous les deux des problèmes à ce niveau-là !), notre connexion est naturelle. Même si ca m'énerve que les gens continuent de penser que tu es mon grand frère parce que tu as plus de barbe, je regarderai toujours avec l'admiration d'un frère ta capacité à ne pas te compromettre.

Pitou, les mystères de ta psychologie ne me font pas oublier que tu es mon grand frère.
Gabrielle et Jef, vous cumulez : trop jeunes pour lire, trop américains pour comprendre le français, et trop hawaiiens pour venir en France. Ca ne m'empêche pas de penser à vous...

Tatamooss, la version la plus pure et directe des gènes Legeay, et c'est ca qu'on aime chez toi.
Anne-Marie, ton calme et ta douceur sont si reposants !
Cathy, la seule à avoir déjà vécu ce rituel de passage. Ta présence compte beaucoup pour moi.

A MES AMIS

Quentin, tes multiples facettes n'ont pas fini de me déstabiliser. Le lundi à l'opéra, le vendredi à Maître Gim's (et tous les autres jours d'astreinte). Sous les airs de it-girl capricieuse que tu t'amuses à te donner parfois, ta générosité est incroyable, et je n'oublierai jamais l'énergie que tu as su mettre dans les moments importants (et même si tu as déclenché une pelade à Zahia). Ta présence dans nos vies est indispensable.
Réré, mon admiration pour toi est éternelle pour ta capacité à gérer la bête. J'envie vos projets incessants de voyage et j'apprécie ta discrétion qui rend ta connaissance lente et subtile, mais fiable et véritable.

Ahlem, ou comment la voisine chelou devient l'amie de toute confiance. Une confiance née d'une fiabilité à toute épreuve : tu es le genre d'amie à qui on peut prêter une méthode Assimil sans arrières pensées, et pas le genre à créer des rumeurs sur l'origine des odeurs de la moquette de tes voisins de chambre. La connexion qu'on a tous les deux est rare parce qu'elle n'a aucun sens et que notre façon de communiquer n'a aucun équivalent. Stay the same, je m'ennuierais tellement sans ton second degré.

JTP, une si belle personne,
votre beauté partout rayonne,
Kaley ne sait pas ce qu'elle perd,
Pour cette thèse et en amour vous êtes un expert.

La famille toulousaine : Rami, notre amitié est restée forte malgré ton départ à l'étranger. Je voulais écrire une envolée lyrique sur ton inébranlable bonne humeur et toutes tes incroyables qualités, puis j'ai réalisé que tu ne lirais jamais cette dédicace vu que tu as assouvi ta soif de vengeance de mon absence à ta thèse.
Boobz, toutes ces années d'externat et surtout pré-externat à philosopher sur le rôle du chiffre 51 dans ta vie et tous nos souvenirs de guerre restent des souvenirs géniaux. Mais toi non plus tu ne liras jamais ce passage.
Mema, ambassadrice de la famille toulousaine. Cette discussion sur les dead sixties assis devant l'entrée des artistes du Gipsy's Paradise me semble être à des années-lumière. On n'est pas d'accord sur tout, et pourtant on peut parler de tout, c'est si rare.

Coline, ton légendaire phlegme est une référence absolue pour moi. Je suis heureux que notre amitié ait donné tort à ton esprit rationaliste de cette fin de stage 2012 (" c'est le moment des promesses de ne pas perdre contact qu'on ne tiendra jamais "). Ta capacité à prendre les choses avec pragmatisme et simplicité m'apporte de la sérénité. Et si seulement je pouvais un jour manipuler comme toi...
Pierre, qui voudra toujours nous faire croire qu'il est écolo mais n'économise ni les barbecues ni les bonnes bouteilles.

Sarra, l'imprévue, apparue mystérieusement au coin d'un Master 2, il aura suffi de quelques blagues stupides pour comprendre que l'année serait bien différente de ce qui était prévu. Ta complexité, ton ouverture d'esprit et ta capacité à t'amuser de tout (notamment de mon arabe imaginaire et de mes côtés autistiques) font de toi une personne rare. Je retourne ta dédicace de thèse, je n'écrirai pas mieux.

Charlie, qui eut cru que notre première discussion (ou pour être exact ton écoute de mon monologue) nous mènerait jusqu'ici. Dans le fond, c'est bien que tu aies écouté mon premier conseil, qui a ouvert la porte à l'entrée de PO dans ta vie (même si ca a impliqué le foot et toutes ces choses bizarres dont tu parles maintenant). Continue ton voyage exploratoire capillaire, il finira bien par te mener quelque part.
PO, la ligue des justiciers de la vérité n'a jamais été aussi forte que quand on en a été les membres les plus illustres.

Sandra et Isma. Wodzi amegbetɔwo katã ablɔɖeviwoe eye wodzena bubu kple gomekpɔkpɔ sɔsɔe. Susu kple dzitsinya le wo dometɔ ɖesiaɖe si eyata wodze be woanɔ anyi le ɖekawɔwɔ blibo me. Je vous aime.

Angie. Pyramide. En trois: Poney, moto, Tillikum. Chaque fois que je te vois c'est une nouvelle surprise, et à chaque fois ca fait tellement plaisir. Tu es juste magique.

Marlène, partie si loin et trop vite. Imprévisible, mystérieuse, et si généreuse. Et surtout la meilleure des chief of staff.

Erwann, une entrée fracassante sur la scène poitevine. Pourvu que ton succès dépasse les frontières et que vous obteniez enfin le visa retour.

Momo, la petite a bien grandi, papa est très fier de ce que tu es devenue. Je sais que les brûlés te manquent, mais promis on s'arrangera quand on sera chefs tous les deux, tu auras ta part du gâteau.

Claire et Micka, par la grâce de Dieu Duc et Duchesse du Languedoc. Que le Royaume vous protège.

Léa, Simon, Marie, Myriam et Manu, la bande infernale, les barbeuk (imaginaires) ne font que commencer, Macklemore peut resurgir à tout moment.

Gillouth et Nico, vous êtes partis bien loin mais on pense toujours à vous en mangeant des sushis et des canelés.

Charlène et Cédric, exilés volontaires dans le monde de la banlieue bordelaise, et pourtant si fiers représentants de la haute société bordelaise. Le bar à whiskies portera votre nom.

Marize, reine des marseillais, princesse de la perfusette, difficile de savoir si je t'ai invitée ou non ou à cette thèse, nos rapports invitationnels sont incompréhensibles.

La famille plasticienne

Elias, je me rappellerai toujours la découverte de ce grand barbu dans les vestiaires du bloc mon premier jour d'internat, que j'ai pris pour le chef de service. J'ai vite apprécié ta compagnie et l'incroyable simplicité des rapports avec toi. J'admire ta capacité à voir le meilleur dans les gens que tu rencontres. Je n'oublierai pas ce que tu as fait pour moi.
Charlotte, continue de nous faire rêver avec tes plats, et ramène ton esprit toulousain à Bordeaux !

Audrey, mummy, mère des plasticiens. Toujours imprévisible, aussi bien capable de rire d'une situation difficile que de devenir toute rouge pour un détail qui te déplait, et c'est toujours aussi génial. Ton rire inimitable nous manque, il est temps que tu reviennes. Et n'oublie pas, pour ceux qui font du mal à un dauphin...

Pivert, que de chemin parcouru depuis *Piverococcus fæcalis* pour devenir Monsieur Lepivert, respecté depuis Miami jusqu'à Roquetaillade. Gros Pivert est devenu Grand Pivert. On le dira jamais assez : " On t'aime Pivert ".
Marie, bravo, car disons-le j'ai tenté de caser Pivert tout le début de mon internat sans succès, et tu es arrivée sans prévenir (et sans mon intervention).

Gaël, mon premier chef de clinique. A tes côtés j'ai appris mes premiers gestes, mais c'est surtout ta "philosophie" (guillemets adaptés tu me l'accorderas) qui ne m'a plus jamais quitté. Grâce à toi j'ai franchi des étapes capitales de ma vie : l'achat d'un costard, et, d'ici 2 ans toutes mes chaussettes seront par paires.

A MES COINTERNES

Une spéciale dédicace à l'équipe novembre 2015 : Paillette (qu'une carrière meilleure et méritée attend à Broadway)– Pallu – Pimax – Leco – Ahmed - Charlotte et bien sûr...

La bergère. Abusé la dédicace, je n'oublierai jamais la première fois où tu m'as relayé sur une garde un samedi matin et où je me suis inquiété toute la journée de ta capacité à gérer 24 heures (" t'inquiète tu connais pas la bergère "). Le début d'une grande série ! Ca va me faire bizarre de plus t'avoir comme cointerne vu qu'on aura passé la moitié de notre formation ensemble. Mais bon, comme tu dirais " voilà, quoi ".
Menez: on a pas vraiment l'occasion de se le dire, sache que ces deux semestres de ta chefferie ont été un plaisir, et que je suis content que ca continue.
Fleury, le front, reste sur ta grande lancée le feu d'artifice commence bientôt !
La bonne héronne, le Jack Johnson de la plastique, à l'aise en toute situation.
Dannepond, encore jeune mais plein de promesses.
Chloé, le futur de la plastique, et la preuve qu'on est pas des racistes on prend même les chtis.

Abad (mmmmh) – Prince Khaled – Tatar (et sa fameuse loi) – Dan – Bartolo – Geoff-X –Maillot – Bonte (la grande tante du Nord) – Camille - Anaïs Chataia – Pierre Moullot - Thicoipe (le bouquineur) - Ce bon MOF et sa gentillesse qui le perdra – Aurélie – Koenig - Alice – Julie - Camille Landric – Carrier – La fifette - Les italiens : Serena (la Maserati des internes italiens) – Enrico - Antonio

AUX EQUIPES DES SERVICES ET DU BLOC

J'en oublie sûrement, mais on a de la chance de vous avoir, car c'est un vrai plaisir de travailler avec vous.
La famille du bloc : Cathy, Elsa, Christophe, Pascal, Olivia, Elo, Aude, Alex, Ursula, Géraldine, Evelyne, Laure, Mina (tu travailles où en ce moment ?), Viviane, Farida, et sans oublier Laeti et Sylvie.
Toute l'équipe du A1 Thibaud, Isabelle, Charlotte, Nathanaëlle, Laurianne, Muriane, Céline, Constance, Judith, Brigitte, Patrick, Dominique, Mme Biasini; des soins externes Bruno, Raphaëlle, Cédrine, Zohra, Aline, Séverine, Isabelle, Sophie, Elisabeth, et sans oublier les kinés (Jef, Pierre et c'est comment déjà ? Marine ?), les brûlés et Cathy la reine du miel.

TABLE DES MATIERES

RESUME **11**
INTRODUCTION **12**
Justification de l'étude 12
Définitions 14
Contexte 15
Hypothèse et objectifs 15
METHODES **16**
Chronologie de l'étude 16
Population d'étude 17
Radiographies et patients 17
Base de données 17
Classement des lésions 19
Reproductibilité inter-observateurs 19
Relecteurs des radiographies 19
Séances de relecture radiographique 21
Obtention du consensus d'experts 23
Reproductibilité intra-observateur 24
Déterminants de la fiabilité 24
Analyse statistique 26
Description 26
Tests de comparaison 26
Concordance inter-observateurs 26
Classement binaire « lésion ou doute » ou « pas de lésion » 26
Classement en familles de diagnostics 28
Concordance intra-observateur 29
Déterminants de la fiabilité 29
Nombre de sujets nécessaires 29
Considérations éthiques 30
Structuration du texte 30
RESULTATS **31**
Patients inclus et données manquantes 31
Reproductibilité inter-observateurs 31
Construction du consensus d'experts 31
Classement des malades 32
Classement binaire « lésion ou doute » ou « pas de lésion » 32
Classement en familles de diagnostics 34
Concordance entre médecins des urgences et consensus d'experts 35
Classement binaire « lésion ou doute » ou « pas de lésion » 35
Classement en familles de diagnostics 36
Concordance entre les relecteurs 37
Concordance entre les experts 37
Concordance entre experts et référent 37
Reproductibilité intra-observateur 38
Déterminants de la fiabilité 38
Caractéristiques du patient et du traumatisme 38
Caractéristiques des relecteurs 40
Réunion entre les relecteurs 42
DISCUSSION **44**
Principaux résultats 44
Validité interne de l'estimation du coefficient κ 46
Limites liées à l'absence de méthode de référence 48
Autres limites principales 49
Comparaisons avec les données de la littérature 53
Implications des résultats 57
REFERENCES **61**
ANNEXES **69**
SERMENT D'HIPPOCRATE **80**

RESUME

Contexte : Les traumatismes du poignet sont fréquents ; les diagnostics des lésions qui en découlent reposent sur la radiographie standard, mais les erreurs diagnostiques sont fréquentes. L'objectif de cette étude était d'évaluer la fiabilité de cet examen pour caractériser les traumatismes fermés récents du poignet.
Méthodes : Nous avons réalisé une étude de la concordance entre le diagnostic posé par plusieurs observateurs d'une même radiographie (concordance inter-observateurs : entre les médecins des urgences et un consensus d'experts puis entre les experts entre eux) ; et entre les diagnostics du même observateur d'une même radiographie à un an d'intervalle (concordance intra-observateur). Les déterminants de la fiabilité (associations entre les caractéristiques du patient, de son traumatisme ou du relecteur) ont également été recherchés. Les patients ayant consulté au Centre Hospitalier Universitaire de Bordeaux en 2014 pour un traumatisme fermé de moins de sept jours ont été inclus rétrospectivement. L'étude de la concordance a été faite avec le coefficient κ de Cohen et son intervalle de confiance à 95 % (IC95 %).
Résultats : Cent-soixante-quatorze radiographies ont été relues. Le coefficient κ pour le classement binaire « lésion ou doute sur une lésion » contre « pas de lésion », entre médecins des urgences et consensus d'experts était de 0,50 (IC95 % : 0,37 - 0,62). Le coefficient κ pour la concordance au sein du groupe d'experts pour ce classement était de 0,47 (IC95 % : 0,34 - 0,60). Le coefficient κ pour la concordance intra-observateur était de 0,12 (IC95 % 0,01 – 0,23). Il n'a pas été mis en évidence de caractéristiques du patient ou du traumatisme associées à la présence d'une discordance.
Conclusion : La fiabilité de la radiographie standard du poignet n'était pas satisfaisante, quels que soient les observateurs comparés et également en comparant les observateurs à eux-mêmes. Cet examen ne peut être recommandé seul dans ces traumatismes.

Abstract

Introduction: Blunt wrist traumas are frequent. The diagnosis of related injuries relies on standard radiographs, but injuries are frequently misdiagnosed. The objective of this study was to assess the reliability of standard radiographs to characterize blunt wrist traumas.
Methods: This was a study of rater agreement between the diagnoses of different observers of the same radiograph (inter-rater agreement: between Emergency Room physicians and a consensus of experts, and afterwards between the experts); and between the diagnoses made by the same observer of the same radiograph after a one-year period (intra-rater agreement). Factors associated with the reliability were sought (characteristics of the patients, the trauma, or the rater). Patients who were managed in the ER of the teaching hospital of Bordeaux for a recent (less than seven days) blunt wrist trauma during 2014 were retrospectively included. The raters' agreement was assessed with Cohen's κ coefficient and its 95% confidence interval (95%CI).
Results: One hundred and seventy-four radiographs were included in the study. The κ coefficient for the binary rating "injury or doubt on injury" versus "no injury" between ER physicians and experts consensus was 0.50 (95%CI: 0.37 - 0.62). The κ coefficient for the agreement between the two experts who rated the same radiograph was 0.47 (95%CI: 0.34 - 0.60). The κ coefficient for the intra-rater agreement was 0.12 (95%CI 0.01 – 0.23). No association between disagreement and characteristics of the patient or his trauma reached statistical significance.
Conclusion: The reliability of standard radiographs does not seem satisfying, whoever are the observers, and even when observers are compared to themselves. This test should not be used alone to diagnose these injuries.
Mots-clés : Poignet, Traumatisme, Concordance, Kappa, Radiographies
Key Words : Wrist, Trauma, Reliability, Kappa, Radiographs

INTRODUCTION

Justification de l'étude

Les traumatismes fermés du poignet représentent un motif fréquent de consultation dans les services d'urgence. Leur incidence aux États-Unis était évaluée à 181 pour 100 000 habitants (1) en 2012, dont 64 pour les entorses simples du poignet. Une étude européenne (2) évaluait ce taux entre 280 et 440 pour 100 000 habitants, uniquement pour les fractures du radius distal ; et à 29 pour 100 000 habitants pour les fractures du scaphoïde (3).

Lors de ces traumatismes, deux entités nosologiques sont au premier plan (4), d'une part les lésions ligamentaires (entorses), d'autre part les fractures. On distingue classiquement les entorses simples ou bénignes, de cicatrisation spontanée, et les entorses graves en cas de déchirure ligamentaire complète, aux possibilités de récupération faibles en l'absence de prise en charge adaptée (5). Les fractures quant à elles intéressent principalement deux os : le radius et le scaphoïde.

La prise en charge initiale standard (6–8) de ces traumatismes consiste en un examen clinique du poignet, ainsi qu'en la réalisation de clichés radiographiques (poignet de face et de profil). Hormis les patients présentant une fracture ou une lésion ligamentaire patente, la majorité des malades, pour lesquels l'examen clinique est peu spécifique et les radiographies interprétées comme normales, rentre à domicile avec le diagnostic d'entorse simple du poignet. La prise en charge consiste alors en un traitement antalgique, le repos articulaire quelques jours et parfois une immobilisation temporaire.

Cependant, les différentes lésions potentielles dans cette région sont de diagnostic difficile. En effet, l'anatomie du poignet est complexe (9,10). Dix os de structure spatiale irrégulière s'articulent entre eux, reliés par de très nombreux ligaments. De plus, le motif de consultation, la douleur, rend l'examen clinique difficile voire impossible et peu spécifique. La radiographie est donc au centre du diagnostic, en ce qu'elle permet la visualisation directe des lésions et n'est théoriquement pas contrainte par la douleur. Mais son interprétation est réputée difficile, y compris par des opérateurs expérimentés. Les deux lésions principales, les fractures et les déchirures ligamentaires, sont particulièrement difficiles à visualiser (Annexe 1). La petite taille des os rend les fractures peu visibles, et ce d'autant que la position du poignet lors de l'examen modifie leur configuration spatiale. Ceci est particulièrement vrai pour le scaphoïde, qui passe d'une position horizontale à une position

verticale selon que le poignet est en flexion / extension ou en inclinaison radiale / ulnaire. Par ailleurs, les lésions ligamentaires ne sont pas visibles directement mais seulement par des arguments indirects (modification des espaces interosseux ou des angles entre les os du carpe) subtils, subjectifs, et parfois décelable seulement lors de procédures dynamiques. De plus, dans les services d'urgence souvent surchargés et devant hiérarchiser les temps de consultation selon la gravité des situations, ces traumatismes, qui ont la réputation d'être bénins, sont parfois traités hâtivement.

En conséquence, la fréquence des erreurs diagnostiques dans la région du poignet est plus importante que pour toutes les autres articulations (11). On retrouve ainsi dans la littérature des proportions notables (jusqu'à 80 % (11–15)) de patients initialement identifiés « entorse simple du poignet », mais présentant en réalité des lésions plus graves.

Ces erreurs diagnostiques ne sont pas sans conséquences. En l'absence de traitement, un certain nombre de ces lésions évoluent progressivement vers des complications, au premier rang desquelles se trouvent l'instabilité du poignet et l'arthrose post-traumatique (8). On individualise dans ce dernier cas plus particulièrement les lésions (5) dites " SLAC-wrist " (*Scapho-Lunate Advanced* Collapse), faisant suite aux ruptures du ligament scapho-lunaire non traitées, et les lésions dites " SNAC-wrist " (*Scaphoid Nonunion Advanced Collapse*), conséquences des pseudarthroses de fractures du scaphoïde. Cette évolution défavorable peut se faire de façon silencieuse. Le handicap qui en découle peut être majeur, *a fortiori* car ces traumatismes touchent en majorité des sujets jeunes, actifs et souvent travailleurs manuels (8). Par ailleurs, il est probable que les répercussions liées à l'errance diagnostique, aux arrêts de travail prolongés (155 jours en moyenne dans les traumatismes du carpe (16)), à la longue rééducation et au reclassement professionnel soient notables d’un point de vue médico-économique.

Plusieurs solutions ont été proposées pour diminuer l'incidence de ces erreurs. Elles peuvent être classées en quatre groupes :

- ✓ Certains auteurs réalisent une immobilisation systématique d'une dizaine de jours, suivie d'une réévaluation à l'ablation du dispositif (6). Deux limites rendent cette solution non consensuelle : un risque élevé de perdre de vue les malades et l'absence de consensus pour la prise en charge à l'ablation du dispositif (nouvelles radiographies ? examen plus sensible ? avis spécialisé ?), ce qui revient à repousser de quinze jours la problématique initiale ;
- ✓ D'autres suggèrent de réaliser une Imagerie par Résonance Magnétique (IRM) ou une tomodensitométrie chez tous les patients (13), examens dont la sensibilité et la spécificité sont bien supérieures aux radiographies (15). Devant la fréquence de ces traumatismes, il n'est néanmoins pas envisageable de proposer une stratégie d'examens complémentaires lourds chez tous les malades. Ceci pour des raisons logistiques, économiques mais également éthiques, car la tomodensitométrie est un examen irradiant ;
- ✓ On peut également proposer un avis d'expert pour toutes les radiographies, par un radiologue ou un chirurgien ;
- ✓ Enfin, la dernière solution consisterait à la mise en place d'une règle de décision, sur le modèle des critères dits d'Ottowa (17), utilisés pour la prise en charge diagnostique des traumatismes de la cheville, pour décider dès la prise en charge initiale si le malade relève ou non d'un examen complémentaire plus sensible.

Ces solutions se fondent toutes sur le même constat : l'interprétation des radiographies, alors qu'elle est au cœur du processus décisionnel, n'est peut-être pas fiable. Paradoxalement, il n'existe à notre connaissance pas d'étude portant sur cette fiabilité.

Définitions

La fiabilité d'un examen diagnostique se définit comme sa capacité à donner le même résultat (ici l'interprétation de la radiographie) dans des conditions d'utilisation différentes (ici par des observateurs différents). Les termes « reproductibilité » et « fiabilité » sont synonymes et donc utilisés de manière équivalente dans cette étude. La concordance entre des observations est la méthode d'étude de la fiabilité.

La fiabilité est complémentaire de l'autre indicateur de qualité d'un outil diagnostique : la validité (capacité de l'examen à donner un résultat juste), qui est mesurée notamment par la sensibilité et la spécificité.

Contexte

Ce travail s'inscrit dans le cadre d'une étude préliminaire à la réalisation d'une cohorte prospective, inspirée du modèle de Stiell (18) utilisé pour le développement des critères d'Ottawa.

Hypothèse et objectifs

L'hypothèse de cette étude est que l'interprétation des radiographies standard du poignet n'est pas fiable dans les traumatismes fermés récents.

L'objectif principal de cette étude était d'évaluer la fiabilité inter-observateurs de l'interprétation des radiographies standard pratiquées aux urgences pour les traumatismes fermés récents du poignet, par une étude de concordance entre d'une part le classement binaire « lésion ou doute sur une lésion » contre « pas de lésion » posé aux urgences, et, d'autre part, un consensus d'experts.

Notre étude comportait les objectifs secondaires suivants :

- Estimer la fiabilité inter-observateurs entre médecins des urgences et consensus d'experts pour les lésions classées en familles de diagnostics (fractures de l'avant-bras, fractures des os du carpe, lésions ligamentaires, doute, entorse simple) d'une part et au sein du groupe d'experts d'autre part (à la fois pour le classement binaire et le classement en familles de diagnostics) ;
- Estimer la fiabilité intra-observateur ;
- Décrire les caractéristiques du patient, du traumatisme ou des relecteurs, associées à la présence d'une discordance d'interprétation ;
- Étudier les éléments radiographiques ou cliniques déterminant le choix des relecteurs pour leur interprétation, pour tenter de mettre en évidence les principaux points de consensus ou de faiblesse dans la démarche diagnostique.

METHODES

Il s'agissait d'une étude de fiabilité diagnostique par étude de la concordance entre interprétations par différents observateurs de radiographies standard du poignet aux Urgences adultes du Centre Hospitalier Universitaire (CHU) de Bordeaux, pour les traumatismes du poignet fermés et récents (délai depuis le traumatisme inférieur ou égal à sept jours) chez les patients ayant consulté pour ce motif durant l'année 2014, à l'exception des fractures extra-articulaires de l'extrémité inférieure du radius.

Trois aspects de la fiabilité ont été étudiés :

- Reproductibilité inter-observateurs : Correspond à la reproductibilité de l'interprétation radiographique entre deux observateurs différents. Dans cette étude, ont été étudiées les concordances :
  - ✓ Entre les médecins des urgences et un consensus d'experts ;
  - ✓ Entre les experts eux-mêmes.
- Reproductibilité intra-observateur : Correspond à la reproductibilité de l'interprétation de la même radiographie, par le même observateur, après un intervalle de temps suffisant pour éviter toute mémorisation.
- Déterminants de la fiabilité : Correspond à l'étude des variables pouvant expliquer des discordances entre interprétations.

Chronologie de l'étude

Cette étude a été réalisée en trois phases. La première phase était construite pour l'étude de la reproductibilité inter-observateurs. Cette partie a été réalisée entre les mois de février et juin 2015. La seconde phase consistait en l'étude de la reproductibilité intra-observateur, avec une nouvelle relecture des radiographies. La troisième phase consistait en la réunion des relecteurs, pour mettre en commun les éléments de réflexion sur les radiographies. Ces deux dernières phases ont été réalisées entre février et mai 2016.

Population d'étude

Radiographies et patients

Pour être incluses, les radiographies devaient concerner des patients âgés d'au moins 18 ans à leur admission, ayant été pris en charge entre le 1er janvier et le 31 décembre 2014 pour un traumatisme fermé du poignet datant de moins de sept jours et dont le diagnostic de sortie était celui d'entorse simple du poignet, d'une lésion, ou de doute nécessitant une prise en charge plus poussée.

L'absence de radiographie dans le dossier médical ou la présence d'une fracture extra-articulaire du radius distal constituaient des critères d'exclusion. Dans les cas où un patient présentait un traumatisme bilatéral, les radiographies des deux poignets étaient incluses séparément.

Les radiographies ont toutes été réalisées sur le même appareil de radiographie du service des urgences du CHU de Bordeaux. Il s'agit d'un Capteur Plan Kodak DR 9500 mis en service en juin 2008. Les clichés étaient réalisés par un manipulateur en radiologie. Toutes les radiographies comprenaient une incidence postéro-antérieure de face et une incidence de profil.

Les notions « pas de lésion (radiographique) » et « entorse simple (du poignet) » sont utilisées de façon synonyme dans cette étude. Ces deux notions sont l'expression respectivement radiographique et clinique de la même entité nosologique dans ce contexte traumatique.

Base de données

Les données ont été recueillies à l'aide de l'outil Business Objects du logiciel DXCare (2013, Medasys SA) du CHU de Bordeaux, qui permet de lister les patients admis en fonction de leur service d'accueil et des codes du Programme de Médicalisation des Systèmes d'Information (PMSI) associés à la sortie du patient. Ceux-ci s'appuient sur la Classification Internationale des Maladies dans sa dixième version (CIM-10). Les patients admis durant la période d'étude dans le service cité et dont le code PMSI de sortie correspondait aux entités des traumatismes fermés du poignet ont ainsi été réunis dans la base de données. Cette extraction a été faite par un praticien hospitalier des Urgences habitué à l'utilisation du logiciel. L'Annexe 2 présente les codes PMSI/CIM-10 utilisés pour cette étude.

Toutes les données ont été enregistrées dans un tableau Excel (2010, Microsoft Corporation) sécurisé par mot de passe. Les informations d'identité n'étaient pas enregistrées. Les patients étaient identifiés par leur numéro d'admission et un numéro d'anonymat propre à l'étude. Les informations sociodémographiques et les données de l'examen clinique étaient réunies et classées de la façon suivante :

- Séjour
  - Quantitatif : Numéro d'admission CHU, Numéro d'anonymat, Heure d'admission ;
  - Date : Date d'admission ;
- Anamnèse
  - Quantitatif : Âge au traumatisme (années), Délai depuis le traumatisme (jours), Évaluation numérique de la douleur ;
  - Binaire : Antécédent traumatique au poignet homolatéral, Sexe, Profession manuelle, Latéralité, Traumatisme de la main dominante, Type de chute (avec ou sans cinétique; exemple de chute sans cinétique associée : patient butant contre un trottoir ; exemple de chute avec cinétique : chute en patinant), Accident de sport, Craquement lors du traumatisme, Intervalle libre entre le traumatisme et les symptômes, Prise d'antalgiques, Localisation de la douleur spontanée, Irradiation de la douleur ;
  - Catégoriel à plus de deux modalités : Mécanisme, Hauteur en cas de chute, Position du poignet lors du traumatisme ;
  - Texte : Antécédents ;
- Examen clinique
  - Binaire : Attitude des traumatisés du membre supérieur, Impotence fonctionnelle (Limitation des amplitudes articulaires), Œdème du poignet, Déformation du poignet, Ecchymose(s), Hématome(s), Dermabrasion(s), Paresthésies, Troubles sensitifs autres que paresthésies ;
  - Catégoriel à plus de deux modalités : Axe de mobilité déclenchant la douleur, Site de douleur exquise à la palpation ;
- Radiographies
  - Quantitatif : Numéro de radiographie (propre à l'étude) ;
  - Binaire : Présence de radiographie dans le dossier ;
- Prise en charge
  - Quantitatif : Durée de prise en charge en cas d'entorse simple (jours) ;
  - Catégoriel à plus de deux modalités : Prise en charge en cas d'entorse simple, Attitude en cas de doute, Diagnostic final.

La définition retenue pour le « Diagnostic final » des urgences était la suivante : diagnostic présenté dans le dossier médical comme le dernier posé par le médecin des urgences :

- Soit à la sortie du malade des urgences ;
- Soit avant le transfert dans une unité spécialisée, que cela soit pour une prise en charge ou pour un avis ;
- Soit avant la demande d'un avis spécialisé aux Urgences (à l'exception de l'interne de radiologie des urgences).

Les définitions et la transformation des variables sont présentées en Annexe 3.

Classement des lésions

La reproductibilité a été étudiée avec deux classements différents. Dans un premier temps, les interprétations ont été classées de façon binaire en deux groupes : « lésion ou doute » et « pas de lésion » (c'est-à-dire entorse simple). Ce choix binaire correspond à une réalité pratique : les patients du groupe « lésion ou doute » voient leur prise en charge continuer, même si le diagnostic n'est pas définitif, même s'il est erroné : ils ne sortent donc pas du système de soins. En revanche, les patients du groupe « pas de lésion » retournent à domicile avec un traitement antalgique et sont donc perdus de vue. C'est dans ce dernier groupe que se trouvent les malades avec une potentielle erreur diagnostique qui risquent d'évoluer péjorativement.

Dans un second temps, les lésions ont été regroupées en cinq familles de diagnostics pour les analyses secondaires : « entorse simple », « doute », « lésion ligamentaire », « fracture d'un ou plusieurs os du carpe », « fracture d'un ou des os de l'avant-bras ».

Reproductibilité inter-observateurs

L'estimation de la reproductibilité inter-observateurs a été faite par l'étude de la concordance entre le diagnostic fait par les médecins des urgences à la prise en charge initiale et un consensus d'experts (objectif principal) ; et également entre le diagnostic fait par les différents relecteurs d'une même radiographie (concordance entre les relecteurs, objectif secondaire).

Relecteurs des radiographies

Les radiographies ont été relues en insu par un groupe d'experts constitué par des radiologues et des chirurgiens spécialistes du membre supérieur. Pour analyser la variabilité potentielle des observations, deux experts de chaque spécialité et de niveaux d'expérience différents ont été sollicités. Il y avait ainsi deux radiologues : un chef de clinique-assistant (CCA) et un praticien hospitalier (PH) ; et deux chirurgiens : un CCA et un PH. Leur

expérience en radiographie traumatique du poignet était respectivement de sept ans pour les CCA et douze ans pour les PH.

Chaque radiographie était revue par deux experts dans des conditions indépendantes (séances de relecture séparées). Les experts étaient aussi en insu de l'interprétation faite par les praticiens des urgences. Toutes les paires possibles (combinaisons entre spécialités et expériences) ont été utilisées, soit six paires au total. Chacune des paires possibles était dotée du même nombre d'observations, c'est-à-dire un sixième du nombre total de radiographies (Tableau 1).

Tableau 1. Répartition des radiographies à relire pour chaque combinaison d'observateurs ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

Expert	**Paire 1**	**Paire 2**	**Paire 3**	**Paire 4**	**Paire 5**	**Paire 6**	**Total**
Radiologue CCA	X	X	X				*n*/2
Radiologue PH	X			X	X		*n*/2
Chirurgien CCA		X		X		X	*n*/2
Chirurgien PH			X		X	X	*n*/2
Total	*2 x n*/6	*2 x n*/6	*2 x n*/6	*2 x n*/6	*2 x n*/6	*2 x n*/6	*2 x n*

Chaque expert était assigné à trois paires, pour être combiné à tous les autres experts.
Toutes les radiographies d'une paire étaient relues par les deux experts de la paire.
n = nombre total de radiographies incluses.

Un cinquième relecteur, désigné comme « référent » dans cette étude, professeur des universités - praticien hospitalier (PU-PH) en chirurgie de la main et du poignet a été consulté (vingt-cinq ans d'expérience en radiographie traumatique du poignet). Le référent a relu l'intégralité des radiographies, en insu des interprétations des autres experts et des praticiens des urgences, au cours d'une séance indépendante.

Dans la suite de l'étude, l'expression « les experts » désigne les quatre relecteurs experts et n'inclut pas le référent. L'expression « consensus d'experts » désigne le résultat du consensus à la fois entre les experts et le référent (cf. paragraphe correspondant). L'ensemble constitué par les quatre experts plus le référent est désigné par l'expression « les relecteurs ».

Séances de relecture radiographique

Cette relecture a été faite au cours de séances séparées standardisées, à raison d'une par relecteur. Les relecteurs ne savaient pas à quelles paires ils appartenaient, ni dans quel ordre les radiographies étaient affichées. Les radiographies étaient affichées par moi-même sur le logiciel utilisé aux Urgences du CHU de Bordeaux, pour contrôler que le relecteur n'avait accès ni au compte-rendu de radiologie, ni aux autres examens réalisés. L'utilisation du logiciel des Urgences permettait de garantir la comparabilité et autorisait les relecteurs à affiner leur interprétation à l'aide des outils informatiques (zoom, variation de contraste, mesures de distances et d'angles). Ces outils sont les mêmes que ceux dont disposent les praticiens aux Urgences.

L'ordre de lecture était déterminé en amont, par génération informatique de listes au hasard. Une nouvelle liste des radiographies contenant tous les numéros d'anonymat, mais suivant un ordre aléatoire, a été générée. Puis cette liste a été découpée en six parts égales, soit une par paire. Chacune des six listes a ensuite été redistribuée deux fois aléatoirement, de telle façon que chacun des experts de chaque paire relise les mêmes radiographies que l'autre, mais dans un ordre différent. Douze listes au total ont donc été générées. Le processus de génération aléatoire est détaillé dans l'Annexe 4. Ces listes ont été créées à l'aide de la fonction *sample* du logiciel R. Pour la relecture par le référent, une autre liste aléatoire a été générée, comprenant l'ensemble des radiographies.

Pour assurer une comparabilité optimale dans la prise de décision, les experts disposaient des éléments cliniques renseignés dans les dossiers médicaux (19). Ils étaient affichés au relecteur sous la forme d'une diapositive par patient, au sein d'un diaporama PowerPoint (version 14, Microsoft Corporation). Chaque patient était présenté de façon standardisée (Figure 1). En haut de la diapositive figuraient le sexe et l'âge, puis l'ensemble des éléments cliniques en rapport avec le traumatisme qui figuraient dans le dossier. Pour éviter des erreurs de classement, le numéro d'anonymat du malade était signalé sur la diapositive.

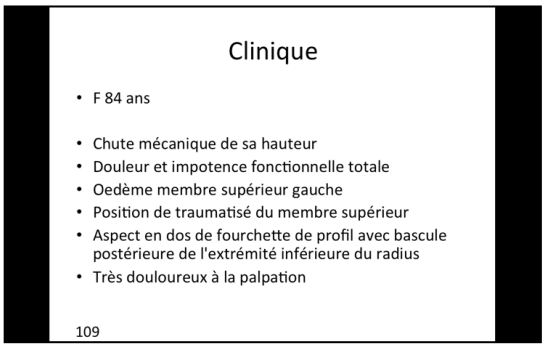


Figure 1. Exemple de diapositive standardisée pour la présentation des éléments cliniques au cours des séances de relecture ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

À l'exception de la correction des fautes d'orthographe (dans un souci de lisibilité) et des abréviations qui étaient remplacées par leur signification, aucune modification n'était apportée. Pour éviter que les diapositives ne soient trop chargées, les notions d'état cutané, de vascularisation et d'examen sensitif ont été enlevées, sauf dans le cas où une anomalie était signalée. Les relecteurs étaient informés que, sauf indication contraire, l'ensemble des malades présentait un traumatisme fermé, sans trouble vasculaire ni neurologique.

Je me suis chargé du défilement du diaporama au fur et à mesure de l'affichage des radiographies. J'ai également procédé à l'enregistrement des différentes variables résumant les conclusions du relecteur.

Une variable « Diagnostic » enregistrait le diagnostic du relecteur. Ce diagnostic pouvait être « doute ». Les experts étaient informés au début des séances qu'il leur était possible de ne pas poser un diagnostic précis et pouvaient indiquer qu'ils avaient un doute. Le doute pouvait être évoqué soit sur critères radiographiques, soit pour une discordance entre les éléments cliniques et radiographiques.

La qualité de la radiographie était évaluée par chaque relecteur. Nous n'avons pas trouvé dans la littérature d'échelle validée pour l'évaluation de la qualité des radiographies du poignet. Il existe un certain nombre de critères objectifs (20–22) de qualité, mais dont le recueil est lourd (mesures d'angles notamment). Le bénéfice d'une échelle non validée, créée pour l'étude à partir de cette dizaine de critères ne nous semblait pas contrebalancer la lourdeur logistique qu'elle impliquait. Une variable de substitution a donc été utilisée, catégorielle à trois modalités : « 1 : Non interprétable », « 2 : Mauvaise qualité mais radiographie interprétable » et « 3 : Bonne qualité ».

Pour ne pas influencer le relecteur, toutes les questions étaient posées de façon ouverte et identique pour chaque radiographie. Les relecteurs n'étaient pas informés des classements de leur diagnostic (binaire et en cinq familles) qui étaient faits par la suite. En revanche, l'échelle d'évaluation de la qualité leur était détaillée. La grille de recueil est détaillée dans le Tableau 2.

Tableau 2. Données recueillies lors des séances de relecture ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

Variable	Type*	Modalités	Question posée
Qualité radiographie	C	- Non interprétable - Mauvaise qualité mais interprétable - Bonne qualité	Quel score donnez-vous à la qualité de la radiographie ?
Interprétation	B	- Lésion ou doute - Pas de lésion	Pouvez-vous tirer une interprétation de cette radiographie et si oui laquelle ? (Question commune avec les variables « Conclusion » et « Famille »)
Conclusion	C	- Ensemble des diagnostics possibles	
Famille (variable créée à partir de la réduction à cinq catégories de la variable « Conclusion »)	C	- Entorse simple - Doute - Lésion ligamentaire - Fracture du carpe - Fracture de l'avant-bras	

*C: Catégoriel à plus de deux modalités. B : Binaire.

Obtention du consensus d'experts

L'objectif principal de l'étude a nécessité l'obtention d'un consensus d'experts. Ce consensus se faisait sur le classement binaire « lésion ou doute » contre « pas de lésion ». Il était obtenu de la façon suivante (Figure 2) :

- Soit les deux experts classaient de façon concordante : leur diagnostic était alors retenu comme consensus ;
- Soit les deux experts étaient discordants : c'est alors le diagnostic choisi par le référent qui était retenu comme consensus.

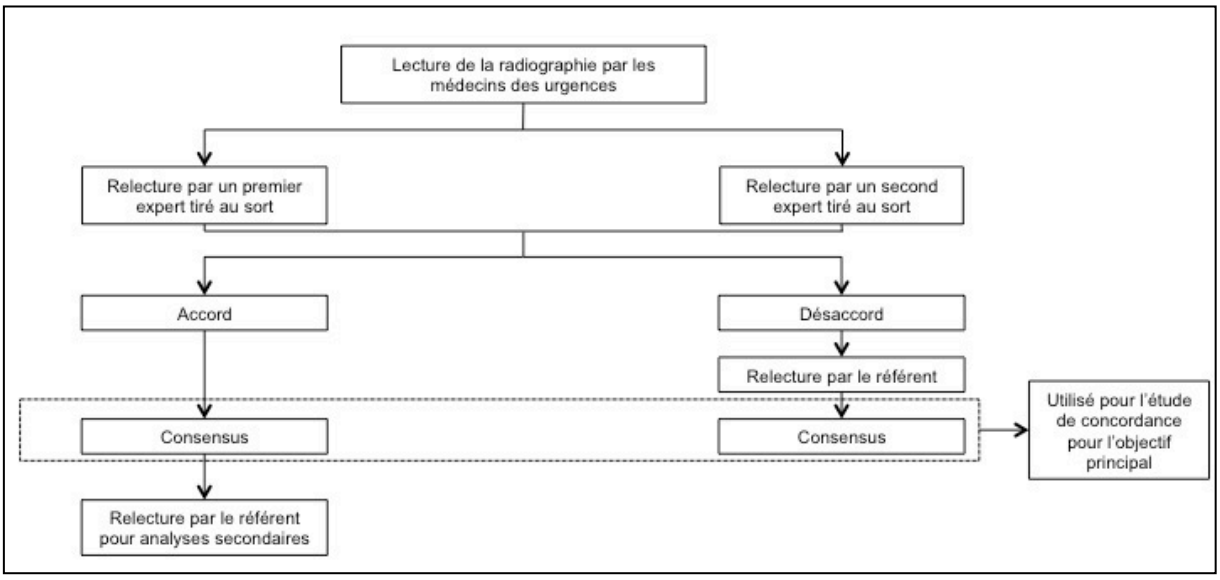


Figure 2. Réalisation du consensus d'expert ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

Reproductibilité intra-observateur

La reproductibilité intra-observateur a été étudiée par une seconde interprétation des mêmes radiographies par les experts. Cette relecture a été faite en insu de leur première interprétation. Un délai de un an a été respecté entre les deux interprétations, pour contrôler le risque de biais d'information lié à une mémorisation. Chaque relecteur a relu la moitié des radiographies initialement interprétées. Ces radiographies ont été sélectionnées par tirage au sort parmi celles qu'ils avaient interprétées lors de leur première séance. Pour réduire la durée des séances de relecture, les radiographies ont été rassemblées avant la séance dans un diaporama Power Point. À leur demande, les relecteurs pouvaient disposer des radiographies dans le logiciel initial afin de disposer des outils d'aide à l'interprétation comme lors des séances initiales.

Déterminants de la fiabilité

Une association entre les différentes variables sociodémographiques, cliniques, et radiographiques des patients et la présence d'une discordance d'interprétation a été recherchée. L'association avec les caractéristiques des relecteurs (spécialité, niveau d'expérience) a été étudiée.

Nous avons également réalisé une réunion de tous les relecteurs, à distance de la première interprétation (une année plus tard), avec pour objectif d'identifier les éléments radiographiques ou cliniques déterminant l'interprétation des relecteurs, pour mettre en évidence d'éventuels points de consensus ou de faiblesse dans la démarche diagnostique.

Ces éléments peuvent en effet expliquer les déterminants de la fiabilité ou non de l'interprétation d'une radiographie du poignet. Toutes les radiographies pour lesquelles il existait au moins une discordance entre l'interprétation du premier relecteur, du deuxième relecteur, ou du référent ont été incluses pour cette réunion. La discordance était jugée sur le classement binaire « pas de lésion » contre « lésion ou doute ».

La réunion a été organisée selon un protocole établi à l'avance. L'ordre de relecture a été déterminé de façon aléatoire à l'aide de la fonction *sample* du logiciel R. Les radiographies étaient affichées, puis les diagnostics faits initialement par les relecteurs étaient rappelés. Le nom des relecteurs initiaux n'était pas donné. Les relecteurs présentaient leur diagnostic puis discutaient de façon ouverte des raisons de leur interprétation. Le premier relecteur à prendre la parole était différent à chaque radiographie pour ne pas créer l'influence excessive d'un relecteur, liée à sa position hiérarchique notamment. Pour la première radiographie, le premier relecteur a été choisi par tirage au sort. Puis une rotation selon le sens inverse des aiguilles d'une montre déterminait le premier relecteur de la radiographie suivante. Les éléments de débat étaient enregistrés sous la forme de notes par moi-même. Un enregistrement audio des discussions a également été réalisé pour contrôler *a posteriori* que des informations n'avaient pas été omises. Un animateur des débats (LRS) contrôlait les prises de parole et le bon déroulement de la réunion.

Pour la fluidité de la réunion, les radiographies avaient été rassemblées *a priori* dans un diaporama PowerPoint qui était affiché par moi-même. Si un relecteur souhaitait utiliser les outils de modifications graphiques (contraste, zoom), la radiographie était affichée sur le logiciel habituel. En parallèle, les informations cliniques de chaque radiographie étaient présentées dans un document distribué à chaque relecteur.

Analyse statistique

L'ensemble des analyses a été réalisé à l'aide du logiciel R (RStudio, version 0.98).

Description

La description pour les variables quantitatives a été faite par des moyennes et des écart-types. Pour les variables catégorielles, les effectifs par modalité ainsi que les proportions ont été calculés.

Tests de comparaison

Pour les variables quantitatives, les comparaisons ont été faites avec un test de Student pour les comparaisons de deux groupes et un test de Kruskal-Wallis pour les comparaisons de quatre groupes (en l'absence de vérification de l'hypothèse de normalité pour les groupes de faible effectif). Pour les variables catégorielles, le test du χ^2 a été utilisé quand les effectifs théoriques étaient supérieurs à cinq, avec correction de Yates s'ils étaient compris entre 2,5 et cinq (uniquement pour les tableaux 2 x 2). Dans les cas où les effectifs théoriques étaient inférieurs à 2,5 (ou inférieurs à cinq pour les tableaux de taille supérieure à 2 x 2), le test exact de Fisher a été utilisé. Pour les comparaisons de classement entre les médecins des urgences et les experts dans le cas où l'ensemble des effectifs était comparé, le test du χ^2 de McNemar a été utilisé pour respecter la notion d'appariement. Tous les tests ont été réalisés dans le cas bilatéral. Les tests statistiques étaient considérés significatifs lorsque la valeur *p* était inférieure à 0,05.

Concordance inter-observateurs

Classement binaire « lésion ou doute » ou « pas de lésion »

La concordance entre les observations a été estimée par le coefficient kappa (κ) de Cohen (23,24) et son intervalle de confiance à 95 % (IC95 %) estimé selon la méthode asymptotiquement exacte de Fleiss (24). Ce coefficient est la référence pour l'étude de la reproductibilité d'un examen dont le résultat est qualitatif. Il permet de mesurer la concordance au-delà de la concordance attendue par hasard. En effet, quel que soit le comportement des observateurs, ils obtiendront parfois le même diagnostic par hasard. Or, si les observateurs sont concordants par hasard, cela ne dit rien de la fiabilité de la démarche diagnostique. Seule la concordance au-delà de celle attendue par hasard peut être considérée comme une « vraie concordance ». C'est cette vraie concordance que mesure le coefficient κ de Cohen. Il est obtenu en divisant l'accord véritable au-delà de l'accord lié au hasard par l'accord possible au-delà de l'accord lié au hasard (Annexe 5). Il s'étend d'une valeur proche de - 1 (discordance parfaite : lorsque le premier observateur dit

A, le second dit toujours B) à 1 (concordance parfaite : les deux observateurs ont toujours le même avis). Un coefficient à 0 correspond à la concordance qui serait obtenue si les observateurs cotaient au hasard. Pour interpréter le résultat avec ce coefficient, Landis et Koch ont proposé une classification (25) qui est la plus utilisée (Tableau 3).

Tableau 3. Classification de Landis et Koch pour l'interprétation du coefficient κ de Cohen (25)

Coefficient κ	Interprétation de la concordance*
≤ 0	Pauvre
0,01 - 0,2	Légère
0,21 - 0,40	Juste
0,41 - 0,60	Modérée
0,61 - 0,80	Substantielle
0,81 - 1	Presque parfaite

* Traduction littérale de la classification anglaise en l'absence de traduction validée

Certains auteurs ont proposé des outils d'aide à l'interprétation du coefficient κ (26,27) que nous avons utilisés. L'index de prévalence (*Prevalence Index,* PI) évalue le biais dans l'estimation du coefficient κ lié à la prévalence de la lésion étudiée dans l'échantillon. Le PI est compris entre 0 (absence d'effet de la prévalence) et 1. L'index de biais (*Bias Index*, BI) évalue le biais dans l'estimation du coefficient κ lié à des différences dans l'attitude globale adoptée par les observateurs. Le BI varie également entre 0 (pas de biais) et 1. Un coefficient κ corrigé à la fois par la prévalence et le biais entre observateurs a également été calculé : le PABAK (Prevalence-Adjusted and Bias-Adjusted Kappa). Les calculs de ces estimateurs et du coefficient κ sont détaillés en Annexe 5.

L'estimation du coefficient κ, dans le cas où le classement est binaire, est optimale quand la prévalence de chaque classement est de 50 % (26,27). Nous avons donc conduit une analyse de robustesse en réduisant l'effectif du groupe « pas de lésion » pour le rendre de taille égale au groupe « lésion ou doute », et donc créer les conditions d'une prévalence à 50 % expérimentalement. Cette réduction a été réalisée par tirage au sort de patients au sein du groupe « pas de lésion ». L'effectif utilisé prenait pour base le nombre de patients diagnostiqués « pas de lésion » dans le consensus d'experts. Une fois cette prévalence rétablie, les estimations des coefficients et des index étaient reproduites. La procédure a été réalisée de façon répétée (1 000 fois) avec remise, pour augmenter la variabilité. La moyenne des estimations à chaque tirage au sort a été ensuite calculée pour obtenir un coefficient κ, un index de biais et de prévalence et un PABAK synthétiques.

La concordance au sein du groupe des relecteurs a été estimée par le coefficient κ selon les mêmes méthodes que l'analyse principale. La concordance a été étudiée de façon globale entre experts, puis paire par paire pour étudier l'homogénéité des paires. Les concordances des paires ont été comparées entre elles graphiquement. Les IC95 % des coefficients κ ont été estimés. L'étude de la concordance entre les experts et le référent incluait les radiographies pour lesquelles les experts étaient concordants (comparaison entre le diagnostic consensuel des experts et le diagnostic fait par le référent).

Classement en familles de diagnostics

La concordance pour les familles de diagnostics a été réalisée par l'étude d'une variable à cinq modalités : « entorse », « doute », « lésion ligamentaire », « fracture d'un ou plusieurs os du carpe », « fracture d'un ou des os de l'avant-bras ». Bien qu'entraînant une perte partielle d'information, ce regroupement en familles était nécessaire pour les besoins de l'analyse, pour ne pas diluer les résultats de façon trop importante avec des modalités multiples. Le coefficient κ pondéré et son IC95 % a été estimé. Celui-ci a été calculé à l'aide d'une méthode de *bootstrap* à 1 000 occurrences grâce à une extension de la fonction *wkappa* du complément *psy* du logiciel R.

Pour cette analyse, un système de pondération quadratique (24) a été utilisé. Cette méthode permet de donner plus de poids à des discordances si les diagnostics donnés par les deux observateurs sont très éloignés, ou moins de poids si les deux diagnostics sont proches. Les familles ont été classées dans l'ordre suivant : « entorse », « doute », « lésion ligamentaire », « fracture du carpe », « fracture de l'avant-bras ». Ainsi, un poids maximal était donné si l'un des observateurs concluait « fracture de l'avant-bras » et l'autre « entorse », et un poids minimal si les deux familles se suivaient dans le classement. Les os de l'avant-bras étant les plus volumineux, leurs fractures sont les plus visibles, la différence avec une entorse simple est donc la plus importante. De plus, il y a moins de différence entre une lésion ligamentaire et une fracture du carpe qu'entre une lésion ligamentaire et une fracture de l'avant-bras, puisque les lésions ligamentaires touchent également le carpe.

Pour étudier la présence de différences en fonction des diagnostics, une variable binaire a été créée pour chaque famille. Par exemple, pour « lésion ligamentaire », une nouvelle variable, de valeur « zéro » si un autre diagnostic qu'une lésion ligamentaire était posé, et de valeur « un » si le diagnostic était une lésion ligamentaire a été étudiée. Les coefficients κ et leur IC95 % ont été estimés, ainsi que les index de biais et prévalence et le PABAK.

Concordance intra-observateur

Pour étudier la reproductibilité intra-observateur, nous avons également utilisé le coefficient κ de Cohen et son IC95 %. Il est utilisable dans ce contexte, bien que l'hypothèse d'indépendance des observations ne soit pas respectée (27), dans la mesure où l'intervalle entre les observations est suffisamment important, que l'objet étudié est stable (ce qui est le cas pour une radiographie, qui n'est pas modifiée dans le temps), et que les radiographies sont relues en insu et selon un ordre aléatoire. Les modalités étaient identiques aux analyses inter-observateurs, dans le cas binaire et en familles de diagnostics.

Déterminants de la fiabilité

Pour l'étude des facteurs associés à la présence d'une discordance, les rapports de cotes et leur IC95 % ont été estimés par régression logistique. La variable dépendante était la présence ou non d'une discordance entre le consensus d'experts et les médecins des urgences pour le classement binaire. Pour les variables quantitatives, l'hypothèse de log-linéarité a été vérifiée par la méthode manuelle de comparaison des intervalles de confiance après catégorisation. Pour respecter l'hypothèse d'indépendance des observations, il n'était pas possible de conserver les radiographies pour les cas où le traumatisme était bilatéral et les deux radiographies incluses. Pour ces patients, un tirage au sort a été réalisé pour ne conserver qu'une des deux radiographies dans l'analyse.

Nombre de sujets nécessaires

Le calcul de la taille de l'échantillon a été réalisé à partir des travaux de Rotondi et Donner (28), qui proposent une approche fondée sur la précision souhaitée de l'IC95 %. Les auteurs ont également développé un outil additionnel pour le logiciel R (*package kappaSize*) qui a été utilisé pour le calcul de la taille de l'échantillon de notre étude. Le nombre de sujets a été calculé pour obtenir un IC95 % d'une précision de plus ou moins 0,15 autour du coefficient κ estimé, avec une proportion attendue de classement « lésion ou doute » de 30 % (4,12–14). La taille a été calculée pour des coefficients kappa entre 0,5 et 0,7. Pour le cas où le coefficient kappa valait 0,5, 172 sujets étaient nécessaires. Pour un coefficient kappa de 0,7, 140 sujets étaient nécessaires.

Le nombre de sujets nécessaire a également été estimé pour l'analyse secondaire de la concordance pour le regroupement des lésions en cinq familles. Cinquante-trois sujets étaient nécessaires pour obtenir un IC95 % d'une précision de plus ou moins 0,2 autour d'une valeur attendue de 0,5, avec les proportions attendues suivantes : Fracture des os de l'avant-bras 50 % - Doute 30 % - Lésions ligamentaires 10 % - Fracture du carpe 10 %.

Considérations éthiques

Cette étude a été réalisée en respect de la Charte d'utilisation des données du PMSI du CHU de Bordeaux, en assurant l'anonymat complet et l'accès au dossier par un médecin autorisé (29). Les dossiers n'ont été ouverts et les informations extraites que par moi-même, qui bénéficie de cette autorisation.

Structuration du texte

Nous nous sommes appuyés au fil de la construction de l'étude et de la rédaction du texte sur les recommandations du groupe d'étude QAREL (Quality Appraisal of Diagnostic Reliability). Il s'agit d'un groupe d'experts ayant défini une liste de points de contrôle (30,31) pour assurer la validité interne d'une étude de fiabilité diagnostique.

RESULTATS

Patients inclus et données manquantes

Cent-soixante-quatorze radiographies ont été incluses dans l'étude (correspondant à 169 patients, cinq patients ayant présenté un traumatisme bilatéral). L'âge moyen des patients était de 38,1 ans (écart-type 17,3). Il y avait 56,3 % d'hommes et 43,7 % de femmes. La majorité des patients (65,4 %) a consulté le jour même du traumatisme, 23,5 % le lendemain et 11,1 % deux jours après ou plus. Le mécanisme traumatique le plus fréquent était la chute (54,9 %), suivie des accidents de la voie publique (AVP) en véhicule motorisé et des chocs directs ou compressions (14 % chacun), puis des AVP à vélo (6,7 %). Les autres mécanismes (rixe, torsion contrariée, port de charge lourde) concernaient 10,4 % des patients.

Il n'y avait pas de données manquantes pour les variables liées au séjour du malade. Cependant, le nombre de données manquantes pour les données de l'anamnèse et de l'examen clinique était élevé, jusqu'à 93,1 % (Annexe 6). Ce chiffre élevé n'a pas rendu possible la réalisation de techniques d'imputation des données manquantes. La proportion de données manquantes était significativement plus élevée dans le groupe « lésion ou doute » (classement par les médecins des urgences) pour les variables site de douleur exquise, impotence fonctionnelle ou limitation des amplitudes articulaires, ecchymose, hématome, dermabrasion, paresthésies et troubles sensitifs autres que paresthésies.

Reproductibilité inter-observateurs

Construction du consensus d'experts

Après relecture de toutes les radiographies selon le protocole, les deux experts relecteurs d'une même radiographie étaient concordants (classement en « lésion ou doute » ou « pas de lésion ») pour 129 radiographies (74,1 %) sur 174. Pour ces radiographies, le consensus était donc obtenu d'emblée et ne nécessitait pas la prise en compte de l'avis du référent. Parmi ces 129 radiographies, 52 (40,3 %) ont été classées « lésion ou doute » et 77 (59,7 %) « pas de lésion ».

Pour les 45 autres radiographies, l'obtention du consensus a nécessité l'avis du référent. Le référent en a classé 26 (57,8 %) en « lésion ou doute » et 19 (42,2 %) en « pas de lésion ». Parmi les radiographies classées par le référent en « lésion ou doute », les familles de

diagnostics étaient réparties en « doute » : 20,0 %, « lésion ligamentaire » : 20,0 %, « fracture du carpe » : 8,9 % et « fracture de l'avant-bras » : 8,9 %.

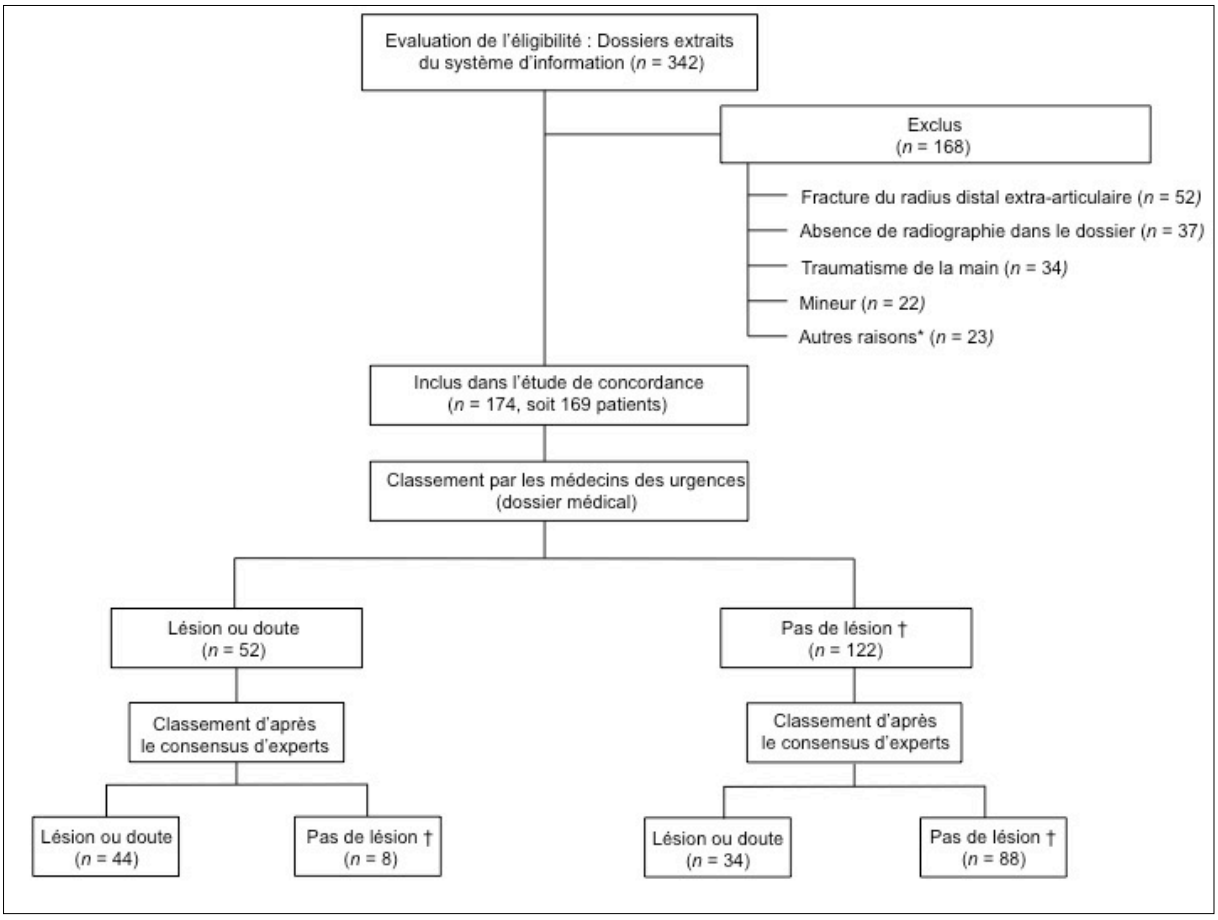
À l'issue de l'avis du référent, toutes les radiographies ont donc pu être classées de façon consensuelle : 78 (44,8 %) en « lésion ou doute » et 96 (55,2 %) en « pas de lésion ».

Classement des malades

Classement binaire « lésion ou doute » ou « pas de lésion »

Le classement par les médecins des urgences et le consensus d'experts (Figure 3) a permis de définir quatre groupes :

- « Accord sur lésion ou doute » : médecins des urgences et consensus d'experts concordaient pour le classement « lésion ou doute » ;
- « Accord sur absence de lésion » : médecins des urgences et consensus d'experts concordaient pour le classement « pas de lésion » ;
- « Désaccord - lésion ou doute pour experts » : désaccord, « lésion ou doute » pour le consensus d'experts; « pas de lésion » pour les médecins des urgences ;
- « Désaccord - lésion ou doute pour urgences » : désaccord, « lésion ou doute » pour les médecins des urgences ; « pas de lésion » pour le consensus d'experts.



* Dont : Non traumatique (*n* = 9) - Ancienneté supérieure à sept jours (*n* = 3) - Atteintes du membre supérieur mais pas du poignet (*n* = 6) - Traumatisme ouvert (*n* = 5) ; † Entorse simple du poignet ; les chiffres correspondent au nombre de radiographies

Figure 3. Diagramme de flux de l'étude de concordance inter-observateurs entre médecins des urgences et consensus d'experts ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014

Les âges moyens et les écart-types (ET) respectifs pour ces groupes étaient de 45,6 ans (accord sur lésion ou doute, ET = 20,5) ; 35,4 ans (accord sur absence de lésion, ET = 14,8); 37,7 ans (désaccord - lésion ou doute pour experts, ET = 16,2) et 32,7 ans (désaccord - lésion ou doute pour urgences, ET = 17,6) ; les groupes différaient significativement sur ce critère (p = 0,04, test de Kruskal-Wallis). Les quatre groupes ne présentaient pas de différence statistiquement significative sur la répartition des sexes, le délai depuis le traumatisme, la saison, l'heure d'admission, le mécanisme ou le caractère lié ou non au sport (Tableau 4).

Tableau 4. Description des patients en fonction du classement fait par les médecins des urgences et le consensus d'experts ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

Variable	Accord sur lésion ou doute *n* = 41		Accord sur absence de lésion *n* = 87		Désaccord - lésion ou doute pour experts *n* = 34		Désaccord - lésion ou doute pour urgences *n* = 7		*p**
	Effectif	%	Effectif	%	Effectif	%	Effectif	%	
Sexe									0,91
Femme	18	43,9	39	44,8	14	41,2	2	28,6	
Homme	23	56,1	48	55,2	20	58,8	5	72,4	
Date du traumatisme (*n* = 148)									0,60
Le jour même	26	74,3	49	62,0	17	58,6	4	80,0	
La veille	6	17,1	22	27,9	8	27,6	0	0	
Deux jours ou plus	3	8,6	8	10,1	4	13,8	1	20,0	
Saison									0,64
Printemps	14	34,2	25	28,7	11	32,4	3	42,9	
Été	13	31,7	19	21,8	9	26,5	3	42,9	
Automne	6	14,6	22	25,3	9	26,5	0	0	
Hiver	8	19,5	21	25,1	5	14,7	1	14,2	
Heure de consultation									0,73
8h30-12h29	9	22,0	25	29,4	5	14,7	2	28,6	
12h30-18h29	14	34,1	28	32,9	16	47,1	2	28,6	
18h30-00h29	13	31,7	20	23,5	8	23,5	1	14,2	
00h30-8h29	5	12,2	14	14,1	5	14,7	2	28,6	
Mécanisme (*n* = 164)									0,22
Chute	24	63,1	44	52,4	17	54,8	4	66,7	
AVP véhicule motorisé	7	18,4	10	11,9	3	9,7	0	0	
AVP vélo	2	5,3	8	9,5	1	3,3	0	0	
Choc†	2	5,3	17	20,2	3	9,7	1	16,7	
Autre‡	3	7,9	5	6,0	7	22,6	1	16,7	
Accident de sport (*n* = 169)									0,44
Oui	34	89,5	75	87,2	27	81,8	5	71,4	
Non	4	10,5	11	12,8	6	18,2	2	28,5	

n = 169 radiographies ; AVP : Accident de la Voie Publique ; * Test exact de Fisher ;
† Choc direct ou compression ; ‡ Rixe, torsion contrariée, port de charge lourde

Classement en familles de diagnostics

Aucun patient ne présentait une lésion correspondant à l'association de deux familles différentes (Tableau 5). Les lésions ligamentaires étaient plus souvent identifiées par le consensus d'experts que par les médecins des urgences. Toutes les lésions ligamentaires diagnostiquées étaient des lésions du ligament scapho-lunaire, dont une associée à une lésion du ligament triquétro-lunaire. Le consensus d'expert a diagnostiqué significativement plus de lésions (c'est-à-dire « fracture de l'avant-bras » ou « fracture du carpe » ou « lésion ligamentaire » contre « doute » ou « pas de lésion ») que les médecins des urgences (58 contre 33, $p < 0{,}001$, test du χ^2 de McNemar). Le consensus d'experts aurait significativement plus souvent abouti à une poursuite de la prise en charge (modalité « lésion ou doute ») que les médecins des urgences.

Tableau 5. Effectifs de chaque famille de diagnostics, pour l'interprétation par les médecins des urgences et le consensus d'experts ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

	Urgences		**Consensus d'experts**		***p****
	Effectif	%	Effectif	%	
Lésion ou doute	52	29,9	78	44,8	<0,001
Fracture des os de l'avant-bras	27	15,5	30	17,2	0,66
Fracture du carpe, dont	6	3,5	10	5,8	0,34
Fracture du scaphoïde	3	1,7	6	3,5	
Fracture du triquetrum	2	1,1	0	0	
Fracture de l'hamatum	1	0,6	2	1,1	
Fracture du lunatum	0	0	1	0,6	
Fracture du trapèze	0	0	1	0,6	
Lésion ligamentaire	0	0	18	10,3	< 0,001
Doute	19	10,9	20	11,5	1
Pas de lésion/Entorse simple	122	70,1	96	55,2	†

n = 174 radiographies ; * Test du χ^2 de McNemar; † Test équivalent à celui pour « Lésion ou doute »

Concordance entre médecins des urgences et consensus d'experts

Classement binaire « lésion ou doute » ou « pas de lésion »

Le coefficient κ de concordance entre médecins des urgences et consensus d'experts (objectif principal) était de 0,50 (IC95 % : 0,37 à 0,62). La proportion d'observations concordantes était de 76 %, la concordance attendue par hasard de 52 %.

L'index de prévalence valait 0,25, l'index de biais 0,15 et le PABAK 0,52 (Tableau 6). Après tirage au sort pour rétablir une prévalence de 50 % des entorses simples du poignet réalisé 1 000 fois, le coefficient κ était de 0,48 (IC95 % : 0,35 à 0,61).

Tableau 6. Concordance entre le classement binaire « lésion ou doute » ou « pas de lésion » fait par les médecins des urgences et le consensus d'experts ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

	Urgences					
	Lésion ou doute		Pas de lésion		Total	
Consensus d'experts	Effectif	%	Effectif	%	Effectif	%
Lésion ou doute	44	25	34	19	78	44
Pas de lésion	8	5	88	51	96	56
Total	52	30	122	70	174	100

n = 174 radiographies

Classement en familles de diagnostics

Lorsque les diagnostics étaient regroupés en cinq familles (Tableau 7), le coefficient κ pondéré pour la concordance entre médecins des urgences et consensus d'experts était de 0,65 (IC95 % : 0,54 à 0,76).

Tableau 7. Concordance du classement en familles de diagnostics fait par les médecins des urgences et le consensus d'experts ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

Consensus d'experts	**Urgences**										Total	
	Fracture avant-bras		Fracture du carpe		Ligament		Doute		Pas de lésion			
	Effectif	%	Effectif	%	Effectif	%	Effectif	%	Effectif	%	Effectif	%
Fracture de l'avant-bras	18	10	2	1	0	0	5	3	5	3	30	17
Fracture du carpe	0	0	3	2	0	0	3	2	4	2	10	6
Ligament	8	4	0	0	0	0	1	1	9	5	18	10
Doute	0	0	1	1	0	0	3	2	16	9	20	12
Pas de lésion	1	1	0	0	0	0	7	4	88	50	96	55
Total	27	15	6	4	0	0	19	12	122	69	174	100

n = 174 radiographies

Lorsque les diagnostics étaient étudiés séparément, les coefficients κ différaient selon la famille de diagnostic (Figure 4). Le plus faible concernait la modalité « lésion ligamentaire » (0) et le plus élevé la modalité « fracture de l'avant-bras » (0,56). Les coefficients κ par famille ainsi que leurs PI, BI et PABAK sont présentés en Annexe 7.

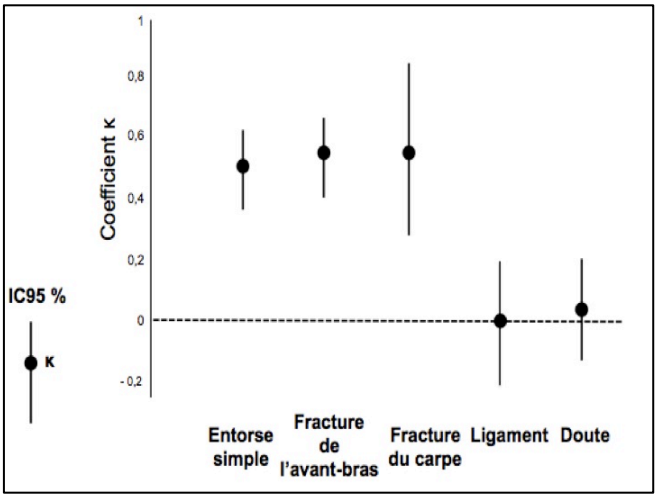


Figure 4. Concordance pour chaque famille de diagnostics entre médecins des urgences et consensus d'experts ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014

Concordance entre les relecteurs

Concordance entre les experts

Le coefficient κ pour la concordance entre les experts (sans le référent) était de 0,47 (IC95 % : 0,34 - 0,60) pour le classement binaire « lésion ou doute » ou « pas de lésion » (Tableau 8). La concordance observée était de 74 %, la concordance attendue par hasard de 51 %. Le PI avait pour valeur 0,14, le BI 0,03 et le PABAK 0,48.

Tableau 8. Concordance entre deux experts relisant la même radiographie pour le classement « lésion ou doute » ou « pas de lésion » ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

	Premier relecteur					
	Lésion ou doute		Pas de lésion		Total	
Second relecteur	Effectif	%	Effectif	%	Effectif	%
Lésion ou doute	52	30	20	12	72	42
Pas de lésion	25	14	77	44	102	58
Total	77	44	97	56	174	100

n = 174 radiographies

L'estimation du coefficient κ avec pondération quadratique pour la concordance entre les experts pour le classement en cinq familles était de 0,70 (IC95 % : 0,57 à 0,78 ; détails en Annexe 7).

Concordance entre experts et référent

Les experts étaient concordants pour le classement binaire « lésion ou doute » ou « pas de lésion » pour 129 radiographies. La concordance avec le diagnostic du référent pour ces 129 radiographies était de 0,56 (IC95 % : 0,42 à 0,69). La concordance observée était de 78 % ; la concordance attendue par chance de 49 % ; le PI de 0,06; le BI de 0,13 et le PABAK de 0,55 (Tableau 9).

Tableau 9. Concordance entre les deux experts relecteurs et le référent pour le classement binaire « lésion ou doute » ou « pas de lésion », pour les radiographies pour lesquelles les deux experts relecteurs concordaient ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014

	Duo d'experts relecteurs					
	Lésion ou doute		Pas de lésion		Total	
Référent	Effectif	%	Effectif	%	Effectif	%
Lésion ou doute	46	35	23	18	69	53
Pas de lésion	6	5	54	42	60	47
Total	52	40	77	60	129	100

n = 129 radiographies

Pour les cinq familles de diagnostics, le coefficient κ avec pondération quadratique était de 0,75 (IC95 % : 0,64 à 0,84, détails en Annexe 7). Lorsque ces cinq familles étaient étudiées individuellement, les coefficients κ étaient de 0,81 (IC95 % 0,68 - 0,94) pour les fractures de l'avant-bras, 0,76 (IC95 % 0,49 - 1,00) pour les fractures du carpe, 0,23 (IC95 % 0,04 - 0,42) pour les lésions ligamentaires, 0,13 (IC95 % -0,11 - 0,38) pour le doute et 0,56 (IC95 % 0,42 - 0,69) pour les entorses simples (Annexe 7).

Reproductibilité intra-observateur

Chacun des quatre experts a relu 43 radiographies, soit un total de 172 relectures. Le coefficient κ pour la concordance intra-observateur globale (ensemble des relecteurs) pour le classement binaire « lésion ou doute » ou « pas de lésion » était de 0,12 (IC95 % 0,01 - 0,23). Les concordances intra-observateur par relecteur ne permettaient pas d'éliminer une concordance par hasard (Tableau 10). Parmi les diagnostics faits par les relecteurs lors de la deuxième séance, six lésions ont été diagnostiquées (3,5 %).

Tableau 10. Coefficients κ pour la concordance intra-observateur globale et pour chaque relecteur ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014

Relecteur	κ	IC 95 %	p_o	p_e	PI	BI	PABAK
Global	0,12	0,01 - 0,23	0,61	0,56	0,37	0,13	0,23
Chirurgien CCA	0,15	- 0,17 - 0,47	0,62	0,56	0,35	0,10	0,24
Chirurgien PH	0,15	- 0,16 - 0,47	0,65	0,58	0,43	0,14	0,30
Radiologue CCA	0,26	-0,08 - 0,59	0,71	0,60	0,47	0,12	0,41
Radiologue PH	0,05	-0,18 - 0,28	0,47	0,44	0,22	0,41	-0,06

n = 172 radiographies ; p_o : proportion observée d'observations concordantes ; p_e : proportion de concordance attendue par hasard ; PI : Prevalence Index ; BI : Bias Index ; PABAK : Prevalence-Adjusted Bias-Adjusted Kappa

Le coefficient κ pondéré pour la concordance intra-observateur globale pour le classement en cinq familles de diagnostics était de 0,04 (IC95 % -0,04 – 0,16). Aucune des concordances par relecteur n'était significativement différente de 0 : chirurgien CCA 0,17 (IC95 % - 0,12 – 0,41), chirurgien PH 0,17 (IC95 % -0,12 – 0,48), radiologue CCA 0,28 (IC95 % - 0,02 – 0,58), radiologue PH -0,04 (IC95 % -0,15 – 0,09).

Déterminants de la fiabilité

Caractéristiques du patient et du traumatisme

Les variables comportant plus de 30 % de données manquantes n'ont pas été incluses dans l'analyse. Seules les variables « Conclusion des experts » et « Discordance entre experts » présentaient une association au seuil $p < 0,05$ (Tableau 11). Étant donnés ces résultats en analyse univariable, il a été décidé de ne pas réaliser de modèle multivariable.

Tableau 11. Association entre caractéristiques du patient ou de son traumatisme et présence d'une discordance d'interprétation des radiographies entre les médecins des urgences et le consensus d'experts. Régression logistique univariable ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

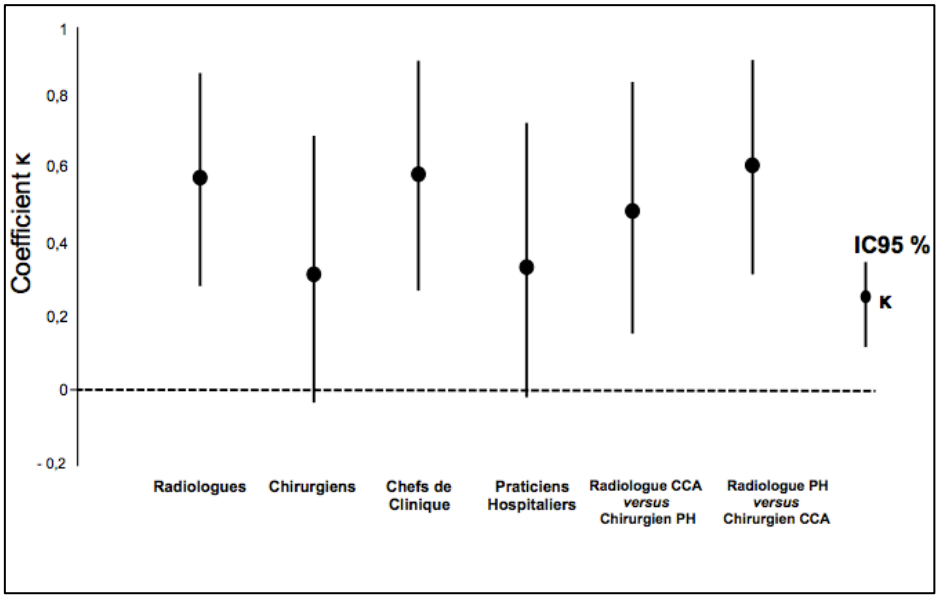
Variable	Rapport de cotes	IC95 %	*p*
Age*			0,21
Moins de 20 ans (*n* = 77)	1		
De 21 à 40 ans (*n* = 45)	1,26	0,49 - 3,55	
De 41 à 60 ans (*n* = 29)	0,48	0,14 - 1,63	
61 ans ou plus (*n* = 23)	1,38	0,40 - 4,79	
Sexe			0,54
Homme	1	-	
Femme	0,80	0,38 - 1,62	
Date du traumatisme (*n* = 153)			0,72
Le jour même	1		
La veille	1,02	0,39 - 2,50	
Deux jours ou plus	1,62	0,47 - 5,02	
Saison d'admission			0,71
Printemps	1	-	
Été	1,04	0,42 - 2,58	
Automne	0,90	0,33 - 2,34	
Hiver	0,58	0,18 - 1,62	
Heure de consultation			0,43
8h30 - 12h29	1	-	
12h30 - 18h29	2,08	0,80 - 5,89	
18h30 - 00h29	1,32	0,44 - 4,10	
00h30 - 8h29	2,00	0,60 - 6,77	
Intervalle libre entre le traumatisme et les symptômes (*n* = 118)			0,82
Non	1	-	
Oui	0,88	0,27 - 2,51	
Mécanisme (*n* = 159)			0,08
Chute	1	-	
AVP véhicule motorisé	0,57	0,12 - 1,91	
AVP vélo	0,32	0,01 - 1,84	
Choc†	0,68	0,18 - 2,06	
Autre‡	3,24	1,07 - 9,86	
Hauteur de chute si mécanisme de chute (*n* = 89)			0,61
Hauteur du patient	1	-	
Surélévation jusqu'à 1m50	1,47	0,19 - 8,20	
Surélévation de plus de 1m50	0,42	0,02 - 2,59	
Type de chute§ (*n* = 87)			0,19
Sans cinétique	1	-	
Avec cinétique	0,51	0,18 - 1,38	
Accident de sport (*n* = 164)			0,22
Non	1	-	
Oui	1,82	0,69 - 4,59	
Antécédent de traumatisme homolatéral (*n* = 152)			0,93
Non	1	-	
Oui	1,05	0,32 - 2,94	
Impotence fonctionnelle (*n* = 127)			0,07
Non	1	-	
Oui	2,90	0,91 - 12,96	
Qualité de la radiographie			0,80
Bonne qualité	1	-	
Mauvaise mais interprétable	0,76	0,29 - 2,12	
Non interprétable	0,66	0,18 - 2,35	
Conclusion des experts			<0,001
Pas de lésion/Entorse simple	1	-	
Doute	40,78	11,88 - 173,35	
Lésion ligamentaire	10,88	3,41 - 36,60	
Fracture du carpe	7,25	1,59 - 31,39	
Fracture de l'avant-bras	2,47	0,69 - 8,18	
Paire de relecteurs			0,32
Radiologues	1	-	
Chirurgiens	0,96	0,24 - 3,87	
CCA	3,16	0,96 - 11,72	
PH	1,00	0,25 - 4,05	
Radiologue CCA *versus* PH	1,53	0,42 - 5,89	
Radiologue PH *versus* Chirurgien CCA	1,75	0,50 - 6,60	
Discordance entre experts			<0,001
Non	1	-	
Oui	7,62	3,53 - 17,02	

n = 169 radiographies ; Variables qualitatives à deux modalités : test de Wald. Variables qualitatives à plus de deux modalités : test du rapport de vraisemblances ; AVP : Accident de la Voie Publique ;
* Variable quantitative catégorisée du fait du non-respect de l'hypothèse de log-linéarité ;
† Choc direct ou compression ; ‡ Rixe, torsion contrariée, port de charge lourde ; § Exemple de chute sans cinétique associée : Patient butant contre un trottoir. Exemple de chute avec cinétique : chute en patinant.

Caractéristiques des relecteurs

Parmi les radiographies où il existait au moins une discordance entre le premier expert, le deuxième expert et le référent (71 radiographies) sur le classement binaire, le nombre de discordances n'était pas différent entre les quatre experts (31, 34, 36 et 37 radiographies selon le relecteur, p = 0,79, test exact de Fisher). Parmi les interprétations en familles de diagnostics de ces 71 radiographies, 54 (76,1 %) étaient partiellement discordantes (accord entre deux relecteurs et désaccord du troisième, quel qu'il soit), et 17 (23,9 %) étaient totalement discordantes (chaque relecteur avait donné un diagnostic différent).

La concordance au sein des paires de relecteurs était variable, de 0,31 pour la paire constituée par les deux chirurgiens, à 0,59 pour la paire « chirurgien CCA / radiologue PH » (Figure 5), mais aucune de ces différences n'était statistiquement significative. Deux IC95 % ne permettaient pas de conclure à une concordance au-delà du hasard (paire de chirurgiens et paire de PH).



CCA : Chef de Clinique – Assistant ; PH : Praticien Hospitalier ; coefficients κ (points) et IC95 % (barre verticale) ; pour la concordance globale, n = 174 radiographies ; pour la concordance au sein de chaque paire, n = 29 radiographies

Figure 5. Concordance pour chaque paire d'experts pour le classement binaire « lésion ou doute » ou « pas de lésion » ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

Parmi les radiographies où les deux experts relecteurs discordaient entre eux (45 radiographies), la majorité des discordances entre les experts était retrouvée lorsque les deux experts appartenaient à une spécialité ou un niveau d'expérience différents. Ainsi, si l'on regarde les spécialités, 29 discordances sur 45 (64,4 %) apparaissaient lorsque l'un des experts était chirurgien et l'autre radiologue (contre dix entre chirurgiens et six entre radiologues, $p < 0{,}001$, test du $\chi 2$). De même, 29 discordances ont eu lieu dans une paire où se trouvaient à la fois un chef de clinique et un praticien hospitalier (contre six discordances entre chefs de clinique et dix entre praticiens hospitaliers, $p < 0{,}001$, test du $\chi 2$). Dans la majorité des cas, le couple de diagnostics discordants était « entorse bénigne » / « doute » (34 radiographies, 75,6 %). Dans tous les cas, l'un des deux relecteurs avait conclu « entorse bénigne » (contre « lésion ligamentaire » dans un cas, « fracture du carpe » dans un cas et « fracture d'un os de l'avant-bras » dans neuf cas).

Pour les 129 autres radiographies (c'est-à-dire celles où les deux experts concordaient), la concordance avec le diagnostic du référent était également différente en fonction des spécialités et des niveaux d'expérience, mais cette différence n'était pas significative (Figure 6). La meilleure concordance sur le classement binaire avec le référent était obtenue pour la paire de praticiens hospitaliers (κ estimé à 0,79, IC95 % : 0,52 à 1) et la plus faible pour la paire de chefs de clinique (κ estimé à 0,37, IC95 % : 0,05 à 0,67).

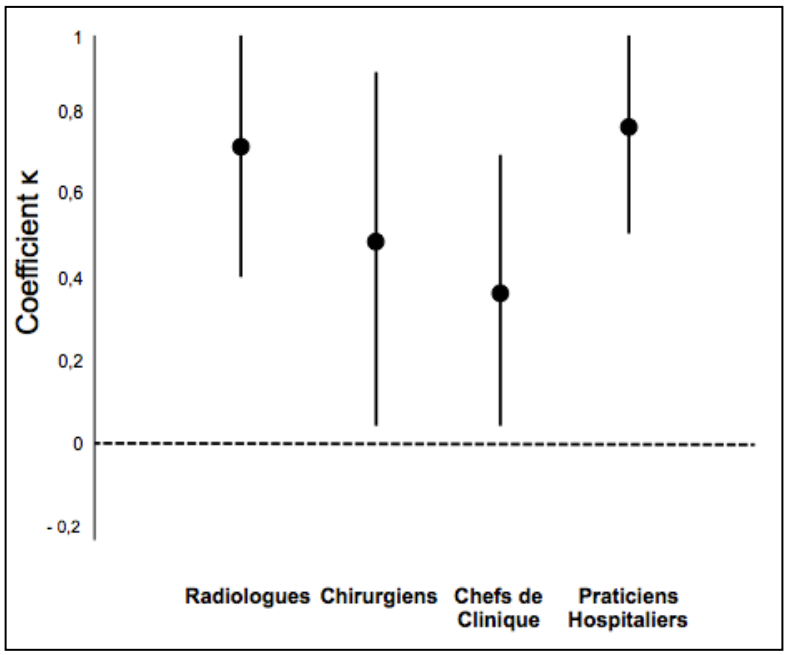


Figure 6. Concordance entre le référent et quatre des paires de relecteurs pour le classement binaire « lésion ou doute » / « pas de lésion », pour les radiographies pour lesquelles les deux experts relecteurs concordaient ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

Réunion entre les relecteurs

Soixante-et-onze radiographies présentaient au moins une discordance. Pour 45 radiographies, il s'agissait d'une discordance entre les deux experts. Pour les 27 autres, il s'agissait d'une discordance entre les experts et le référent alors que les experts avaient donné un avis identique. Pour 44 (62,0 %) radiographies, les interprétations divergentes étaient « Entorse simple » pour deux relecteurs et « Doute » pour le troisième. Il est apparu rapidement au début de la réunion, après la relecture des cinq premières radiographies présentant ce profil de divergence que la confrontation n'était pas informative dans ce cas précis. Les radiographies suivantes qui correspondaient à ce profil n'ont donc pas été discutées. Au final, 32 (45,1 %) radiographies ont été discutées. L'un des relecteurs n'a pas pu assister physiquement à la réunion du fait d'une mutation géographique. Sa participation à la réunion a pu se faire par le biais d'une connexion téléphonique. Il disposait du diaporama avec les radiographies comme les autres relecteurs.

Les relecteurs étaient d'accord sur le même diagnostic en fin de discussion pour six (18,8 %) radiographies. Le diagnostic évoqué était dans deux cas une rupture du ligament scapho-lunaire, dans deux cas une entorse simple, dans un cas une pseudarthrose du scaphoïde et dans un cas le doute des relecteurs avec demande d'examens complémentaires. Pour toutes les autres (81,2 %), il existait au moins un des relecteurs en désaccord avec les autres.

La mauvaise qualité des radiographies a été notée pour 19 (59,4 %) des clichés. Dans 15 cas la qualité était jugée insuffisante pour une interprétation valable et dans quatre cas il s'agissait de l'absence d'une incidence. Pour les 15 radiographies jugées d'insuffisante qualité, la totalité était en prono-supination non neutre ; associé ou non dans sept cas à la présence d'une inclinaison radiale horizontalisant le scaphoïde et dans un cas à une inclinaison ulnaire. À noter que pour une des radiographies, le côté indiqué pour la face était différent de celui du profil (alors que le traumatisme était unilatéral).

L'insuffisance d'informations cliniques (souhait du relecteur d'examiner le malade) a été évoquée dans six (18,8 %) cas, toujours par les cliniciens. La présence d'un élément radiographique objectif a été évoquée pour douze (37,5 %) radiographies. Ces éléments étaient la présence d'une déformation du lunatum en « DISI » (*Dorsal Intercalated Segmental Instability*), d'un décroché cortical, l'aspect ascensionné du lunatum, l'augmentation de l'espace interosseux scapho-lunaire (sans mesure objective), le signe de l'anneau du scaphoïde, la rupture des arcs de Gilula, la présence d'un trait transversal trans-

styloïdien au niveau du radius ou transversal au corps du scaphoïde et la disparition du triangle graisseux péri-scaphoïdien.

L'immobilisation temporaire avec réévaluation entre le dixième et le quinzième jour a été évoquée par les chirurgiens dans six cas (18,8 %). Un examen complémentaire a été évoqué par les relecteurs pour dix radiographies (31,3 %). Il s'agissait dans six cas de la réalisation de clichés dynamiques ou poing fermé et dans quatre cas d'une tomodensitométrie.

DISCUSSION

Principaux résultats

Notre étude apporte cinq enseignements principaux :

- Nous confirmons que la radiographie du poignet de face et de profil n'est pas un examen fiable dans le diagnostic des lésions traumatiques fermées du poignet.
- Cette mauvaise fiabilité est retrouvée quels que soient les observateurs comparés et lorsque les observateurs sont comparés à eux-mêmes.
- La radiographie du poignet est un examen d'interprétation difficile, avec des images souvent ambiguës. La majeure partie des discordances était liée à des radiographies classées « doute » par au moins un des observateurs.
- Les lésions ligamentaires sont les lésions dont le diagnostic est le moins fiable.
- Il ne semble pas exister de caractéristiques du patient ou du traumatisme déterminant la fiabilité d'une interprétation. En revanche, la qualité de la radiographie influe sa fiabilité.

La concordance entre observateurs est mauvaise et cet effet est stable, quels que soient les observateurs comparés. Ainsi, la concordance au sein du groupe des experts n'est pas différente de celle entre consensus d'experts et médecins des urgences (chevauchement des IC95 %) et pourrait même être légèrement inférieure. On observe également une mauvaise concordance entre les experts et le référent, y compris pour les cas où les experts concordaient.

Cette mauvaise concordance se retrouve également en comparant les observateurs à eux-mêmes. La stabilité de l'interprétation d'une même radiographie du poignet par le même observateur n'est pas satisfaisante. En l'absence de reproductibilité, la radiographie standard du poignet ne peut pas être considérée comme fiable pour décider de la poursuite ou non de la prise en charge en urgence.

Le classement « doute » a joué un rôle central dans les résultats observés. Cette propension à douter illustre bien la difficulté d'interprétation des images. Dans les deux groupes où médecins des urgences et consensus d'experts discordaient (groupes « Désaccord - lésion ou doute pour experts » et « Désaccord - lésion ou doute pour urgences »), les proportions de classement « doute » (par comparaison aux quatre autres familles de diagnostic) étaient

respectivement de 47 et 88 %. On observe la même tendance dans le groupe où médecins des urgences et consensus d'experts concordaient sur « lésion ou doute ». Ainsi, lorsque l'on regarde la famille de diagnostics choisie au sein de ce groupe, on remarque que 45 % des discordances se portaient sur le classement « doute » par les médecins des urgences. La majorité des problèmes de discordance, quel que soit leur sens, est donc portée par le doute, plus que par des discordances entre un diagnostic patent (une fracture par exemple) et le diagnostic d'entorse simple, même si cette situation existait.

La concordance pour le classement en « doute » était très faible (0,05). Cela signifie que les experts et les praticiens des urgences ne doutaient pas sur les mêmes radiographies, et donc probablement pas pour les mêmes raisons. La perte d'information liée au caractère rétrospectif n'a pas permis d'étudier ces causes. Mais les conséquences étaient similaires, puisque la proportion de patients classés « doute » n'était pas significativement différente. On remarque la même tendance au sein du groupe d'experts. Parmi les zones de discordances (un expert dit « lésion ou doute », l'autre « pas de lésion »), la modalité « doute » était choisie par un des deux relecteurs dans 73 % des cas. Au total, on voit le rôle central joué par le doute dans les résultats, comme cela est souvent le cas pour les modalités intermédiaires dans les études de concordance. En pratique, cela signifie que les médecins sont trop peu confiants en cet examen pour trancher dans de nombreux cas.

Le diagnostic de lésion ligamentaire n'a jamais été évoqué par les médecins des urgences (contre 10,3 % des radiographies pour le consensus d'experts). Sur ces lésions ligamentaires, la moitié avait cependant été classée « doute » ou « fracture de l'avant-bras » par les médecins des urgences ; ces patients n'étaient donc pas sortis du système de soins. Il n'empêche que ces lésions, parmi les plus pourvoyeuses d'arthrose (8), sont à risque élevé de passer inaperçues puisque la fiabilité de leur diagnostic est nulle. La difficulté du diagnostic est en cause et l'on note que le référent, plus expérimenté, est celui qui a diagnostiqué le plus de ces lésions. Mais il est également possible que le manque de formation à l'existence et au diagnostic de ces lésions puisse être incriminé. En effet, même en cas de doute, ce diagnostic n'a jamais été évoqué par les médecins des urgences. De plus, on observe une tendance similaire chez les experts. Ainsi, sur les radiographies où le référent n'était pas en accord avec les experts alors qu'ils concordaient entre eux, 62 % correspondaient à ce diagnostic pour le référent.

Le nombre de lésions diagnostiquées (c'est-à-dire « fracture de l'avant-bras » ou « fracture du carpe » ou « lésion ligamentaire » contre « doute » ou « pas de lésion ») était croissant par niveau d'expérience en radiographie du poignet. De 33 pour les médecins des urgences, à 58 pour le consensus d'experts et enfin 78 pour le référent (à noter cependant qu'une partie des observations du consensus d'experts correspondait à l'avis du référent). Il semble donc que l'expérience joue un rôle important dans la capacité à détecter des lésions. Cela semble intuitif, mais l'amplitude du phénomène est notable (supérieure à un doublement du nombre de lésions diagnostiquées). On observe le même phénomène pour le classement des radiographies douteuses. Lorsque l'urgentiste disait « doute », les experts diagnostiquaient une lésion dans la moitié des cas. A l'inverse, lorsque les experts disaient « doute », les urgentistes concluaient à « entorse simple » dans la plupart des cas. Ces résultats montrent donc une probable meilleure capacité de classement par les experts. Ces différents constats confirment la difficulté d'interprétation.

Nous n'avons pas retrouvé dans notre étude de facteurs associés significativement à la présence d'une discordance entre les relecteurs. Les caractéristiques des relecteurs semblent peu influer la fiabilité (pas de différences significatives), mais l'on dégage une tendance des radiologues à être plus fiables (meilleure concordance entre les diagnostics des deux radiologues). La réunion des relecteurs entre eux n'a pas non plus permis de mettre en évidence de réels déterminants de meilleure fiabilité. Le seul facteur retrouvé de façon presque systématique et consensuelle était la mauvaise qualité des radiographies réalisées aux Urgences.

Validité interne de l'estimation du coefficient κ

Les outils d'aide à l'interprétation (PI, BI, PABAK et *bootstrap*) utilisés dans l'étude semblent assurer une bonne validité de l'estimation des coefficients κ. Néanmoins, les concordances observées étaient élevées mais les coefficients κ moyens. Il s'agit d'un paradoxe décrit dans la littérature (32–34), lié à des déséquilibres symétriques dans les effectifs marginaux (plus de jugements A que B chez les deux observateurs), ce qui est le cas dans notre étude. Il s'agit donc du reflet de la prévalence (35) observée dans l'échantillon. Plus elle s'éloigne de 50 %, plus le coefficient κ sera sous-estimé. Cet effet est d'autant plus important que la valeur du κ est grande (26). Le *Prevalence Index* a été proposé pour estimer l'amplitude de cet effet. La prévalence ne semble cependant pas avoir induit de biais pour l'analyse de l'objectif principal. Notre analyse de sensibilité rétablissant un équilibre 50 % / 50 % entre « lésion ou doute » et « pas de lésion » montrait un faible effet de la prévalence, le coefficient κ restant stable, avec une concordance observée quasiment identique. Cette stratégie a ses limites du fait d'un biais de classement potentiel lié à l'incertitude du

diagnostic, car nous avons pris pour référence pour le tirage au sort le diagnostic du consensus d'experts et non un classement par une méthode de référence.

En revanche, la prévalence semble avoir biaisé les estimations pour les objectifs secondaires. Les familles de diagnostics avaient effectivement des effectifs faibles et leurs prévalences étaient donc faibles. Les PI pour la concordance pour chaque famille de diagnostics entre médecins des urgences et consensus d'experts étaient élevés et les PABAK étaient éloignés des coefficients κ.

Les effectifs marginaux étaient déséquilibrés (nombre de classements « pas de lésion » différent du nombre de classements « lésion ou doute ») et de façon symétrique (dans le même sens pour chacun des deux observateurs : plus de classements « pas de lésion » que « lésion ou doute »). Mais cette symétrie n'était pas parfaite (les deux observateurs faisaient plus de classements « pas de lésion » que « lésion ou doute », mais l'un beaucoup plus que l'autre). Cette situation entraîne un second paradoxe (32–34) : les coefficients κ sont supérieurs à ceux que l'on observerait sans ces déséquilibres, à concordance observée égale. Le *Bias Index* permet d'estimer l'amplitude de cet effet, qui semble faible dans notre étude (valeur maximale observée : 0,15).

Une vision synthétique de l'effet de ces deux paradoxes est complexe, car ils agissent en sens inverse. Dans notre étude, les deux paradoxes sont vraisemblablement présents mais de façon modérée. On observe en effet un PI peu élevé, un BI faible et un PABAK (index composite corrigeant le κ à la fois par la prévalence et les différences de comportement) peu différent du κ. Malgré les paradoxes entre la concordance observée et le coefficient κ, la validité interne semble donc assurée dans l'analyse pour l'objectif principal. Mais des différences notables entre κ et PABAK apparaissent dans certaines analyses secondaires où la prévalence a joué un rôle plus important (concordance par famille de diagnostics). L'effet des différences entre observateurs était faible (BI peu élevés) mais la prévalence a pu biaiser les résultats pour les lésions plus rares (PI élevés). Le premier paradoxe a joué un rôle plus important que le second et il existait donc un biais à l'origine d'une probable sous-estimation du coefficient κ.

Limites liées à l'absence de méthode de référence

Il est capital de clarifier que, à l'inverse d'une étude de validité diagnostique (sensibilité, spécificité), une étude de fiabilité ne nécessite pas de comparaison à une méthode de référence (« *gold standard* »). En effet l'objectif est de mesurer la capacité de l'examen à donner des résultats (les interprétations) stables. Peu importe qu'elles soient justes ou fausses, car ce n'est pas la question posée ici et que cela ne change pas les résultats.

En revanche, pour l'interprétation pratique des résultats, cette comparaison à une méthode de référence aurait été très utile. En effet, même l'observateur sensé être le plus sensible (le référent) peut faire des erreurs de classement. Et d'ailleurs, il n'est pas possible d'affirmer qu'il s'agissait réellement de l'observateur le plus sensible. La tendance observée à une augmentation du nombre de lésions diagnostiquées avec le niveau d'expérience pourrait en réalité être liée à une augmentation des faux positifs. Par ailleurs, lors de la réunion entre relecteurs, l'étude des éléments radiographiques analysés par les relecteurs aurait pris plus de sens si le diagnostic réel avait été connu. De plus, l'utilisation de la méthode de référence avant l'étude aurait permis d'inclure des patients aux lésions sélectionnées et donc de proposer d'emblée aux relecteurs un échantillon où les prévalences de « lésion ou doute » et « pas de lésion » était de 50 % chacun. En effet, des auteurs (33) ont proposé cette méthode pour contrôler l'influence de la prévalence sur le coefficient κ et donc le biais potentiel qui en découle. Cette dernière proposition a cependant été critiquée (35).

La méthode de référence pour le diagnostic des lésions traumatiques du poignet est la chirurgie par arthroscopie (36,37). Cette méthode permet la visualisation directe des lésions, mais il s'agit d'une méthode invasive, présentant les risques d'une intervention chirurgicale, et notamment des risques septiques. Une justification forte est donc nécessaire. Les études et la pratique clinique utilisent donc plutôt des examens paracliniques de validité et fiabilité proches de l'arthroscopie (36,38,39). Il peut s'agir d'IRM ou de tomodensitométrie, ainsi que de la scintigraphie osseuse.

Ces méthodes n'ont pas été utilisées dans l'étude pour des raisons éthiques mais également pour des problèmes de faisabilité (caractère rétrospectif compliquant fortement la réalisation avec un risque élevé de perdus de vue). Il est probable que les praticiens aient eu plus tendance à diagnostiquer par excès que par défaut (c'est-à-dire qu'il y ait eu plus de faux positifs que de faux négatifs), du fait de l'hypothèse même de l'étude et de leur mise en situation d'expert. On peut aussi faire l'hypothèse que certains relecteurs pouvant se sentir en compétition avec le second relecteur (malgré les conditions d'indépendance et

d'anonymat), aient pu craindre de rater des diagnostics et ainsi classer de façon plus large qu'ils ne l'auraient fait en situation réelle.

Autres limites principales

La taille d'échantillon n'était probablement pas suffisante pour les analyses pour les objectifs secondaires. Il en a résulté un manque de puissance, mais également un risque de biais dans l'estimation de certains paramètres. De plus, cette faible taille a parfois imposé l'utilisation de tests non paramétriques (test exact de Fisher, test de Kruskall-Wallis), pour lesquels la mise en évidence d'une différence est moins aisée. Cet effet se voit notamment à travers les IC95 % dont l'étendue est large pour de nombreux paramètres. Il est particulièrement marqué pour les concordances intra-observateur par relecteur qui incluent la valeur 0 dans leur IC95 %. Il y a lieu de penser que le manque de puissance a ici joué un rôle, car il semble peu envisageable qu'un observateur ne concorde pas avec lui-même au-delà du hasard.

Les coefficients pondérés apparaissaient élevés. La sous-représentation des modalités autres que « pas de lésion » a pu jouer un rôle. En effet, la proportion importante de la modalité « pas de lésion » a pu diluer des différences potentielles, même avec l'utilisation d'une pondération. Il aurait été nécessaire d'inclure plus de patients des autres familles. Le manque de puissance a donc probablement conduit à une surestimation des coefficients κ pondérés. Cet effet est classique des études de concordance pour les variables qualitatives à plus de deux modalités (27).

La concordance intra-observateur apparaît très faible et à la limite de la significativité. Les relectures ayant eu lieu peu de temps avant la réunion entre les relecteurs, leur comportement a pu en être modifié, notamment avec une tendance excessive à l'utilisation du doute. De plus, à cette phase de l'étude, les relecteurs avaient eu connaissance des résultats de la concordance inter-observateurs. La sachant faible, cela a dû augmenter leur propension à remettre leur diagnostic en question et préférer douter là où précédemment ils avaient fait un diagnostic lésionnel. On note ainsi une proportion très faible de lésions diagnostiquées. Une autre hypothèse serait que tous les observateurs se soient lassés au bout d'une année d'implication dans cette étude. Par ailleurs, la comparabilité a pu être compromise par la présentation des radiographies dans un diaporama PowerPoint. Même si ce choix a été fait suite au constat que les relecteurs avaient peu utilisé les outils d'aide diagnostique lors de la première séance, il est possible que la qualité des images en ait été modifiée. Un renouvellement de l'analyse de la concordance intra-observateur dans un protocole dédié serait intéressant pour confirmer ou infirmer ces résultats. En pratique, il

semble que tous ces éléments aient conduit à une sous-estimation des coefficients κ pour la concordance intra-observateur.

Les données manquantes ont représenté un problème important. Leur proportion très élevée pour certaines variables a induit un manque de puissance pour les études des associations entre les caractéristiques des malades et leur classement ou la présence d'une discordance. L'absence de variables cliniques associées significativement à une discordance y est sûrement liée. Nous n'avons donc pas pu réaliser de modèle multivariable, ce qui ne nous a pas permis d'étudier la présence de facteurs de confusion ou d'ajuster les analyses. Nous n'avons également pas pu étudier les variables associées à la présence d'une lésion.

Pour les cas où les données étaient manquantes de façon informative, c'est-à-dire qu'il existait une association statistiquement significative entre présence d'une donnée manquante et classement de la radiographie, les patients étaient plus souvent classés « lésion ou doute » que « pas de lésion ». Cela peut s'expliquer par le fondement même du diagnostic d'entorse simple du poignet, qui est un diagnostic d'élimination. Les médecins des urgences avaient donc sûrement tendance à étoffer la description de leur examen clinique pour montrer qu'ils avaient éliminé les diagnostics différentiels. Cela signifie que ces variables étaient sûrement recherchées même en cas de lésion, mais non enregistrées dans le dossier. La perte d'information pour ces variables est donc sûrement plus liée au manque de temps qu'au fait qu'elles ne sont pas recherchées. Cependant, la place trop importante (au vu de nos résultats) donnée à la radiographie standard du poignet en urgence est également mise en exergue ici. Notamment en cas de lésion ou de doute, les observations médicales se limitaient souvent à recopier le résultat de l'interprétation radiographique. L'examen clinique n'était alors pas renseigné et on ne peut exclure qu'il n'ait simplement pas été fait, car la radiographie n'était pas remise en question.

La méthode de construction du consensus d'experts présente également plusieurs faiblesses. D'une part, le nombre de relectures était faible (deux ou trois selon l'accord direct ou non des experts). D'autre part, il n'y avait pas de discussion en direct des relecteurs qui aurait pu aboutir à un consensus plus affiné. Cependant, la séance de relecture à distance a montré que le consensus n'était obtenu qu'assez rarement. Il aurait été possible de réaliser une étude de la concordance entre les quatre observations de chaque radiographie (médecin des Urgences, premier et deuxième expert, référent). Mais les méthodes statistiques qu'une telle étude implique sont moins fiables et imposent des échantillons plus larges.

La constitution de la base de données par des codes PMSI a pu créer un problème de représentativité. Pour être fiable, ce système implique que les codes aient été rentrés correctement par les médecins des urgences. Pour des raisons pratiques (complexité du système, manque de temps), cela n'est pas toujours le cas. De plus, en cas de lésions multiples, il arrive que seule une des lésions soit codée (par exemple uniquement « plaie de la face » chez un patient avec une plaie de la face et un traumatisme du poignet) puisque seul le « diagnostic principal » est obligatoire pour la validation administrative du dossier. Dans ces situations, le patient n'aurait pas été identifié dans notre étude. Les patients potentiellement manquants par mauvais codage le sont probablement de façon aléatoire (absence d'association avec une caractéristique du patient). D'un point de vue statistique, ce processus, bien que diminuant le nombre de sujets donc la puissance, ne crée pas de biais de sélection. En revanche, dans les cas où les motifs d'admission étaient multiples, un processus de sélection pourrait s'engager, car les sujets ne seraient pas manquants au hasard (patients présentant des lésions multiples dans le cadre de traumatismes plus graves).

Le sens de la distorsion de l'estimation de la concordance qui en résulte est difficile à anticiper. On peut penser que les radiographies non incluses seraient celles de traumatismes plus graves, donc de plus haute énergie. Dès lors, les lésions auraient tendance à être plus graves et radiologiquement plus marquées. La concordance serait donc plus élevée si ces malades étaient inclus. D'un autre côté, pour les malades présentant des lésions engageant le pronostic vital, la gestion des problèmes non vitaux est différée. Les lésions fermées du poignet sont alors mises au second plan. Dans ce cas, la concordance aura tendance à être plus faible, car les relecteurs ne se concentreront que sur le poignet, à l'inverse du lecteur initial aux Urgences. Ils détecteront donc plus souvent des lésions que les médecins des urgences.

Parmi les patients exclus, 37 l'étaient par absence de radiographie dans le dossier médical. Soit les patients avaient réalisé leurs clichés dans un autre centre (prescrits par leur médecin traitant), soit le médecin des urgences a jugé qu'il n'y avait pas d'indication à une radiographie. Les patients présentant des entorses simples ont alors été peut-être exclus plus souvent que des patients avec des lésions. En effet, les patients sortis sans radiographie avaient sûrement des symptômes peu marqués, donc potentiellement plutôt des entorses simples. De même pour les patients ne consultant pas aux urgences d'emblée. La prévalence observée des entorses simples est donc peut-être sous-estimée dans l'étude.

Nous avons fait le choix d'exclure les fractures extra-articulaires du radius distal. En effet, ces fractures sont très fréquentes (40) (incidence estimée de 165 pour 100 000) et représentent un profil spécifique de malades (chute de la personne âgée ostéoporotique). La surreprésentation de cette catégorie particulière risquait donc de biaiser certaines estimations, notamment les associations entre les caractéristiques du malade et la présence d'une discordance. Par ailleurs, ces fractures représentent une problématique un peu différente de celle qui nous intéresse (lésions à risque de passer inaperçues et d'évolution péjorative). En effet, leur diagnostic radiographique est rarement problématique et leur caractère non articulaire rend leur potentiel arthrogène faible et limité aux déplacements majeurs qui ne seraient pas pris en charge par défaut de consultation du malade (situation en pratique assez rare). Leur exclusion a donc probablement diminué la concordance estimée.

Le centre où a été réalisée l'étude présente deux spécificités en radiologie, qui peuvent avoir compromis la représentativité. D'une part les radiographies sont toutes interprétées par un radiologue. Cette pratique, fréquente dans les CHU, n'est pas toujours la règle dans les hôpitaux périphériques ou dans les structures privées. Cela peut également avoir entraîné un biais en surestimant la concordance avec les experts, notamment radiologues. D'autre part, la relecture est rarement faite par un radiologue expert. La majorité des radiologues présents aux Urgences sont des internes, souvent en début de formation. Ainsi, deux des trois radiologues de jour et les deux radiologues présents la nuit sont des internes. Dès lors, bien que la présence de spécialistes ait pu augmenter la concordance, cet effet est probablement limité. Par ailleurs, cette spécificité locale n'est pas représentative des centres qui n'ont pas d'internes.

Le caractère monocentrique était donc une limite de l'étude. Il aurait été nécessaire, pour obtenir une meilleure représentativité, d'inclure plusieurs centres. Pour limiter l'apparition d'un effet centre dans les relectures et le consensus d'experts, nous avons sélectionné des praticiens de profils divers (civils et militaire, formés dans des villes différentes).

Comparaisons avec les données de la littérature

La prise en charge globale des traumatismes fermés du poignet (c'est-à-dire sans orientation diagnostique *a priori*) est peu documentée dans la littérature. En revanche, la fracture occulte du scaphoïde a fait l'objet de nombreux travaux.

La reproductibilité des radiographies dans les traumatismes du poignet n'a à notre connaissance fait l'objet que de peu d'études (41,42), toutes centrées sur les fractures du scaphoïde et dans lesquelles les coefficients κ étaient faibles (de 0,14 à 0,39). Pour les radiographies osseuses tous sites confondus en traumatologie (43), les études de fiabilité disponibles retrouvent une meilleure concordance que notre étude (κ = 0,70) chez des experts. La concordance observée pour les lésions ligamentaires est également meilleure (κ = 0,44) que celle retrouvée ici (44). De même, la fiabilité des classifications radiographiques utilisées notamment pour les fractures du radius distal est variable mais souvent moyenne (maximum retrouvé : 0,6) (45–47). Il a aussi été démontré que la fiabilité de l'évaluation des angles sur les radiographies était mauvaise (48). Ces résultats vont donc dans le même sens de notre étude, en montrant une mauvaise reproductibilité pour les radiographies du poignet.

Il semble que la connaissance des éléments cliniques améliore la reproductibilité (19) dans l'interprétation des radiographies. Il existe un score clinique pour la suspicion des fractures du scaphoïde (*Clinical Scaphoid Score*) (49). Ce score, bien que validé par une étude publiée dans une revue de qualité, n'est pas diffusé en pratique. On ne trouve d'ailleurs pas d'études y faisant référence de façon explicite. Sa fiabilité n'a également pas été estimée. Les signes cliniques pour le diagnostic de fracture du scaphoïde sont les seuls étudiés et validés (50,51) (douleur à la compression de la colonne du pouce, du tubercule du scaphoïde et de la tabatière anatomique). La sensibilité de ces signes associés serait proche de 100 % mais cela n'est pas consensuel (52,53). Dans notre étude, les données manquantes étaient trop nombreuses pour étudier ces associations. Une méta-analyse récente (54) retrouvait une faible valeur prédictive positive (21 %) de l'examen clinique dans les fractures du scaphoïde.

Les proportions de lésions passant inaperçues que l'on trouve dans la littérature sont très variables et discutables. Dans une cohorte rétrospective en 2002 (11), 2 % de lésions non diagnostiquées initialement étaient retrouvées. Cependant, il s'agissait uniquement de patients ayant consulté à distance dans le même hôpital. Il y a fort à penser qu'un patient présentant une évolution défavorable ne consultera pas dans le même service, du fait d'une

possible perte de confiance. Une sous-évaluation est donc probable. Une autre série retrouvait en 2010 (14) 63 % de lésions occultes. Mais ces patients étaient sélectionnés (examen clinique « évocateur d'une fracture ») et seuls 38 de ces 63 % correspondaient à de véritables fractures. Les autres lésions étiquetées « osseuses » étaient en réalité des œdèmes intra-médullaires qui, bien qu'à l'origine de douleurs prolongées et justifiant une immobilisation, sont sans potentiel arthrogène documenté (55). En 2012, une équipe norvégienne (13) a réalisé de façon systématique pendant un an une IRM à tous les patients admis dans un service d'urgence pour un traumatisme fermé du poignet de moins d'une semaine, avec radiographies considérées sans lésions. Ils retrouvaient 80 % de lésions occultes. Cependant, si l'on écarte les œdèmes et contusions simples, ce chiffre était de 41 %. Une étude de l'équipe de Besançon (56) retrouvait également 41 % de fractures occultes chez des patients avec une suspicion clinique de fracture et des radiographies considérées normales, à qui étaient réalisés une scintigraphie en urgence. Le chiffre observé dans notre étude est donc compatible avec ces données.

Plusieurs stratégies d'imagerie ont été évaluées pour la prise en charge des traumatismes du poignet où existe un doute, mais surtout pour la recherche de lésions précises. La lésion la plus étudiée semble être la fracture occulte du scaphoïde. Trois examens principaux sont utilisés en pratique : la tomodensitométrie, l'IRM et la scintigraphie osseuse (SO).
La tomodensitométrie est la plus réalisée en pratique pour des raisons d'accessibilité. Ses performances diagnostiques sont bonnes (sensibilité et spécificité respectivement jusqu'à 86 % et 100 % pour les fractures du scaphoïde (57,58)). Il semble cependant qu'elle soit surtout adaptée pour la mise en évidence de lésions corticales (59). La tomodensitométrie a pour inconvénient principal de générer des faux positifs liés au réseau vasculaire (60), ce qui signifie qu'elle est plus utile pour éliminer une fracture que pour en diagnostiquer une (61). Ainsi, la reproductibilité de l'interprétation des tomodensitométries dans le diagnostic des fractures du scaphoïde est assez modérée (62). Par ailleurs, en l'absence de la réalisation d'un arthro-scanner (impliquant donc une injection intra-articulaire), elle ne permet pas de diagnostiquer les lésions ligamentaires. Il est à noter que l'irradiation reçue lors d'une tomodensitométrie du poignet est très faible, d'environ 0,03 milli Sievert, soit l'équivalent de moins de 40 % d'une radiographie du thorax (63) (en comparaison, une tomodensitométrie du rachis lombaire correspond à une dose d'environ 20 milli-Sievert). L'IRM est globalement retrouvée comme examen avec les meilleures performances diagnostiques (38,39,64–66), mais également le plus coûteux. Sa fiabilité est très bonne (κ entre 0,87 et 0,95 (65,67)). Une étude (68) concluait que dans les traumatismes du poignet, l'IRM augmentait la valeur prédictive positive de la nécessité d'un traitement, mais n'avait pas d'intérêt dans l'identification de malades pouvant être renvoyés à domicile sans suivi. La SO, dont

l'accessibilité est bonne, présente une validité diagnostique très proche des deux autres examens, mais peut souffrir de son manque de précision dans la localisation exacte de la lésion (69) qui justifie son association systématique en cas de positivité à une tomodensitométrie. Il a cependant été relevé par certains auteurs que cet examen entraînait plus de faux positifs que les autres (70). En effet, des traumatismes simples (tel que les œdèmes intra-osseux) peuvent entraîner une fixation.

Ces trois examens ont été comparés entre eux dans plusieurs études. Des auteurs (71) concluaient récemment que la validité diagnostique de l'IRM et de la SO étaient proches, mais les données insuffisantes pour la tomodensitométrie. Pour d'autres auteurs, les performances de l'IRM et la tomodensitométrie sont équivalentes (57). Une méta-analyse concluait que pour les fractures du scaphoïde, l'IRM était l'examen le plus précis, que les radiographies à 15 jours et le scanner étaient tous les deux moins sensibles, et la scintigraphie moins spécifique (61). Une récente revue de la *Cochrane Database* (72) sur les stratégies d'imagerie pour la suspicion de fracture du scaphoïde à radiographies normales concluait que la SO était le meilleur examen mais en insistant sur ses inconvénients, dont une dose significativement supérieure aux autres techniques.

L'échographie a également été proposée, notamment pour les lésions ligamentaires (73,74). Mais son apport pour le diagnostic des fractures est moindre et elle présente une reproductibilité médiocre. Par ailleurs, les chirurgiens n'apprécient pas cet examen qui, à la différence des autres, n'est interprétable que par celui qui le fait, est dépendant de l'opérateur et ne donne pas de repères anatomiques en vue d'une intervention. D'autres auteurs ont également proposé de réaliser trois incidences radiographiques plutôt que deux, avec une augmentation significative du coefficient κ (75) ; d'autres suggèrent même quatre incidences, en incluant des clichés en pronation et déviation ulnaire (76,77). Ces solutions simples présentent l'inconvénient d'une irradiation supplémentaire, mais qui est assez modérée, la dose reçue pour une radiographie standard du poignet étant très faible (78). En revanche, il n'est pas démontré que cette stratégie augmente la reproductibilité ni la validité des interprétations.

En ce qui concerne les lésions ligamentaires, leur diagnostic est plus complexe car elles impliquent souvent une opacification articulaire (arthro-scanner, arthro-IRM). Il ne semble pas y avoir de supériorité d'une des techniques par rapport à l'autre (79). L'IRM a de bonnes performances diagnostiques (80) : une sensibilité et une spécificité de 88 % pour les ruptures du ligament scapho-lunaire, allant jusqu'à 100 % de spécificité pour l'arthro-IRM. De

même, sans opacification, la sensibilité pour les lésions du complexe fibro-cartilagineux triangulaire est d'environ 90 % et la spécificité entre 74 et 100 %.

Une des stratégies classiques pour les traumatismes du poignet avec radiographies normales est l'immobilisation temporaire par attelle avec réévaluation dans les deux semaines (81), comme cela a été souvent proposé par les relecteurs lors de leur réunion. Une large étude scandinave (82) appuyait l'intérêt de radiographies répétées (avec cinq incidences) au quatorzième jour en cas de suspicion de fracture du scaphoïde. Pour la même lésion, des auteurs ont démontré qu'il n'y avait pas plus de complications tant que le traitement était réalisée dans les quatre semaines qui suivent le traumatisme (83). Cependant, l'intérêt de cette stratégie est contesté (82,84,85), plusieurs études ayant démontré une très faible augmentation du nombre de fractures diagnostiquées avec les radiographies répétées à distance, avec des indicateurs de validité et fiabilité diagnostiques faibles (86). Par ailleurs, la valeur prédictive négative d'une présomption clinique forte d'une fracture du scaphoïde (74 % dans une méta-analyse (54)) signifie que les trois quarts des patients auront une immobilisation injustifiée pendant quinze jours, avec ce que cela implique en termes d'arrêt de travail et de risque de raideur. D'autres auteurs ont proposé une réévaluation radiographique à six semaines, mais cette attitude n'est pas consensuelle (42).

Le rapport coût-efficacité semble être en faveur de la réalisation d'une IRM en urgence. Des auteurs ont démontré que la réalisation d'une IRM immédiate était équivalente, en termes de coût (54), à une immobilisation pendant 15 jours suivie de radiographies répétées pour les suspicions de fracture du scaphoïde. Deux études récentes (87,88) ont comparé les rapports coût-efficacité des différentes stratégies diagnostiques pour les suspicions de fractures du scaphoïde (ce qui implique donc déjà qu'elles soient évoquées), et concluaient que la meilleure était la tomodensitométrie ou l'IRM au moment du traumatisme. De même, une étude réalisée dans un service d'urgences en milieu rural concluait que la réalisation d'une IRM en urgence systématique pour les suspicions de fracture du scaphoïde était moins coûteuse que la prise en charge standard en termes de coûts globaux de soins (89). Une autre étude concluait à des coûts équivalents entre ces deux stratégies, avec une durée d'immobilisation et d'arrêt de travail deux fois moindre chez les patients présentant une entorse simple (90). D'autres auteurs, avec les mêmes résultats en ce qui concerne les coûts, insistaient sur l'avantage que présente l'IRM de diagnostiquer un panel large d'autres lésions, notamment ligamentaires (91).

Il existe peu d'études sur les conséquences à long terme d'une absence de prise en charge aiguë. Il a été montré, toujours dans les fractures du scaphoïde, que l'immobilisation rapide (dans la semaine), et donc le diagnostic rapide, prévenait le risque de pseudarthrose (92). Les taux de pseudarthrose pour les fractures du scaphoïde prises en charge au-delà de quatre semaines sont élevés (de 20 à 88 % (83,93)).

Implications des résultats

La radiographie standard du poignet ne semble pas être un examen à recommander de façon isolée en pratique courante dans les traumatismes fermés récents du poignet. Notre étude doit donc avant tout encourager les praticiens à la prudence. Il est nécessaire de garder à l'esprit que les conclusions faites par un observateur ne sont pas toujours reproductibles. Les praticiens, y compris experts, ne devraient pas hésiter à demander l'avis de leurs confrères ou poursuivre la prise en charge par des examens plus sensibles. La meilleure fiabilité de l'interprétation par les radiologues qui semble se dégager de nos résultats doit ainsi inciter les chirurgiens à faire appel à leur relecture au moindre doute. De même, les médecins prenant en charge les traumatismes du poignet doivent être en mesure d'identifier les situations douteuses et accepter que cette situation intermédiaire, le doute, doit être considérée comme une entité à part entière, impliquant une prise en charge spécifique. Ils ne devraient également pas faire preuve d'excès de confiance, puisqu'il semble que la stabilité des interprétations chez un même observateur soit mauvaise. Une remise en question systématique des diagnostics est nécessaire, surtout lorsqu'il s'agit de conclure « entorse simple » et donc de sortir le malade du système de soins.

Dans l'optique d'améliorer la prise en charge actuelle (décision diagnostique centrée sur le couple « examen clinique – interprétation radiographique »), il serait indispensable que l'examen clinique soit systématique et rigoureux. Or, cela n'est pas le cas en pratique : d'une part, l'examen requiert du temps et de l'expérience (94) ; d'autre part, la douleur le limite le plus souvent en urgence. De plus, les études indiquent un faible apport de l'examen clinique au diagnostic. Pour améliorer l'interprétation radiographique, le point majeur soulevé par l'étude est la qualité des clichés en urgence, trop souvent mauvaise. Un travail de sensibilisation auprès des manipulateurs en radiologie semble être un complément indispensable. La systématisation du positionnement du poignet pour les clichés est capitale pour une interprétation valable. Il conviendrait également de standardiser l'interprétation. Des signes radiologiques validés existent, par exemple pour le diagnostic de rupture du ligament scapho-lunaire (95). Mais ces informations sont peu diffusées, même chez les spécialistes. Par ailleurs, elles concernent des troubles précis, qu'il faut connaître pour rechercher. Il n'existe pas à notre connaissance de grille simple et validée pour une analyse

systématique des clichés sans orientation diagnostique *a priori*. Pour envisager une utilisation courante, il serait nécessaire de disposer d'éléments d'orientation simples et compatibles avec la pratique d'urgence. Enfin, il semble impératif d'améliorer la formation sur les lésions ligamentaires. Elles sont de diagnostic difficile, et devraient être évoquées plus fréquemment. Il est indispensable que ces lésions soient prises en charge le plus tôt possible (8,83). Passées les premières semaines, il n'existe pas d'intervention ayant fait la preuve de son efficacité dans la réduction du risque d'arthrose.

Il convient donc d'envisager les stratégies diagnostiques alternatives. Les données disponibles ne sont pas en faveur de l'immobilisation temporaire avec réévaluation à distance. Cette stratégie est également peu applicable aux praticiens généralistes car elle requiert un examen clinique adapté. L'avis d'experts systématique n'est également pas encouragé par nos résultats. De plus, le rapport coût-efficacité de ce choix n'est pas clair. Il n'est pas possible de dire le nombre de consultations spécialisées à réaliser pour diagnostiquer une lésion supplémentaire. Ces consultations sont plus chères et logistiquement lourdes (déplacement du malade, journée non travaillée, mobilisation d'un spécialiste). Pour être justifiées, ces contraintes doivent être inférieures à celles qu'impliquent les lésions non diagnostiquées (arrêts de travail prolongés, lésions dégénératives avec recours itératifs aux soins, situation de handicap avec reclassement professionnel et actes chirurgicaux lourds notamment). Il est aussi probable que les chirurgiens n'accepteraient pas de recevoir les patients en première intention en l'absence sinon de lésion, au moins d'une suspicion de lésion. L'incidence notable de ces traumatismes entraînerait une augmentation importante de l'activité de consultations. Un autre problème se poserait également : à quels spécialistes adresser les malades en première intention ? Les résultats des comparaisons avec le référent feraient plutôt recommander l'avis d'un radiologue que d'un chirurgien. Cependant, les radiologues ne sont pas formés à la prise en charge clinique des urgences et cette solution n'est donc pas envisageable. Quant aux chirurgiens, se pose également le problème de la sur-spécialisation. Au CHU de Bordeaux notamment, les lésions du carpe et les lésions ligamentaires sont prises en charge en chirurgie plastique, mais les fractures de l'avant-bras en chirurgie orthopédique. Des réorientations des malades par ces spécialités sont possibles mais complexifient le parcours de soins. Enfin, tous les centres d'urgences ne disposent pas de chirurgiens sur place et, dans les régions à faible démographie médicale, les patients seront toujours vus en première intention par les médecins des urgences ou leur médecin traitant.

La troisième stratégie, la réalisation systématique d'examens plus poussés chez tous les malades, semble peu envisageable. Des études ont démontré une équivalence ou une diminution des coûts avec l'IRM initiale systématique, mais elles ne concernaient que la fracture du scaphoïde : les malades sont donc sélectionnés. En l'absence de sélection, l'incidence notable des traumatismes du poignet ne permet pas d'envisager des examens coûteux pour tous les malades.

Au total, il apparaît que la solution la plus adaptée est le développement d'une règle de décision clinique pour poser l'indication ou non d'une imagerie complémentaire aux radiographies standard. Il s'agirait donc de sélectionner les malades à risque de présenter une lésion. Pour cela, il conviendra donc de mettre en évidence un certain nombre de caractéristiques du patient, de son traumatisme, de son examen clinique ou de sa radiographie permettant d'identifier le risque de présence d'une lésion. Ces caractéristiques n'ont été que peu étudiées à notre connaissance. Van den Brand *et al* (96) citent deux travaux à ce sujet, mais qui n'ont pas été publiés. Ceux-ci ne retrouvaient pas les mêmes critères et ne s'intéressaient qu'aux fractures (97,98).

L'intérêt d'une règle de décision pour les traumatismes du poignet a été débattue (96,99,100), jusqu'au développement récent d'une règle de décision clinique, les *Amsterdam Wrist Rules* (101). Cependant, ces différents travaux s'inscrivent dans un cadre différent du nôtre, car ils étudiaient l'intérêt d'une règle pour déterminer l'indication ou non d'une radiographie initiale. À notre sens, la radiographie reste indispensable chez tous les malades et c'est d'ailleurs la conclusion d'une étude sur l'intérêt d'une telle règle (96). Par ailleurs, ces études présentent pour faiblesse commune de ne pas prendre en compte les lésions ligamentaires. Or, comme nous l'avons souligné précédemment, le diagnostic et la prise en charge précoce de ces lésions sont capitaux. Une règle de décision clinique valable pour les traumatismes du poignet devra donc impérativement prendre en compte les lésions ligamentaires.

Les limites de l'étude, notamment le manque de puissance et les données manquantes, encouragent donc à réaliser un protocole prospectif avec recueil standardisé pour contrôler au maximum les biais potentiels. Ce travail sera la continuité naturelle de celui réalisé pour cette étude. Pour une mise en œuvre réelle en pratique, il sera indispensable de disposer de critères simples, rapides et sans ambiguïté pour permettre de classer les malades. Le recueil de données devra être le plus large possible pour rechercher les signes les plus prédictifs de lésions. Mais des choix seront malgré tout nécessaires pour ne pas alourdir inutilement la grille de recueil, qui devra être compatible avec les contraintes des médecins des urgences.

On peut envisager au CHU de Bordeaux un questionnaire DXCare, comme cela existe par exemple pour les consultations Orthoplast. Pour les critères d'Ottawa, une phase pilote avait été conduite pour évaluer et optimiser le questionnaire de recueil. Cette phase méritera d'être reproduite dans le protocole prospectif, que nous espérons mettre en œuvre prochainement.

REFERENCES

1. Ootes D, Lambers KT, Ring DC. The epidemiology of upper extremity injuries presenting to the emergency department in the United States. Hand. 2012 Mar;7(1):18–22.
2. Laulan J, Hérard J, Bacle G. Épidémiologie, physiopathologie et classification des traumatismes du poignet. Lett Médecine Phys Réadapt. 2009 Mar 29;25(1):4–8.
3. Duckworth AD, Jenkins PJ, Aitken SA, Clement ND, Court-Brown CM, McQueen MM. Scaphoid fracture epidemiology. J Trauma Acute Care Surg. 2012 Feb;72(2):E41–5.
4. Larsen CF, Lauritsen J. Epidemiology of acute wrist trauma. Int J Epidemiol. 1993 Oct;22(5):911–6.
5. Green D, Hotchkiss R, Pederson W, Wolfe S. Green's operative hand surgery - Fifth edition. Philadelphia (Pennsylvania) : Elsevier; 2005.
6. Graham B. Diagnosis of closed wrist injuries. Lancet. 1996 Jun 22;347(9017):1709.
7. Pin PG, Young VL, Gilula LA, Weeks PM. Wrist pain: a systematic approach to diagnosis. Plast Reconstr Surg. 1990 Jan;85(1):42–6.
8. Merle M, Dautel G. La main traumatique - 2 : Chirurgie secondaire - Le poignet traumatique. Paris : Éditions Masson; 1994.
9. Testut L. Articulation du poignet. In: Traité d'anatomie humaine - Tome premier. Paris : Octave Doin; 1899. p. 512–22.
10. Charpy A, Nicolas A, Prenant A, Jacques P. Articulations du poignet. In: Traité d'anatomie humaine - Tome Premier. Paris : Masson; 1892. p. 608–22.
11. Guly H. Injuries initially misdiagnosed as sprained wrist (Beware the sprained wrist). Emerg Med J. 2002 Jan;19(1):41–2.
12. Jones WA. Beware the sprained wrist. The incidence and diagnosis of scapholunate instability. J Bone Joint Surg Br. 1988 Mar;70(2):293–7.
13. Bergh TH, Lindau T, Bernardshaw SV, Behzadi M, Soldal LA, Steen K, et al. A new definition of wrist sprain necessary after findings in a prospective MRI study. Injury. 2012 Oct;43(10):1732–42.
14. Pierre-Jerome C, Moncayo V, Albastaki U, Terk MR. Multiple occult wrist bone injuries and joint effusions: prevalence and distribution on MRI. Emerg Radiol. 2010

May;17(3):179–84.
15. Fotiadou A, Patel A, Morgan T, Karantanas AH. Wrist injuries in young adults: the diagnostic impact of CT and MRI. Eur J Radiol. 2011 Feb;77(2):235–9.
16. Van der Molen AB, Groothoff JW, Visser GJ, Robinson PH, Eisma WH. Time off work due to scaphoid fractures and other carpal injuries in The Netherlands in the period 1990 to 1993. J Hand Surg Br. 1999 Apr;24(2):193–8.
17. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Ann Emerg Med. 1992 Apr;21(4):384–90.
18. Stiell IG, Wells GA. Methodological standards for the development of clinical decision rules in emergency medicine. Annals of Emergency Medicine. 1999;33:437–47.
19. Tudor GR, Finlay D, Taub N. An assessment of inter-observer agreement and accuracy when reporting plain radiographs. Clin Radiol. 1997 Mar;52(3):235–8.
20. Taourel P. Imagerie des urgences. Paris : Masson; 2001.
21. Dumontier C, Zeitoun F, Laredo J-D, Petrover D. Imagerie rhumatologique et orthopédique - Tome 2 - Rachis, ceinture scapulaire et membre supérieur. Paris : Sauramps Médical; 2013.
22. Bhat AK, Kumar B, Acharya A. Radiographic imaging of the wrist. Indian J Plast Surg. 2011;44(2):186–96.
23. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.
24. Fleiss J, Levin B, Myunghee CP. Statistical methods for rates and proportions, Third Edition. New York : John Wiley & Sons; 2003.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159–74.
26. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol. 1993 May;46(5):423–9.
27. Sim J, Wright CC. The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005 Mar 1;85(3):257–68.
28. Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. J Clin Epidemiol. 2012 Jul;65(7):778–84.
29. Direction du Système d’Information. Charte de l’utilisateur pour l’usage du

système d'information du CHU de Bordeaux. Bordeaux (France) : CHU de Bordeaux; 2014.
30. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J Clin Epidemiol. 2010 Aug;63(8):854–61.
31. Lucas N, Macaskill P, Irwig L, Moran R, Rickards L, Turner R, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). BMC Med Res Methodol. 2013;13:111.
32. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543–9.
33. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol. 1990;43(6):551–8.
34. Zwick R. Another look at interrater agreement. Psychol Bull. 1988 May;103(3):374–8.
35. Vach W. The dependence of Cohen's kappa on the prevalence does not matter. J Clin Epidemiol. 2005 Jul;58(7):655–61.
36. Kitay A, Wolfe SW. Scapholunate instability: current concepts in diagnosis and management. J Hand Surg. 2012 Oct;37(10):2175–96.
37. Hofmeister EP, Dao KD, Glowacki KA, Shin AY. The role of midcarpal arthroscopy in the diagnosis of disorders of the wrist. J Hand Surg. 2001 May;26(3):407–14.
38. Magee T. Comparison of 3-T MRI and arthroscopy of intrinsic wrist ligament and TFCC tears. AJR Am J Roentgenol. 2009 Jan;192(1):80–5.
39. Buijze GA, Wijffels MME, Guitton TG, Grewal R, van Dijk CN, Ring D, et al. Interobserver reliability of computed tomography to diagnose scaphoid waist fracture union. J Hand Surg. 2012 Feb;37(2):250–4.
40. Sigurdardottir K, Halldorsson S, Robertsson J. Epidemiology and treatment of distal radius fractures in Reykjavik, Iceland, in 2004. Comparison with an Icelandic study from 1985. Acta Orthop. 2011 Aug;82(4):494–8.
41. Tiel-van Buul MM, van Beek EJ, Broekhuizen AH, Nooitgedacht EA, Davids PH, Bakker AJ. Diagnosing scaphoid fractures: radiographs cannot be used as a gold standard! Injury. 1992;23(2):77–9.
42. Mallee WH, Mellema JJ, Guitton TG, Goslings JC, Ring D, Doornberg JN, et al. 6-week radiographs unsuitable for diagnosis of suspected scaphoid fractures.

Arch Orthop Trauma Surg. 2016 Jun;136(6):771–8.
43. Robinson PJ, Wilson D, Coral A, Murphy A, Verow P. Variation between experienced observers in the interpretation of accident and emergency radiographs. Br J Radiol. 1999 Apr;72(856):323–30.
44. Gradl G, Neuhaus V, Fuchsberger T, Guitton TG, Prommersberger K-J, Ring D, et al. Radiographic diagnosis of scapholunate dissociation among intra-articular fractures of the distal radius: interobserver reliability. J Hand Surg. 2013 Sep;38(9):1685–90.
45. Kural C, Sungur I, Kaya I, Ugras A, Ertürk A, Cetinus E. Evaluation of the reliability of classification systems used for distal radius fractures. Orthopedics. 2010 Nov;33(11):801.
46. Andersen DJ, Blair WF, Steyers CM, Adams BD, el-Khouri GY, Brandser EA. Classification of distal radius fractures: an analysis of interobserver reliability and intraobserver reproducibility. J Hand Surg. 1996 Jul;21(4):574–82.
47. Plant CE, Hickson C, Hedley H, Parsons NR, Costa ML. Is it time to revisit the AO classification of fractures of the distal radius? Inter- and intra-observer reliability of the AO classification. Bone Jt J. 2015 Jun;97-B(6):818–23.
48. Gaskin JSH, Pimple MK, Wharton R, Fernandez C, Gaskin D, Ricketts DM. How accurate and reliable are doctors in estimating fracture angulation? Injury. 2007 Feb;38(2):160–2.
49. Bergh TH, Lindau T, Soldal LA, Bernardshaw SV, Behzadi M, Steen K, et al. Clinical scaphoid score (CSS) to identify scaphoid fracture with MRI in patients with normal x-ray after a wrist trauma. Emerg Med J. 2014 Aug;31(8):659–64.
50. Parvizi J, Wayman J, Kelly P, Moran CG. Combining the clinical signs improves diagnosis of scaphoid fractures. A prospective study with follow-up. J Hand Surg Br. 1998 Jun;23(3):324–7.
51. Grover R. Clinical assessment of scaphoid injuries and the detection of fractures. J Hand Surg Br. 1996 Jun;21(3):341–3.
52. Waizenegger M, Barton NJ, Davis TR, Wastie ML. Clinical signs in scaphoid fractures. J Hand Surg Br. 1994 Dec;19(6):743–7.
53. Freeland P. Scaphoid tubercle tenderness: a better indicator of scaphoid fractures? Arch Emerg Med. 1989 Mar;6(1):46–50.
54. Dorsay TA, Major NM, Helms CA. Cost-effectiveness of immediate MR imaging versus traditional follow-up for revealing radiographically occult scaphoid

fractures. AJR Am J Roentgenol. 2001 Dec;177(6):1257–63.
55. Thavarajah D, Syed T, Shah Y, Wetherill M. Does scaphoid bone bruising lead to occult fracture? A prospective study of 50 patients. Injury. 2011 Nov;42(11):1303–6.
56. Lepage D, Obert L, Garbuio P, Tropet Y, Paratte B, Runge M, et al. Contribution of quantitative radio-scintigraphy to diagnosis of wrist injuries undetected on plain films: a prospective study of 154 cases. Rev Chir Orthopédique Réparatrice Appar Mot. 2004 Oct;90(6):542–9.
57. Mallee W, Doornberg JN, Ring D, van Dijk CN, Maas M, Goslings JC. Comparison of CT and MRI for diagnosis of suspected scaphoid fractures. J Bone Joint Surg Am. 2011 Jan 5;93(1):20–8.
58. Ilica AT, Ozyurek S, Kose O, Durusu M. Diagnostic accuracy of multidetector computed tomography for patients with suspected scaphoid fractures and negative radiographic examinations. Jpn J Radiol. 2011 Feb;29(2):98–103.
59. Memarsadeghi M, Breitenseher MJ, Schaefer-Prokop C, Weber M, Aldrian S, Gäbler C, et al. Occult scaphoid fractures: comparison of multidetector CT and MR imaging--initial experience. Radiology. 2006 Jul;240(1):169–76.
60. Adey L, Souer JS, Lozano-Calderon S, Palmer W, Lee S-G, Ring D. Computed tomography of suspected scaphoid fractures. J Hand Surg. 2007 Jan;32(1):61–6.
61. Yin Z-G, Zhang J-B, Kan S-L, Wang X-G. Diagnostic accuracy of imaging modalities for suspected scaphoid fractures: meta-analysis combined with latent class analysis. J Bone Joint Surg Br. 2012 Aug;94(8):1077–85.
62. De Zwart AD, Beeres FJP, Kingma LM, Otoide M, Schipper IB, Rhemrev SJ. Interobserver variability among radiologists for diagnosis of scaphoid fractures by computed tomography. J Hand Surg. 2012 Nov;37(11):2252–6.
63. Biswas D, Bible JE, Bohan M, Simpson AK, Whang PG, Grauer JN. Radiation exposure from musculoskeletal computerized tomographic scans. J Bone Joint Surg Am. 2009 Aug;91(8):1882–9.
64. Jenkins PJ, Slade K, Huntley JS, Robinson CM. A comparative analysis of the accuracy, diagnostic uncertainty and cost of imaging modalities in suspected scaphoid fractures. Injury. 2008 Jul;39(7):768–74.
65. Breitenseher MJ, Metz VM, Gilula LA, Gaebler C, Kukla C, Fleischmann D, et al. Radiographically occult scaphoid fractures: value of MR imaging in detection.

Radiology. 1997 Apr;203(1):245–50.
66. Gaebler C, Kukla C, Breitenseher M, Trattnig S, Mittlboeck M, Vécsei V. Magnetic resonance imaging of occult scaphoid fractures. J Trauma. 1996 Jul;41(1):73–6.
67. Hunter JC, Escobedo EM, Wilson AJ, Hanel DP, Zink-Brody GC, Mann FA. MR imaging of clinically suspected scaphoid fractures. AJR Am J Roentgenol. 1997 May;168(5):1287–93.
68. Nikken JJ, Oei EHG, Ginai AZ, Krestin GP, Verhaar JAN, van Vugt AB, et al. Acute wrist trauma: value of a short dedicated extremity MR imaging examination in prediction of need for treatment. Radiology. 2005 Jan;234(1):116–24.
69. De Zwart AD, Beeres FJP, Rhemrev SJ, Bartlema K, Schipper IB. Comparison of MRI, CT and bone scintigraphy for suspected scaphoid fractures. Eur J Trauma Emerg Surg. 2015 Nov 10;
70. Van Beek EJ, van Buul MM, Broekhuizen AH. Diagnostic problems of scaphoid fractures: the value of radionuclide bone scintigraphy. Neth J Surg. 1990 Apr;42(2):50–2.
71. Yin Z-G, Zhang J-B, Kan S-L, Wang X-G. Diagnosing suspected scaphoid fractures: a systematic review and meta-analysis. Clin Orthop. 2010 Mar;468(3):723–34.
72. Mallee WH, Wang J, Poolman RW, Kloen P, Maas M, de Vet HCW, et al. Computed tomography versus magnetic resonance imaging versus bone scintigraphy for clinically suspected scaphoid fractures in patients with negative plain radiographs. Cochrane Database Syst Rev. 2015;6:CD010023.
73. Finlay K, Lee R, Friedman L. Ultrasound of intrinsic wrist ligament and triangular fibrocartilage injuries. Skeletal Radiol. 2004 Feb;33(2):85–90.
74. Bajaj S, Pattamapaspong N, Middleton W, Teefey S. Ultrasound of the hand and wrist. J Hand Surg. 2009 Apr;34(4):759–60.
75. Jackson ME, Henderson JE. Following trauma, should adult wrist radiographic examinations be two or three projections? Emerg Radiol. 2010 Mar;17(2):87–93.
76. Russin LD, Bergman G, Miller L, Griffin WM, Walter M, Bhargavan M, et al. Should the routine wrist examination for trauma be a four-view study, including a semisupinated oblique view? AJR Am J Roentgenol. 2003 Nov;181(5):1235–8.
77. Cheung GC, Lever CJ, Morris AD. X-ray diagnosis of acute scaphoid fractures. J Hand Surg Br. 2006 Feb;31(1):104–9.

78. Institut de Radioprotection et de Sûreté Nucléaire. Doses délivrées aux patients en scanographie et en radiologie conventionnelle - Résultats d'une enquête multicentrique en secteur public - Rapport DRPH/SER n°2010-12. Paris : Ministère de la Santé; 2010.
79. Lee RKL, Ng AWH, Tong CSL, Griffith JF, Tse WL, Wong C, et al. Intrinsic ligament and triangular fibrocartilage complex tears of the wrist: comparison of MDCT arthrography, conventional 3-T MRI, and MR arthrography. Skeletal Radiol. 2013 Sep;42(9):1277–85.
80. Ringler MD. MRI of wrist ligaments. J Hand Surg. 2013 Oct;38(10):2034–46.
81. Sprague B, Justis EJ. Nonunion of the carpal navicular. Modes of treatment. Arch Surg. 1974 May;108(5):692–7.
82. Munk B, Frøkjaer J, Larsen CF, Johannsen HG, Rasmussen LL, Edal A, et al. Diagnosis of scaphoid fractures. A prospective multicenter study of 1,052 patients with 160 fractures. Acta Orthop Scand. 1995 Aug;66(4):359–60.
83. Langhoff O, Andersen JL. Consequences of late immobilization of scaphoid fractures. J Hand Surg Br. 1988 Feb;13(1):77–9.
84. Duncan DS, Thurston AJ. Clinical fracture of the carpal scaphoid--an illusionary diagnosis. J Hand Surg Br. 1985 Oct;10(3):375–6.
85. Jacobsen S, Hassani G, Hansen D, Christensen O. Suspected scaphoid fractures. Can we avoid overkill? Acta Orthop Belg. 1995;61(2):74–8.
86. Low G, Raby N. Can follow-up radiography for acute scaphoid fracture still be considered a valid investigation? Clin Radiol. 2005 Oct;60(10):1106–10.
87. Yin Z-G, Zhang J-B, Gong K-T. Cost-effectiveness of diagnostic strategies for suspected scaphoid fractures. J Orthop Trauma. 2015 Mar 9;
88. Karl JW, Swart E, Strauch RJ. Diagnosis of Occult Scaphoid Fractures: A Cost-Effectiveness Analysis. J Bone Joint Surg Am. 2015 Nov 18;97(22):1860–8.
89. Kelson T, Davidson R, Baker T. Early MRI versus conventional management in the detection of occult scaphoid fractures: what does it really cost? A rural pilot study. J Med Radiat Sci. 2016 Mar;63(1):9–16.
90. Bergh TH, Steen K, Lindau T, Soldal LA, Bernardshaw SV, Lunde L, et al. Costs analysis and comparison of usefulness of acute MRI and 2 weeks of cast immobilization for clinically suspected scaphoid fractures. Acta Orthop. 2015 Jun;86(3):303–9.
91. Patel NK, Davies N, Mirza Z, Watson M. Cost and clinical effectiveness of MRI

in occult scaphoid fractures: a randomised controlled trial. Emerg Med J. 2013 Mar;30(3):202–7.
92. Roolker W, Maas M, Broekhuizen AH. Diagnosis and treatment of scaphoid fractures, can non-union be prevented? Arch Orthop Trauma Surg. 1999;119(7-8):428–31.
93. Eddeland A, Eiken O, Hellgren E, Ohlsson NM. Fractures of the scaphoid. Scand J Plast Reconstr Surg. 1975;9(3):234–9.
94. Young DK, Giachino A. Clinical examination of scaphoid fractures. Phys Sportsmed. 2009 Apr;37(1):97–105.
95. Dornberger JE, Rademacher G, Mutze S, Eisenschenk A, Stengel D. Accuracy of simple plain radiographic signs and measures to diagnose acute scapholunate ligament injuries of the wrist. Eur Radiol. 2015 May 17;
96. Van den Brand CL, van Leerdam RH, van Ufford JHMEQ, Rhemrev SJ. Is there a need for a clinical decision rule in blunt wrist trauma? Injury. 2013 Nov;44(11):1615–9.
97. Turkmen S, Cansu A, Karaca Y, Baki ME, Eroglu O, Tatli O, et al. Determination of radiography requirement in wrist trauma. Am J Emerg Med. 2015 Sep;33(9):1269–72.
98. Duckworth AD, Buijze GA, Moran M, Gray A, Court-Brown CM, Ring D, et al. Predictors of fracture following suspected injury to the scaphoid. J Bone Joint Surg Br. 2012 Jul;94(7):961–8.
99. Walenkamp MMJ, Schep NWL. Re: Is there a need for a clinical decision rule in blunt wrist trauma? Injury. 2014 Nov;45(11):1798–9.
100. Van den Brand CL, Quarles van Ufford JHME, van Leerdam RH, Rhemrev SJ. Comment on “Re: Is there a need for a clinical decision rule in blunt wrist trauma?.” Injury. 2014 Nov;45(11):1803–4.
101. Walenkamp MMJ, Bentohami A, Slaar A, Beerekamp MSH, Maas M, Jager LC, et al. The Amsterdam wrist rules: the multicenter prospective derivation and external validation of a clinical decision rule for the use of radiography in acute wrist trauma. BMC Musculoskelet Disord. 2015;16:389.

ANNEXES

Annexe 1. Exemples de radiographies des lésions principales du poignet 70
Annexe 2. Codes PMSI/CIM-10 utilisés pour la recherche des patients 72
Annexe 3. Définition, codage et transformation des variables..... 73
Annexe 4. Listes aléatoires et distribution des radiographies..... 75
Annexe 5. Notations et formules en rapport avec le coefficient κ 76
Annexe 6. Données manquantes..... 77
Annexe 7. Proportions observées et attendues par hasard, PI, BI, PABAK 78
Annexe 8. Effectifs pour les concordances en cas de classement en cinq familles de diagnostics..... 79

Annexe 1. Exemples de radiographies des lésions principales du poignet

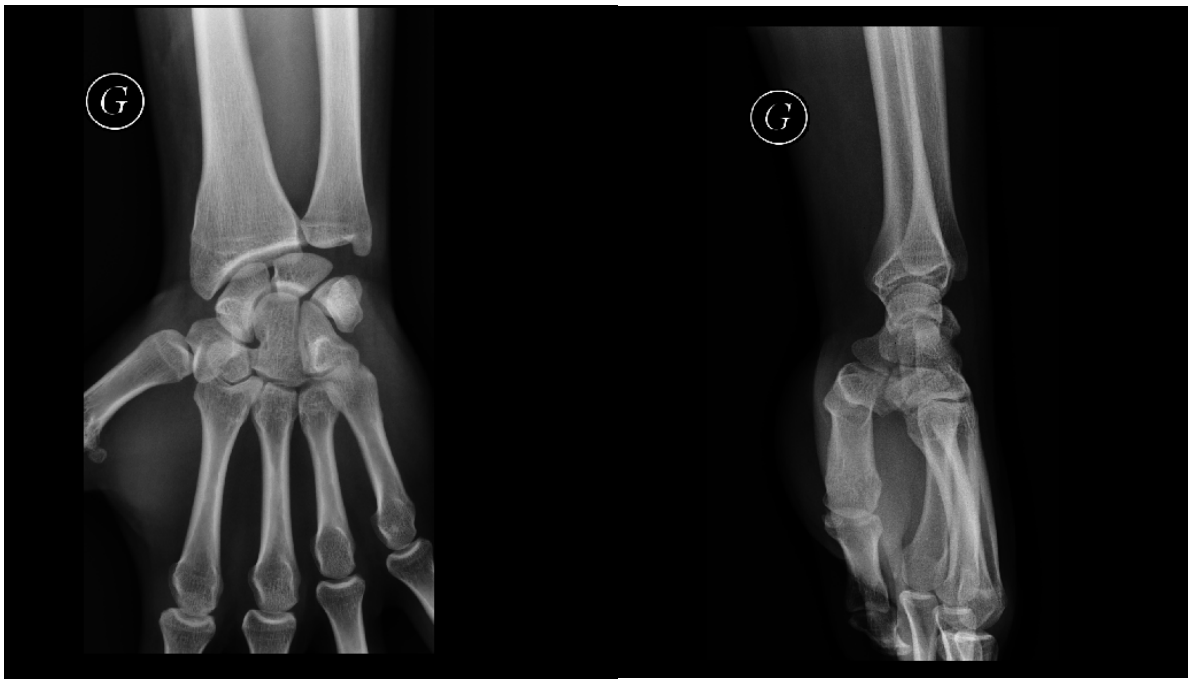


Figure 7. Radiographie standard du poignet de face et de profil sans lésion

Figure 8. Radiographie standard du poignet de face et de profil montrant une fracture articulaire de l'extrémité inférieure du radius

Figure 9. Radiographie standard du poignet de face montrant une fracture du corps du scaphoïde.

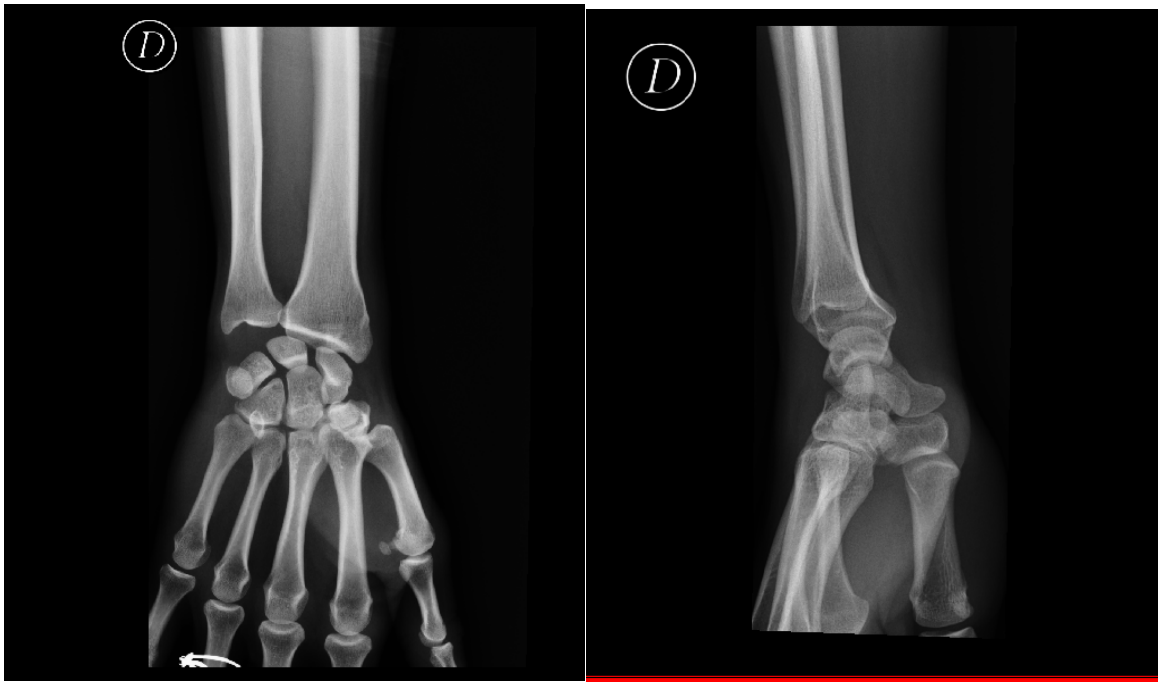


Figure 10. Radiographie standard du poignet de face et de profil montrant une rupture du ligament scapho-lunaire

Annexe 2. Codes PMSI/CIM-10 utilisés pour la recherche des patients

Série des S50 : Lésion traumatique superficielle de l'avant-bras

Série des S52 : Fracture de l'avant-bras

Série des S56 : Lésion traumatique de muscles et de tendons au niveau de l'avant-bras

Série des S57 : Écrasement de l'avant-bras

Série des S59 : Lésions traumatiques de l'avant-bras, autres et sans précision

Série des S60 : Lésion traumatique superficielle du poignet et de la main

Série des S62 : Fracture du poignet et de la main

Série des S63 : Luxation, entorse et foulure d'articulations et de ligaments au niveau du poignet et de la main

Série des S66 : Lésion traumatique de muscles et de tendons au niveau du poignet et de la main

Série des S67 : Écrasement du poignet et de la main

Série des S69 : Lésions traumatiques du poignet et la main, autres et sans précisions

Annexe 3. Définition, codage et transformation des variables

L'heure d'admission a été transformée en variable discrète. Entre 0 et 29 minutes incluses après l'heure, la valeur retenue était l'heure en cours (exemple, pour 17h22 : 17). Au-delà, la valeur retenue était l'heure suivante (exemple, pour 17h30 : 18).

Le caractère manuel se rapportait à la profession. Le caractère manuel était retenu si l'usage de la main était un élément déterminant à la réalisation de la profession. Le Tableau 12 présente le classement des professions recueillies.

La hauteur de la chute était classée en trois modalités : chute de sa hauteur, chute d'une hauteur inférieure à 1,50 m, chute d'une hauteur de plus de 1,50 m. Le type de chute était classé en " sans cinétique " pour les chutes simples de la hauteur du malade (glissade par exemple) et en " avec cinétique " dans les cas où une énergie supplémentaire était appliquée (chute dans les escaliers, chute en mouvement par exemple). La cinétique des accidents de la voie publique était classée en faible/moyen ou haute, tel que rapporté dans le dossier médical.

La présence d'un intervalle libre était définie comme la présence d'une période sans douleur ni œdème entre le traumatisme et l'apparition d'un de ces signes. L'évaluation de la douleur, chiffrée de 0 à 10 selon l'échelle d'évaluation analogique ou l'échelle numérique utilisées en clinique était reportée telles qu'indiquées dans le dossier. La variable « Mouvement douloureux » indiquait dans quel degré de liberté articulaire apparaît la douleur.

Tableau 12. Professions considérées comme manuelles dans l'étude

Manuel	Non manuel
Agent de sécurité	Administratif
Approvisionneur	Assistant en management
Auxiliaire de vie	Commercial
Boucher	Danseur classique professionnel
Charpentier	Direction établissement santé
Couturière	Dirigeant d'entreprise
Femme de ménage	Disc-Jockey
Infirmière	Étudiant commerce
Maçon	Étudiant en école d'ingénieur
Menuisier	Lycéen filière générale
Policier	Retraité sans activité manuelle
Pompier	Sans emploi
Plaquiste	
Porteur de journaux	
Serveur	
Technicien de maintenance	
Technicien en papeterie	
Videur	

Annexe 4. Listes aléatoires et distribution des radiographies

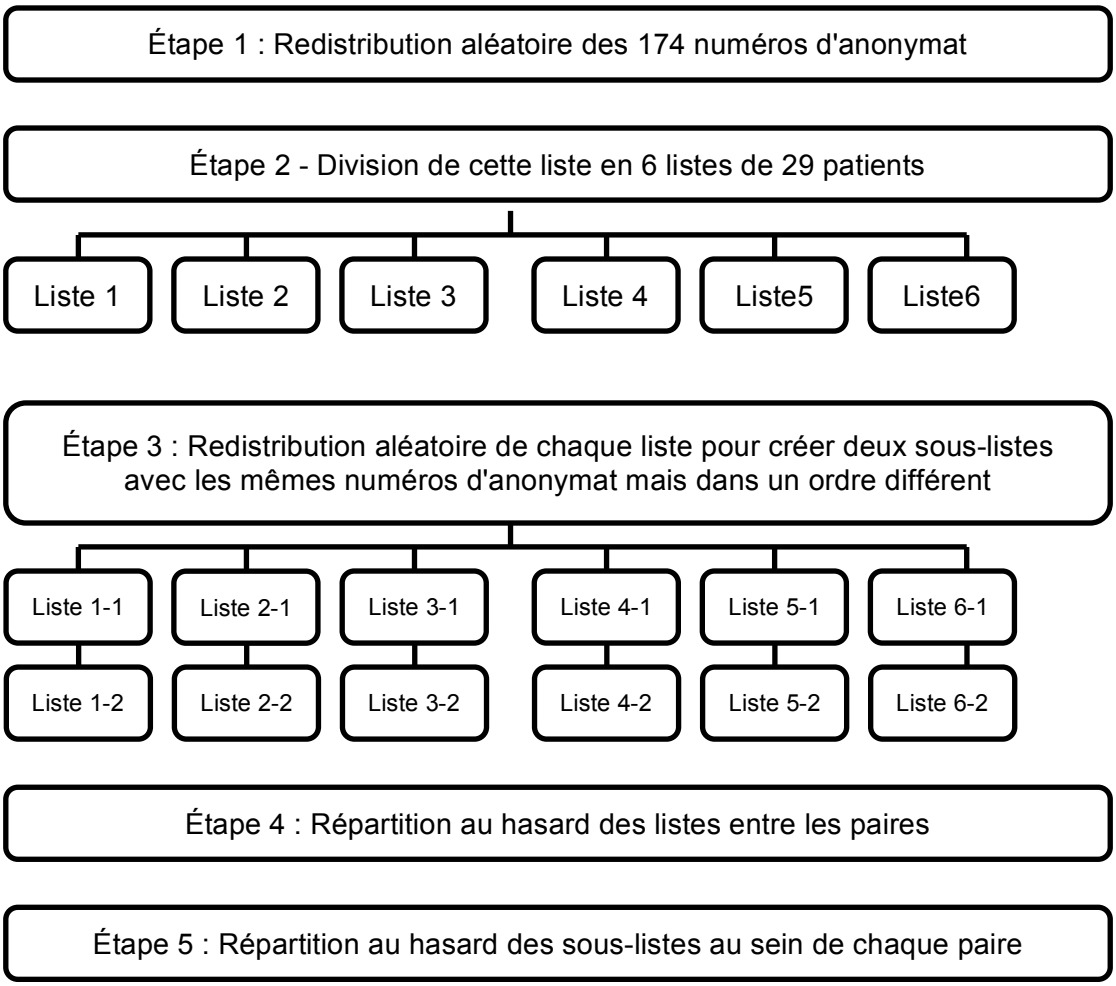


Tableau 13 . Distribution finale des listes aléatoires. Chaque liste incluait 29 radiographies (n/6). Chaque relecteur interprétait 87 radiographies (n/2) ; CCA : chef de clinique assistant ; PH : praticien hospitalier; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

	Paire 1	**Paire 2**	**Paire 3**	**Paire 4**	**Paire 5**	**Paire 6**
Radiologue CCA	4-1	1-2	6-1			
Radiologue PH	4-2			5-2	2-1	
Chirurgien CCA		1-1		5-1		3-1
Chirurgien PH			6-2		2-2	3-2

Annexe 5. Notations et formules en rapport avec le coefficient κ

		Observateur 1		
		Lésion	Pas de lésion	Total
Observateur 2	Lésion	a	b	g_1
	Pas de lésion	c	d	g_2
Total		f_1	f_2	*n*

Concordance observée : $p_o = \frac{a+d}{n}$

Concordance théorique ou par hasard : $p_e = \frac{\frac{f1 \times g1}{n} + \frac{f2 \times g2}{n}}{n}$

Accord véritable au-delà de l'accord lié au hasard : $p_o - p_e$

Accord possible au-delà de l'accord lié au hasard : $1 - p_e$

Coefficient κ = $\frac{po - pe}{1 - pe}$

Index de biais (*Bias Index*) = $\frac{|b-c|}{n}$

Index de prévalence (*Prevalence Index*) = $\frac{|a-d|}{n}\frac{|d-a|}{n}$

Le PABAK est calculé en remplaçant les cellules a et d par leur moyenne, et les cellules b et c par leur moyenne, ce qui donne par simplification : PABAK = $2p_o - 1$

Annexe 6. Données manquantes

Tableau 14. Nombre de données manquantes pour les groupes « lésion ou doute » et « pas de lésion ». Classement à partir des données du diagnostic des médecins des urgences. Étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

Variable	**Lésion ou doute** ***n* = 52**		**Pas de lésion** ***n* = 122**		**Total** ***n* = 174**		***p****
	Effectif	%	Effectif	%	Effectif	%	
Anamnèse							
Délai depuis le traumatisme	8	15,4	13	10,7	21	12,0	0,38
Intervalle libre	16	30,8	35	28,7	51	29,3	0,78
Mécanisme	4	7,7	6	4,9	10	5,7	0,76
Accident de sport	3	5,8	2	1,2	5	2,9	0,16
Antécédent homolatéral†	4	7,8	13	10,7	17	9,7	0,55
Atteinte du côté dominant	46	88,5	107	87,7	153	87,9	0,37
Profession manuelle	44	85,0	93	77	137	78,7	0,30
Position du poignet lors du traumatisme	43	82,7	98	80,0	141	81,0	0,72
Craquement lors du traumatisme	45	86,5	100	82,0	145	84,3	0,46
Évaluation de la douleur	41	78,9	93	76,2	134	77,0	0,71
Prise d'antalgiques	47	90,4	110	90,2	157	90,2	0,96
Localisation douleur spontanée	46	88,5	102	83,6	148	85,1	0,41
Irradiation de la douleur	50	96,2	112	91,8	162	93,1	0,48
Examen clinique							
Attitude des traumatisés du membre supérieur	48	92,3	114	93,4	162	93,0	1
Impotence fonctionnelle / limitation des amplitudes articulaires	22	42,3	22	18,0	44	25,3	<0,001
Axe de mobilité douloureux	42	80,8	91	74,6	133	76,4	0,38
Site de douleur exquise	30	57,7	38	31,2	68	39,0	<0,001
Accident de sport	3	5,8	2	1,2	5	2,9	0,16
Déformation	28	53,9	79	64,7	107	61,5	0,18
Œdème	25	48,0	53	43,4	78	44,8	0,57
Ecchymose	43	82,7	77	63,1	120	69,0	0,01
Hématome	43	82,7	78	63,9	121	69,5	0,01
Dermabrasion	42	80,7	77	63,1	119	68,4	0,02
Paresthésies	30	58,7	48	39,3	78	44,8	0,03
Troubles sensitifs autres que paresthésies	34	65,4	55	45,0	89	51,2	0,01

Les variables non indiquées ne comportaient pas de données manquantes ; * test du Chi-2 avec ou sans correction de Yates, ou test exact de Fisher, en fonction des effectifs théoriques ; † Antécédent de traumatisme du poignet

Annexe 7. Proportions observées et attendues par hasard, PI, BI, PABAK

Concordance inter-observateurs entre médecins des urgences et consensus d'experts pour chaque famille de diagnostics

Tableau 15. Concordance pour chaque famille de diagnostics entre médecins des urgences et consensus d'experts ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

Diagnostic	κ	IC 95 %	p_o	p_e	PI	BI	PABAK
Fracture de l'avant-bras	0,56	0,39-0,73	0,90	0,73	0,67	0,02	0,76
Fracture du carpe	0,55	0,24-0,87	0,97	0,92	0,92	0,02	0,93
Ligament	0	-0,21-0,22	0,90	0,90	0,90	0,10	0,79
Doute	0,05	-0,12-0,21	0,81	0,80	0,78	0,01	0,62
Pas de lésion	0,50	0,37-0,62	0,76	0,52	0,25	0,15	0,52

n = 174 radiographies ; p_o : proportion observée d'observations concordantes ; p_e : proportion de concordance attendue par hasard ; PI : Prevalence Index ; BI : Bias Index ; PABAK : Prevalence-Adjusted Bias-Adjusted Kappa

Concordance inter-observateurs entre les experts et le référent pour chaque famille de diagnostics

Tableau 16. Concordance pour chaque famille de diagnostics, entre le duo d'experts relecteurs d'une même radiographie d'une part, et le référent d'autre part, pour les radiographies pour lesquelles les deux experts relecteurs concordaient ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

Diagnostic	κ	IC 95 %	p_o	p_e	PI	BI	PABAK
Fracture de l'avant-bras	0,81	0,68-0,94	0,94	0,69	0,60	0	0,88
Fracture du carpe	0,76	0,49-1,00	0,98	0,90	0,90	0	0,95
Lésion ligamentaire	0,23	0,04-0,42	0,89	0,85	0,84	0,08	0,77
Doute	0,13	-0,11-0,38	0,88	0,86	0,84	0,02	0,75
Pas de lésion	0,56	0,42-0,69	0,78	0,49	0,06	0,13	0,55

n = 129 radiographies ; p_o : proportion observée d'observations concordantes ; p_e : proportion de concordance attendue par hasard ; PI : Prevalence Index ; BI : Bias Index ; PABAK : Prevalence-Adjusted Bias-Adjusted Kappa

Annexe 8. Effectifs pour les concordances en cas de classement en cinq familles de diagnostics

Concordance inter-observateurs entre les deux experts relecteurs d'une même radiographie pour le classement en cinq familles de diagnostics

Tableau 17. Concordance entre experts relisant la même radiographie pour le classement en cinq familles de diagnostics ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

	Premier relecteur											
	Fracture avant-bras		Fracture du carpe		Ligament		Doute		Pas de lésion		Total	
Second relecteur	Effectif	%	Effectif	%	Effectif	%	Effectif	%	Effectif	%	Effectif	%
Fracture de l'avant-bras	18	10	0	0	4	3	2	1	2	1	26	15
Fracture du carpe	0	0	4	2	0	0	2	1	2	1	8	4
Ligament	3	2	0	0	2	1	0	0	2	1	7	4
Doute	5	3	0	0	0	0	12	7	14	8	31	18
Pas de lésion	3	2	1	1	2	1	19	11	77	44	102	59
Total	29	17	5	3	8	5	35	20	97	55	174	100

n = 174 radiographies

Concordance inter-observateurs entre les experts et le référent pour le classement en cinq familles de diagnostics

Tableau 18. Concordance pour le classement en cinq familles de diagnostics, entre les deux experts relecteurs d'une même radiographie d'une part, et le référent d'autre part, pour les radiographies pour lesquelles les deux experts relecteurs concordaient ; étude de la fiabilité de l'interprétation des radiographies simples du poignet, Bordeaux, 2014.

	Duo d'experts relecteurs											
	Fracture avant-bras		Fracture du carpe		Ligament		Doute		Pas de lésion		Total	
Référent	Effectif	%	Effectif	%	Effectif	%	Effectif	%	Effectif	%	Effectif	%
Fracture de l'avant-bras	22	17	0	0	4	3	0	0	0	0	26	20
Fracture du carpe	0	0	5	4	0	0	0	0	2	2	7	6
Ligament	0	0	0	0	5	4	4	3	18	14	27	21
Doute	3	2	1	1	0	0	2	2	3	2	9	7
Pas de lésion	1	1	0	0	0	0	5	4	54	42	60	47
Total	26	20	6	5	9	7	11	9	77	60	129	100

n = 129 radiographies

SERMENT D'HIPPOCRATE

Au moment d'être admis à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité.

Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences.
Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences.

Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admis dans l'intimité des personnes, je tairai les secrets qui me seront confiés. Reçu à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs.
Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité.

Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré et méprisé si j'y manque.

Fiabilité de l'interprétation radiographique des traumatismes fermés récents du poignet

Résumé

Contexte : Les traumatismes du poignet sont fréquents ; les diagnostics des lésions qui en découlent reposent sur la radiographie standard, mais les erreurs diagnostiques sont fréquentes. L'objectif de cette étude était d'évaluer la fiabilité de cet examen pour caractériser les traumatismes fermés récents du poignet.

Méthodes : Nous avons réalisé une étude de la concordance entre le diagnostic posé par plusieurs observateurs d'une même radiographie (concordance inter-observateurs : entre les médecins des urgences et un consensus d'experts puis entre les experts entre eux) ; et entre les diagnostics du même observateur d'une même radiographie à un an d'intervalle (concordance intra-observateur). Les déterminants de la fiabilité (associations entre les caractéristiques du patient, de son traumatisme ou du relecteur) ont également été recherchés. Les patients ayant consulté au Centre Hospitalier Universitaire de Bordeaux en 2014 pour un traumatisme fermé de moins de sept jours ont été inclus rétrospectivement. L'étude de la concordance a été faite avec le coefficient κ de Cohen et son intervalle de confiance à 95 % (IC95 %).

Résultats : Cent-soixante-quatorze radiographies ont été relues. Le coefficient κ pour le classement binaire « lésion ou doute sur une lésion » contre « pas de lésion », entre médecins des urgences et consensus d'experts était de 0,50 (IC95 % : 0,37 - 0,62). Le coefficient κ pour la concordance au sein du groupe d'experts pour ce classement était de 0,47 (IC95 % : 0,34 - 0,60). Le coefficient κ pour la concordance intra-observateur était de 0,12 (IC95 % 0,01 – 0,23). Il n'a pas été mis en évidence de caractéristiques du patient ou du traumatisme associées à la présence d'une discordance.

Conclusion : La fiabilité de la radiographie standard du poignet n'était pas satisfaisante, quels que soient les observateurs comparés et également en comparant les observateurs à eux-mêmes. Cet examen ne peut être recommandé seul dans ces traumatismes.

Reliability of the interpretation of standard radiographs in recent blunt wrist traumas

Abstract

Introduction: Blunt wrist traumas are frequent. The diagnosis of related injuries relies on standard radiographs, but injuries are frequently misdiagnosed. The objective of this study was to assess the reliability of standard radiographs to characterize blunt wrist traumas.

Methods: This was a study of rater agreement between the diagnoses of different observers of the same radiograph (inter-rater agreement: between Emergency Room physicians and a consensus of experts, and afterwards between the experts); and between the diagnoses made by the same observer of the same radiograph after a one-year period (intra-rater agreement). Factors associated with the reliability were sought (characteristics of the patients, the trauma, or the rater). Patients who were managed in the ER of the teaching hospital of Bordeaux for a recent (less than seven days) blunt wrist trauma during 2014 were retrospectively included. The raters' agreement was assessed with Cohen's κ coefficient and its 95% confidence interval (95%CI).

Results: One hundred and seventy-four radiographs were included in the study. The κ coefficient for the binary rating "injury or doubt on injury" versus "no injury" between ER physicians and experts consensus was 0.50 (95%CI: 0.37 - 0.62). The κ coefficient for the agreement between the two experts who rated the same radiograph was 0.47 (95%CI: 0.34 - 0.60). The κ coefficient for the intra-rater agreement was 0.12 (95%CI 0.01 – 0.23). No association between disagreement and characteristics of the patient or his trauma reached statistical significance.

Conclusion: The reliability of standard radiographs does not seem satisfying, whoever are the observers, and even when observers are compared to themselves. This test should not be used alone to diagnose these injuries.

Mots-clés : Poignet, Traumatisme, Concordance, Kappa, Radiographies

Key Words : Wrist, Trauma, Reliability, Kappa, Radiographs