

VNIVERSITAT E VALÈNCIA

FACULTAD DE MEDICINA Y ODONTOLOGÍA
DEPARTAMENTO DE MEDICINA
PROGRAMA DE DOCTORADO
3139 MEDICINA



**EL ERROR DIAGNÓSTICO EN RADIOLOGÍA. ANÁLISIS
DE VARIABLES EXPLICATIVAS**

TESIS DOCTORAL

Presentada por:

D. Fernando Facal de Castro

Dirigida por:

Dr. D. Estanislao Arana Fernández

Dra. Dña. Delfina Dualde Beltrán

VALENCIA, Enero de 2019

D. Estanislao Arana Fernández, doctor en Medicina y facultativo especialista en Radiodiagnóstico de la Fundación Instituto Valenciano de Oncología.

Dña. Delfina Dualde Beltrán, doctora en Medicina, facultativo especialista en Radiodiagnóstico del Hospital Clínico Universitario de Valencia, y Profesora Asociada del Departamento de Medicina de la Universidad de Valencia.

Como directores del trabajo titulado: **“El error diagnóstico en Radiología. Análisis de variables explicativas”**, realizado por **D. Fernando Facal de Castro**, certificamos que reúne los requisitos necesarios para su presentación y defensa para optar al título de Doctor en Medicina por la Universitat de València.

Y para que conste y obre a los efectos oportunos, expedimos el presente certificado en Valencia a 14 de Noviembre de 2018.

Fdo:

Dr. Estanislao Arana Fernández.

Fdo:

Dra. Delfina Dualde Beltrán

Facultat de Medicina i Odontologia

Departament de Medicina

Av. Blasco Ibàñez, 15

VALÈNCIA

Telèfon (+34) 96 386 41 68

Fax (+34) 96 386 40 08

E-mail: dep.medicina@uv.es

AGRADECIMIENTOS:

Durante estos años son muchas las personas que han participado en este trabajo y a quienes quiero expresar mi gratitud:

Al Doctor Javier Mota Martínez, por su confianza, consejos y ayuda desinteresada. Le agradezco que me haya abierto hace ya cuatro años las puertas de su grupo de trabajo, dándome la oportunidad de tener una visión más amplia del mundo de la Radiología.

A Julio Moreno Parra, por el apoyo y tiempo que me ha dedicado con el manejo de las bases de datos.

A mis Directores de Tesis, los Doctores Estanislao Arana Fernández y Delfina Dualde Beltrán, por brindarme la oportunidad de recurrir a su capacidad y experiencia científica.

A Rebeca Mirón Mombiela, por el entusiasmo y motivación que me ofrece cada día para crecer como profesional y como persona.

A mis padres. De todos los amigos y maestros que he tenido en la vida, habéis sido sin duda los mejores.

Gracias.

A Fernando y Angelina.

RESUMEN:

ANTECEDENTES:

Se ha demostrado que la etiología del error médico es multifactorial, con convergencia de factores humanos activos (errores de percepción, cognición y/o comunicación) y factores latentes del sistema. Los factores latentes del sistema son aquellas variables que pueden predisponer a la aparición de errores activos, o aumentar el impacto negativo de estos. Incluyen una amplia variedad de elementos, entre otros el material de trabajo inadecuado, la sobrecarga de trabajo, la falta de experiencia del radiólogo, la información clínica insuficiente y la no disponibilidad de estudios previos. Nuestra hipótesis de trabajo es que los factores latentes contribuyen significativamente al error diagnóstico en Radiología, sin embargo hasta la fecha no hay estudios que evalúen esta asociación.

OBJETIVOS:

- Determinar la tasa de discrepancia entre la primera lectura y la revisión del caso y comparar los resultados con las tasas de error diagnóstico reflejadas en la bibliografía médica.
- Estimar la significación clínica de los errores detectados en base a la gravedad de los hallazgos de imagen.
- Identificar los factores latentes asociados a una mayor incidencia de error diagnóstico.
- Descubrir puntos débiles en el trabajo radiológico que puedan reforzarse con medidas específicas de formación o de control de calidad.

METODOLOGÍA:

Análisis observacional retrospectivo de las discrepancias diagnósticas en estudios de telerradiología de pacientes adultos de 37 centros sanitarios detectadas tras una revisión de casos por parte de radiólogos realizados en la empresa IBERORAD SL. desde 2014 a 2016.

RESULTADOS PRINCIPALES:

Para el total de estudios existe una tasa de discrepancia diagnóstica del 3,0%, que desciende al 1,4% en los estudios urgentes y se sitúa en el 3,6% en los estudios programados). La tasa media de estudios con discrepancias clínicamente significativas se sitúa en el 23,8% del total de discrepancias. En los estudios programados se observa

asociación con la discrepancia diagnóstica, en orden decreciente de magnitud, de las siguientes variables: deficiente calidad técnica del estudio, estudio de TC frente a estudio de RM, disponibilidad de estudios de imagen previos, concordancia media o baja entre la especialidad del radiólogo y la especialidad del estudio, no utilización de sistemas de reconocimiento de voz, uso de un solo monitor de visualización de las imágenes y estudio informado fuera del horario laboral de un día laborable. En los estudios urgentes se observa asociación estadísticamente significativa con la discrepancia diagnóstica, en orden decreciente de magnitud, de las siguientes variables: disponibilidad de estudios de imagen previos, concordancia media entre la especialidad del radiólogo y del estudio, información clínica insuficiente, no utilización de sistemas de reconocimiento de voz, concordancia baja entre la especialidad del radiólogo y del estudio, estudio de complejidad alta, estudio de Body, uso de un solo monitor de visualización de las imágenes y estudio de MSK/Columna.

CONCLUSIONES:

La tasa de discrepancia diagnóstica en IBERORAD SL. durante el periodo estudiado es similar a la reflejada en la bibliografía médica. La mayoría de las discrepancias no son clínicamente significativas. Existen múltiples factores asociados al error diagnóstico, dependientes tanto de los estudios de imagen como del propio radiólogo, y la intensidad en que cada uno de estos factores se asocia al error diagnóstico es distinta en cada factor y variable en función del resto de características presentes en el proceso diagnóstico.

ABSTRACT

BACKGROUND:

It has been shown that the etiology of medical error is multifactorial, with convergence of active human factors (errors of perception, cognition and/or communication) and latent system factors. Latent system factors are those variables that may predispose to the arising of active errors or increase the negative impact of these. They include a wide variety of elements such as inadequate work material, work overload, inexperience of the radiologist, insufficient clinical information, and unavailability of previous studies, among others. Our working hypothesis is that the latent factors contribute significantly to diagnostic error in Radiology, however, to date there are no studies that evaluate this association.

OBJECTIVES:

- To determine the discrepancy rate between the first reading and the case review and to compare the results with the diagnostic error rates reflected in the medical literature.
- To estimate the clinical significance of errors detected based on the severity of imaging findings.
- To identify those latent factors associated with a higher incidence of diagnostic error.
- To discover weak points in the radiological work that can be solved by specific training or quality control measures.

METHOD:

Retrospective observational analysis of diagnostic discrepancies in teleradiology studies of adult patients from 37 health centres detected after a review of cases performed by radiologists at the company IBERORAD SL from 2014 to 2016.

MAIN RESULTS:

For the total number of studies, there is a diagnostic discrepancy rate of 3%, which drops to 1.4% in emergency studies and stands at 3.6% in scheduled studies. The average rate of studies with clinically significant discrepancies is 23.8% of total discrepancies. In the programmed studies, an association was found with the diagnostic discrepancy, in

decreasing order of magnitude, of the following variables: poor technical quality of the study; CT study versus MR study; availability of previous imaging studies; average or low level of concordance between the field of expertise of the radiologist and modality of the study; non-use of voice recognition systems; use of a single monitor to visualize images; and studies reported outside working hours of a working day. In the emergency studies, a statistically significant association was noticed with the diagnostic discrepancy, in decreasing order of magnitude, of the following variables: availability of previous imaging studies; average level of concordance between area of expertise of the radiologist and modality of the study; insufficient clinical information; non-use of speech recognition systems; low level of concordance between area of expertise of the radiologist and modality of the study; study of high complexity; body studies; use of a single monitor to visualize images; and MSK/spine studies.

CONCLUSIONS:

The diagnostic discrepancy rate in IBERORAD SL during the period analysed is similar to that reflected in the medical literature. Most discrepancies are not clinically significant. There are multiple factors associated with diagnostic error, depending on both the imaging studies and the radiologists themselves. The intensity in which each of these factors is associated with diagnostic error is different and variable determined by the rest of the characteristics present in the diagnostic process.

RESUM

ANTECEDENTS:

S'ha demostrat que l'etiologia de l'error mèdic és multifactorial, amb convergència de factors humans actius (errors de percepció, cognició y/o comunicació) i factors latents del sistema. Els factors latents del sistema són aquelles variables que poden predisposar a l'aparició d'errors actius, o augmentar l'impacte negatiu d'estos. Inclouen una àmplia varietat d'elements, entre altres el material de treball inadequat, la sobrecàrrega de treball, la falta d'experiència del radiòleg, la informació clínica insuficient, i la no disponibilitat d'estudis previs. La nostra hipòtesi de treball és que els factors latents contribueixen significativament a l'error diagnòstic en Radiologia, no obstant això fins a la data no hi ha estudis que avaluen esta associació."

OBJECTIUS:

- Determinar la taxa de discrepància entre la primera lectura i la revisió del cas i comparar els resultats amb les taxes d'error diagnòstic reflectides en la bibliografia mèdica.
- Estimar la significació clínica dels errors detectats basant-se en la gravetat de les troballes d'imatge.
- Identificar els factors latents associats a una major incidència d'error diagnòstic. Descobrir punts dèbils en el treball radiològic que puguin reforçar-se amb mesures específiques de formació o de control de qualitat.

METODOLOGIA:

Anàlisi observacional retrospectiu de les discrepàncies diagnòstiques en estudis de telerradiologia de pacients adults de 37 centres sanitaris detectades després d'una revisió de casos per part de radiòlegs realitzats en l'empresa IBERORAD SL. des de 2014 a 2016.

RESULTATS PRINCIPALS:

Per al total d'estudis hi ha una taxa de discrepància diagnòstica del 3,0%, que descendix al 1,4% en els estudis urgents i se situa en el 3,6% en els estudis programats). La taxa mitjana d'estudis amb discrepàncies clínicament significatives se situa en el 23,8% del total de discrepàncies. En els estudis programats s'observa associació amb la discrepància diagnòstica, en orde decreixent de magnitud, de les variables següents: deficient qualitat

tècnica de l'estudi, estudi de TC enfront d'estudi de RM, disponibilitat d'estudis d'imatge previs, concordança mitjana o baixa entre l'especialitat del radiòleg i l'especialitat de l'estudi, no utilització de sistemes de reconeixement de veu, ús d'un sol monitor de visualització de les imatges i estudi informat fora de l'horari laboral d'un dia laborable. En els estudis urgents s'observa associació estadísticament significativa amb la discrepància diagnòstica, en orde decreixent de magnitud, de les variables següents: disponibilitat d'estudis d'imatge previs, concordança mitjana entre l'especialitat del radiòleg i de l'estudi, informació clínica insuficient, no utilització de sistemes de reconeixement de veu, concordança baixa entre l'especialitat del radiòleg i de l'estudi, estudi de complexitat alta, estudi de Body, ús d'un sol monitor de visualització de les imatges i estudi de MSK/Columna.

CONCLUSSIONS:

La taxa de discrepància diagnòstica en IBERORAD SL. durant el període estudiat és semblant a la reflectida en la bibliografia mèdica. La majoria de les discrepàncies no són clínicament significatives. Hi ha múltiples factors associats a l'error diagnòstic, dependents tant dels estudis d'imatge com del propi radiòleg, i la intensitat en què cada un d'estos factors s'associa a l'error diagnòstic és distinta en cada factor i variable en funció de la resta de característiques presents en el procés diagnòstic.

ÍNDICE.

Abreviaturas

Índice de tablas.

Índice de figuras.

1. INTRODUCCIÓN	1
1.1. EL ERROR EN RADIOLOGÍA	1
1.2. RELEVANCIA DEL ERROR DIAGNÓSTICO.....	5
1.3. FRECUENCIA DEL ERROR DIAGNÓSTICO.....	12
1.4. CLASIFICACIÓN DEL ERROR DIAGNÓSTICO	20
1.4.1. Errores activos.....	24
1.4.1.1. Error de percepción.....	24
1.4.1.2. Error cognitivo	27
1.4.1.3. Error de comunicación	32
1.4.2. Errores latentes.....	35
1.5. PATOGENIA DEL ERROR DIAGNÓSTICO	41
1.6. REVISIÓN POR PARES (PEER REVIEW).....	46
1.7. EL ERROR EN MEDICINA, NO ES LO QUE PARECE.....	51
2. HIPÓTESIS Y OBJETIVOS	55
2.1. HIPÓTESIS	55
2.2. OBJETIVOS	55
3. MATERIAL Y MÉTODOS	57

3.1. DESCRIPCIÓN DEL ENTORNO. CARACTERÍSTICAS DEL CENTRO	57
3.2. DISEÑO DEL ESTUDIO.....	58
3.3. ANÁLISIS ESTADÍSTICO	62
4. RESULTADOS	69
4.1. CARACTERÍSTICAS DE LOS RADIÓLOGOS QUE VALORARON INICIALMENTE LOS ESTUDIOS	69
4.2. CARACTERÍSTICAS DE LOS ESTUDIOS DE TELERRADIOLOGÍA INCLUIDOS EN EL ESTUDIO.....	70
4.3. ESTIMACIÓN DE LA TASA DE DISCREPANCIAS DIAGNÓSTICAS	80
4.4. EVALUACIÓN DE FACTORES RELACIONADOS CON LA DISCREPANCIA DIAGNÓSTICA EN RADIOLOGÍA.....	83
4.4.1. Evaluación de los factores relacionados con discrepancia diagnóstica en los estudios de telerradiología programados	83
4.4.1.1. Análisis univariante de factores significativos en los estudios programados.....	84
4.4.1.2. Análisis multivariante de factores significativos en los estudios programados.....	84
4.4.2. Evaluación de los factores relacionados con discrepancia diagnóstica en los estudios de telerradiología urgentes	88
4.4.2.1. Análisis univariante de factores significativos en los estudios urgentes	88
4.4.2.2. Análisis multivariante de factores significativos en los estudios urgentes	89
4.4.3. Comparación de los modelos ajustados para ambos estratos....	91
4.4.4. Evaluación de los factores relacionados con discrepancia diagnóstica sobre todos los estudios de telerradiología	92

4.4.4.1. Análisis univariante de factores significativos.....	93
4.4.4.2. Análisis multivariante de factores significativos	93
4.4.5. Evaluación sobre una submuestra de los estudios de telerradiología muestreados.....	96
4.4.5.1. Análisis univariante de factores significativos.....	96
4.4.5.2. Análisis multivariante de factores significativos	96
5. DISCUSIÓN	101
5.1. ASPECTOS GENERALES	101
5.2. COMPLEJIDAD DEL ESTUDIO	104
5.3. MODALIDAD DEL ESTUDIO (TC/RM)	105
5.4. CALIDAD TÉCNICA DEL ESTUDIO.....	107
5.5. DISPONIBILIDAD DE ESTUDIOS PREVIOS.....	109
5.6. ADECUACIÓN DE LA INFORMACIÓN CLÍNICA	111
5.7. TIPO DE VISOR.....	115
5.8. NÚMERO DE MONITORES	116
5.9. USO DE SISTEMAS DE RECONOCIMIENTO DE VOZ.....	117
5.10. EDAD Y EXPERIENCIA DEL RADIÓLOGO	120
5.11. GÉNERO DEL RADIÓLOGO	122
5.12. HORARIO DE LA PRIMERA LECTURA. FATIGA LABORAL	123
5.13. ESPECIALIDAD DEL ESTUDIO. ESPECIALIDAD DE RADIÓLOGO. CONCORDANCIA ENTRE AMBAS.....	126
5.14. TIPO DE ESTUDIO (URGENTE/PROGRAMADO).....	131

6. LIMITACIONES DEL ESTUDIO.....	133
7. CONCLUSIONES	135
8. BIBLIOGRAFÍA	137
9. ANEXOS.....	153
9.1. ANEXO 1. Análisis univariante de factores relacionados con discrepancia diagnóstica en Radiología.....	153
9.1.1. Análisis univariante de factores significativos en los estudios programados.....	153
9.1.2. Análisis univariante de factores significativos en los estudios urgentes	162
9.1.3. Análisis univariante de factores significativos en la totalidad de los estudios (programados y urgentes)	174
9.1.4. Análisis univariante de factores significativos en la submuestra de los estudios de telerradiología muestreados.....	176
9.2. ANEXO 2. Resumen de resultados.....	183
9.3. ANEXO 3. Documentación del Comité Ético de Investigación en Humanos de la Comisión de Ética en Investigación Experimental de la Universitat de València	184

ABREVIATURAS.

ACR	Colegio Americano de Radiología.
AngioRM	Angiografía por RM.
Body	Radiología de Tórax-Abdomen-Pelvis.
EE.UU	Estados Unidos de América.
Est	Estimado.
GLAD	<i>Glenolabral Articular Disruption</i> . Disrupción articular glenolabral.
HL7	<i>Health Level Seven</i> .
IBM PC	Ordenador Personal IBM o compatible.
IOM	Instituto de Medicina.
MIR	Médico Interno Residente.
MSK	Radiología Musculoesquelética.
Neuro	Neurorradiología.
Obs	Observado.
OR	<i>Odds Ratio</i> . Razón de probabilidades.
PACS	Sistemas de archivo y transmisión de imágenes.
RIS	Sistema de Información Radiológica.
ROC	<i>Receiver Operating Characteristic</i> . Característica Operativa del Receptor.
RM	Resonancia Magnética.
TC	Tomografía Computarizada.

ÍNDICE DE TABLAS.

- Tabla 1.** Estudios retrospectivos de revisión de eventos adversos.
- Tabla 2.** Relación de estudios que analizan la tasa de discrepancia diagnóstica.
- Tabla 3.** Comparativa de factores latentes y activos.
- Tabla 4.** Características generales de los radiólogos que valoraron inicialmente los estudios.
- Tabla 5.** Características generales de las estaciones de trabajo.
- Tabla 6.** Características de los estudios de telerradiología según tipología.
- Tabla 7.** Presencia de discrepancias diagnósticas en los estudios de telerradiología según tipología.
- Tabla 8.** Estimación de las características de todos los estudios de telerradiología del periodo.
- Tabla 9.** Estimación del porcentaje de discrepancias estimadas según los diferentes factores recogidos.
- Tabla 10.** Porcentaje de discrepancias según modalidad y tipo de estudio.
- Tabla 11.** Modelo multivariante del efecto de los factores latentes considerados sobre la discrepancia diagnóstica en los estudios programados.
- Tabla 12.** Modelo multivariante del efecto de los factores latentes considerados sobre la discrepancia diagnóstica en los estudios urgentes.
- Tabla 13.** Modelos multivariantes para los estudios programados y urgentes.
- Tabla 14.** Modelo multivariante del efecto de los factores latentes considerados sobre la discrepancia diagnóstica.
- Tabla 15.** Modelo multivariante del efecto de los factores latentes considerados sobre la discrepancia diagnóstica.
- Tabla 16.** Relación entre la complejidad y la discrepancia diagnóstica en los estudios programados.
- Tabla 17.** Relación entre la modalidad y la discrepancia diagnóstica en los estudios programados.
- Tabla 18.** Relación entre calidad técnica de las imágenes y discrepancia diagnóstica en los estudios programados.
- Tabla 19.** Relación entre disponibilidad de imágenes previas y discrepancia diagnóstica en los estudios programados.
- Tabla 20.** Relación entre información clínica de la solicitud y discrepancia diagnóstica en los estudios programados.
- Tabla 21.** Relación entre tipo de visor utilizado y discrepancia diagnóstica en los estudios programados.

- Tabla 22.** Relación entre número de monitores utilizados y discrepancia diagnóstica en los estudios programados.
- Tabla 23.** Relación entre el uso de un sistema de reconocimiento de voz y la discrepancia diagnóstica en los estudios programados.
- Tabla 24.** Relación entre edad del radiólogo y discrepancia diagnóstica en los estudios programados.
- Tabla 25.** Relación entre experiencia radiólogo y discrepancia diagnóstica en los estudios programados.
- Tabla 26.** Relación entre género radiólogo y discrepancia diagnóstica en los estudios programados.
- Tabla 27.** Relación entre horario valoración inicial y discrepancia diagnóstica en los estudios programados.
- Tabla 28.** Relación entre la especialidad del estudio y discrepancia diagnóstica en los estudios programados.
- Tabla 29.** Relación entre la especialidad del estudio recodificada y la discrepancia diagnóstica en los estudios programados.
- Tabla 30.** Relación entre la especialidad del radiólogo y la discrepancia diagnóstica en los estudios programados.
- Tabla 31.** Relación entre la especialidad del radiólogo y la especialidad de los estudios programados.
- Tabla 32.** Relación de la concordancia de especialidades estudio-radiólogo y la discrepancia diagnóstica en los estudios programados.
- Tabla 33.** Relación entre la complejidad y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 34.** Relación entre la modalidad y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 35.** Relación entre calidad técnica de las imágenes y discrepancia diagnóstica en los estudios urgentes.
- Tabla 36.** Relación entre la disponibilidad de imágenes de estudios previos y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 37.** Relación entre la suficiencia de la información clínica de la solicitud y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 38.** Relación entre el tipo de visor utilizado en la primera lectura y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 39.** Relación entre el número de monitores utilizados en la primera lectura y la discrepancia diagnóstica en los estudios urgentes.

- Tabla 40.** Relación entre el uso de un sistema de reconocimiento de voz y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 41.** Relación entre edad del radiólogo y discrepancia diagnóstica en los estudios urgentes.
- Tabla 42.** Relación entre la experiencia del radiólogo y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 43.** Relación entre el género del radiólogo que ha valorado la primera lectura y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 44.** Relación entre el horario de la valoración inicial y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 45.** Relación entre la especialidad del estudio y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 46.** Relación entre la especialidad del estudio recodificada y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 47.** Relación entre especialidad del radiólogo y discrepancia diagnóstica en los estudios urgentes.
- Tabla 48.** Relación entre la especialidad del radiólogo y la de los estudios urgentes.
- Tabla 49.** Efecto de la relación de especialidades sobre la discrepancia diagnóstica en los estudios urgentes.
- Tabla 50.** Relación de la concordancia de especialidades estudio-radiólogo y la discrepancia diagnóstica en los estudios urgentes.
- Tabla 51.** Modelos univariantes ajustados para los estudios programados y urgentes.
- Tabla 52.** Relación de los factores latentes considerados con la discrepancia diagnóstica.
- Tabla 53.** Relación de los factores latentes considerados con la discrepancia diagnóstica en la submuestra.
- Tabla 54.** Relación entre la especialidad del radiólogo y la especialidad del estudio valorado en la submuestra.
- Tabla 55.** Efecto de la relación de especialidades sobre la discrepancia diagnóstica en la submuestra.
- Tabla 56.** Relación entre la concordancia de especialidades estudio-radiólogo y la discrepancia diagnóstica en la submuestra.

ÍNDICE DE FIGURAS.

- Figura 1.** Número de muertes anuales en 2014 de las principales causas de muerte en EE.UU.
- Figura 2.** Imágenes de TC (izquierda) y PET (derecha) de un paciente oncológico, en las que se identifica una adenopatía mesentérica que modifica el tratamiento.
- Figura 3.** Imágenes axial y coronal de TC que muestran un hematoma subdural en el tentorio.
- Figura 4.** Imágenes sagitales de RM de columna vertebral cervical en mujer de 35 años con cervicalgia, que muestran un hematoma epidural espontáneo.
- Figura 5.** Imagen axial de TC abdominal por dolor y fiebre en paciente de 17 años, que presentaba derrame pleural.
- Figura 6.** Imagen axial de TC abdominal que muestra neumoperitoneo por diverticulitis.
- Figura 7.** Imágenes de RM de columna vertebral torácica con una metástasis vertebral. El localizador muestra un carcinoma de pulmón.
- Figura 8.** Imágenes coronales de TC de columna vertebral que muestran un error de búsqueda satisfactoria. Además de fractura vertebral hay una tumoración renal derecha.
- Figura 9.** Imagen sagital de TC de columna vertebral que muestra un error de búsqueda satisfactoria. Además de una fractura dorsal se observa una fractura sacra.
- Figura 10.** Imágenes axiales y sagitales de RM de columna vertebral cervical en paciente de 40 años con cervicobraquialgia izquierda deficitaria. Muestran dos hernias discales foraminales de las que en la primera lectura sólo se detectó una, debido al fenómeno de búsqueda satisfactoria.
- Figura 11.** Imagen axial de TC de columna vertebral torácica que ilustra un ejemplo de error cognitivo. Las venas basivertebrales que transcurren por dentro del cuerpo vertebral pueden ser interpretadas como fractura.
- Figura 12.** Imágenes coronal y sagital de RM de rodilla potenciadas en densidad protónica con saturación grasa. Muestran un error cognitivo en un paciente con osteoporosis transitoria que presentaba gonalgia progresiva. El edema óseo no siempre representa contusión.
- Figura 13.** Lesión GLAD (disrupción articular glenolabral), visible sólo en la artroRM, no en la RM convencional.
- Figura 14.** RM de rodilla de un varón de 27 años deportista profesional. Comparación de imágenes adquiridas en equipos de 3 Tesla y 1,5 Tesla para detección de lesiones condrales.
- Figura 15.** Imagen sagital de TC (derecha), que muestra fractura del extremo de la apófisis coronoides, no visible en la imagen sagital T1 de RM de bajo campo.

- Figura 16.** Imagen coronal de angioRM, que muestra una falsa imagen de trombosis de la vena yugular debido a artefacto.
- Figura 17.** Imágenes axiales de TC de cráneo con artefacto en diana visible sólo en 2 cortes. Se debe a fallo de calibración del scanner.
- Figura 18.** Imagen sagital T2 con saturación grasa de RM de cadera en la que se aprecia un artefacto por osteosíntesis que simula necrosis. Las imágenes coronales T1 muestran un hueso normal.
- Figura 19.** Ejemplos de factores latentes en Radiodiagnóstico.
- Figura 20.** Modelo de queso suizo de Reason de los contribuyentes activos y latentes a los errores.
- Figura 21.** Defensas y filtros de seguridad en el modelo de queso suizo de Reason.
- Figura 22.** Filtros de seguridad insuficientes en el modelo de queso suizo de Reason.
- Figura 23.** Imágenes coronales de RM de rodilla que muestran una rotura horizontal del menisco medial con un colgajo meniscal desplazado al receso coronario.
- Figura 24.** Selección de estudios para el análisis de factores latentes en la discrepancia diagnóstica.
- Figura 25.** Estudios incluidos en el análisis según su tipología.
- Figura 26.** Estudios incluidos en el análisis según hora y país de valoración inicial.
- Figura 27.** Estudios incluidos en el análisis según hora y tipo de día de la valoración inicial.
- Figura 28.** Modalidad de los estudios programados y urgentes.
- Figura 29.** Discrepancia diagnóstica en los estudios programados y urgentes.
- Figura 30.** Distribución de las características de interés para los estudios urgentes, programados y para el total de estudios.
- Figura 31.** Evolución mensual de la tasa de discrepancias diagnósticas estimada.
- Figura 32.** Tasa de discrepancias diagnósticas estimada según características del estudio.
- Figura 33.** Curva ROC asociada al modelo predictivo obtenido para los estudios programados.
- Figura 34.** Curva ROC asociada al modelo predictivo obtenido para los estudios urgentes.
- Figura 35.** Curva ROC asociada al modelo predictivo obtenido en la submuestra.
- Figura 36.** OR estimado e IC95% univariante de los factores de riesgo considerados para los estudios programados.
- Figura 37.** OR estimado e IC95% en el análisis univariante de los factores de riesgo considerados para los estudios urgentes.

**EL ERROR DIAGNÓSTICO EN RADIOLOGÍA. ANÁLISIS DE
VARIABLES EXPLICATIVAS**

1. INTRODUCCIÓN.

1.1. EL ERROR EN RADIOLOGÍA.

“Errar es humano, perseverar en el error es de necios”.
Cicerón.

“Errar es humano, pero más lo es culpar de ello a otros”.
Baltasar Gracián.

En Radiología, como en todas las ramas de la Medicina, existe el riesgo inevitable de exponer al paciente al error humano. Se define el error médico como “el fallo de una actuación planificada para ser llevada a cabo conforme a lo previsto o el uso de un plan incorrecto para conseguir un objetivo” (Kohn, Corrijan, & Donaldson, 2000) (Moskop, Geiderman, Hobgood, & Larkin, 2006), independientemente de que cause o no un daño (Brook et al., 2010). En Radiología se considera error diagnóstico al diagnóstico que se ha omitido, que es incorrecto o que se ha retrasado, de acuerdo con el resultado de una prueba definitiva o de un hallazgo a posteriori (Graber M., 2005).

Es fundamental distinguir entre el error (un proceso) y el evento adverso o daño al paciente (un resultado). Un efecto adverso es una lesión causada por un error médico y no por la condición médica del paciente. Algunos ejemplos de efectos adversos son el neumotórax tras colocación de un catéter venoso central, la anafilaxis por penicilina, o el desarrollo de trombosis venosa profunda durante la estancia hospitalaria (Bhimani & Bruno, 2018). Sin embargo, no todos los errores diagnósticos producen daño en el paciente. De hecho el término “incidente” hace referencia a un error que no conduce a un resultado negativo para el paciente, pero que en circunstancias distintas podría haber sido un efecto adverso. También puede definirse como un hecho que no descubierto o corregido a tiempo podría haber implicado daños para el paciente (Bhimani & Bruno, 2018). A su vez, no todos los daños que se producen en el paciente son consecuencia de un error diagnóstico, sino que pueden estar relacionados con la propia evolución de la enfermedad (Newman-Toker & Pronovost, 2009).

La Comisión Conjunta (*Joint Commission*) define un evento centinela como un suceso inesperado que implica muerte o lesión física o psicológica grave, o el riesgo de que se produzcan. Estos eventos se denominan centinela porque indican que puede existir un problema grave en el sistema que requiera una investigación inmediata y una respuesta (Bhimani & Bruno, 2018).

En Radiología los errores pueden ocurrir desde que el clínico considera la solicitud de un estudio de imagen hasta el momento en que tanto los médicos destinatarios del informe como el paciente comprenden plenamente el resultado de la prueba y las recomendaciones del radiólogo (Allen, 2014).

No siempre está claro qué es propiamente un “error diagnóstico” (Van Such, Lohr, Beckman, & Naessens, 2017). A veces es posible ser concluyente en un diagnóstico radiológico, pero generalmente la interpretación radiológica está influenciada por las circunstancias del paciente, sus antecedentes y estudios de imagen previos, y otros múltiples factores, incluyendo sesgos de los que los radiólogos pueden no siempre ser conscientes (Dekker, 2006) (Conklin, 2012). Existe gran variabilidad tanto en la anatomía normal como en la patología. Además, los estudios radiológicos no incluyen marcas indicando las alteraciones más significativas, e interpretarlos no es un proceso binario (normal versus patológico, cáncer versus “sin patología”) (Brady A. P., 2017). En este sentido, se define “opinión” como “una conclusión que se alcanza tras la consideración de la evidencia, pero abierta a debate o a sugerencias”, sin esperar que sea incontrovertible (Robinson, 1997). Esta situación se ejemplifica en estudios como el de Armato et al., el cual muestra una importante variabilidad entre cuatro radiólogos torácicos experimentados al establecer “la verdad” en la determinación del número y localización de nódulos pulmonares de más de 3 mm en 25 estudios de tomografía computarizada) (TC) (Armato et al., 2009).

En 2004, el escritor Malcolm Gladwell en un artículo en *The New Yorker*, hizo la siguiente observación en relación a la falta de certeza diagnóstica, estrechamente relacionada con el error diagnóstico en Radiología: “ la razón por la que un radiólogo se ve obligado a asumir que la abrumadora cantidad de hallazgos dudosos son normales es, en otras palabras, que dicha abrumadora cantidad de hallazgos dudosos son en realidad normales (Gladwell, 2004). “Los radiólogos actúan, por ilustrarlo con un ejemplo gráfico, de forma similar a los vigilantes de seguridad de equipajes en los aeropuertos. Lo más

probable es que un objeto negro en el medio de una maleta no sea una bomba debido a que han visto cientos de objetos negros como ese en maletas anteriormente, y ninguno de ellos era una bomba. Pero eso, por supuesto, no significa que no sea una bomba. Todo con lo que cuentan es con el aspecto que muestra en la pantalla de rayos X –y la pantalla rara vez aporta suficiente información-“ (Gladwell, 2004).

Para articular esta falta de certeza diagnóstica en Radiología surge el concepto de variabilidad, de la cual existen dos tipos: interobservador e intraobservador. La variabilidad interobservador ocurre cuando diferentes radiólogos no alcanzan un consenso en el diagnóstico, pero en una situación en que ninguna de las opiniones es errónea. Por el contrario, la variabilidad intraobservador existe cuando un mismo radiólogo no emite el mismo diagnóstico ante varios pacientes diferentes que muestran el mismo aspecto radiológico. La variabilidad puede ocurrir en relación con la detectabilidad de una lesión (por ejemplo, cardiomegalia –presente o ausente-), o al gradar la gravedad de la lesión (por ejemplo, cardiomegalia –leve o moderada-). Se considera en general que es más difícil realizar un diagnóstico radiológico con alta certeza en aquellas patologías en las que la variabilidad inter e intraobservador es alta (Robinson, 1997). Por tanto es importante considerar la variabilidad ya que puede modificar la sensibilidad y especificidad diagnósticas (Hansson et al., 2009).

En definitiva, sólo puede considerarse error cuando no hay variabilidad (Kaewlai & Abujudeh, 2012). La variabilidad en un diagnóstico por el contrario puede reflejar un error pero también puede corresponder simplemente a una genuina diferencia de opinión (Berlin, 2018).

Una línea poco definida entre variabilidad y error ocurre cuando no puede concluirse una respuesta definitiva al motivo de solicitud de la prueba diagnóstica por otros medios, ya que la comprobación patológica o quirúrgica no está disponible (Kaewlai & Abujudeh, 2012). La Radiología ofrece un diagnóstico de presunción, no un diagnóstico histológico ni microbiológico. Por tanto, y debido a la subjetividad de la interpretación de las imágenes radiológicas, la definición de error depende con frecuencia de una “opinión experta”, es decir, un observador comete un error si no alcanza la misma conclusión que haría un grupo de observadores expertos (Fitzgerald, 2001). Si, por el contrario, los expertos no coinciden en su opinión al respecto, las diferentes interpretaciones de un hallazgo de imagen se consideran una discrepancia o variabilidad interobservador

(Huntley et al., 2015). A efectos prácticos en nuestro trabajo emplearemos el término discrepancia diagnóstica entre dos radiólogos como sinónimo de error diagnóstico, pues es la aproximación más exacta de que disponemos al carecer de medios de confirmación definitiva, sin desconocer por ello que la variabilidad es parte inevitable de los procesos de diagnóstico por la imagen.

1.2. RELEVANCIA DEL ERROR DIAGNÓSTICO.

La evaluación de los errores diagnósticos en Radiología comenzó con la revelación pionera de Garland en 1949 de que los estudios radiológicos están expuestos a cierto error inherente, y de que el grado de error o variación es significativo (Garland, 1949). Garland centró su atención en un hecho hoy conocido y aceptado por todos, pero que en ese momento todavía no se había reconocido sin importantes reticencias: existe una prevalencia muy importante de error de interpretación en los informes radiológicos.

Fue a partir de 1999 cuando empezó a aumentar el grado de vigilancia hacia los errores, coincidiendo con la publicación por el Instituto de Medicina de EE.UU. del artículo “Error es humano: construyendo un sistema de salud más seguro” (Kohn et al., 2000). Este trabajo revelaba el alarmante dato de que hasta 100.000 personas morían cada año en EE.UU. como resultado de errores médicos (Kohn et al., 2000). Sucesivos estudios posteriores recogidos por medios de comunicación confirmaban la magnitud del problema e indicaban que la asistencia médica no estaba siendo óptima (Newman-Toker & Pronovost, 2009) (Drozda et al., 2011) (James J.T, 2013) (Abujudeh, Kaewlai, Shaqdan, & Bruno, 2017). Los estudios más recientes elevan incluso el número de muertes anuales por errores médicos en EE.UU. a más de 151.000 por año, situando a los errores médicos como la tercera causa de muerte en EE.UU. por delante de los accidentes de tráfico y el SIDA y sólo por detrás del cáncer y las enfermedades cardíacas (Figura 1) (Makary & Daniel, 2016). Hay que considerar que estas cifras representan tan sólo aquellos pacientes cuyo diagnóstico ha tenido resultado de muerte, y no incluyen la discapacidad, que puede ser igual de frecuente (Saber Tehrani et al., 2013). Según la ley de Heinrich, en un lugar de trabajo por cada error que causa una lesión grave, 29 errores causan lesiones menores y 300 accidentes no causan lesiones (Heinrich, 1931). Un estudio observacional de Singh et al estimaba que el 5% de los adultos de los EE.UU, unos 12 millones de personas cada año, recibían un diagnóstico incorrecto (Singh et al., 2014). Para estimar la incidencia de error diagnóstico a nivel mundial debe tenerse en cuenta que se realizan anualmente en torno a un billón de exploraciones radiológicas, y que la mayoría de las mismas son interpretadas por radiólogos (Krupinski & Samei, 2010). Asumiendo una tasa media de error del 4% (en el extremo inferior de las estimaciones que se han hecho para el error diagnóstico en Radiología), se estarían cometiendo anualmente 40 millones de errores diagnósticos en Radiología (Bruno, Walker, & Abujudeh, 2015).

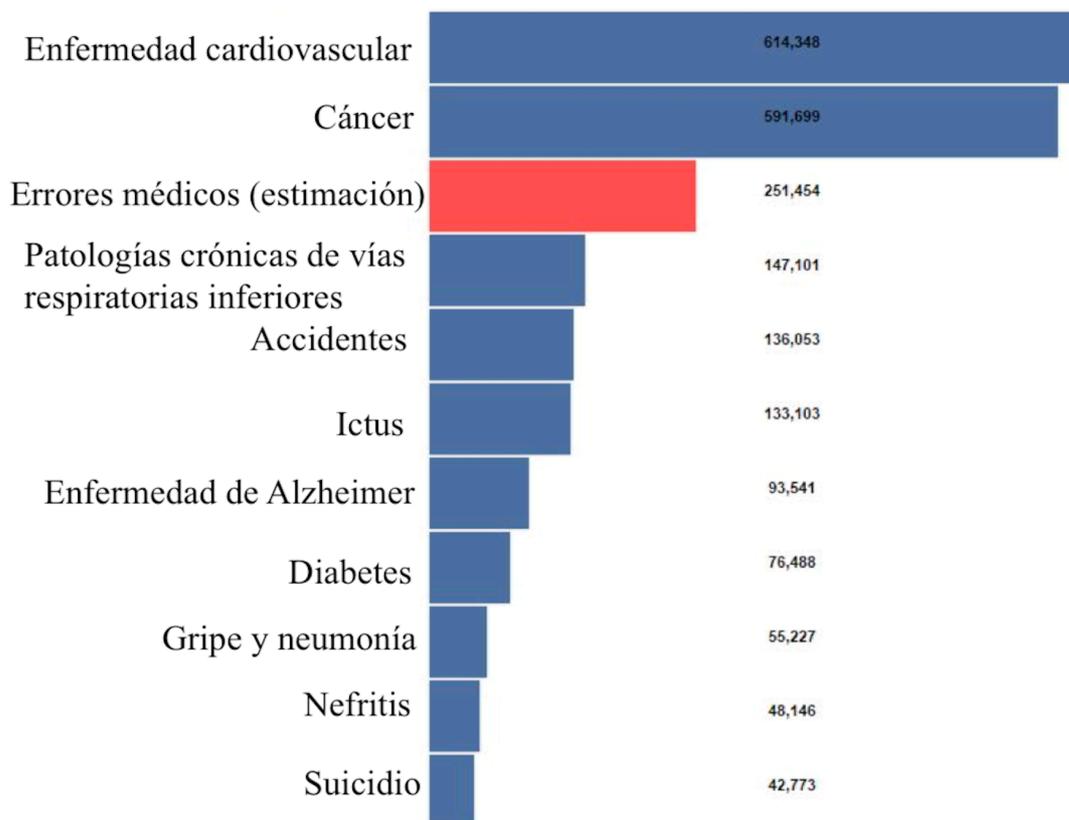


Figura 1. Número de muertes anuales en 2014 de las principales causas de muerte en EE.UU. Fuente CDC. Centro Nacional de Estadísticas Sanitarias.

Dentro del conjunto de errores médicos, los errores diagnósticos no han empezado a estudiarse hasta hace poco. Por ejemplo, en el informe “Errar es humano” el término “errores de medicación” se mencionaba con una frecuencia 35 veces superior al término “error diagnóstico” (Wachter, 2010). Es el reflejo de que los esfuerzos sistemáticos para medir y analizar las causas de los errores diagnósticos han sido limitados en comparación con otros esfuerzos de mejora en el ámbito de la calidad y la seguridad en la asistencia sanitaria (Wachter, 2010) (Croskerry, 2012). Esta asimetría en la gestión de calidad no se corresponde con la realidad de la práctica médica; en un estudio realizado en la Universidad de Harvard se observó que los errores médicos que condujeron a eventos adversos eran con más frecuencia diagnósticos que en relación con el tratamiento (14% vs 9%) (Leape et al., 1991). Se sabe que estos errores pueden desembocar en retrasos en el diagnóstico y/o el tratamiento, en no reconocer complicaciones del tratamiento, en realizar estudios no indicados o en no supervisar o monitorizar a un paciente (Figura 2) (Brenner et al., 1998). Además ciertos errores

diagnósticos (como por ejemplo no diagnosticar una enfermedad infecciosa como la tuberculosis) no afectan solamente a un paciente aislado, sino a muchas otras personas con las que el paciente en cuestión entra en contacto (Khullar et al., 2015).

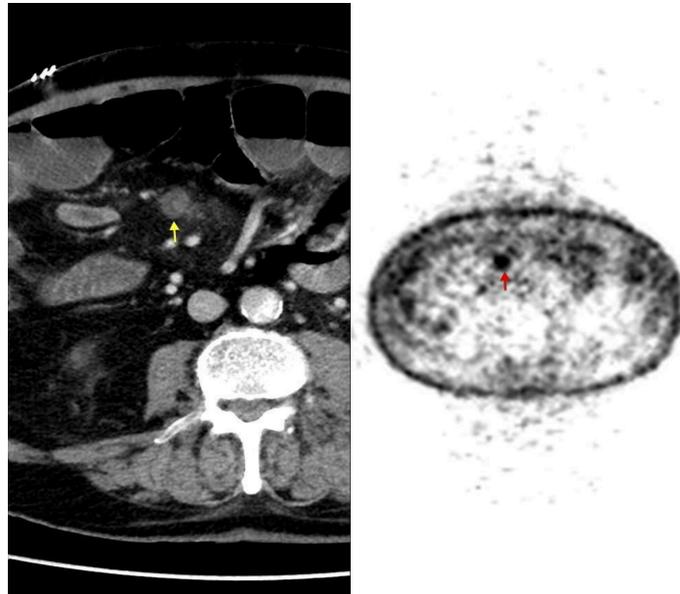


Figura 2. Imágenes de TC (izquierda) y PET (derecha) de un paciente oncológico, en las que se identifica una adenopatía mesentérica que modifica el tratamiento.

En 2015 el Instituto de Medicina (IOM) publicó un nuevo informe sobre la prevalencia y el alcance del error diagnóstico, titulado “Mejorando el Diagnóstico en la Asistencia Sanitaria” (*Committee on Diagnostic Error in Health Care, 2015*). Este informe afirma que:

- a) La mayoría de las personas experimentarán un error diagnóstico a lo largo de su vida.
- b) El 5% de los pacientes ambulatorios de los EE.UU. experimentan un error diagnóstico cada año.
- c) Los errores diagnósticos contribuyen al 10% de las muertes de pacientes y entre el 6% y el 17% de los eventos adversos en los hospitales.
- d) Los errores diagnósticos son la principal causa de indemnización por demandas de mala praxis médica.

Dicho informe del Instituto de Medicina se basa en grandes estudios retrospectivos de revisión de registros de eventos adversos que han confirmado un alto porcentaje de errores diagnósticos (Tabla 1).

ESTUDIO	AÑO	PAÍS	PORCENTAJE DE EVENTOS ADVERSOS RELACIONADOS CON ERROR DIAGNÓSTICO
Baker et al. <i>The Canadian adverse events study: The incidence of adverse events among hospital patients in Canada.</i>	2004	Canadá	10,5%
Davis et al. <i>Adverse events in New Zealand public hospitals II: Preventability and clinical context.</i>	2003	Nueva Zelanda	8%
Thomas et al. <i>Incidence and types of adverse events and negligent care in Utah and Colorado.</i>	2000	EE.UU.	6,9%
Leape et al. <i>The nature of adverse events in hospitalized patients. Results of the Harvard medical practice study II.</i>	1991	EE.UU.	17%

Tabla 1. Estudios retrospectivos de revisión de eventos adversos.

Los errores diagnósticos (en sentido amplio, no únicamente en el ámbito del Diagnóstico por la Imagen) constituyen además el evento adverso más frecuente referido por los residentes de Medicina (Weingart et al., 2000) (Wu et al., 2003). Blendon et al encuestaron a pacientes y médicos sobre hasta qué nivel ellos o algún miembro de su familia habían experimentado errores diagnósticos, definidos como fallos de diagnóstico que “desembocan en daño grave, como muerte, discapacidad, o tratamiento adicional o de mayor duración”. Encontraron que el 35% de los médicos y el 42% de los pacientes refirieron dichos errores (Blendon et al., 2002). Una encuesta más reciente de 2.201 adultos de los EE.UU. encontró resultados similares (*YouGov survey of medical misdiagnosis. Isabel healthcare-clinical decision support system.2005*). En dicha encuesta, el 35% experimentaron un error médico (de diagnóstico o de tratamiento) en los 5 años previos afectándoles a sí mismos, a sus familias o a sus amigos; la mitad de dichos fallos se describieron como errores de diagnóstico. Es interesante el hecho de que el 55% de los encuestados afirmaron que un diagnóstico incorrecto era su mayor preocupación cuando asistían al médico en consultas ambulatorias, mientras que el 23% afirmaron que este era el tipo de error que más les preocupaba en un contexto hospitalario. También mostraron preocupación por los errores médicos el 38% de los pacientes que habían visitado

recientemente un Servicio de Urgencias; de ellas, la preocupación más frecuente era un diagnóstico incorrecto (22%) (Burroughs et al., 2005). En la Figura 3 se ilustra un error diagnóstico por omisión que con cierta frecuencia se produce en Urgencias, al no detectarse un hematoma subdural adyacente a la tienda del cerebelo.

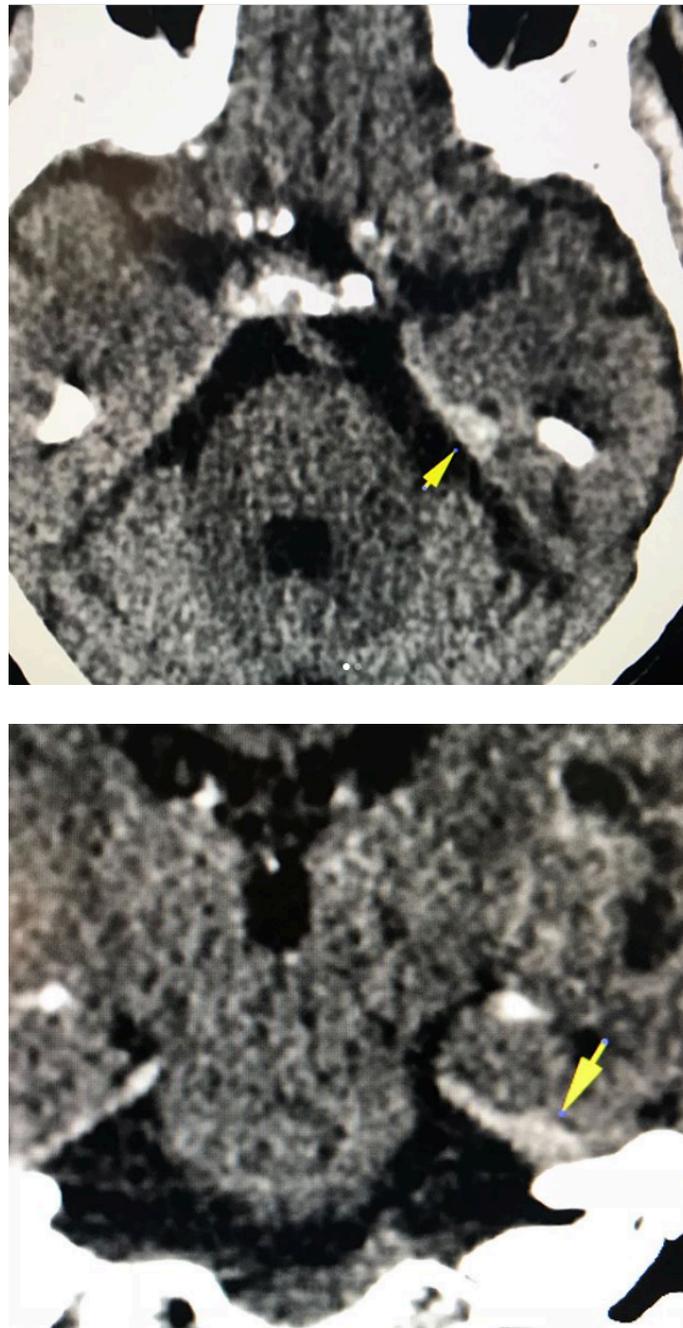


Figura 3. Imágenes axial y coronal de TC que muestran un hematoma subdural en el tentorio (flechas), omisión frecuente en Urgencias.

Los errores diagnósticos son típicamente la primera o la segunda causa de demandas por mala praxis (Balsamo & Brown, 1998) (*Physician Insurers Association of America &*

Rockville., 2002) (Berlin, 2004a) (*Failure to diagnose. Misdiagnosis of conditions and diseases. Medical malpractice lawyers and attorneys online*, 2006), las cuales a su vez son: el doble de frecuentes que las demandas por errores de medicación, las que con mayor frecuencia alcanzan un veredicto desfavorable para el médico (Berlin, 2018) y las que se asocian a mayores indemnizaciones (Weeks et al., 2001). En EE.UU. un tercio de los radiólogos son demandados y, a su vez, un tercio de estos son acusados de negligencia médica (Whang et al., 2013). En Radiología los problemas más frecuentes que conducen a demandas por mala praxis se deben a diagnósticos erróneos (Pinto & Brunese, 2010), entendiéndose como tales tanto las omisiones diagnósticas como la interpretación equivocada de imágenes radiológicas (Berlin, 1986) (Berlin & Berlin, 1995). Y lo que agrava la situación para los radiólogos es que, a diferencia del resto de especialidades (con excepción de Anatomía Patológica), no sólo cometemos errores sino que almacenamos las imágenes para permitir una revisión retrospectiva de las mismas, lo cual también es causa de observaciones desfavorables de políticos y abogados por igual (Fitzgerald, 2001) (Goddard et al., 2001).

La importancia clínica y económica de los errores diagnósticos aumenta en paralelo al progreso de la Medicina. En el pasado el sistema sanitario tenía menor capacidad (y posiblemente menor necesidad) de afrontar este problema. Al ser las opciones de tratamiento más limitadas existía una menor posibilidad de iatrogenia por diagnósticos equivocados. Las herramientas disponibles para controlar y prevenir los errores diagnósticos estaban menos desarrolladas, y la presión para estudiar y abordar este problema también era menor. También conforme las opciones de tratamiento se han vuelto más efectivas y caras, los costes clínicos y económicos de errar el diagnóstico de una patología fácilmente tratable son sustancialmente mayores. Téngase en cuenta por ejemplo cómo ha variado en las dos últimas décadas el coste de tratar un cáncer de colon en estadio precoz. Hoy en día tratar un cáncer de colon en estadio IV cuesta más del triple que tratar un estadio I, debido en parte a la aparición de nuevos tratamientos potencialmente efectivos aunque de coste elevado. Antes de que estos tratamientos estuvieran disponibles, los costes clínicos del error diagnóstico eran altos, pero no lo eran tanto los costes económicos. Conforme aumente la disponibilidad de tratamientos caros para enfermedades avanzadas, los costes de los errores diagnósticos probablemente aumenten (Khullar et al., 2015). Las estimaciones de incidencia anual de errores médicos en EE.UU. están entre 44.000 y 400.000 (Kohn et al., 2000) (Makary & Daniel, 2016). En términos económicos se ha calculado que los errores médicos pueden

tener un coste de aproximadamente entre 29 y 38 billones de dólares anualmente tan sólo en EE.UU. (Johnson C. D. et al., 2009) (Kruskal, Anderson, Yam, & Sosna, 2009).

En definitiva, los errores diagnósticos en Radiología tienen costes en términos de vidas humanas, pérdida de ingresos y productividad, efectos físicos, psicológicos y económicos en los pacientes y sus familias y pérdida de confianza en el sistema sanitario. Los profesionales sanitarios implicados también padecen consecuencias físicas y psicológicas cuando los pacientes sufren un daño como consecuencia del error (“síndrome de estrés por mala praxis” o “síndrome de la segunda víctima”) (Wu, 2000) (West et al., 2006) (Sadigh et al., 2015) y tienden a modificar su comportamiento profesional hacia prácticas de medicina defensiva (Studdert et al., 2005). Es por tanto crucial para la comunidad radiológica encontrar formas de mitigar y prevenir errores que potencialmente causen resultados perjudiciales. Además de las necesidades de la práctica radiológica en sí misma, los hospitales, las organizaciones, los estados, las entidades económicas, y lo más importante, los pacientes, demandan mayor transparencia de los médicos. Reducir los errores significa no sólo mejorar el cuidado al paciente sino también aumentar la eficiencia de los servicios de Radiología, reducir los costes y evitar el estrés y el aislamiento en caso de demanda al radiólogo.

1.3. FRECUENCIA DEL ERROR DIAGNÓSTICO.

Al investigar la frecuencia del error diagnóstico en Radiología, y pese a la distinción anteriormente explicada entre los términos “error” y “discrepancia”, observamos que con frecuencia se produce implícitamente una asimilación de ambos conceptos, por la imposibilidad de confirmar de forma definitiva el diagnóstico más allá de la falta de concordancia entre dos lecturas. En este sentido hacemos una consideración importante para el desarrollo posterior de nuestros razonamientos: con frecuencia se usa el término “discrepancia diagnóstica” como sinécdoque (metonimia) del término “error diagnóstico”.

Las tasas de discrepancia diagnóstica en Radiología recogidas en la bibliografía varían ampliamente, oscilando entre un 0,8% hasta un 58% según las series y criterios estudiados (Siegle et al., 1998) (Borgstede et al., 2004) (Soffa et al., 2004) (Berlin, 2007a) (Abujudeh et al., 2010) (Pinto & Brunese, 2010) (Lee et al., 2013).

Garland, en su artículo sobre error diagnóstico (“Sobre la evaluación científica de los procedimientos diagnósticos”) del *American Journal of Roentgenology* en 1939 (Garland, 1949) (Garland, 1959) describió que cuando radiólogos experimentados interpretaban radiografías patológicas existía una tasa de omisiones diagnósticas (discrepancia interobservador) del 30%. Esto significa que si 100 estudios radiológicos con hallazgos patológicos son distribuidos de forma ciega a los radiólogos, una media del 30% de las alteraciones pasarán inadvertidas. Garland también observó que el 2% de las radiografías normales fueron sobrediagnosticadas (falsos positivos), y que el 20% de los radiólogos discrepaban consigo mismos (discrepancia intraobservador) en una segunda lectura de los estudios (Garland, 1959).

Un estudio de 2010 en el que tres radiólogos experimentados especialistas en Radiología abdominal revisaron 90 estudios de TC abdominopélvica y, en una fecha posterior, sin conocer sus interpretaciones previas hubieron de reinterpretar estudios radiológicos tanto de sí mismos como de sus compañeros, la tasa de discrepancia interobservador era del 26% y la discrepancia intraobservador del 32% (Abujudeh et al., 2010), prácticamente idénticas a las observadas por Garland 60 años antes. Y existen otros múltiples trabajos posteriores a Garland que han reproducido la misma tasa media de omisión diagnóstica

del 30% en la evaluación retrospectiva de radiografías convencionales, tomografía computarizada, resonancia magnética (RM) y ecografía, sin ninguna mejoría apreciable a pesar de los avances en la tecnología radiológica y en el conocimiento clínico (Stevenson, 1969) (Lehr et al., 1976) (Berlin, 1977) (Markus J B et al., 1989) (Renfrew et al., 1992) (Berlin, 1994) (Berlin, 1996a) (Siegle et al., 1998) (Robinson, Wilson, Coral, Murphy, & Verow, 1999) (Berlin, 2001) (Fitzgerald, 2001) (Soffa et al., 2004) (FitzGerald, 2005) (Potchen, 2006) (Bruno et al., 2015). En cuanto a la patología de Urgencias, por ejemplo, se han evidenciado tasas de error en torno al 35% entre radiólogos interpretando estudios de pacientes que habían sufrido un traumatismo (Janjua et al., 1998).

Algunos trabajos muestran cifras incluso superiores. Un estudio realizado por radiólogos de la Universidad de Harvard reveló que los radiólogos discrepan en las interpretaciones de las radiografías de tórax hasta el 56% de las ocasiones (Herman PG et al, 1975). Otro estudio reveló que hasta el 90% de todos los carcinomas de pulmón no eran detectados por los radiólogos cuando interpretaban las radiografías (Muhm et al., 1983). En la misma línea, las tasas de error diagnóstico observadas en mamografía son alarmantes (Baines et al., 1990) (Beam et al., 1996) (Kerlikowske et al., 1998) (Berlin, 1999b) (Berg et al., 2000) (Sickles, 2000). Un estudio de la Universidad de Arizona comprobó que en el 75% de las mamografías inicialmente interpretadas como normales se observaban carcinomas de mama retrospectivamente (Harvey et al., 1993).

AÑO	AUTOR	TÉCNICA/PATOLOGÍA	OBSERVACIÓN
1939	Garland LH.	Radiología convencional. Múltiples patologías.	En estudios patológicos: discrepancia interobservador del 30%, 2% de falsos positivos y discrepancia interobservador del 20%.
1975	Herman PG et al.	Radiología convencional de tórax. Múltiples patologías.	Tasa de discrepancia del 56%.
1983	Muhm JR et al.	Radiología convencional de tórax. Cáncer de pulmón.	90% de falsos negativos.
1989	Markus JB et al.	Radiología convencional de abdomen. Múltiples patologías.	Concordancia interobservador buena-excelente para 11 signos y diagnósticos radiológicos, y pobre para otros 8 signos/diagnósticos radiológicos.
1999	Robinson PJ et al.	Radiología convencional en Urgencias. Múltiples patologías.	Discrepancia entre el 8 y el 19% de los casos. Incidencia media de error por observador en el rango de 3-6%.
1998	Janjua KJ et al.	Radiología de Urgencias. Múltiples técnicas y patologías.	Discrepancia interobservador de hasta el 35%
2010	Abujudeh et al.	Múltiples técnicas y patologías.	2,9% de discrepancias en casos difíciles, 0,8% de discrepancias en casos sin especial dificultad.
1993	Harvey JA et al.	Mamografías. Cáncer de mama.	75% de falsos negativos.
1996	Beam CA et al.	Mamografías. Cáncer de mama.	Variabilidad del 40% de sensibilidad diagnóstica entre los radiólogos.
2000	Berg WA et al.	Mamografías. Cáncer de mama.	45% de variabilidad interobservador y 15% de variabilidad intraobservador en el manejo de las lesiones.
1998	Siegle RL et al.	Múltiples técnicas y patologías.	Tasa media de discrepancia del 4,4%
2004	Borgstede JP et al.	Múltiples técnicas y patologías.	2,9% de discrepancias en casos difíciles, 0,8% de discrepancias en casos sin especial dificultad.
2004	Soffa DJ et al.	Múltiples técnicas y patologías.	Tasa global de discrepancia del 3,48%: 3,03% en radiología general, 3,61% en mamografía diagnóstica, 5,79% en mamografía de screening y 4,07% para ecografía. Tasas de discrepancia por radiólogo oscilan entre 2,04% y 6,90%.

Tabla 2. Relación de estudios que analizan la tasa de discrepancia diagnóstica.

Los estudios anteriormente mencionados, sin embargo, no reflejan la práctica diaria de la Radiología, en la que el número de estudios normales excede con creces aquellos que son patológicos. Al analizar la tasa de error, Garland usó exclusivamente estudios patológicos, es decir, aquellos con una prevalencia de enfermedad del 100%. Debido a que, como ya se ha mencionado, en un entorno clínico habitual hay un importante número de estudios sin hallazgos patológicos, Garland emitió la hipótesis de que la tasa de error radiológico esperable en la práctica diaria era próxima al 5% (Garland, 1949).

Por tanto las tasas de error pueden calcularse de dos formas diferentes, dependiendo del denominador que se utilice (Garland, 1959). Pueden computarse solamente estudios patológicos o por el contrario una mezcla de estudios patológicos y normales, siendo este último caso el aplicable a la práctica asistencial diaria. Si el denominador consiste sólo en estudios con patología, la tasa media de error se sitúa en torno al 30%. Si el denominador consiste en una mezcla de casos patológicos y no patológicos representativos de la práctica asistencial diaria, la tasa de error se sitúa entre el 3,5% y el 4% (Berlin, 2014).

Es importante además considerar el grado de fiabilidad de un resultado falso negativo cuando los hallazgos clínicos son altamente indicativos de que existe una patología. Esto es así porque la “probabilidad a posteriori” (post-prueba) de que dicha patología esté presente se calcula según parámetros bayesianos en base tanto a los hallazgos clínicos por un lado como a los resultados de la prueba diagnóstica por otro (Sackett et al, 1985). Por eso en aquellos entornos con presumiblemente mayor prevalencia de enfermedad, como los centros médicos universitarios, las tasas de error tienden a ser mayores que en los entornos con más cantidad de estudios normales como los hospitales no universitarios (Borgstede et al., 2004) (Berlin, 2007a) (Berlin, 2007b) (Zan et al, 2010). Un ejemplo ilustrativo de esta situación es que aunque la sensibilidad globalmente estimada de la angiografía por RM (angioRM) para los aneurismas iguales o mayores de 6 mm es del 95%, cuando la probabilidad pre-prueba es alta (por ejemplo del 90%) (es decir, cuando la clínica del paciente es muy sugestiva de la presencia de un aneurisma), la tasa de falsos negativos es del 33%. En estas circunstancias no es razonable por tanto considerar que un resultado negativo de una angio-RM descarta el diagnóstico (Johnson M R. et al., 2001).

Un reciente metaanálisis de revisión de las tasas de discrepancia de 58 estudios diferentes a lo largo de un periodo de 66 años, todos ellos analizando las discrepancias en la interpretación radiológica de TC de adultos, reveló una tasa global de discrepancia del 7,7%, con un 2,4% de discrepancia grave (Wu M Z, McInnes, Macdonald, Kielar, & Duigenan, 2014). Había una considerable heterogeneidad de la menor muestra. Los estudios de menor tamaño muestral mostraban mayores tasas de discrepancia y mayor variabilidad de las mismas, probablemente debidos al efecto causado por el menor tamaño de las muestras. No se observaron discrepancias significativas en las tasas de discrepancia entre adjuntos y residentes. Este metaanálisis también mostró que la tasa

de discrepancia grave era seis veces menor en los estudios en que el radiólogo revisor tenía acceso al informe inicial que cuando este era oculto.

Se han descrito también tasas de error similares en estudios de imagen urgentes. Un estudio realizado por Robinson et al que evaluaba el error diagnóstico en las radiografías simples del Servicio de Urgencias mostró discrepancias importantes entre dos observadores en el 5 a 9% de los casos, y estimó una incidencia de error por observador de entre el 3 y el 6% (Robinson et al., 1999). Y en un estudio de TC realizadas a pacientes con traumatismos, Geyer et al observaron una tasa de error diagnóstico del 10,4%, con impacto clínico en el 6,7% de los casos (Geyer et al., 2013).

En definitiva, los estudios sobre el trabajo de los radiólogos en condiciones habituales (cuando el conjunto de estudios interpretados incluye una mezcla de casos normales y patológicos con una prevalencia de enfermedad representativa de la realidad) revelan una tasa media de error entre el 3% y el 4% (Borgstede et al., 2004). Debido a que muchos radiólogos interpretan más de 100 estudios en una jornada laboral habitual, esto se traduce en aproximadamente de tres a cuatro errores al día por cada radiólogo, de media (Bhimani & Bruno, 2018). Afortunadamente, sólo una pequeña porción de estos errores generan daño al paciente (Berlin, 2007a), pero es preocupante que la mayoría pasen inadvertidos; generalmente los radiólogos nunca llegan a ser conscientes de la mayoría de sus propios errores y sólo reciben observaciones útiles y a tiempo para actuar de una fracción muy pequeña de la totalidad real de sus errores (Bhimani & Bruno, 2018).

Las diferencias de resultados entre los distintos estudios probablemente representen variabilidad en la metodología, incluyendo disparidad en las definiciones de discrepancia, tipos de técnicas radiológicas evaluadas, selección de casos, poblaciones de pacientes y radiólogos individuales (Abujudeh et al., 2010). Por ejemplo, algunos estudios pueden partir del análisis de los errores diagnósticos (no todos los cuales conducen a un efecto adverso) y algunos otros estudios pueden estar dirigidos a evaluar los eventos adversos (los cuales pasarán por alto aquellos errores diagnósticos que no producen daño o incapacidad significativos). Otro posible sesgo es incluir sólo los casos con patologías difíciles, seleccionadas generalmente de conferencias y archivos personales (Renfrew et al., 1992) (Kim Y. W. & Mansfield, 2014). Además pueden existir conclusiones sesgadas sobre el alcance de los diagnósticos incorrectos al considerar sólo aquellos pacientes que fallecieron y a los que se les realizó autopsia, o aquellos que interpusieron demandas por

mala praxis, o aquellos que sufrían de una enfermedad grave; todas estas situaciones conducen a sobreestimar la tasa de errores, porque dichos ejemplos no son representativos de la mayoría de pacientes que atienden los médicos. Por otra parte en los pacientes ambulatorios dada la fragmentación de la asistencia sanitaria, la dificultad del seguimiento de los pacientes y el tiempo que tarda una enfermedad en manifestarse de forma clara, también puede existir una infraestimación de la tasa real de error (Berner, Miller, & Graber, 2007).

Otro factor que influye en la tasa de error diagnóstico es el tipo de sistema sanitario y sus coberturas. Por ejemplo, un estudio mostró que las mujeres en los EE.UU. tienen el doble de posibilidades de tener una biopsia de mama negativa que en el Reino Unido (*Working Group chaired by Professor Sir Patrick Forrest. Breast Cancer Screening, 1986*) (Henson et al., 1996) (Ballard-Barbash R. et al, 1997) (May D. Et al, 1998) (*UK National Health Service Breast Screening Web site. The NHS Breast Screening Programme, 2003*). Esto se explica debido a que los médicos en EE.UU solicitan el doble de estudios de seguimiento a las pacientes (estudios de imagen, biopsia, o seguimiento clínico estrecho). Sin embargo, las tasas de detección de cáncer en ambos países eran comparables (Smith-Bindman et al., 2003).

En relación al resto de especialidades médicas, aunque los errores diagnósticos se producen en todas, generalmente son más bajos en las basadas en la interpretación visual: Radiología, Dermatología y Anatomía Patológica (Berner & Graber, 2008).

En Dermatología la mayoría de los estudios evalúan el diagnóstico de lesiones pigmentarias (por ejemplo descartar melanoma). En un estudio de 5.136 biopsias, se encontró un cambio significativo de diagnóstico en el 11% de las segundas revisiones. Aproximadamente el 1% de los diagnósticos se cambió de benigno a maligno, aproximadamente el 1% de los diagnósticos se cambió de maligno a benigno y en aproximadamente el 8% de los casos se cambió el grado tumoral suficientemente como para modificar el tratamiento (McGinnis et al., 2002).

Tasas de error diagnóstico similares se han descrito en Anatomía Patológica. En el año 2000, la Sociedad Americana de Patólogos Clínicos convocó una conferencia de consenso para revisar las segundas lecturas en Anatomía Patológica (Tomaszewski et al., 2000). En dicho estudio, el Servicio de Anatomía Patológica del Hospital John Hopkins solicitó una

segunda lectura de cada una de las 6.171 muestras obtenidas durante un periodo de 18 meses; en 1,4% de los casos se encontraron discrepancias que implicaban cambios importantes de tratamiento o de pronóstico (Kronz et al., 1999). Un estudio similar en el centro médico Harsley de Pensilvania identificó una incidencia de cambios clínicamente significativos del 5,8% (Tomaszewski et al., 2000). El rango de discrepancia variaba con el tipo de muestra, desde 1,3% en las muestras de próstata al 5% en los tejidos del sistema reproductor femenino y el 10% en pacientes con cáncer. Algunos tejidos eran notoriamente difíciles, por ejemplo las tasas de discrepancia variaban entre el 20 y el 25% en los linfomas y sarcomas (Kim H. et al., 1982) (Harris et al., 1991). Un artículo más reciente del Wall Street Journal (Laura Landro, 2006) mostró una tasa de error de entre el 25% y el 49% en las interpretaciones que realizaban los patólogos de las muestras de biopsia, y un segundo artículo reveló una tasa de error del 24% en las pruebas de laboratorio (Plebani, 2006).

Las tasas de error en las especialidades clínicas son superiores a las de las especialidades de percepción, algo coherente con la mayor necesidad de recopilación y de síntesis de información (Berner & Graber, 2008). Un estudio de los ingresos en hospitales británicos mostró que el 6% de los diagnósticos al ingreso eran incorrectos (Neale et al., 2001). Hay que considerar que la actividad de un Servicio de Urgencias implica decisiones complejas en condiciones de incertidumbre y estrés por encima de la media. La tasa de error diagnóstico en este contexto varía entre el 0,6% y el 12% (O'Connor et al., 1995) (Chellis et al, 2001).

Ya en un artículo de 1959 Garland referenció estudios que mostraban una tasa de error del 34% en el diagnóstico de infarto de miocardio, una tasa de concordancia de sólo el 15% entre ocho internistas experimentados sobre la presencia de “los más simples signos de enfisema” al examinar el tórax de pacientes afectados por dicha patología, una discrepancia muy importante en la evaluación clínica de indicación de amigdalectomía en 1.000 escolares, una tasa de concordancia de sólo el 7% entre pediatras experimentados al determinar clínicamente si los niños sufrían o no malnutrición, una tasa de error del 20% en la interpretación de electrocardiogramas, una tasa de error del 28% entre 59 laboratorios clínicos hospitalarios diferentes al informar de los resultados de análisis químicos, y una tasa de error del 28% entre laboratorios clínicos al medir el recuento eritrocitario (Garland, 1959). Investigaciones posteriores han confirmado y ampliado estas revelaciones de Garland sobre los errores en la medicina clínica. La

concordancia entre médicos de centros universitarios al realizar exámenes físicos para esplenomegalia (Grover et al., 1993), hepatomegalia (Naylor, 1994), ascitis (Williams & Simel, 1992), otitis media aguda (Pichichero & Poole, 2001) y otra serie de hallazgos físicos (Elliot, 1987) (Sackett, 1992) ha mostrado ser baja. Estudios de autopsias con grandes muestras de casos han revelado frecuentes errores clínicos, con tasas que alcanzan el 47% (Anderson et al., 1989) (Roosen et al., 2000).

En resumen, los estudios muestran de forma constante una tasa de error diagnóstico que oscila entre valores inferiores al 5% en las especialidades “perceptuales” (Anatomía Patológica, Radiología, Dermatología) hasta entre el 10 y el 15% en la mayoría de las otras especialidades (Berner & Graber, 2008).

Para valorar si el porcentaje de error diagnóstico ha variado o no con el tiempo puede recurrirse a los datos de las autopsias. Estas revelan una tasa de omisión diagnóstica importante en el 25% de los casos; a pesar de las técnicas más modernas de diagnóstico, el número de diagnósticos importantes omitidos no ha variado a lo largo de 30 años. Un metaanálisis de 53 artículos de autopsias a lo largo de un periodo de 40 años reveló una media de error en patologías importantes del 23,5%, con un rango de entre 4% y 50%. “La tasa de error permanece sorprendentemente invariable”, enfatizaron los autores (Shojania, Burton, McDonald, & Goldman, 2003).

Esta uniformidad en la tasa de error diagnóstico mostrada por diferentes estudios sugiere que el problema es muy básico, que no es el resultado de “malos médicos” (en nuestro caso, “malos radiólogos”) y que la solución no puede basarse simplemente en una intervención tecnológica o educativa (Bhimani & Bruno, 2018). Por ello surge la necesidad de profundizar en el conocimiento de la naturaleza del error diagnóstico, los tipos de errores y los diferentes mecanismos que explican su aparición.

1.4. CLASIFICACIÓN DEL ERROR DIAGNÓSTICO.

Los errores diagnósticos en Radiología son tan frecuentes como en cualquier otro campo de la Medicina y pueden ocurrir en múltiples fases de la actividad. Se han cometido errores en la identificación del paciente, en el diseño o la ejecución de los protocolos de estudio, en la dosis de radiación, en la interpretación de la imagen, en la comunicación de los hallazgos, y también en las intervenciones radiológicas (Jones D N. Et al., 2010). De todos ellos puede decirse que el más importante es el error en la interpretación de las imágenes porque tiene una fuerte asociación con el desenlace final de la salud del paciente (Filatova & Bruno, 2018).

Es importante clasificar los errores para comprender por qué ocurren y también para poder desarrollar sistemas de detección y análisis que minimicen tanto su aparición como el grado de daño resultante (Brook et al., 2010).

La clasificación más general de los errores es aquella que distingue entre errores activos y errores latentes.

Los errores activos ocurren en un punto de contacto entre el ser humano y el sistema, como por ejemplo al pulsar un botón incorrecto (Berlanstein, 2018). Generalmente se asocian con actividades ejercidas por trabajadores “en primera línea”, en la interfaz ser humano-entorno operativo. En Radiología son los errores producidos por el propio radiólogo y consisten en un diagnóstico omitido o incorrecto, en no recomendar seguimiento u otras pruebas de imagen adicionales o en comunicar deficientemente los hallazgos críticos o inesperados al médico prescriptor de la prueba (Renfrew et al., 1992) (Donnelly & Strife, 2005) (van Beuzekom, Akerboom, & Boer, 2007) (Pinto & Brunese, 2010) (Kim & Mansfield, 2014). En la Figura 4 se ilustra un ejemplo de omisión diagnóstica, de un hematoma epidural espontáneo en la columna cervical.

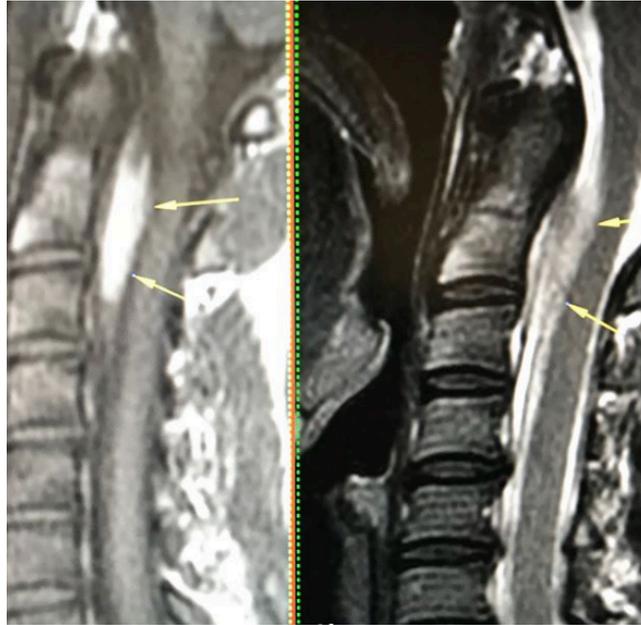


Figura 4. Imágenes sagitales de RM T1 (izquierda) y STIR (derecha) en mujer de 35 años con cervicalgia, que muestran un hematoma epidural espontáneo (flechas). Los casos espontáneos son raros (0,1 de cada 100.000), y requieren descompresión quirúrgica si hay déficit. Se omiten fácilmente, pues no siempre son tan hiperintensos.

El 70% de los errores diagnósticos activos en Radiología son de percepción, es decir, aquellos en los que los radiólogos no detectan una alteración presente en el estudio (Figuras 5 y 6). El 30% restante de los errores son cognitivos, en los cuales el radiólogo identifica la alteración pero le atribuye un significado incorrecto, bien por falta de conocimientos o por un razonamiento erróneo (Berlin, 1996b) (Kabadi & Krishnaraj, 2017).

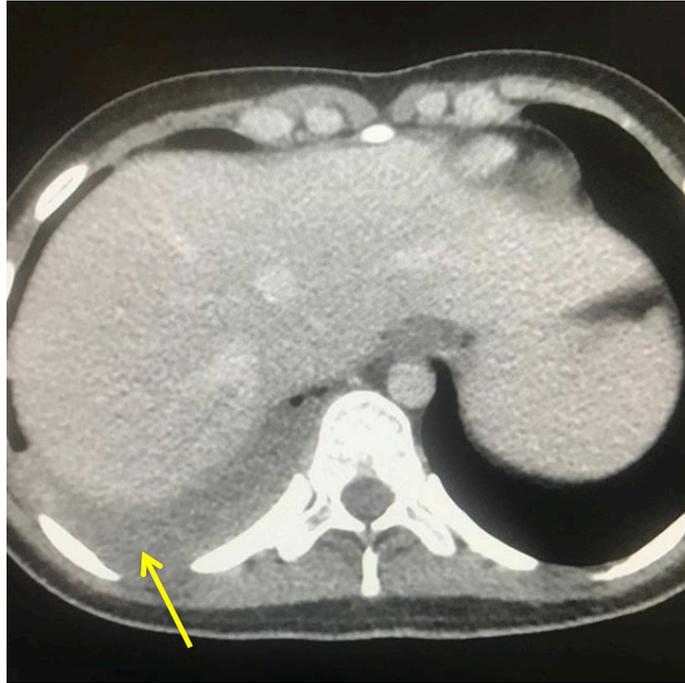


Figura 5. Imagen axial de TC abdominal por dolor y fiebre en paciente de 17 años, que presentaba derrame pleural (flecha) y neumonía derecha. En la lista de verificación de abdomen deben incluirse las bases pulmonares. En este caso se omitió el derrame pleural (flecha).



Figura 6. Imagen axial de TC abdominal que muestra neumoperitoneo por diverticulitis. Hasta en el 17% de los pacientes se omite, por no usar un adecuado nivel de ventana.

Los errores latentes son, por el contrario, aquellas variables que pueden predisponer a la aparición de errores activos, o aumentar el impacto negativo de estos (Renfrew et al., 1992) (Brook et al., 2010) (Pinto & Brunese, 2010) (van Beuzekom, Boer, Akerboom, &

Hudson, 2010). Se consideran “accidentes a la espera de ocurrir” (Berlanstein, 2018). Incluyen una amplia variedad de elementos, entre otros el material de trabajo inadecuado, la sobrecarga de trabajo, la falta de experiencia del radiólogo, la información clínica insuficiente, y la no disponibilidad de estudios previos (Renfrew et al., 1992) (Brook et al., 2010) (Pinto & Brunese, 2010) (van Beuzekom, Boer, Akerboom, & Hudson, 2010).

Factores activos	Factores latentes
Actos imprudentes que están directamente relacionados con un error	Condiciones previas a actuaciones no seguras
Inmediatamente evidentes	Menos evidentes. Pasan inadvertidos durante largo tiempo
Relacionados con los individuos	Relacionados con el sistema
Implican en sí mismos un alto riesgo	Predisponen a error

Tabla 3. Comparativa de factores latentes y activos (Brook et al., 2010).

Los errores activos son más fácilmente identificables porque son el resultado de acciones de individuos concretos y producen un efecto inmediato (por ejemplo un retraso en el diagnóstico y el tratamiento) (Brenner et al., 1998). Por el contrario, los errores latentes normalmente no son atribuibles a una única persona sino que son el resultado de fallos inherentes al diseño del sistema, que generan daño involuntario, aunque prevenible, al paciente (Bhimani & Bruno, 2018). Sus causas no son tan fáciles de identificar, en parte debido a que pueden permanecer inactivos durante largos periodos de tiempo, sólo haciéndose evidentes cuando se combinan con factores desencadenantes. Su rasgo definitorio es que estaban presentes dentro del sistema tiempo antes del inicio de la secuencia de producción del evento adverso (Reason, 1990).

En el 90% de los casos los dos tipos de error ocurren conjuntamente, es decir, un error activo ocurre en combinación con uno o varios errores latentes (Irish, 2008) (van Beuzekom et al., 2010).

1.4.1. ERRORES ACTIVOS.

1.4.1.1. Error de percepción.

La Radiología difiere del funcionamiento normal de la mayoría del resto de las especialidades médicas en que depende por completo de la percepción visual. Eso explica los resultados de múltiples estudios, según los cuales las omisiones diagnósticas o errores de percepción (falsos negativos) constituyen entre el 60% y el 80% de los errores cometidos por los radiólogos (Renfrew et al., 1992). Estos errores pueden tener drásticos efectos negativos retrasando el diagnóstico y el manejo de la enfermedad (Figura 7). Un ejemplo de esta situación puede observarse en el trabajo de Quekel et al, quienes observaron que el 19% de los cánceres de pulmón que se presentaban como un nódulo con un diámetro medio de 16 mm en las radiografías de tórax habían pasado inadvertidos en el diagnóstico radiológico (Quekel et al., 1999).

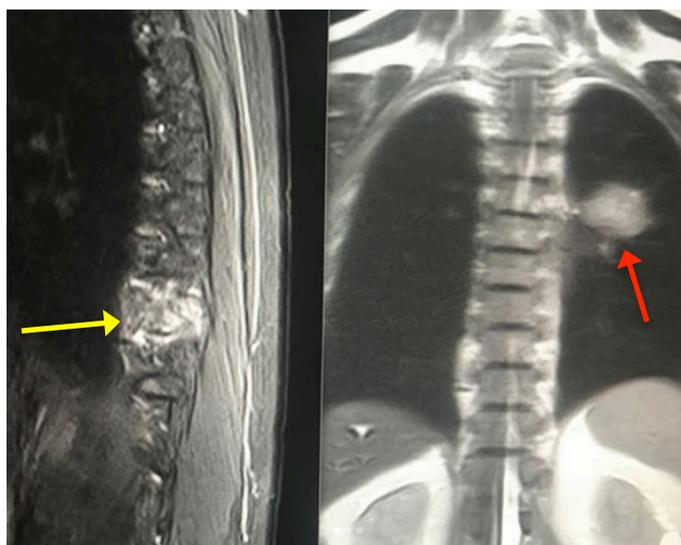


Figura 7. Imágenes de RM de columna vertebral torácica. El localizador (derecha) contiene información valiosa, pues muestra un carcinoma de pulmón (flecha roja), origen de la metástasis vertebral (flecha amarilla).

Probablemente no hay suceso más frustrante para los radiólogos que el darse cuenta de que han cometido un error de percepción, de que no “vieron” en un estudio de Radiología una alteración que valorada a posteriori es evidente. A pesar de la abundante bibliografía que se ha publicado sobre los errores de percepción (Garland, 1949) (Tuddenham, 1969) (Anderson et al., 1989) (Renfrew et al., 1992) (Leape, 1994), todavía no se sabe exactamente la respuesta a la pregunta que se plantean los radiólogos: ¿cómo y por qué no vi ese hallazgo? (Berlin, 2014). Aparentemente están implicados fenómenos

de búsqueda satisfactoria, fallos de reconocimiento, distracciones, fatiga visual, y tiempo de visualización inadecuado (Alpert & Hillman, 2004).

El fenómeno de búsqueda satisfactoria se produce en estudios de imagen que contienen varias alteraciones patológicas; el radiólogo detecta una o quizás dos alteraciones pero tiende a detener la búsqueda de hallazgos adicionales, su “búsqueda” ha sido “satisfecha” prematuramente (Figuras 8 a 10) (Robinson, 1997) (Rogers, 2000) (Kundel, 2006). En un trabajo en el que a los radiólogos se les mostraron en orden aleatorio una serie de estudios con una alteración y el mismo número de estudios con dos o más alteraciones, el 75% de las alteraciones fueron reseñadas en los casos con una o dos alteraciones. Sin embargo, en los estudios que contenían tres o más alteraciones, sólo se detectaron el 41% (Ashman et al., 2000). También pueden producirse errores de percepción por inexperiencia. Si un radiólogo nunca ha oído hablar del síndrome de Gardner por ejemplo, puede que no haga una búsqueda de tumores desmoides en la pared abdominal anterior (Berlanstein, 2018).



Figura 8. Imágenes coronales de TC de columna vertebral que muestran un error de búsqueda satisfactoria. Además de fractura vertebral (flecha amarilla) hay una tumoración renal derecha (flecha blanca).



Figura 9. Imagen sagital de TC de columna vertebral. El radiólogo detectó la fractura dorsal (flecha amarilla) y paró de buscar hallazgos adicionales, omitiendo la fractura sacra (flecha roja).

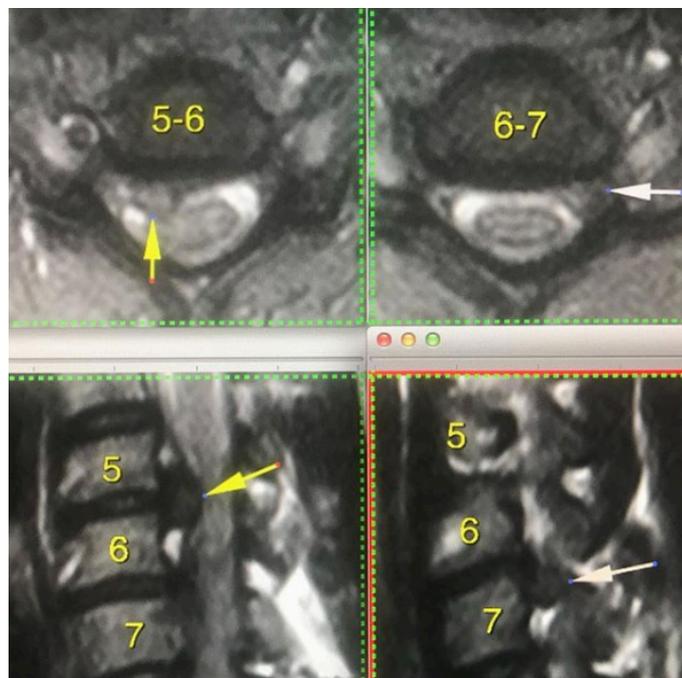


Figura 10. Imágenes axiales (superiores) y sagitales (inferiores) de RM de columna vertebral cervical en paciente de 40 años con cervicobraquialgia izquierda deficitaria. Debido a un error de búsqueda satisfactoria se omitió la hernia del disco C6-C7 (flecha blanca) debido a que se terminó prematuramente la búsqueda tras detectar la hernia del disco C5-C6 (flecha amarilla). Las hernias discales foraminales se omiten fácilmente, y con frecuencia son quirúrgicas.

No obstante, a pesar de haberse descrito algunos factores que pueden participar en la aparición de los errores de percepción estos no pueden ser completamente eliminados; son un riesgo inevitable de la condición humana al depender del procesamiento visual, que en sí mismo es imperfecto. Aunque la tecnología ha hecho enormes progresos en el último siglo, no hay evidencias de que se haya producido un avance similar en el funcionamiento del cerebro y la vista de los seres humanos (Robinson, 1997).

1.4.1.2. Error cognitivo.

Los errores cognitivos, de interpretación o de razonamiento, surgen cuando una alteración es identificada visualmente, pero su implicación diagnóstica o la importancia del hallazgo no se comprende de forma correcta y por tanto no se transmite de un modo efectivo al informe (Bruno et al., 2015).

Un caso relativamente frecuente de error cognitivo consiste en considerar estructuras anatómicas normales o variantes anatómicas de la normalidad como hallazgos patológicos (Figura 11). Otro ejemplo de error cognitivo sería detectar un nódulo pulmonar en una radiografía de tórax e interpretarlo erróneamente como una lesión benigna debido a sus bordes lisos y a no detectar que el nódulo ha ido aumentando progresivamente de tamaño en base a los estudios de imagen previos (Berlanstein, 2018).



Figura 11. Imagen axial de TC de columna vertebral torácica que ilustra un ejemplo de error cognitivo. Las venas basivertebrales que transcurren por dentro del cuerpo vertebral pueden ser interpretadas como fractura.

La mayoría de los errores cognitivos de diagnóstico ocurren durante el paso de “síntesis”, en el que el médico integra su conocimiento académico con la historia clínica del paciente y los hallazgos (Graber M L, Franklin, & Gordon, 2005). Este proceso es en gran medida subconsciente y automático.

Este tipo de error puede ser secundario a falta de conocimiento, a sesgos de razonamiento, o a información clínica confusa que distorsione la aparente probabilidad pretest de la patología; podría también deberse a que el radiólogo inadvertidamente propague un error cometido por un compañero en un informe radiológico previo (error aliterativo) (Bruno et al., 2015) o a que el diagnóstico diferencial no incluya patologías relevantes en relación a los hallazgos.

Las siguientes son algunas categorías importantes de sesgos cognitivos:

1. **Sesgo de anclaje:** es el efecto que ocurre cuando un radiólogo permanece centrado en su propia impresión inicial sobre el diagnóstico del paciente y no considera aquella información no concordante que apunta hacia el diagnóstico correcto (Croskerry, 2002)

Croskerry, 2003) (Lee C S. et al., 2013) (Waite et al., 2016). Por ejemplo, al conocer que el paciente presenta altos niveles plasmáticos de péptido natriurético cerebral, el radiólogo decide que el engrosamiento intersticial del pulmón representa edema pulmonar, a pesar de la presencia de una masa pulmonar y la consiguiente probabilidad de que dicho engrosamiento represente diseminación linfática del tumor (Bhimani & Bruno, 2018). Este sesgo se basa en el efecto de anclaje, que es la tendencia a asignar erróneamente demasiada importancia a la información aportada inicialmente cuando hay que tomar una decisión (Strack & Mussweiler, 1997). Por ejemplo, si se pregunta a un grupo de personas si la longitud del Golden Bridge es mayor o menor de 2500 m y a continuación se les pide que hagan una estimación de la longitud exacta del puente, las respuestas de dicho grupo estarán agrupadas en torno a 2500 m, independientemente de si esta es o no una medida precisa (Strack & Mussweiler, 1997). Este sesgo es particularmente peligroso cuando se combina con el sesgo de confirmación, en el que el radiólogo busca evidencia confirmatoria de su hipótesis inicial equivocada en vez de evidencia que pueda refutarla (Croskerry, 2002) (Croskerry, 2003) (Gunderman, 2009). Por ejemplo, en un estudio que se solicita por dolor en fosa ilíaca derecha con sospecha de apendicitis, el radiólogo puede detectar un apéndice cecal inflamado con moderada dilatación pero puede pasar por alto el edema y engrosamiento mural del ciego y colon ascendente, hallazgos estos últimos que hacen sospechar colitis con cambios reactivos secundarios en el apéndice (Bhimani & Bruno, 2018).

2. **Sesgo de disponibilidad:** ocurre cuando alguien permite que su experiencia profesional previa ejerza una influencia indebida sobre su razonamiento diagnóstico, porque los eventos frecuentes o recientes son más fáciles de recordar o de concebir que los infrecuentes, como se ilustra en la Figura 12. Por ejemplo, si un radiólogo no detectó un nódulo pulmonar en una radiografía de tórax, puede que tienda a sobrediagnosticar nódulos pulmonares en las radiografías que informe a partir de ese momento (Lee C S. et al., 2013).

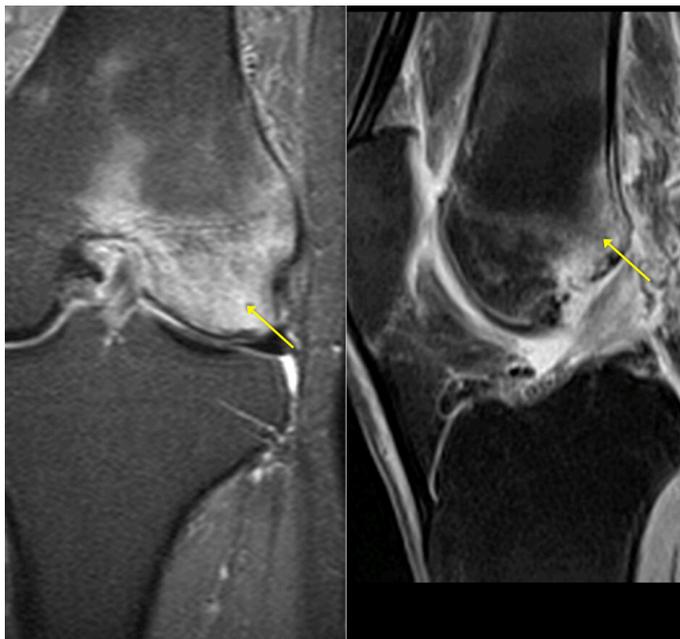


Figura 12. Imágenes coronal (izquierda) y sagital (derecha) de RM de rodilla potenciadas en densidad protónica con saturación grasa. Muestran un error cognitivo en un paciente con osteoporosis transitoria que presentaba gonalgia progresiva. El edema óseo no siempre representa contusión.

3. **Sesgo de resultado:** consiste en optar por un diagnóstico que implicará un mejor pronóstico para el paciente, como por ejemplo cuando el radiólogo es más proclive a diagnosticar un tumor como benigno a pesar de los hallazgos que sugieren malignidad (Berlin, 2004b).

4. **Sesgo de retrospectión:** es la predisposición a atribuir un diagnóstico correcto a unas brillantes habilidades interpretativas personales en lugar de reconocer que el hallazgo era reconocible de una forma mucho más evidente de modo retrospectivo, esto es, tras conocer su desenlace final (Berlin, 2000c). Por ejemplo, si se conoce que el paciente tiene el diagnóstico histológico de una neoplasia, el radiólogo puede tener mayor facilidad para interpretar determinada anomalía en una radiografía de tórax como un tumor primario. Este es un problema común en los juicios por mala praxis, cuando se acusa al radiólogo de haber omitido un hallazgo obvio, cuando en realidad puede ser obvio sólo retrospectivamente (Bhimani & Bruno, 2018).

5. **“Retirada de la cebra”:** es un sesgo particular de radiólogos expertos, en el que un diagnóstico poco frecuente (“cebra” frente a “caballos”) es apoyado por la historia clínica y por los hallazgos de imagen pero el radiólogo es reticente a emitirlo debido a su baja frecuencia (Miller, 2013).

6. **Cierre prematuro:** implica optar por un diagnóstico final muy específico antes de obtener suficiente información que lo confirme. Por ejemplo, en un paciente con miastenia gravis conocida que tiene una masa tímica, un radiólogo podría asumir el diagnóstico de timoma sin insistir en la necesidad de un diagnóstico histológico. Aunque el timoma está típicamente asociado con la miastenia gravis, esta masa tímica podría ser una hiperplasia tímica, un linfoma, o un tumor de células germinales (Lee et al., 2013).

7. **Sesgo de mención/omisión:** ocurre cuando el radiólogo no menciona la falta de certeza en un diagnóstico o la posibilidad de un diagnóstico alternativo más grave debido a las propias dudas en la probabilidad real de este último. Por ejemplo, al informar una TC de un paciente politraumatizado, una zona lineal de alta atenuación en la vecindad de la arteria esplénica puede interpretarse como una placa de ateroma calcificada de la arteria. Otra posibilidad sería que correspondiese a extravasación activa de contraste, pero el radiólogo podría no mencionarla en el diagnóstico diferencial por temor a que el paciente se sometiese a pruebas adicionales o a cirugía. Al citar esta posibilidad, otras pruebas complementarias no invasivas, como por ejemplo repetir la TC sin contraste del abdomen, podrían determinar si el hallazgo corresponde a calcificación o a hemorragia (Bhimani & Bruno, 2018).

8. **Sesgo de composición (“framing bias”):** ocurre cuando un radiólogo presta excesiva atención a un diagnóstico asignado previamente al paciente y no hace énfasis en aquellos detalles adicionales que pudieran apuntar hacia el diagnóstico correcto (Croskerry, 2003). Por ejemplo, al interpretar el radiólogo la TC torácica de una paciente que acude con dolor torácico, fiebre y leucocitosis, decide que la condensación alveolar periférica del pulmón representa una neumonía, a pesar de la presencia de un defecto de repleción en la arteria pulmonar tributaria y la consiguiente probabilidad de que dicha condensación represente un infarto pulmonar secundario a tromboembolismo.

Kim y Mansfield, en su estudio de 2014 publicado en el *American Journal of Roentgenology*, describen una nueva categoría de error radiológico, la **satisfacción del informe o error aliterativo**, que ocurre cuando un hallazgo se omite debido a una confianza excesiva en el informe de un estudio radiológico previo (Kim & Mansfield, 2014). Un ejemplo podría ser si un compañero con alta reputación profesional describió que el hígado era normal, sesgando a un segundo radiólogo en un control posterior a

pasar por alto heterogeneidades del parénquima que podrían en realidad reflejar hepatopatía difusa.

1.4.1.3. Error de comunicación.

Un radiólogo puede cometer un error incluso si diagnostica correctamente un caso. De hecho en el núcleo de muchos errores médicos se observa una comunicación defectuosa (Woolf et al., 2004), la cual supone una importante fuente potencial de daño al paciente.

Aunque en Radiología los hallazgos a veces se comunican directamente de forma oral, generalmente el informe radiológico es el único instrumento del radiólogo para comunicar su interpretación, conclusiones y consejo al médico solicitante. Es crucial por tanto evitar discrepancias entre la interpretación que hacen del contenido del informe el radiólogo y el médico solicitante (FitzGerald, 2005). No es útil para el paciente que una anomalía sea identificada por el radiólogo y correctamente descrita en el informe si ese informe no es suficientemente claro para el médico que lo recibe (Bruno et al., 2015). Mientras que en algunos campos de la Radiología (por ejemplo mamografía) es posible realizar un informe estandarizado, no ocurre lo mismo con gran parte de la especialidad, en que los informes constan tanto de un juicio diagnóstico como de una descripción de los hallazgos (Hobby et al., 2000) (Hayward, 2003). Y aquí es donde surgen una serie de posibles fallos. Encontramos deficiencias en la estructura u organización del texto, pobre elección de vocabulario, errores de gramática o de puntuación, y errores de escritura por funcionamiento subóptimo del sistema automático de reconocimiento de voz.

Dos errores de comunicación característicos son incorrección en la lateralidad de los hallazgos o en el género del paciente. Los errores en la lateralidad de los hallazgos pueden tener un impacto significativo en la asistencia, especialmente cuando el paciente requiere una intervención quirúrgica (Kwaan et al., 2006) (Seiden & Barach, 2006) (Luetmer et al., 2013). Los errores de género (por ejemplo el informe menciona una glándula prostática en una paciente mujer) pueden deberse al aspecto post-histerectomía de la pelvis femenina, inclusión accidental de plantillas de informe con texto predefinido, o errores de transcripción en los sistemas de dictado por reconocimiento de voz. Aunque estos errores con menos frecuencia impactan en el cuidado del paciente, generan confusión innecesaria y reducen la credibilidad de los radiólogos (Minn et al., 2015).

Con frecuencia los errores de comunicación en el informe radiológico están relacionados con los sistemas de reconocimiento automático de voz que los radiólogos usan para el dictado. Una de las principales desventajas de estos sistemas son los errores de transcripción, que incluyen entre otros deleción y transcripción errónea de palabras, generando frases confusas e imprecisas (Sangwaiya et al., 2009) (Hawkins, Hall, Hardin, Salisbury, & Towbin, 2012) (Larson, Towbin, Pryor, & Donnelly, 2013) (Hawkins, Hall, Zhang, & Towbin, 2014). Algunos errores como la sustitución de palabras por otras con un sonido similar, como por ejemplo “renal” por “adrenal”, “lateral” por “bilateral” o “hiperintenso” por “hipointenso” pueden tener un impacto clínico no tan relevante (Chang et al., 2011). Pero el riesgo de iatrogenia es mayor cuando se confunden unidades de medida (por ejemplo centímetros por milímetros), o cuando se omite o añade la palabra “no” como en “se observan signos de malignidad” por “no se observan signos de malignidad”. La importancia de este tipo de errores no radica únicamente en que puedan ser clínicamente significativos, sino también en que transmiten al lector la impresión de que el radiólogo no presta suficiente atención a los detalles, es descuidado y negligente, independientemente de cuán acertado sea el contenido del informe (Hawkins et al., 2012) (Kuzminski, 2016).

También constituyen errores de comunicación el no realizar un informe escrito del estudio, vaguedad o ambigüedad en el contenido y omisión del grado de certeza en relación a los hallazgos (Alpert & Hillman, 2004). Además de proporcionar un informe radiológico escrito, el radiólogo es responsable de comunicar los hallazgos directamente y a tiempo al médico solicitante cuando estos sean inesperados, críticos, o modifiquen el tratamiento inmediato (Kline & Kline, 1992), para evitar consecuencias de un retraso en el tratamiento. Dicha comunicación oral directa de los hallazgos, cuando se realice, deberá ser además documentada incluyendo la fecha, hora y nombre del interlocutor, así como el contenido de la comunicación (Raskin, 2006).

Asimismo, es un error de comunicación el no sugerir la siguiente técnica de imagen a realizar. En la Figura 13 se ilustra la importancia de una correcta elección del estudio de imagen para evitar errores diagnósticos. Los radiólogos deben asegurarse de realizar, cuando proceda, recomendaciones o sugerencias sobre cualquier procedimiento radiológico adicional que pueda añadir información relevante para clarificar, confirmar o descartar la impresión diagnóstica inicial (Pinto & Brunese, 2010).

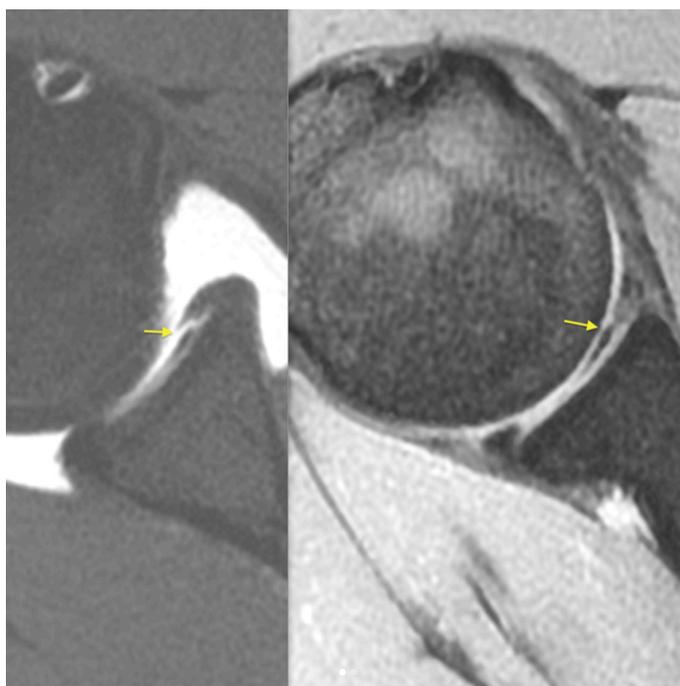


Figura 13. Lesión GLAD (disrupción articular glenolabral), visible sólo en la artrografía por RM (izquierda), no en la RM convencional (derecha).

Los errores de comunicación anteriormente descritos hacen referencia a la transmisión de los resultados del estudio de imagen. Sorprendentemente, suponen menos del 50% de los errores de comunicación en Radiología. El resto de errores ocurren durante la solicitud, la planificación y la realización del estudio (Siewert B. et al., 2015) (Siewert B. et al., 2016). Incluyen los procedimientos de consentimiento informado y comunicación con el paciente, documentación inadecuada (por ejemplo, confundir un paciente con otro, confundir la lateralidad de la patología, no documentar antecedentes de alergia al medio de contraste...) y en general todas las ausencias o defectos en los trámites pre-procedimiento (Brook et al., 2010). Un ejemplo sería un paciente con riesgo de aspiración cuya TC se solicita sin contraste oral y en el que, a través de una sucesión de errores de comunicación el paciente acaba recibiendo una dosis estándar de contraste oral, sufre una broncoaspiración, y requiere ser atendido en la UCI con prolongación del tiempo de estancia hospitalaria (Siewert B. et al., 2015).

Se ha observado que los errores de comunicación son la causa de hasta el 52% de todos los incidentes hospitalarios graves (Rabol et al., 2011) y del 56% de los eventos adversos en Radiología (*The Big Picture: Radiology Quality Improvement in the Hospital QA Setting: Communication Is (almost) Everything*, 2013). En un trabajo que analizaba casos

con múltiples errores médicos (en general, no limitados al ámbito de la Radiología), se detectó que en el 80% de los casos con una “cascada de errores” el desencadenante de la misma fue un error en la comunicación de información o en la comunicación interpersonal (Woolf et al., 2004). Woolf et al. observaron que aunque el 92% de los errores médicos implicaban errores en diagnóstico y tratamiento, dos tercios (67%) de estos errores fueron desencadenados por errores de comunicación.

Los errores de comunicación son el cuarto motivo más frecuente de reclamación legal contra los radiólogos en los juicios por mala praxis médica (Harrigal & Erly, 2007). Whang et al observaron que los errores de comunicación estaban implicados en el 4% de las demandas a radiólogos por mala praxis en las que podía determinarse un error primario (Whang et al., 2013). En la serie de Renfrew et al de demandas por mala praxis los errores de comunicación alcanzaban el 10% de los casos revisados (Renfrew et al., 1992), y Brenner et al mostraron una tasa de errores de comunicación del 20% en los casos revisados por la Asociación de Seguros Médicos de América. En un 10% de las demandas revisadas, no se remitió ningún informe radiológico al médico solicitante ni al paciente, y en otro 10% adicional el envío del informe ocurrió con retraso (Brenner et al., 1998).

En definitiva, la interacción entre los médicos solicitantes, los técnicos y los radiólogos conforma un circuito de transmisión de información altamente susceptible de sufrir errores, y que por ello debe estar sujeto a escrutinio y documentación constantes (Brook et al., 2010).

1.4.2. ERRORES LATENTES.

Los errores latentes son condiciones predisponentes a la materialización de un error activo. Estos errores pueden permanecer inadvertidos durante largos periodos de tiempo y puede que no se corrijan adecuadamente si no se realizan los procedimientos necesarios de vigilancia, la plantilla de trabajadores no informa de aquellos de los que es consciente, o si no hay sistemas de vigilancia continua en el entorno laboral. Hay 2 tipos de factores latentes: los técnicos y los relacionados con el sistema (Brook et al., 2010).

Los *factores latentes técnicos* incluyen aquellos relacionados con los equipos y material (defectos de diseño y de construcción, facilidad de uso, aspectos de seguridad),

diseño departamental, organización de flujos de trabajo, hardware, software, y fallos en los equipos, sistema de archivo y comunicación de imágenes (PACS) y la integridad del entorno digital, los materiales y su manejo (dispositivos, medios de contraste), protocolos, políticas, normativas, regulaciones, y mantenimiento rutinario de todos los sistemas involucrados. Los factores técnicos son particularmente relevantes en los servicios de Radiología, en los que pueden minimizarse con mantenimiento rutinario de los equipos y adherencia a guías de seguridad, los cuales deben ser claros y fáciles de llevar a cabo para evitar ambigüedades potencialmente peligrosas. Se encuadrarían dentro de los factores latentes aquellas características técnicas de la imagen derivadas de las características de los equipos, artefactos o protocolos de adquisición, que pueden condicionar una mayor incidencia de error diagnóstico, como se ilustra en las Figuras 14 a 18.

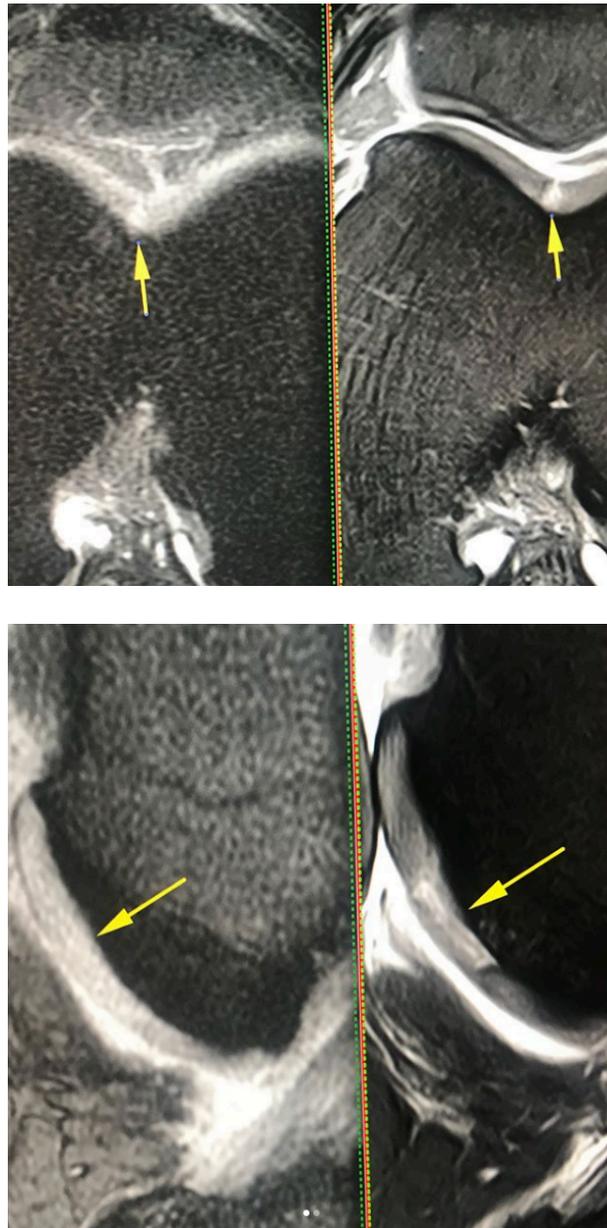


Figura 14. RM de rodilla de un varón de 27 años deportista profesional. Las imágenes adquiridas en equipo de 3 Tesla (derecha) muestran lesiones condrales (flechas amarillas) que pasan desapercibidas en equipos de 1,5 Tesla (izquierda).



Figura 15. Imagen sagital de TC (derecha), que muestra fractura del extremo de la apófisis coronoides (flecha), no visible en la imagen sagital T1 de RM de bajo campo (izquierda).



Figura 16. Imagen coronal de angioRM, que muestra una falsa imagen de trombosis de la vena yugular (flecha) debido a artefacto.

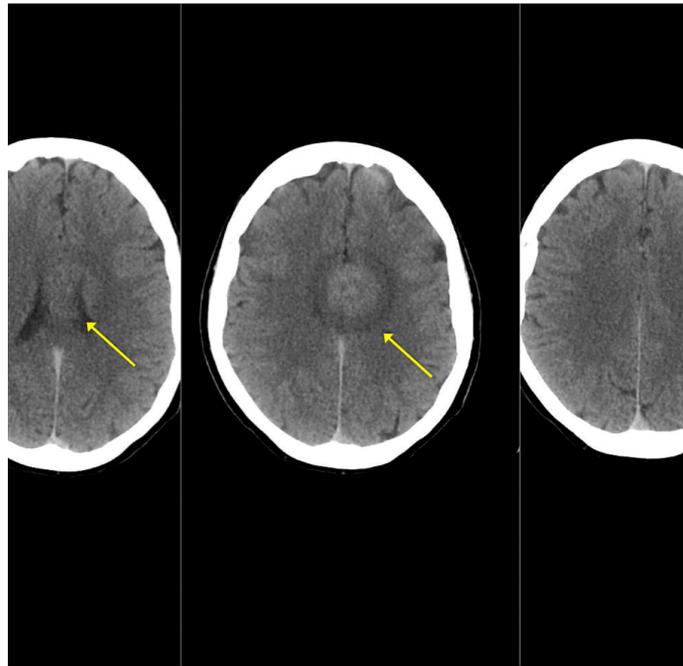


Figura 17. Imágenes axiales de TC de cráneo con artefacto en diana visible sólo en 2 cortes. Se debe a fallo de calibración del scanner. Puede interpretarse como tumor o hematoma.

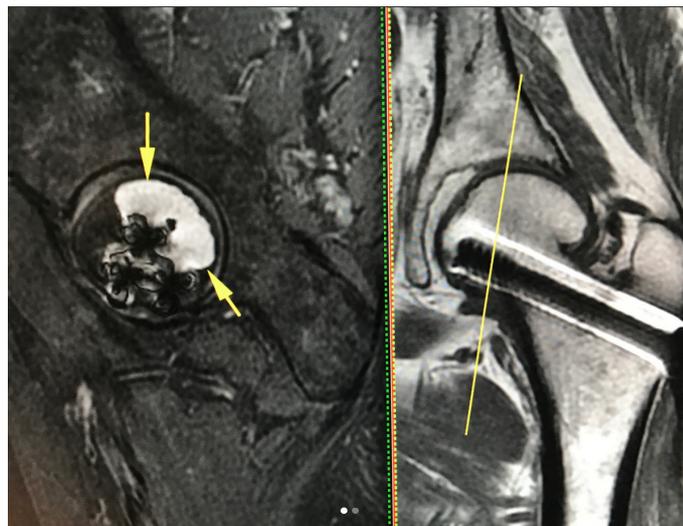


Figura 18. Imagen sagital T2 con saturación grasa de RM de cadera en la que se aprecia un artefacto por osteosíntesis que simula necrosis (flechas). Las imágenes coronales T1 (derecha) muestran un hueso normal.

Los *factores latentes del sistema* incluyen el tamaño y formación de la plantilla de trabajadores, los horarios de trabajo, la ergonomía, la cultura de seguridad y entrenamiento en liderazgo del departamento, supervisión y gerencia. Los errores del sistema están determinados por personas que toman decisiones a un nivel superior (gerentes, supervisores, jefes de departamento...). Por ejemplo, si la escasez de radiólogos

o la mala gestión de los recursos humanos conduce a los radiólogos a extender sus jornadas laborales, el cansancio físico general y la fatiga mental o visual (vista cansada o astenopía) pueden en último término conducir a estos radiólogos a cometer más errores de todo tipo. En tal escenario, estos errores adicionales sobre la tasa de error basal serían atribuibles al propio sistema laboral más que a fallos individuales o al error humano (Bhimani & Bruno, 2018). Aunque estén alejados física y temporalmente del error en sí mismo, los errores del sistema generalmente tienen una gran influencia.

<i>Disponibilidad de estudios de imagen previos.</i>	<i>Problemas en la entrega de los informes al médico solicitante.</i>	<i>"Pérdida" o no realización de informes.</i>	<i>Cancelación de estudios.</i>	<i>Estudios de imagen no indicados o no correctamente programados para la indicación clínica.</i>
<i>Codificación incorrecta de los datos del estudio.</i>	<i>Personal insuficiente debido a vacaciones.</i>	<i>Mantenimiento de las estaciones de trabajo y del resto del material.</i>	<i>Lentitud en el suministro del material.</i>	<i>Equipamiento insuficiente. Falta de espacio.</i>
<i>Material de baja calidad.</i>	<i>No disponibilidad de medios de contraste.</i>	<i>Falta de experiencia de los técnicos de Radiodiagnóstico.</i>	<i>Falta de experiencia de los radiólogos.</i>	<i>Factores que producen distracción de los radiólogos o del resto de la plantilla.</i>
<i>Protocolos o indicaciones incorrectos, imprecisos o subóptimos.</i>	<i>Listas de verificación de procedimientos demasiado complejas o redundantes.</i>	<i>Problemas en la calendarización y el manejo de los recursos humanos.</i>	<i>Flujos de trabajo subóptimos.</i>	<i>Dificultades para usar los equipos de trabajo.</i>
<i>Deficiente calidad de las imágenes.</i>	<i>Lentitud en la conexión informática.</i>	<i>Problemas con el PACS.</i>	<i>Tiempos de exploración prolongados.</i>	<i>Insuficiente número de equipos para el flujo de trabajo.</i>

Figura 19. Ejemplos de factores latentes en Radiodiagnóstico.

En la Figura 19 se enumeran varios de los múltiples posibles factores latentes que pueden existir en Radiodiagnóstico. Aunque no es un tipo específico de error, la presencia de un **gradiente de autoridad** es también un factor de riesgo conocido de error en Medicina. Fue descrito inicialmente en el contexto de la aviación cuando los copilotos sentían dificultades para comunicar errores a los pilotos a tiempo para prevenir un daño (como por ejemplo una colisión del avión), debido al sentimiento por parte del copiloto de inferioridad jerárquica. De un modo similar hay muchos niveles jerárquicos en un equipo médico (por ejemplo médicos y enfermería, adjuntos y residentes, médicos y farmacéuticos) en los que este gradiente de autoridad es aplicable. Esta es una causa significativa de error médico y una causa bien documentada de errores de aviación que han conducido a numerosas pérdidas de vidas en accidentes (Cosby & Croskerry, 2004).

1.5. PATOGENIA DEL ERROR DIAGNÓSTICO.

Para explicar cómo actúan los errores latentes en la aparición de un error médico se ha hecho una analogía de aquellos con los “patógenos residentes” dentro del cuerpo humano, que se combinan con factores externos (estrés, agentes tóxicos, etc) para producir la enfermedad (Reason, 1990).

Otro modelo utilizado con frecuencia para explicar la patogenia del error médico es el modelo de queso suizo de Reason (Reason, 1990). Para aumentar la seguridad en los sistemas se establecen múltiples barreras de defensa (Reason, 2008) (Brook et al., 2010). En Medicina algunas de estas defensas podrían ser por ejemplo verificar el consentimiento informado, la identificación del paciente, los antecedentes alérgicos, el estado de la coagulación... Cada defensa tiene agujeros (factores latentes) que en ocasiones permiten que un error la atraviese. Aisladamente ninguno de estos factores latentes produce un evento adverso. Mientras los agujeros no se alineen todos entre sí, los fallos humanos (errores activos) pueden ser bloqueados en algún punto antes de que lleguen a producir un evento adverso. Sin embargo, cuando coincide una combinación adecuada de factores latentes con un error activo, este sí puede atravesar todas las defensas y llegar finalmente a materializarse en daño al paciente (Larson, Kruskal, Krecke, & Donnelly, 2015) (Figuras 20 a 22).

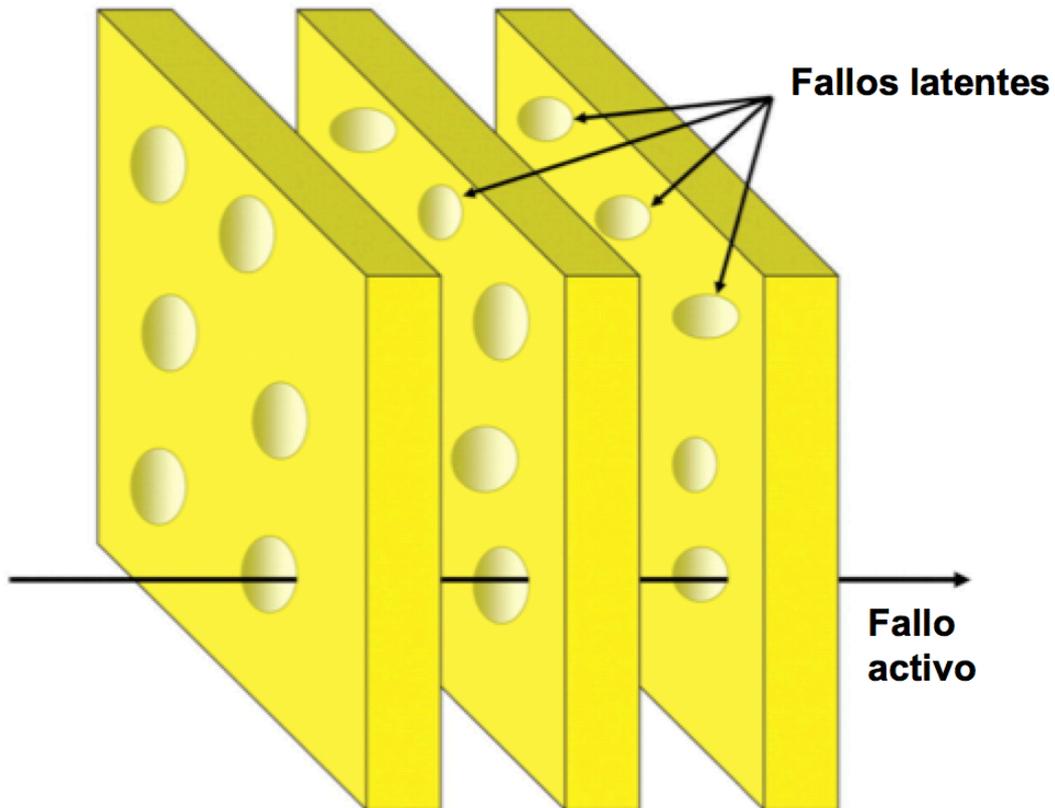


Figura 20. Modelo de queso suizo de Reason de los contribuyentes activos y latentes a los errores. Una serie de filtros de seguridad (rebanadas de queso suizo) existen en un sistema o se instauran en él para minimizar la oportunidad de que ocurra un error. En cada filtro de seguridad existen fallos latentes (agujeros en el queso) que predisponen al sistema a errores. Un único fallo latente es improbable que produzca un error; sin embargo, con la alineación adecuada de fallos latentes y en presencia de un fallo activo, es probable que se manifieste un error. Cuando se toman medidas para prevenir errores, la atención debe centrarse no sólo en los fallos humanos sino también en instaurar filtros de seguridad para prevenir que los fallos latentes se alineen. Traducido del inglés al español y reimpresso con permiso de la *Radiological Society of North America (RSNA)*. La RSNA no se hace responsable de cualquier inexactitud o error que surja de la traducción del texto. Brook, O. R., O'Connell, A. M., Thornton, E., Eisenberg, R. L., Mendiratta-Lala, M., & Kruskal, J. B. (2010). Quality initiatives: Anatomy and pathophysiology of errors occurring in clinical radiology practice. *Radiographics : A Review Publication of the Radiological Society of North America, Inc*, 30(5), 1401-1410. doi:10.1148/rg.305105013 [doi]

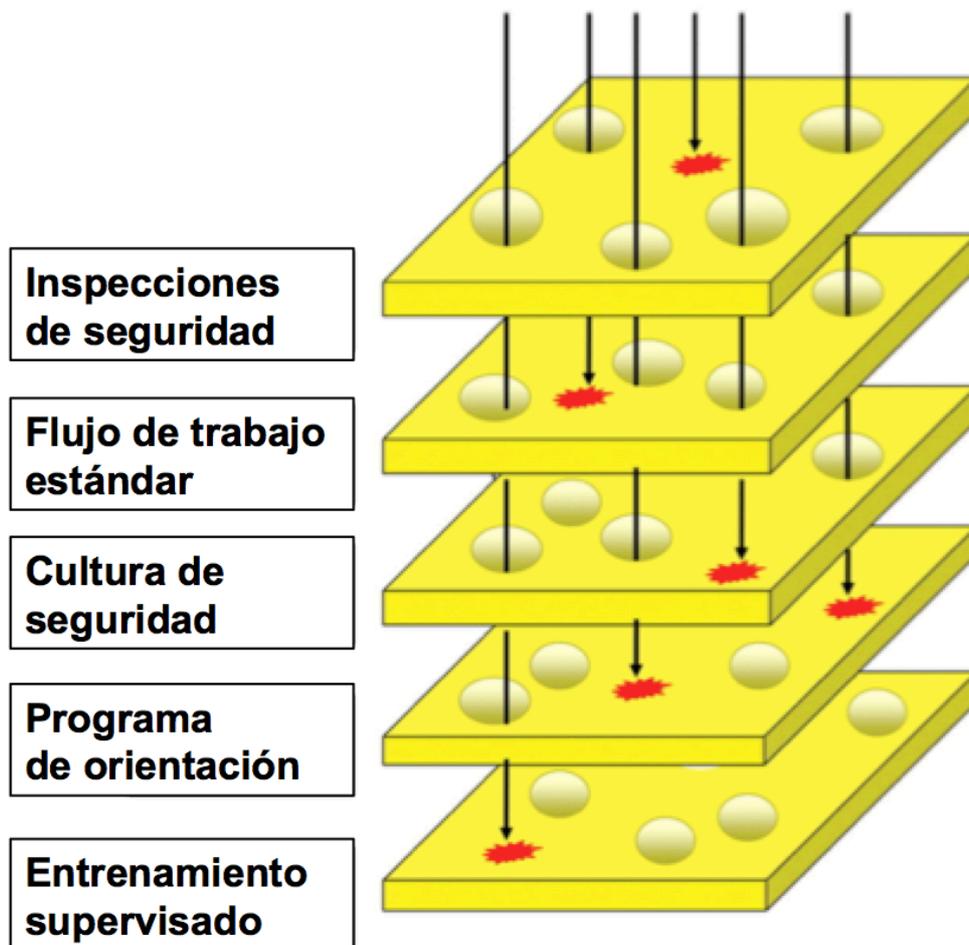


Figura 21. Defensas y filtros de seguridad. Se han introducido una serie de defensas y filtros de seguridad para minimizar la probabilidad de que los fallos latentes se alineen para producir un error. Al realizar inspecciones de seguridad frecuentes, introducir flujos de trabajo estandarizados, fomentar una cultura de seguridad a lo largo de un servicio, entrenar y orientar a todo el personal recién contratado, y asegurarse de que los procedimientos se supervisan adecuadamente, los efectos o complicaciones relativos a los contribuyentes latentes que residen dentro de un sistema pueden ser minimizados o incluso prevenidos. Traducido del inglés al español y reimpresso con permiso de la *Radiological Society of North America (RSNA)*. La RSNA no se hace responsable de cualquier inexactitud o error que surja de la traducción del texto. Brook, O. R., O'Connell, A. M., Thornton, E., Eisenberg, R. L., Mendiratta-Lala, M., & Kruskal, J. B. (2010). Quality initiatives: Anatomy and pathophysiology of errors occurring in clinical radiology practice. *Radiographics : A Review Publication of the Radiological Society of North America, Inc*, 30(5), 1401-1410. doi:10.1148/rg.305105013 [doi]

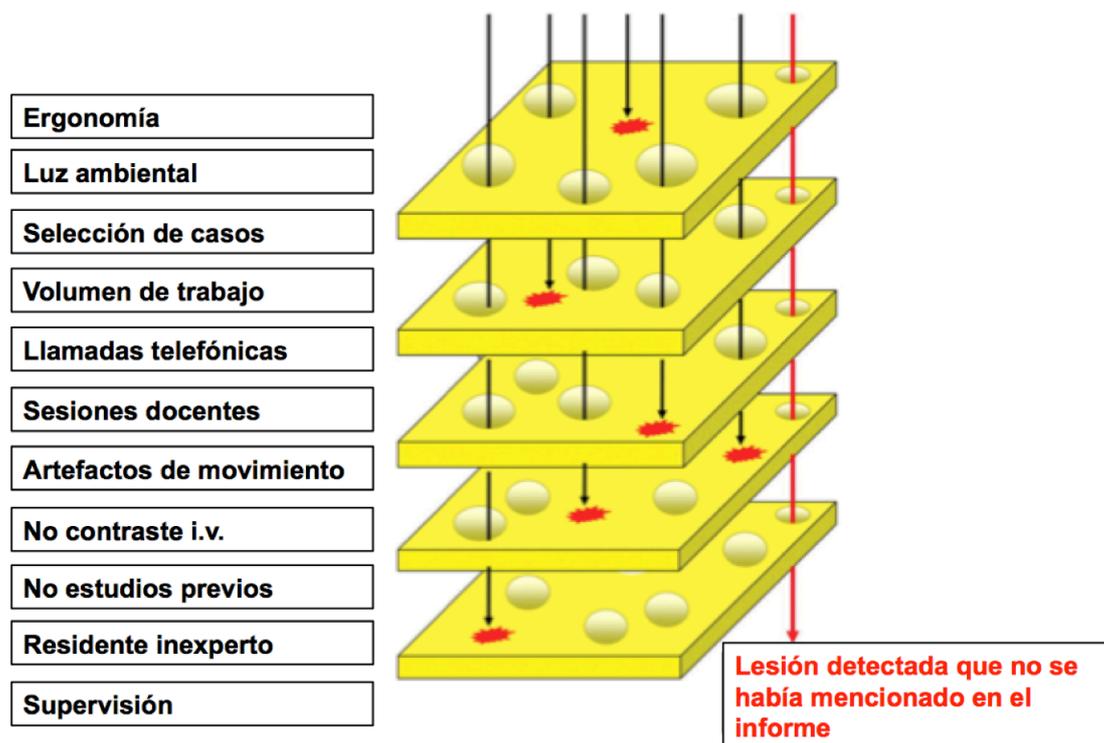


Figura 22. Filtros de seguridad insuficientes. Si las defensas o filtros de seguridad existentes son insuficientes, es probable que finalmente ocurra un error. En este ejemplo, a pesar de la demostración de que algunas defensas son eficaces para prevenir o impedir que ocurran errores, se presentará una ocasión en la que una serie de contribuyentes latentes se alinean y ocurra un error. Aquí, la ergonomía deficiente y la iluminación ambiental, frecuentes interrupciones por llamadas de teléfono, trabajar con un residente inexperto, y otros factores enumerados en el lado izquierdo se alinean para permitir que ocurra un error. Una lesión detectada en una sesión clínica no había sido mencionada en el informe, posiblemente con consecuencias importantes para el manejo del paciente. Dichos contribuyentes deben ser identificados a través del proceso de análisis de causa-raíz, y tomarse medidas para evitar que este error ocurra de nuevo. IV = intravenoso. Traducido del inglés al español y reimpresso con permiso de la *Radiological Society of North America (RSNA)*. La RSNA no se hace responsable de cualquier inexactitud o error que surja de la traducción del texto. Brook, O. R., O'Connell, A. M., Thornton, E., Eisenberg, R. L., Mendiratta-Lala, M., & Kruskal, J. B. (2010). Quality initiatives: Anatomy and pathophysiology of errors occurring in clinical radiology practice. *Radiographics : A Review Publication of the Radiological Society of North America, Inc*, 30(5), 1401-1410. doi:10.1148/rg.305105013 [doi].

La probabilidad de un accidente está en función del número total de factores latentes dentro del sistema. Todos los sistemas tienen un cierto número. Pero conforme más abundantes son, mayor es la probabilidad de que un determinado conjunto de factores latentes encuentren el desencadenante necesario para completar la secuencia del accidente (Reason, 1990). Conforme más complejo, interactivo, rígido y opaco sea el sistema, mayor será el número de fallos latentes (*Normal accidents: Living with high-risk technologies / Charles Perrow, 1984*). Sin embargo, es probable que los sistemas más simples requieran un menor número de factores latentes para producir un evento adverso, pues tienen menos defensas. Hay que considerar también que conforme más

alta es la posición de un individuo dentro de una organización, mayor es la capacidad que tiene de generar factores latentes (Reason, 1990).

Las barreras defensivas no son ni estáticas ni constantes y pueden o no ser independientes una de otra. Por lo tanto, mantener la seguridad depende de tener un número adecuado de barreras defensivas, haciéndolas tan infalibles como sea posible, y monitorizando constantemente aquellos fallos que puedan haber traspasado una o más barreras de defensa (Larson et al., 2015).

Hay que considerar que todas las organizaciones sanitarias deben asignar recursos limitados a dos objetivos diferentes: la productividad y la seguridad. A largo plazo, son claramente compatibles. Pero a corto plazo surgen conflictos de interés en los que los recursos asignados a la productividad podrían ir en detrimento de la seguridad y viceversa (Reason, 1990). Y en este contexto de recursos limitados los factores latentes tienen una propiedad muy interesante. A diferencia de los factores activos, que pueden ser difíciles de prever, los factores latentes sí pueden identificarse y tratarse antes de que ocurra un evento adverso (Reason, 1990) (Reason, 2000). La consecuencia es que los esfuerzos de los profesionales en seguridad podrían ser dirigidos con mayor rentabilidad hacia la identificación proactiva y la neutralización de los factores latentes asociados al error, más que a la prevención de los factores activos (Reason, 1990) al contrario de lo que con frecuencia se ha hecho en el pasado.

1.6. REVISIÓN POR PARES (*PEER REVIEW*).

En la descripción de la patogenia del error médico se realizó una analogía de los sistemas complejos con el funcionamiento de un organismo, y de los errores latentes con “patógenos residentes” en el mismo. Este planteamiento genera las preguntas de ¿cómo puede medirse la “morbilidad” de los sistemas de alto riesgo? ¿Tienen los sistemas indicadores generales, comparables al recuento leucocitario o un registro de presión arterial, de los que sea posible obtener una visión panorámica de su estado general de salud? (Reason, 1990).

La evaluación de la competencia profesional en Medicina ha sido durante largo tiempo una parte integral de la práctica médica profesional. Sirve para mantener altos estándares de calidad no solamente ante los profesionales individuales, los compañeros de trabajo y los pacientes, sino también ante el conjunto de la plantilla médica, la organización sanitaria, y las entidades políticas y administrativas (Kaewlai & Abujudeh, 2012).

En Radiología, un medio importante y aceptado por múltiples instituciones de observar y evaluar de forma continua, sistemática y crítica la competencia médica es la revisión de casos. Consiste en una revisión profesional de los casos registrados que pueden contener potenciales errores y que se realiza, generalmente por consenso, con un compañero de la misma o de otra disciplina, a través de procedimientos estructurados (Kaewlai & Abujudeh, 2012). Surge como respuesta al hecho de que la comprobación patológica o quirúrgica normalmente no está disponible.

Debe distinguirse entre los conceptos “revisión de casos”, “doble lectura” e “informe conjunto” (Strickland, 2015). La revisión de casos consiste en que un segundo radiólogo revisa un informe ya cerrado por otro radiólogo y decide el grado de discrepancia entre su propia opinión sobre los hallazgos y la opinión expresada previamente en el informe escrito del primer radiólogo. La doble lectura ocurre cuando dos radiólogos interpretan de forma independiente un mismo estudio y realizan sendos informes independientes. El grado de discrepancia entre estos dos informes independientes se evalúa por un tercer radiólogo y cualquier incertidumbre se resuelve por consenso. Esta doble lectura consume mucho tiempo y no es práctica para el uso clínico diario. Por último, el informe

conjunto implica que la interpretación del estudio y la redacción del informe se realiza al unísono por dos o más radiólogos (los cuales interactúan entre sí durante la interpretación) y, si hace falta, se alcanza una opinión de consenso; todos los radiólogos asumen la responsabilidad del contenido del informe. Esta última modalidad de informe no se considera adecuada por la revista *Radiology* (Levine, 2009).

Diariamente, en la práctica radiológica hospitalaria tiene lugar una revisión de casos con retroalimentación inmediata, en un entorno informal, a través de las consultas que se realizan a compañeros de trabajo sobre casos determinados. Es inmensamente útil en el desarrollo de las posturas de cada uno en relación con el caso y para construir una opinión diagnóstica más consistente que en última instancia acaba beneficiando al paciente (Strickland, 2015).

En un nivel mucho más estructurado y uniforme encontramos los programas RADPEER y e-RADPEER, iniciados por el Grupo de Trabajo en Seguridad del Paciente del Colegio Americano de Radiología (ACR). Funcionan bajo la premisa de que cuando un radiólogo interpreta un nuevo estudio de imagen conjuntamente con otro estudio previo de comparación de otro radiólogo, el segundo radiólogo puede evaluar la corrección del informe previo (Borgstede et al., 2004) (Halsted, 2004) (Jackson et al., 2009) (Abujudeh et al., 2010).

Un punto clave que consigue el sistema RADPEER del ACR es la estandarización de los tipos de error o de su gravedad (Jackson et al., 2009). Su clasificación para las discrepancias diagnósticas detectadas en las revisiones de casos consta de 4 niveles: 1 (de acuerdo con la interpretación), 2 (discrepancia de interpretación difícil o hallazgo sutil – omisión comprensible-), 3 (discrepancia de interpretación que generalmente debería hacerse), 4 (discrepancia de interpretación que casi siempre debería hacerse/equivocación de paciente). A su vez, las categorías 2, 3 y 4 se subclasifican en “a” (sin probable significación patológica) y “b” (con probable significación patológica). En general, las categorías 1 y 2 no requieren ninguna intervención adicional, pero las categorías 3 y 4 sí requieren una revisión interna por el comité de revisión para confirmar o modificar dicha puntuación.

Algunos ejemplos de las categorías expuestas anteriormente serían (Jackson et al., 2009):
Discrepancia 2 A:

- Rotura pequeña, sutil, de uno de los ligamentos colaterales de la rodilla.
- Adenopatía mesentérica de 7 mm en una TC abdominal.

Discrepancia 2 B:

- Sutil realce meníngeo en TC o RM.
- Sutil diastasis escafolunar.

Discrepancia 3 A:

- Hemangioma vertebral en una RM de columna.
- Masa tiroidea de 3 cm en una TC de tórax.

Discrepancia 3 B:

- Pequeño hematoma subdural en TC de cráneo.
- Rotura horizontal del menisco medial con un colgajo meniscal desplazado al receso coronario (Figura 23).

Discrepancia 4 A:

- Ganglio pélvico de 4 mm en un paciente que comienza una quimioterapia por linfoma.
- Colelitiasis calcificada de 2 cm visible en la TC de un paciente con dolor en fosa ilíaca izquierda y diverticulitis.

Discrepancia 4 B:

- Fractura desplazada del quinto metatarsiano.
- Neumotórax a tensión.



Figura 23. Imágenes coronales de RM de rodilla que muestran una rotura horizontal del menisco medial con un colgajo meniscal desplazado al receso coronario. El deslizamiento de una porción meniscal hacia el receso es causa frecuente de errores de omisión.

En la versión revisada de 2016 del sistema RADPEER se simplifican las categorías, eliminando la categoría 4 y fusionando las categorías 3 y 4 de la clasificación previa en una única categoría 3 (discrepancia de interpretación que generalmente debería hacerse); el motivo es que la distinción entre las categorías 3 y 4 era subjetiva, no reproducible, consumía mucho tiempo y carecía de efectos prácticos (Goldberg-Stein et al., 2017).

La revisión de casos es sólo una de las múltiples herramientas existentes para el control de calidad (Kaewlai & Abujudeh, 2012), pero su importancia es fundamental porque analiza, entre otros aspectos, la precisión diagnóstica, fuertemente asociada con el resultado clínico (Alpert & Hillman, 2004).

En una visión más amplia, la revisión de casos tiene 4 funciones principales: mejorar la formación académica de los individuos involucrados (“entrenar”), identificar a los individuos con un rendimiento inferior a un umbral determinado (“juzgar”), reorientar el tratamiento de los pacientes (“tratar”) y crear conocimiento extrapolable sobre la calidad y seguridad en Radiología (“investigar”) (Foteno & Nagy, 2012). Se potenciará un objetivo u otro según la metodología de la misma. Por ejemplo, si la selección de los casos a revisar es aleatoria se favorece una finalidad evaluadora de los radiólogos (“juzgar”), mientras que si la selección de los casos es en base a un especial interés académico de los mismos se potencia la finalidad educadora (“entrenar”). Según el tiempo transcurrido desde la realización del informe inicial y su revisión, si los estudios revisados son los previos de cada paciente cuando acude a revisión la finalidad es evaluadora (“juzgar”) mientras que si la revisión se hace poco después de la firma de los informes se posibilita el ajuste del tratamiento del paciente (“asistencia”). Según la riqueza del contenido de la revisión, si esta es una simple categoría numérica se cumple el objetivo de “juzgar” al profesional, mientras que si se realiza un informe independiente alternativo o se evalúa la concordancia con un estándar de referencia se permiten respectivamente la docencia y la investigación. Si la revisión acontece de un modo cerrado, puramente administrativo, se potencia el objetivo de juzgar al profesional concreto, siendo este el único que se beneficia del aprendizaje de la evaluación, mientras que si las discrepancias se debaten en reuniones conjuntas se facilita la formación académica del conjunto de los profesionales de la empresa (Foteno & Nagy, 2012).

Además de las funciones descritas anteriormente, la revisión de casos tiene un impacto directo en el rendimiento diagnóstico. Por una parte aumenta la probabilidad de detectar pequeñas lesiones (Robinson, 1997). Y por otra existe el efecto Hawthorne, también conocido como el efecto del observador, consistente en un fenómeno psicológico por el cual el simple conocimiento por parte de los trabajadores de que están siendo observados o evaluados modifica su comportamiento e incrementa su productividad. Este efecto se ha observado en Medicina y, específicamente, en Radiología (Abujudeh et al., 2017). Por ejemplo, podría explicar en parte el acortamiento de las exploraciones de fluoroscopia cuando el profesional está monitorizado (Vehmas, 1997) y también podría constituir uno de los efectos positivos de la revisión de casos.

Debe quedar claro no obstante que el propósito principal de la revisión de casos no es medir la tasa de error individual o del servicio y que no se usa para identificar a los peores profesionales. Este procedimiento sólo evalúa algunas partes de la complejidad de la práctica radiológica individual a través de una muestra relativamente reducida de casos; por lo tanto, puede que no sea representativa de las capacidades globales de los radiólogos (Cascade, 1990). Por otra parte, puede existir reticencia de los radiólogos a reseñar un error diagnóstico de un compañero por temor a provocar efectos punitivos. Es por tanto necesario eliminar la cultura de la culpabilización y crear una cultura más colaborativa (Berlanstein, 2018). Si la revisión de casos se realiza correctamente debe fomentar que los radiólogos sean transparentes en relación con el error y promover la discusión sobre las causas (Donald & Barnard, 2012).

1.7. EL ERROR EN MEDICINA, NO ES LO QUE PARECE.

“ El personal médico y de enfermería debe aceptar la idea de que el error es un compañero inevitable de la condición humana, incluso en profesionales meticulosos de alto nivel. Los errores deben considerarse una evidencia de fallos de los sistemas, no de fallos de los individuos. Hasta que eso ocurra, no es probable que exista ningún progreso significativo en la reducción de los errores médicos”.

Lucian Leape (Leape, 1994)

Una concepción equivocada y muy extendida es que los errores son intrínsecamente malos. Sin embargo, pueden existir aspectos positivos en los errores, como por ejemplo el descubrimiento que Alexander Fleming hizo de la penicilina en una placa de Petri que se había desechado por estar contaminada por un hongo. En este sentido, la detección y análisis de los errores pueden posibilitar actuaciones que mejoren el rendimiento futuro, minimizando la incidencia o el impacto de los errores subsiguientes, y previniendo la aparición de errores más graves (Brook et al., 2010).

Garland (Garland, 1949) no se centró en evaluar los métodos para reducir el error radiológico más allá de escribir que “los errores son corregibles siendo cuidadosos, con entrenamiento y experiencia, aumento de las interconsultas entre compañeros, e intentos continuos de descubrir y corregir los factores implicados”. Quedaba fuera del alcance de este análisis el discutir en detalle los métodos para reducir el error radiológico, pero debemos ser conscientes de que si sólo medimos las tasas de error diagnóstico y no las variables explicativas asociadas al mismo no podemos tener la certeza de que el rendimiento diagnóstico vaya a mejorar.

Con frecuencia se cree que los errores ocurren de forma aleatoria. Aunque esto pueda parecer cierto a primera vista, los errores con frecuencia ocurren según determinados patrones y reflejan prácticas subóptimas de largo tiempo de evolución (normalmente reconocidas de forma retrospectiva, pero no en el momento en el que ocurren), conjuntamente con fallos latentes del sistema (Halsted et al., 2004) (Kim & Mansfield, 2014).

Durante mucho tiempo se estuvo adoptando un abordaje dirigido hacia el omnipresente “factor humano”, que responsabilizaba a los individuos de cometer los errores. Se consideraba a los errores, por tanto, el resultado de descuidos, falta de atención, falta de motivación y negligencia (Reason, 1990) (Leape, 1994) (Shojania, Wald, & Gross, 2002). Pero a pesar del hecho de que los accidentes son en gran medida causados por fallos humanos, los abordajes dirigidos a los individuos casi nunca resolvían el problema (Cooper J B. et al., 1978) (Reason, 1995).

Una de las conclusiones principales del informe “Errar es humano, construyendo un sistema de salud más seguro” (2000) fue que el 90% de los errores médicos se derivaban de problemas del sistema más que de factores individuales (Kohn et al., 2000). Los fallos humanos suponen tan sólo la punta del iceberg del conjunto de factores causales (Reason, 1990). Esto implica a su vez que, al contrario de la cultura dominante en la profesión médica, la gran mayoría de los errores médicos no se deben a individuos ignorantes, incompetentes, descuidados, o negligentes sino que ocurren por fallos prevenibles en los sistemas, procesos y condiciones en los que operan profesionales competentes y cuidadosos con su trabajo (Irish, 2008) (Kohn et al., 2000) (van Beuzekom et al., 2010). Por tanto, y aunque culpar a los individuos es emocionalmente más satisfactorio que abordar las instituciones en las que se cometen los errores (Reason, 1990) (Reason, 2000), hay que identificar aquellos factores latentes “inactivos” que permiten que los errores activos lleguen a manifestarse.

Los errores son además difíciles de medir, no sólo debido a las variadas definiciones de error sino también porque la mayoría de los incidentes de error no son actos aislados sino una cadena de sucesos (Robinson, 1997) . Los errores iniciales que producen a su vez un error distal más evidente pueden ser infraestimados si sólo se mide el error distal. Por ejemplo, omitir el diagnóstico de una fractura puede considerarse un error de percepción, pero dicho error radiológico puede haber ocurrido debido a insuficiente información clínica en la solicitud, y esta última a su vez debido a fatiga del médico solicitante por turnos de trabajo excesivamente prolongados. Los administradores y los gestores que ignoren esta complejidad pueden producir estadísticas sesgadas y propagar nociones imprecisas sobre la anatomía, las causas y las consecuencias de los errores. Esta imprecisión afecta a los pacientes, clínicos y políticos, porque focaliza la atención y la culpabilidad en lugares erróneos alejándose de los problemas fundamentales y las

modificaciones del sistema que pudieran mejorar de forma más efectiva la salud del paciente (Moray, 1994) (Nolan, 2000).

2. HIPÓTESIS Y OBJETIVOS.

2.1. HIPÓTESIS.

- El error diagnóstico en Radiología tiene una etiología multifactorial.
- Los factores latentes contribuyen significativamente al error diagnóstico.
- Dentro de los factores latentes no todos muestran el mismo grado de asociación con el error diagnóstico.

Hasta la fecha no hay estudios que cuantifiquen con un análisis multivariante esta asociación de múltiples variables explicativas con el error diagnóstico. Planteamos que afrontar los factores latentes del sistema mejorará las condiciones de seguridad de la organización al reducir los problemas derivados de fallos de seguridad antes de que aparezcan (Hudson, 2009).

2.2. OBJETIVOS.

- Determinar la tasa de discrepancia entre la primera lectura y la revisión del caso y comparar los resultados con las tasas de error diagnóstico reflejadas en la bibliografía médica.
- Estimar la significación clínica de los errores detectados en base a la gravedad de los hallazgos de imagen.
- Identificar los factores latentes asociados a una mayor incidencia de error diagnóstico. Descubrir puntos débiles en el trabajo radiológico que puedan reforzarse con medidas específicas de formación o de control de calidad.

3. MATERIAL Y MÉTODOS.

Para realizar el estudio, se solicitó y fue concedida la correspondiente autorización por parte del Comité Ético de Investigación en Humanos de la Comisión de Ética en Investigación Experimental de la Universitat de València.

3.1. DESCRIPCIÓN DEL ENTORNO. CARACTERÍSTICAS DEL CENTRO.

IBERORAD SL. es una empresa de telerradiología que proporciona asistencia 24 horas al día 365 días al año a 37 centros en España.

El equipo humano de la empresa se compone de un director médico, un responsable de Tecnologías de la Información y la Comunicación (TIC), dos técnicos superiores especialistas en Radiodiagnóstico (TER), dos programadores de aplicaciones informáticas especialistas en imagen médica y homologados en Health Level Seven (HL7), tres coordinadores, veintinueve radiólogos con sistema de formación MIR (Médico Interno Residente) (doce radiólogos generales, ocho radiólogos dedicados preferentemente a Radiología Musculoesquelética, cinco dedicados preferentemente a Neurorradiología y cuatro radiólogos dedicados preferentemente a Radiología de Tórax-Abdomen-Pelvis) y un contable.

Todos los radiólogos de la empresa tienen el español como lengua materna.

La actividad de la empresa incluye la realización de informes radiológicos de radiología convencional, TC y RM (programados y urgentes), control de calidad (revisiones de los informes radiológicos que realizan otros radiólogos distintos de la misma empresa), instalación de PACS, integraciones informáticas y establecimiento de sistemas de recogida online de informes e imágenes.

En el momento del estudio la empresa realiza cada año aproximadamente 30.000 informes de estudios programados y 11.000 informes de estudios urgentes.

3.2. DISEÑO DEL ESTUDIO.

Se ha realizado un estudio observacional retrospectivo en el cual se analizan las discrepancias diagnósticas en estudios de telerradiología de pacientes adultos de 37 centros sanitarios detectadas tras una revisión de informes radiológicos de la empresa IBERORAD SL durante un periodo de dos años (entre el 04/06/2014 y el 03/06/2016). Estas revisiones de casos forman parte de la actividad ordinaria de control de calidad de la empresa, las realiza siempre un radiólogo distinto al autor del informe inicial, y se llevan a cabo en el 100% de los estudios urgentes (21.882 estudios) y en torno a un 10% de los estudios programados seleccionados de forma aleatoria (6.378 estudios), lo que supone un total de estudios revisados de 28.260. Los radiólogos revisores son parte de la propia plantilla de radiólogos de la empresa y realizan las revisiones de casos como actividad complementaria a la realización de informes radiológicos ordinarios, siendo estos últimos a su vez revisados por otros revisores distintos al autor inicial. No muestran por tanto como conjunto un perfil profesional diferenciado del grupo de radiólogos cuyos informes revisan. Tampoco existe subespecialización en la revisión de casos, por ejemplo, radiólogos generales pueden realizar revisión de casos de los diferentes aparatos y sistemas, y radiólogos con formación preferente en un área de conocimiento pueden ocasionalmente realizar revisiones de estudios radiológicos fuera de su campo de especialización.

- Los criterios de inclusión son las revisiones de estudios de TC o RM realizados a pacientes mayores de 18 años e informados por radiólogos de IBERORAD S.L., independientemente de si el contenido de las mismas es concordante (acuerdo diagnóstico) o discordante (discrepancia diagnóstica) con el informe radiológico inicial.
- Como criterios de exclusión se consideran los estudios radiológicos de radiografía convencional, los de mamografía, los estudios de TC o RM realizados a pacientes menores de 18 años, y aquellos en los cuales no se obtenga información de los factores latentes que se valoran en el presente estudio.
- Se ha llevado a cabo un análisis descriptivo de los radiólogos que valoraron inicialmente los estudios, incluyendo tanto características generales de los mismos como de las estaciones de trabajo utilizadas:

- Edad del radiólogo.
- Género del radiólogo.
- Años de experiencia del radiólogo tras la obtención de la titulación.
- Si el radiólogo informa estudios a) de varios aparatos/sistemas o b) de un solo aparato sistema. El carácter general o subespecialista del radiólogo (Neurorradiología, Radiología Musculoesquelética y Radiología de Tórax-Abdomen). Se considerará radiólogo especialista en un aparato-sistema a aquel con actividad profesional dedicada preferentemente a ese aparato-sistema.
- La especialidad del estudio de imagen analizado (Neurorradiología, Radiología Musculoesquelética, Columna o Radiología de Tórax-Abdomen-Pelvis). Aunque los estudios de imagen de columna no constituyen en sí mismos una especialidad radiológica, se han segregado del resto a efectos de análisis estadístico de la concordancia con la formación de los radiólogos, pues dependiendo de los centros sanitarios se consideran que pertenecen indistintamente a los ámbitos de la Neurorradiología y de la Radiología Musculoesquelética.
- Modalidad de imagen del estudio analizado (TC, RM).
- Naturaleza urgente/ordinaria del estudio.
- Estación de trabajo del radiólogo: número de monitores auxiliares y tamaño en pulgadas de los mismos, tipo de visor radiológico utilizado (Clear Canvas® (PC)/ OsiriX® (Mac®), y si se utiliza o no sistema de reconocimiento de voz (SpeechExec Pro PSE4400/4550 de Philips®/ Nuance Speech Magic®).
- Hora local real (local) de la primera lectura realizada por el radiólogo, pues los radiólogos trabajan desde España y Latinoamérica con husos horarios diferentes.
- División del estudio en 3 grupos horarios:
 - Horario laboral normal (8:00-17:59 h) en días laborables.
 - Otro horario: (18:00-7:59 h) en días laborables.
 - Cualquier hora de fin de semana o festivo.
- Calidad técnica de las imágenes: correcta, deficiente. Se considera calidad técnica deficiente aquella que debido a los parámetros del estudio (por ejemplo, artefactos, grosor de corte, número de secuencias, insuficiente campo de visión...) es susceptible de favorecer la aparición de un error diagnóstico.
- Disponibilidad o no de estudios de imagen previos.
- Complejidad del estudio (en áreas anatómicas y reconstrucciones de imagen): normal o alta. Se consideran estudios complejos los estudios de hipófisis,

peñascos, cuello, aorta, angiografía (por TC o RM) de cuello-aorta-extremidades, RM de los dedos, TC de cuerpo completo (pacientes politraumatizados).

- Adecuación de la información clínica de la solicitud del estudio (adecuada, deficiente, ausente). Se entiende por información clínica deficiente la que es muy genérica u omite datos esenciales (por ejemplo, “gonalgia” en una RM de rodilla sin indicar en qué compartimento de la rodilla se sospecha la patología, o “contusión” en una TC de cuerpo completo sin indicar la localización de la misma).

También se registran las discrepancias diagnósticas según la escala RADPEER del ACR vigente antes de la nueva clasificación de 2016 que unifica las categorías 3 y 4 (Jackson et al., 2009):

- 1: de acuerdo con la interpretación.
 - 2: discrepancia de interpretación difícil o hallazgo sutil – omisión comprensible.
 - 3: discrepancia de interpretación que generalmente debería hacerse.
 - 4: discrepancia de interpretación que casi siempre debería hacerse/equivocación de paciente.
- Cada una de las discrepancias 2 a 4 se subclasifican en “a” (poco probable que tenga significación clínica) y “b” (probable que tenga significación clínica). La significación clínica de los errores se determina de forma subjetiva en base a la gravedad de los hallazgos de imagen (sin seguimiento evolutivo del paciente ni verificación quirúrgica o patológica).
 - El radiólogo revisor puede añadir texto libre en los comentarios de la revisión.
 - Los errores detectados se también se clasifican en 5 grupos:
 1. Infralectura (omisión de un hallazgo).
 - 1.1. Localización: el hallazgo se localiza fuera del área primaria de interés
 - 1.2. Búsqueda satisfactoria: otro hallazgo del estudio causa finalización prematura de la búsqueda diagnóstica.
 - 1.3. Descripción no suficientemente exhaustiva (en profundidad) del hallazgo radiológico.
 2. Sobrelectura.
 - 2.1. Error cognitivo: hallazgo patológico detectado pero interpretado erróneamente.
 - 2.2. El diagnóstico diferencial no incluye patologías relevantes en relación a los hallazgos.

- 2.3. Variante anatómica interpretada como patología.
3. Comunicación defectuosa (en el informe o en la comunicación directa con el clínico).
 - 3.1. Paciente o estudio equivocado.
 - 3.2. Error tipográfico.
 - 3.3. Error de lateralidad.
 - 3.4. Error de nivel.
 - 3.5. Error de género.
 - 3.6. Recomendaciones insuficientes o equivocadas (completar estudio o seguimiento).
 - 3.7. Fallo en indicar que la técnica ha sido deficiente.
 - 3.8. Fallo en la comunicación directa con el clínico y/o en el registro de esta en el informe.
4. Información clínica insuficiente.
5. Técnica defectuosa.
 - 5.1. Parámetros incorrectos o insuficientes (contraste, FOV, fases, secuencias, planos...).
 - 5.2. Artefactos de la imagen (por ejemplo, artefactos de movimiento, artefactos metálicos en pacientes portadores de prótesis...etc).

· La base de datos es confidencial y sólo el investigador principal, los codirectores del trabajo y el estadístico encargado de analizar los datos tienen acceso a la misma.

3.3. ANÁLISIS ESTADÍSTICO.

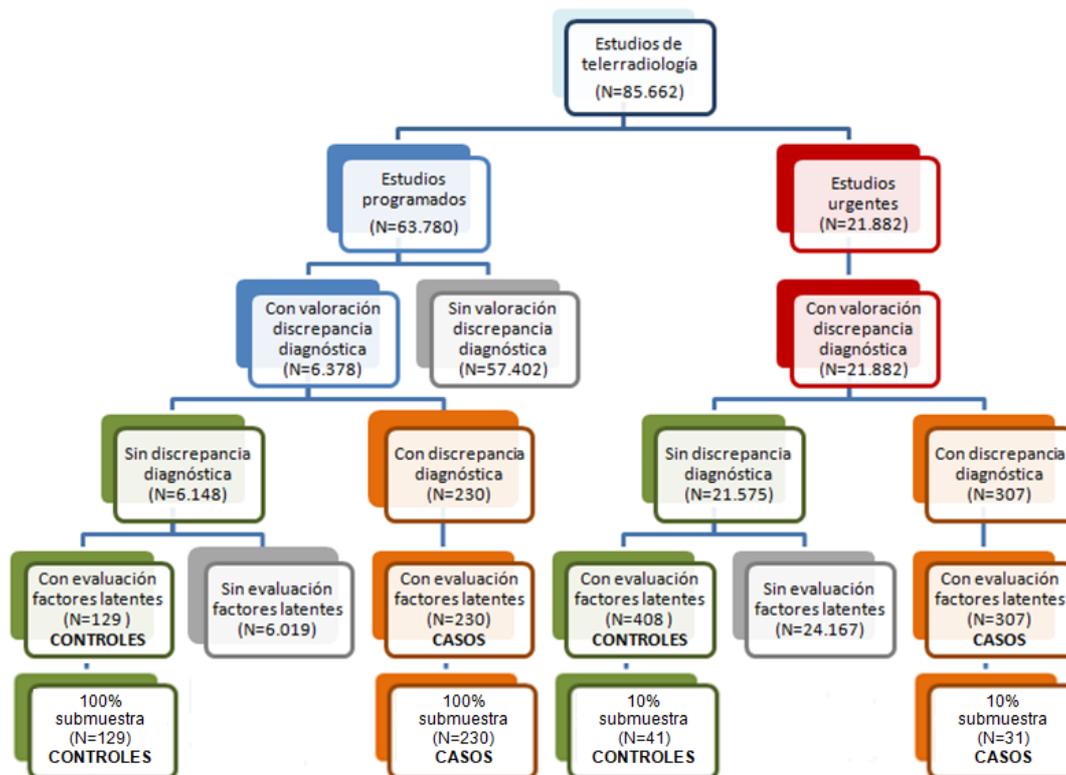


Figura 24. Selección de estudios para el análisis de factores latentes en la discrepancia diagnóstica.

La población objeto de análisis son todos los estudios de telerradiología realizados en la empresa durante un periodo de dos años. En concreto, está formada por 21.882 estudios urgentes y aproximadamente 63.780 estudios programados, es decir, un total de 85.662 estudios.

No disponemos de la valoración de la discrepancia diagnóstica sobre todos los elementos de la población objetivo, sino que se ha evaluado esta característica sobre todos los estudios urgentes (21.882 estudios) y sobre una muestra aleatoria del 10% de estudios programados (6.378 estudios). Por tanto, la población accesible sería un subconjunto de la población objetivo formada por aquellos 28.260 elementos sobre los que se ha observado la característica de interés (presencia o ausencia de discrepancia diagnóstica), y que tendrá una mayor proporción de elementos urgentes que la población total.

De los 21.882 estudios urgentes, se ha observado discrepancia diagnóstica en 307 de ellos, mientras que de los 6.378 estudios programados se ha observado discrepancia

diagnóstica en 230. Por tanto, existen en total un conjunto de 537 estudios con discrepancia diagnóstica y 27.723 estudios que no presentan discrepancia diagnóstica.

Durante el periodo estudiado se han incluido, como se ha indicado, 28.260 revisiones de estudios de telerradiología realizadas partiendo de 85.662 estudios totales con un 24,6% de TC y un 75,4% de RM. La presencia de discrepancias diagnósticas se estima en un 3,0% para el total de estudios de telerradiología durante el periodo de análisis, con un número estimado de $230 \times 10 + 307 = 2.607$ estudios con discrepancia diagnóstica de los $6.378 \times 10 + 21.882 = 85.662$ estudios totales estimados.

Teniendo en cuenta que únicamente se dispone de un 3,0% estimado de estudios con segunda valoración discrepante diagnósticamente, se ha planteado un estudio de casos–controles formado por los 537 estudios radiológicos con alguna discrepancia diagnóstica (casos) y una muestra aleatoria de 537 estudios de todos los que no presentan ninguna discrepancia diagnóstica (controles). Estos 1.074 estudios de telerradiología han sido analizados de forma individualizada, añadiendo todos los factores de interés para los objetivos de la presente tesis. No se ha podido trabajar con toda la muestra inicial (todos los estudios urgentes y un 10% de los estudios programados), ya que algunas de las variables estudiadas no aparecen en la base de datos y requieren de un análisis individualizado de los estudios. En concreto, los factores latentes de interés a añadir individualizadamente en cada estudio son la calidad técnica de las imágenes (adecuada o deficiente), la adecuación de la información clínica de la solicitud del estudio (adecuada o deficiente/ausente), la disponibilidad de estudios de imagen previos (sí o no), la complejidad del estudio (sí o no) y la especialidad del estudio (Neurorradiología, Radiología Musculoesquelética, Columna o Radiología de Tórax-Abdomen-Pelvis). Con este objetivo se selecciona una muestra aleatoria de todos los estudios con segunda lectura sobre los que se cumplimenta esta información adicional.

En la Figura 24 se aprecia el esquema de selección de los elementos incluidos en este análisis.

No se ha seleccionado una muestra aleatoria directamente sobre toda la población de estudios con segunda valoración, ya que ello nos llevaría a una presencia muy reducida de casos (estudios con discrepancia diagnóstica), lo que daría lugar a ajustes con escasa

potencia y sería difícil evaluar la asociación de los diferentes factores pronóstico de interés con la discrepancia diagnóstica.

Sobre la población formada por todos los estudios de telerradiología de carácter programado existe una muestra de casos ($n_1=230$) y una muestra de controles ($n_2=129$) que proceden y son representativos de todos los estudios programados. Hay que tener en cuenta que este diseño no permite obtener estimaciones de las tasas de discrepancias diagnósticas asociadas a cada uno de los factores de interés, ya que el número de estudios sin discrepancia diagnóstica ha sido fijado arbitrariamente. No obstante, puede medirse el grado de asociación entre cada uno de los factores de interés y la presencia o ausencia de discrepancia diagnóstica obteniendo la Razón de Probabilidades (Odds Ratio - OR) asociado a la presencia de un factor respecto a la no presencia. Por tanto, el análisis de los casos y controles de la población de estudios programados nos permite estimar los OR asociados a cada factor latente, e inferir aquellos que están asociados a la discrepancia diagnóstica desde el punto de vista univariante y multivariante.

Del mismo modo, sobre la población formada por todos los estudios de telerradiología de carácter urgente se obtiene una muestra de casos ($n_1=307$) y una muestra de controles ($n_2=408$) que proceden y son representativos de todos los estudios urgentes. Por tanto, pueden estimarse los OR asociados a cada factor latente, e inferir aquellos que están asociados a la discrepancia diagnóstica desde el punto de vista univariante y multivariante.

Algunos de los factores tienen una presencia muy reducida en los estudios de telerradiología considerados, por lo que la estimación de los OR asociados es muy poco precisa o imposible. Sería necesario incrementar el número de sujetos incluidos en el estudio de casos-contróles para poder determinar el efecto de estos factores de forma satisfactoria.

Hay que tener en cuenta también que algunos de los factores de interés, registrados por defecto en la base de datos de la empresa sin necesidad de introducirlos manualmente caso por caso, han sido observados para todos los estudios de la población accesible y no únicamente para la muestra de casos-contróles, por lo que podrían haberse estimado los OR asociados de forma más precisa. No obstante, como para el análisis multivariante únicamente pueden incluirse aquellos estudios en los que se han registrado todos los

factores de interés, se ha optado por analizar sobre esta muestra de estudios el efecto de estos factores.

Por otra parte, debemos considerar que tanto los casos como los controles son una muestra aleatoria de la población de estudios programados con y sin discrepancia diagnóstica. Para este análisis se ha llevado a cabo, por una parte para el estrato de estudios programados, y por otra para el estrato de estudios urgentes, un ajuste univariante para cada factor de interés mediante regresión logística que permite modelizar la probabilidad de discrepancia diagnóstica en función de los valores del factor analizado. A continuación se ha construido un modelo de regresión logística multivariante utilizando un procedimiento de construcción por pasos hacia atrás donde se modeliza el efecto conjunto de los factores latentes significativos sobre la probabilidad de una discrepancia diagnóstica en un estudio de telerradiología.

También se ha realizado un estudio global incluyendo todos los casos y controles seleccionados. Hay que tener en cuenta que la muestra de casos y controles es representativa de la población accesible (formada por todos los estudios con una segunda valoración), por lo que las conclusiones que se extraigan de este análisis en particular serán válidas para esta población. Para que sean también válidas para la población total (formada por todos los estudios, con independencia de si tienen o no segunda valoración), el factor que determina la realización de la segunda valoración (el carácter del estudio: programado o urgente) no tendrá que estar relacionado con la relación entre el factor de análisis y la presencia o ausencia de discrepancia diagnóstica. Deberemos por tanto ser muy cautos con las estimaciones de los OR asociados a cada factor en este análisis global debido a la presencia de un sesgo de selección, al realizarse la segunda valoración con una mayor probabilidad en los estudios urgentes (100% de probabilidad) que en los estudios programados (únicamente con un 10% de probabilidad). En cualquier caso se ha llevado a cabo un estudio multivariante de los factores significativos desde el punto de vista univariante incluyendo el carácter del estudio (urgente o programado) para tratar de controlar el efecto de este último sobre cada factor.

Como las revisiones de casos de la empresa se realizan sobre todos los estudios urgentes y únicamente un 10% de los estudios programados, la muestra obtenida presenta un sesgo de verificación al contar con más estudios urgentes de los que constan en la

población de referencia. Para tratar de evitar este problema se selecciona una submuestra en la población de estudios urgentes del 10% de elementos, de forma que las muestras de casos y controles sean representativas de todos los casos y controles de la población.

A su vez, dentro de la submuestra del 10% de estudios urgentes, se realiza una estratificación por la presencia de discrepancia diagnóstica, obteniendo un subconjunto de 41 estudios urgentes sin discrepancia diagnóstica y 31 estudios urgentes con discrepancia diagnóstica. De esta forma, el conjunto de 261 estudios con discrepancia diagnóstica (230 programados y 31 urgentes) son una muestra aleatoria de los 2.607 estudios estimados con discrepancia diagnóstica (2.300 programados y 307 urgentes) con la misma distribución según carácter del estudio (programado o urgente). Del mismo modo, el conjunto de 170 estudios sin discrepancia diagnóstica (129 programados y 41 urgentes) son una muestra aleatoria de los 83.055 estudios estimados sin discrepancia diagnóstica (61.480 programados y 21.575 urgentes) con la misma distribución según carácter del estudio (programado o urgente). Por tanto, existe una muestra de 431 casos – controles representativa de los casos y controles del total de estudios. No obstante, hay que tener en cuenta que de esta forma se añade una fuente de variabilidad adicional en el análisis (si el estudio es programado o urgente) sin un incremento significativo del tamaño muestral, lo que llevará a estimaciones de peor calidad que el análisis estratificado individual de cada uno de los estratos.

Aunque en todos los análisis el objetivo principal es principalmente explicativo (determinar el efecto de cada uno de los factores sobre la presencia de discrepancia diagnóstica), se ha evaluado también la capacidad predictiva de los modelos obtenidos estimando la curva ROC (*Receiver Operating Characteristic*, o Característica Operativa del Receptor) asociada a las probabilidades estimadas por los modelos.

Para las variables cualitativas se han utilizado tablas de frecuencias, mientras que para las variables cuantitativas se han utilizado estadísticos de resumen, tanto de localización (medias como estimadores no robustos y medianas como estimadores robustos) como de dispersión (desviaciones típicas (DT) como estimadores no robustos y rangos intercuartílicos (RI) como estimadores robustos).

Para las variables cualitativas se han construido las tablas de contingencia asociadas, utilizando el estadístico Chi-cuadrado de Pearson para aquellas variables con más de 2 categorías y el estadístico exacto de Fischer para aquellas variables con 2 categorías. Para las variables cuantitativas, se han utilizado pruebas no paramétricas U de Mann-Whitney de comparación de dos muestras independientes, ya que son más conservadoras en situaciones donde no es posible asumir la normalidad de la variable analizada.

El análisis estadístico se ha llevado a cabo utilizando los programas IBM© SPSS© Statistics 22 y R 3.4.2 para Microsoft© Windows. En todos los contrastes se han considerado pruebas bilaterales y un nivel de significación del 5%.

4. RESULTADOS.

4.1. Características de los radiólogos que valoraron inicialmente los estudios.

En la Tabla 4 se describen las características de los 29 radiólogos que evaluaron inicialmente estos estudios. La mayoría son radiólogos generales (41,4%), seguidos de los dedicados preferentemente a Radiología Musculoesquelética (MSK) (27,6%), a Neurorradiología (Neuro) (17,2%) y a Radiología de Tórax-Abdomen-Pelvis (Body) (13,8%). Respecto al género de los radiólogos, se observa que la mayoría son varones (69,0%). Estos radiólogos presentan una media de edad de 44,8 años, con una mediana de 45 años y un rango entre 31 y 64 años de edad. La mayoría ha realizado sus valoraciones desde España (72,4%), seguido de Colombia (13,8%), Chile (6,9%) y Venezuela (6,9%). Respecto a los años de experiencia del radiólogo en el momento de la valoración del estudio de telerradiología, se observa en la Tabla 4 que existe una mediana de 9 años de experiencia, con un rango intercuartílico de 2 a 19 años de experiencia.

	Total		Total
Total	29 (100%)	Total	29 (100%)
Especialidad radiólogo		Género del radiólogo	
Neuro	5 (17,2%)	Masculino	20 (69,0%)
Body	4 (13,8%)	Femenino	9 (31,0%)
MSK	8 (27,6%)	Edad del radiólogo	
General	12 (41,4%)	Media (DT)	44,8 (9,19)
País		Mediana (RI)	45 (37-51)
España	21 (72,4%)	Años de experiencia	
Colombia	4 (13,8%)	Media (DT)	11,5 (9,55)
Chile	2 (6,9%)	Mediana (RI)	9 (2-19)
Venezuela	2 (6,9%)		

Tabla 4. Características generales de los radiólogos que valoraron inicialmente los estudios.

En la Tabla 5 se describen las principales características de las estaciones de trabajo de los radiólogos que valoraron inicialmente estos estudios. Se aprecia que 22 radiólogos (75,9%) tienen un único monitor como estación de trabajo (18 de 21", 1 de 24" y 3 de 27"), mientras que 6 radiólogos disponen de dos monitores (3 con los dos de 21", 1 con uno de 21" y otro de 27", y 2 con dos de 27") y 1 radiólogo dispone de 3 monitores (los tres de 21"). Se observa también que la mayoría no usa sistemas de reconocimiento de voz (89,7%). Respecto al tipo de visor radiológico, un 34,5% usa Clear Canvas® (que opera sobre ordenador personal IBM -IBM PC-) y un 65,5% usa OsiriX® (que opera sobre Mac®).

	Total		Total
Total	29 (100%)	Total	29 (100%)
Número de monitores		Sistema operativo	
Monitor único	22 (75,9%)	Mac	19 (65,5%)
Dos monitores	6 (20,7%)	PC	10 (34,5%)
Tres monitores	1 (3,4%)	Pulgadas máximas monitor	
Sistema de reconocimiento de voz		21"	22 (75,9%)
No	26 (89,7%)	24"	1 (3,4%)
Sí	3 (10,3%)	27"	6 (20,7%)
Tipo de visor radiológico			
Clear Canvas®	10 (34,5%)		
OsiriX®	19 (65,5%)		

Tabla 5. Características generales de las estaciones de trabajo.

4.2. Características de los estudios de telerradiología incluidos en el estudio.

En la Figura 25 se ilustra la evolución mensual del número de estudios incluidos en el análisis según su tipología. Puede apreciarse que, por término medio, existen 1.178 estudios de telerradiología mensuales incluidos en el estudio (todos los urgentes y aproximadamente un 10% de los programados).

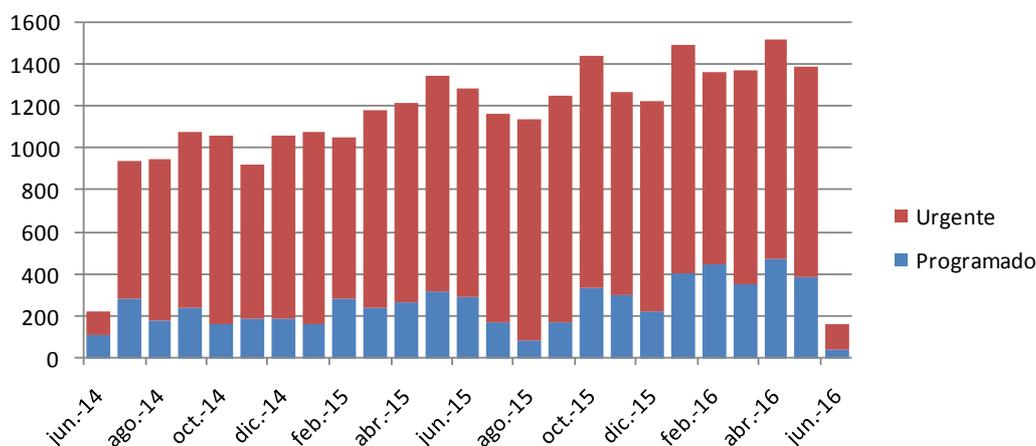


Figura 25. Estudios incluidos en el análisis según su tipología.

Una de las características recogidas en la base de datos para cada uno de los informes de telerradiología es la hora en que se ha producido la primera valoración. En las valoraciones realizadas en España hemos considerado la hora que aparece en el sistema, mientras que en las realizadas en Colombia y Venezuela hemos restado 6 horas, y en las realizadas en Chile hemos restado 5 horas. En la Figura 26 se representa el número de estudios incluidos en el análisis según la hora y país de valoración, evidenciándose que

las 18h es la hora en que se han valorado más estudios de telerradiología, independientemente del lugar de trabajo del radiólogo.

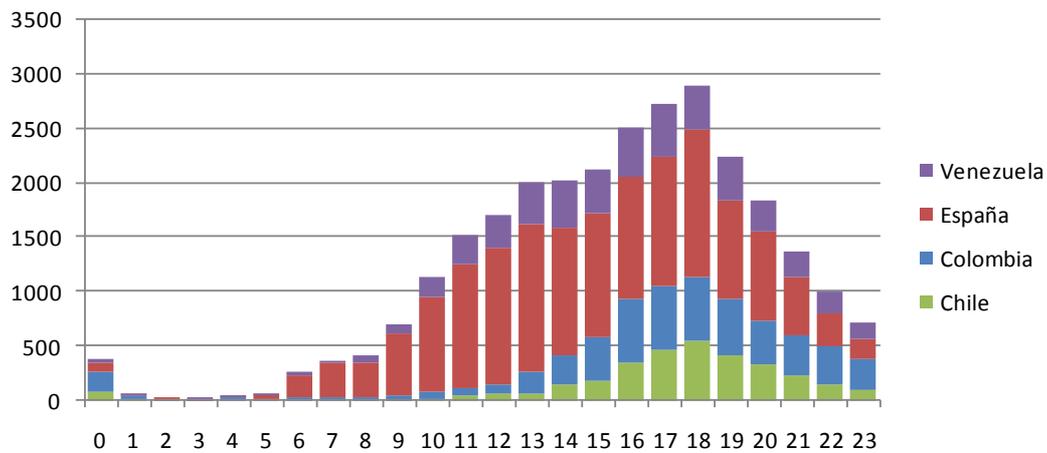


Figura 26. Estudios incluidos en el análisis según hora y país de valoración inicial.

Además de la hora en que se ha realizado la valoración inicial, se ha recogido también el tipo de día. De los 28.260 estudios con segunda valoración registrados durante el periodo, 19.203 (68,0%) fueron valorados inicialmente un día laborable, mientras que 9.057 (32,0%) fueron valorados inicialmente un día festivo. En la Figura 27 se ilustra la distribución horaria para ambos tipos de días.

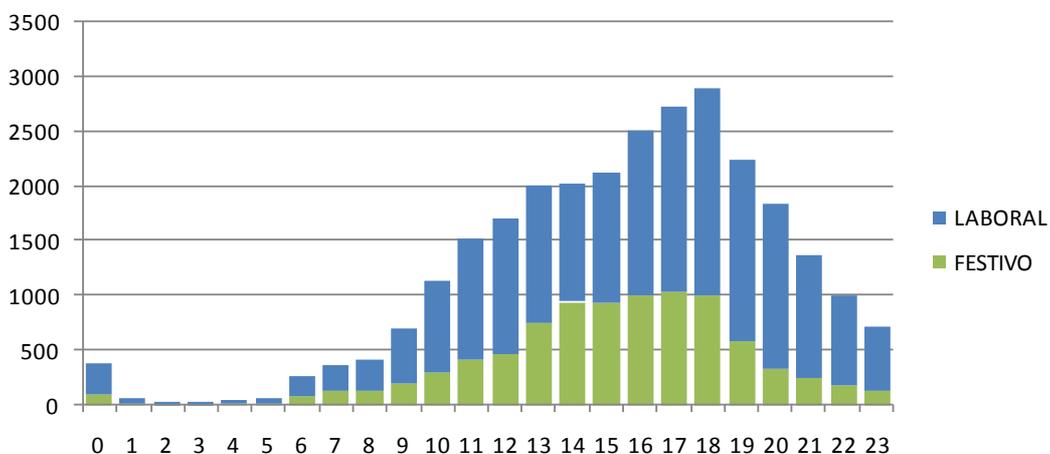


Figura 27. Estudios incluidos en el análisis según hora y tipo de día de la valoración inicial.

A partir de la información del tipo de día y hora de la valoración inicial se construirá una variable de horario con tres categorías: laborable de 8:00h a 17:59h, laborable de 18:00h a 7:59h y festivo. De los 28.260 estudios considerados en el presente análisis, 10.707 (37,9%) fueron evaluados inicialmente en día y horario laboral, 8.496 (30,1%) fueron evaluados inicialmente en día laborable pero horario no laboral, y 9.057 (32,0%) fueron evaluados inicialmente en día festivo.

En la Tabla 6 pueden apreciarse las principales características observadas sobre este conjunto de estudios de telerradiología según la tipología del mismo (si es de carácter programado o urgente).

	Urgente	Programado	p-valor
Total	21.882 (100%)	6.378 (100%)	
Características del estudio			
Modalidad del estudio			<0,001
Tomografía Computarizada	18.697 (86,9%)	191 (3,1%)	
Resonancia Magnética	2.826 (13,1%)	6.047 (96,9%)	
Horario de la valoración			
Laborable 8:00-17:59h	8.428 (38,5%)	2.279 (35,7%)	
Laborable 18:00-7:59 h.	5.420 (24,8%)	3.076 (48,2%)	
Festivo	8.034 (36,7%)	1.023 (16,0%)	
Características del radiólogo			
Género del radiólogo			<0,001
Masculino	14.340 (65,5%)	3.976 (62,3%)	
Femenino	7.542 (34,5%)	2.402 (37,7%)	
Edad del radiólogo			0,172
Media (DT)	45,3 (10,4)	44,6 (7,6)	
Mediana (RI)	43 (34-56)	46 (38-51)	
Años de experiencia			
Media (DT)	10,3 (10,5)	12,2 (8,9)	<0,001
Mediana (RI)	4 (1-19)	10 (2-20)	
Especialidad radiólogo			
General	12.742 (58,2%)	2.300 (36,1%)	
Especialista	9.140 (41,8%)	4.078 (63,9%)	
Neuro	3.018 (13,8%)	1.059 (16,6%)	
Body	4.096 (18,7%)	83 (1,3%)	
MSK	2.026 (9,3%)	2.936 (46,0%)	
Características de la estación de trabajo			
Número de monitores			<0,001
Monitor único	14.292 (65,3%)	4.519 (70,9%)	
Más de un monitor	7.590 (34,7%)	1.859 (29,1%)	
Dos monitores	5.808 (26,5%)	890 (14,0%)	
Tres monitores	1.782 (8,1%)	969 (15,2%)	
Sistema reconocimiento de voz			
No	21.016 (96,0%)	5.623 (88,2%)	<0,001
Sí	866 (4,0%)	755 (11,8%)	
Tipo de visor radiológico			
Clear Canvas®	7.069 (32,3%)	678 (10,6%)	<0,001
OsiriX®	14.813 (67,7%)	5.700 (89,4%)	

Tabla 6. Características de los estudios de telerradiología según tipología.

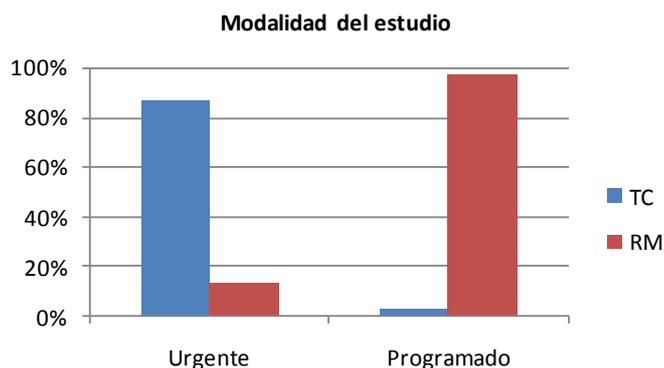


Figura 28. Modalidad de los estudios programados y urgentes.

Respecto a las características de los estudios de telerradiología analizados, en la Tabla 6 se aprecia que el porcentaje de RM en los estudios programados (96,9%) (IC95% 96,5-97,4%) es significativamente muy superior (p-valor < 0,001) que en los urgentes (13,1%) (IC95% 12,7-13,6%), tal y como puede apreciarse en la Figura 28. También se observan diferencias significativas en los patrones horarios de ambos tipos de estudios (programados y urgentes) (p-valor < 0,001), con un porcentaje superior de valoraciones en día y horario laborable en los estudios urgentes (38,5%) (IC95% 37,9-39,2%) respecto a los programados (35,7%) (IC95% 34,6-36,9%), un porcentaje muy inferior de valoraciones en día laborable pero horario no laborable en los urgentes (24,8%) (IC95% 24,2-25,3%) respecto a los programados (48,2%) (IC95% 47,0%-49,5%), y un porcentaje muy superior de valoraciones en día festivo en los urgentes (36,7%) (IC95% 36,1-37,4%) respecto a los programados (16,0%) (IC95% 15,1-16,9%).

También se han descrito las características del radiólogo que ha realizado la valoración inicial en ambos tipos de estudios. En la Tabla 6 se observan diferencias significativas en el género del radiólogo que ha realizado la valoración inicial (p-valor < 0,001), con un porcentaje superior de radiólogas del género femenino en los estudios programados (37,7%) (IC95% 36,5%-38,9%) que en los estudios urgentes (34,5%) (IC95% 33,8%-35,1%). También existe una edad mediana ligeramente superior de los radiólogos que valoraron inicialmente los estudios programados (46 años) que los que valoraron inicialmente los estudios urgentes (43 años), aunque sin detectarse diferencias estadísticamente significativas (p-valor = 0,172). Respecto a los años de experiencia, los radiólogos que han valorado los estudios programados presentan una mediana (10 años) muy superior a los que han valorado los estudios urgentes (4 años) (p-valor < 0,001). También se ha recogido la especialidad del radiólogo que valora el estudio inicialmente, observándose

de nuevo diferencias estadísticamente significativas entre los estudios programados y urgentes (p -valor $< 0,001$), con un mayor porcentaje de radiólogos especialistas en los estudios programados (63,9%) (IC95% 62,8-65,1%) que en los urgentes (41,8%) (IC95% 41,1-42,2%).

En la Tabla 6 también se comparan las principales características asociadas a la estación de trabajo del radiólogo que ha realizado la valoración inicial para ambos tipos de estudios. Se observan diferencias significativas respecto al número de monitores utilizados en la valoración inicial entre ambos tipos de estudios (programados y urgentes) (p -valor $< 0,001$), con un porcentaje superior de valoraciones con más de un monitor en los estudios urgentes (34,7%) (IC95% 34,1-35,3%) que en los programados (29,1%) (IC95% 28,0-30,3%). Se observan diferencias significativas en el uso de sistemas de reconocimiento de voz por parte del radiólogo que ha realizado la valoración inicial (p -valor $< 0,001$), con un porcentaje superior de uso del reconocimiento de voz en los estudios programados (11,8%) (IC95% 11,0-12,6%) que en los estudios urgentes (4,0%) (IC95% 3,7-4,2%). Por último, también existen diferencias estadísticamente significativas en el tipo de visor radiológico utilizado en la valoración inicial (p -valor $< 0,001$), con un porcentaje superior de uso del visor *OsiriX*[®] en los estudios programados (89,4%) (88,6-90,1%) que en los urgentes (67,7%) (67,1-68,3%).

A partir de la segunda valoración realizada por otro radiólogo se ha evaluado la presencia o ausencia de discrepancias diagnósticas entre ambas valoraciones. En la Tabla 7 se muestra que existen importantes diferencias en la presencia de discrepancias diagnósticas (p -valor $< 0,001$) entre estos dos tipos de estudios, con un porcentaje de discrepancias superior en los estudios programados (3,6%) (IC95% 3,1-4,1%) que en los urgentes (1,4%) (IC95% 1,2-1,6%), aunque en ambos casos estas cifras se sitúan en valores bastante reducidos. Hay que tener en cuenta las diferencias existentes entre ambos tipos de estudios, tanto en la modalidad de los mismos como en las características del radiólogo y estación de trabajo de la valoración inicial. En la Figura 29 se muestra gráficamente la tasa de discrepancias diagnósticas de ambos tipos de estudios.

	Urgente	Programado	p-valor
Total	21.882 (100%)	6.378 (100%)	
Discrepancia diagnóstica			<0,001
Sin discrepancias	21.575 (98,6%)	6.148 (96,4%)	
Con discrepancias	307 (1,4%)	230 (3,6%)	
Con discrepancias diagnósticas	307 (100%)	230 (100%)	
Clasificación escala RADPEER			<0,001
Interpretación difícil o hallazgo sutil	219 (71,3%)	126 (54,8%)	
Interpretación que generalmente debería hacerse	57 (18,6%)	76 (33,0%)	
Interp. que casi siempre debería hacerse / equiv. paciente	31 (10,1%)	28 (12,2%)	
Causa discrepancia diagnóstica			<0,001
Infralectura	225 (74,8%)	132 (57,9%)	
Sobrelectura	43 (14,3%)	40 (17,5%)	
Comunicación defectuosa	19 (6,3%)	27 (11,8%)	
Datos clínicos insuficientes	11 (3,7%)	24 (10,5%)	
Técnica defectuosa	3 (1,0%)	5 (2,2%)	
Repercusiones significativas			0,102
No	242 (78,8%)	167 (72,6%)	
Sí	65 (21,2%)	63 (27,4%)	

Tabla 7. Presencia de discrepancias diagnósticas en los estudios de telerradiología según tipología.

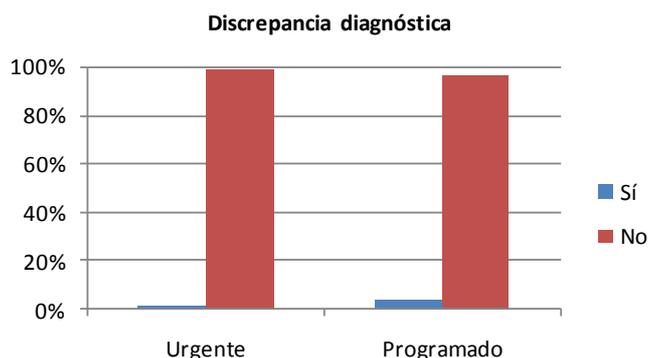


Figura 29. Discrepancia diagnóstica en los estudios programados y urgentes.

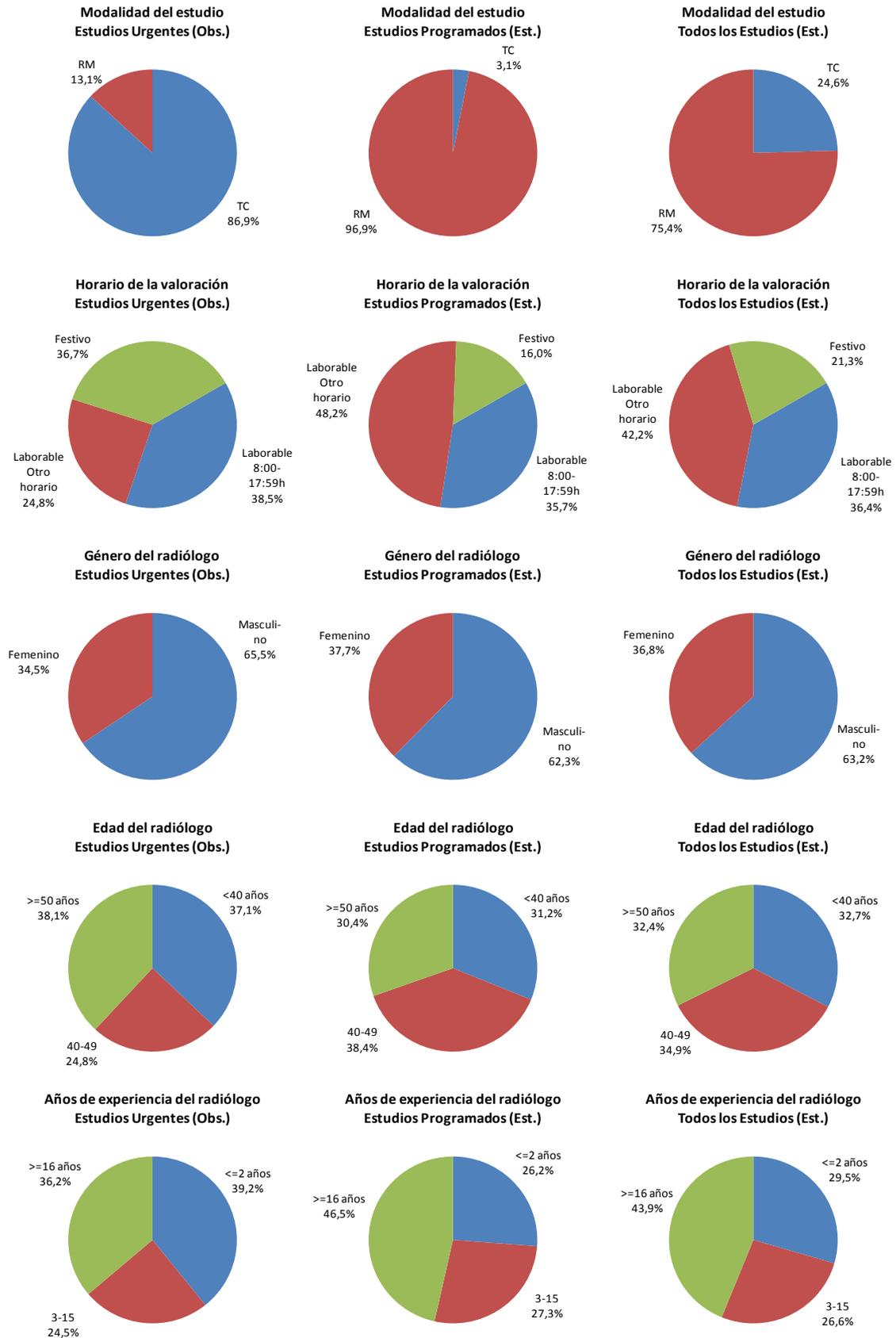
En la Tabla 7 están representadas también las características de estas discrepancias. Puede apreciarse que los pacientes programados presentan un porcentaje superior de discrepancias diagnósticas por interpretaciones que generalmente deberían hacerse (33,0% frente al 18,6% en los urgentes), los urgentes por interpretaciones difíciles o hallazgo sutil (71,3% frente al 54,8% en los programados), y ambos grupos presentan valores similares en interpretaciones que casi siempre deberían hacerse / equivocación de paciente. También se evidencia que los pacientes urgentes presentan un mayor grado de infralectura (74,8% frente al 57,9% de los programados), mientras que los programados presentan un mayor porcentaje del resto de causas, especialmente en el porcentaje de estudios con discrepancia diagnóstica con comunicación defectuosa y con datos clínicos insuficientes. Por último, la tasa media de estudios con discrepancias

clínicamente significativas se sitúa en el 23,8%. Aunque los programados presentan un mayor porcentaje de discrepancias con repercusiones significativas, estas diferencias no son estadísticamente significativas (p-valor = 0,102).

	N. obs	N. est	% est
Total	28.260	85.662	100%
Características del estudio			
Modalidad del estudio			
Tomografía Computarizada	18.888	20.607	24,6%
Resonancia Magnética	8.873	63.296	75,4%
Horario de la valoración			
Laborable 8:00-17:59 h.	10.707	31.218	36,4%
Laborable 18:00-7:59 h.	8.496	36.180	42,2%
Festivo	9.057	18.264	21,3%
Características del radiólogo			
Género del radiólogo			
Masculino	18.316	54.100	63,2%
Femenino	9.944	31.562	36,8%
Edad del radiólogo			
Media (Mediana)	45,2 (44)	44,8 (45)	-
Años de experiencia			
Media (Mediana)	10,8 (5)	11,7 (10)	-
Especialidad radiólogo			
Neuro	4.077	13.608	15,9%
Body	4.179	4.926	5,8%
MSK	4.962	31.386	36,6%
General	15.042	35.742	41,7%
Características de la estación de trabajo			
Número de monitores			
Monitor único	18.811	59.482	69,4%
Dos monitores	6.698	14.708	17,2%
Tres monitores	2.751	11.472	13,4%
Sistema reconocimiento de voz			
No	26.639	77.246	90,2%
Sí	1.621	8.416	9,8%
Tipo de visor radiológico			
Clear Canvas®	7.747	13.849	16,2%
OsiriX®	20.513	71.813	83,8%
Discrepancia diagnóstica			
Sin discrepancias	27.723	83.055	97,0%
Con discrepancias	537	2.607	3,0%

Tabla 8. Estimación de las características de todos los estudios de telerradiología del periodo.

En la Figura 30 se representa la distribución de las características de interés, tanto para los estudios urgentes como para los estudios programados analizados. También se muestra la distribución estimada de las características de interés para el total de estudios realizados en la empresa IBERORAD SL durante el periodo de dos años analizado.



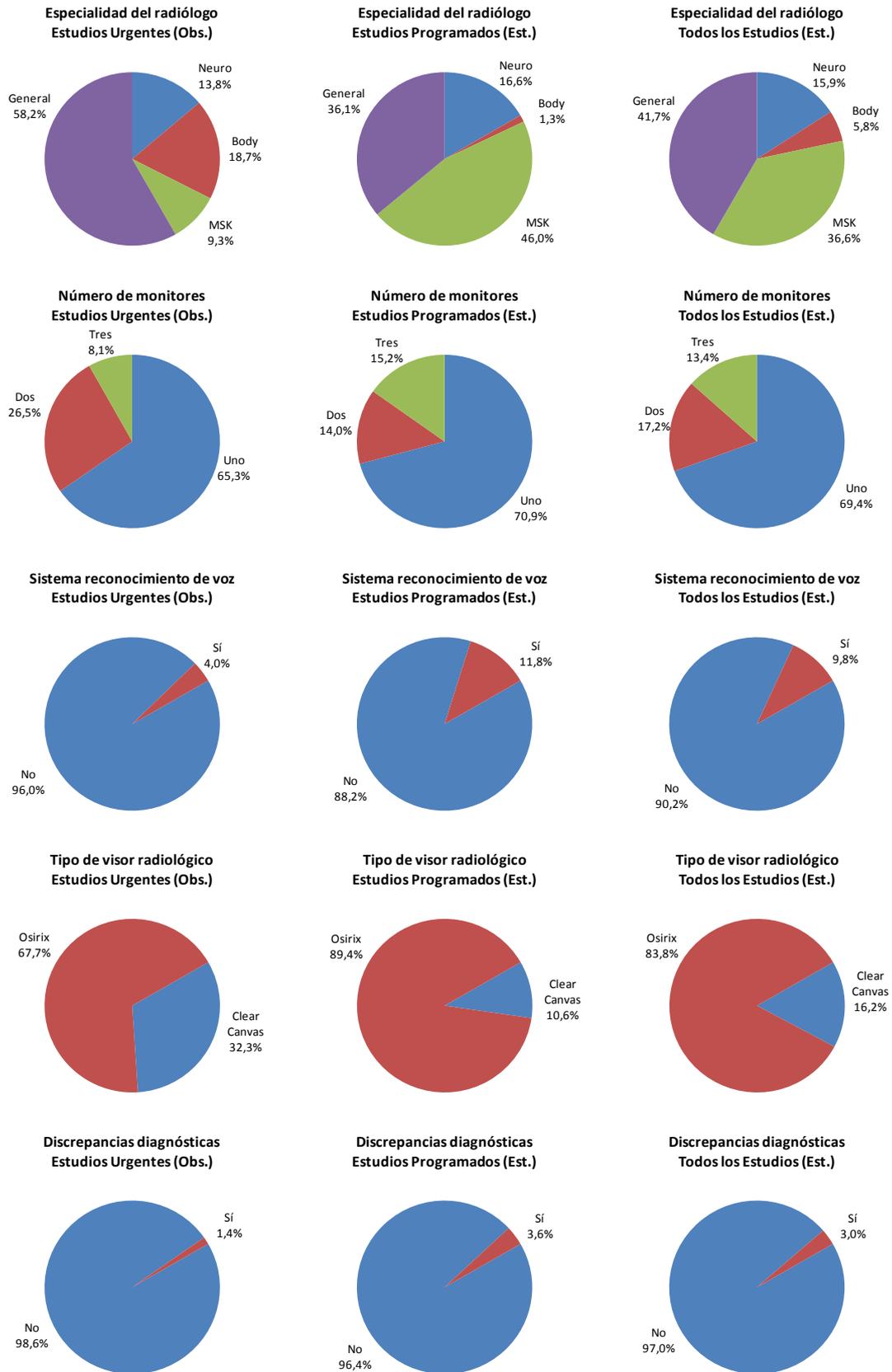


Figura 30. Distribución de las características de interés para los estudios urgentes, programados y para el total de estudios. Obs: observadas. Est: estimadas.

4.3. Estimación de la tasa de discrepancias diagnósticas.

Se aprecia que para el total de estudios existe una tasa de discrepancias diagnósticas del 3,0%, que desciende al 1,4% en los estudios urgentes y se sitúa en el 3,6% en los estudios programados.

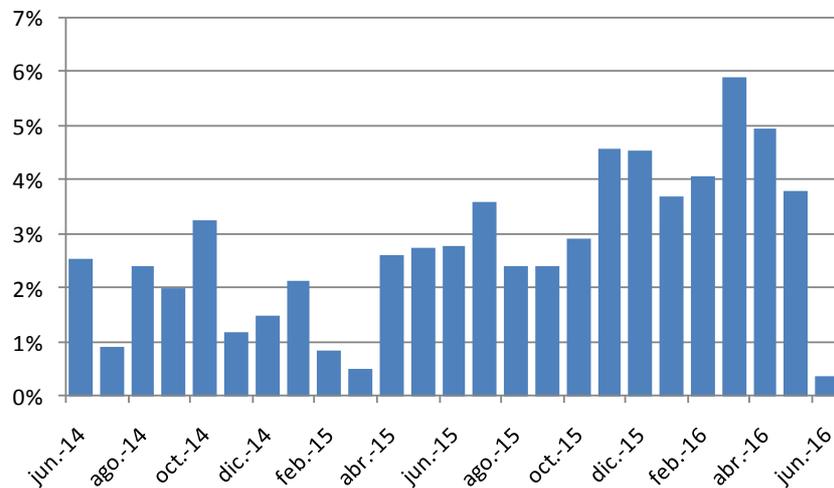


Figura 31. Evolución mensual de la tasa de discrepancias diagnósticas estimada.

En la Figura 31 se observa la evolución mensual de la tasa de discrepancias diagnósticas estimada. Se aprecia que la tasa de discrepancias diagnósticas más alta se produce en marzo de 2016 (5,9%).

	Programado		Urgente		Estimación Total		
	Discrep.	No discrep.	Discrep.	No discrep.	Núm. Estudios	Con discrep.	Tasadis crep.
Total	230	6.148	307	21.575	85.662	2.607	3,0%
Características del estudio							
Modalidad del estudio							
Tomografía Computarizada	19	172	268	18.429	20.607	458	2,2%
Resonancia Magnética	211	5.836	39	2.787	63.296	2.149	3,4%
Horario de la valoración							
Laborable 8am-5pm	60	2.219	119	8.309	31.218	719	2,3%
Laborable Otro horario	120	2.956	90	5.330	36.180	1.290	3,6%
Festivo	50	973	98	7.936	18.264	598	3,3%
Características del radiólogo							
Género del radiólogo							
Masculino	118	3.858	157	14.183	54.100	1.337	2,5%
Femenino	112	2.290	150	7.392	31.562	1.270	4,0%
Edad del radiólogo							
<40	135	1.852	151	7.975	27.996	1.501	5,4%
40-49	71	2.380	67	5.349	29.926	777	2,6%
>=50	24	1.916	89	8.251	27.740	329	1,2%
Años de experiencia							
<=2	115	1.556	158	8.427	25.295	1.308	5,2%
3-15	64	1.678	83	5.282	22.785	723	3,2%
>=16	51	2.914	66	7.866	37.582	576	1,5%
Especialidad radiólogo							
Neuro	18	1.041	22	2.996	13.608	202	1,5%
Body	7	76	38	4.058	4.926	108	2,2%
MSK	51	2.885	9	2.017	31.386	519	1,7%
General	154	2.146	238	12.504	35.742	1.778	5,0%
Características de la estación de trabajo							
Número de monitores							
Monitor único	203	4.316	238	14.054	59.482	2.268	3,8%
Más de un monitor	27	1.832	69	7.521	26.180	339	1,3%
Sistema reconocimiento voz							
No	220	5.403	302	20.714	77.246	2.502	3,2%
Sí	10	745	5	861	8.416	105	1,2%
Tipo de visor radiológico							
Clear Canvas®	15	663	63	7.006	13.849	213	1,5%
OsiriX®	215	5.485	244	14.569	71.813	2.394	3,3%

Tabla 9. Estimación del porcentaje de discrepancias estimadas según los diferentes factores recogidos.

En la Tabla 9 se muestran las estimaciones de las tasas de discrepancias diagnósticas obtenidas a partir de los datos analizados para las principales características de interés recogidas en la base de datos estudiada. Hay que tener en cuenta que en esta tabla se presenta únicamente la tasa de error asociada a cada factor latente analizado de forma individualizada, por lo que no se tiene en cuenta las posibles interacciones entre los diferentes factores. En la sección siguiente se llevará a cabo un estudio univariante y

multivariante para analizar el efecto de cada uno de estos factores sobre la probabilidad de una discrepancia diagnóstica en la segunda valoración.

Por ejemplo, como se indica en la Tabla 9, se han observado 19 discrepancias en los 191 estudios programados de TC (9,9% de discrepancias) y 268 discrepancias en los 18.697 estudios urgentes (1,4% de discrepancias) con esta modalidad, por lo que se obtiene una estimación de 458 estudios con discrepancia de los 20.607 estimados con esta modalidad, lo que representa una tasa de discrepancia global del 2,2%. En cambio, se observa para la RM un porcentaje de discrepancia del 3,5% en los estudios programados y 1,4% en los urgentes, por lo que la estimación para el total de los estudios será de 2.149 discrepancias de un total de 63.296 estudios, que se traduce en un 3,4% de tasa de discrepancias estimada. Por lo tanto, la tasa de discrepancias es mayor en las RM que en las TC. En la Tabla 10 se muestran las tasas de discrepancias diagnósticas estimados para ambas modalidades de estudio.

TASAS DE DISCREPANCIA	PROGRAMADOS	URGENTES	GLOBAL (est.)
TC	9,9%	1,4%	2,2%
RM	3,5%	1,4%	3,4%
GLOBAL (est.)	3,6%	1,4%	3%

Tabla 10. Porcentaje de discrepancias según modalidad y tipo de estudio.

Del mismo modo, se observa que las tasas de discrepancias más elevadas se dan en los estudios valorados en día laborable fuera del horario de 18:00 a 7:59 h (3,6%), seguido de los estudios valorados en festivo (3,3%), mientras que los estudios valorados en día laborable de 8:00 a 17:59h presentan una tasa menor (2,3%). También se muestran los valores estimados para el resto de factores registrados en la base de datos analizada. En la Figura 32 se muestran las tasas de discrepancias diagnósticas estimadas para estos factores, donde hemos marcado en rojo aquella categoría con mayor tasa de discrepancia diagnóstica y en verde la categoría con menor tasa de discrepancia diagnóstica.

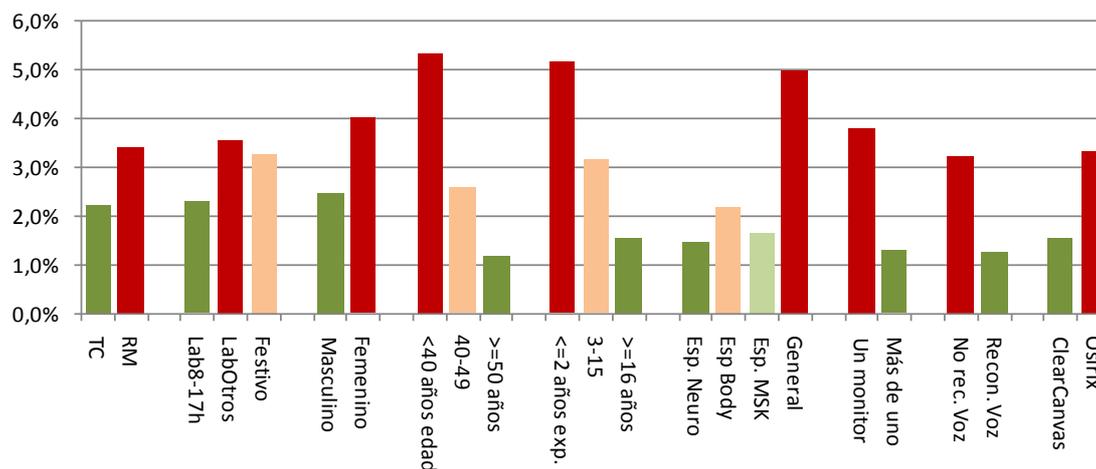


Figura 32. Tasa de discrepancias diagnósticas estimada según características del estudio. Verde: variable con menor tasa de discrepancia. Rojo: variable con mayor tasa de discrepancia. Rosado: variable con tasa de discrepancia intermedia.

4.4. Evaluación de factores relacionados con la discrepancia diagnóstica en Radiología.

4.4.1. Evaluación de los factores relacionados con la discrepancia diagnóstica en los estudios de telerradiología programados.

Como se observa en la Figura 24, en este análisis se han incluido los 6.378 estudios de telerradiología de carácter programado realizados por radiólogos en la empresa IBERORAD SL durante un periodo de dos años que han recibido una segunda lectura, y por tanto una evaluación de la posible discrepancia diagnóstica. Estos estudios representan una muestra aleatoria de aproximadamente el 10% del total de estudios programados realizados.

Para el análisis del efecto de los diferentes factores de interés considerados sobre la presencia de discrepancia diagnóstica en los estudios de telerradiología programados se han seleccionado los 230 estudios programados que presentaron discrepancias diagnósticas (casos) y una muestra de 129 estudios de los 6.148 estudios programados que no presentaron discrepancias diagnósticas (controles), en total 359 estudios programados.

En primer lugar se ha analizado el efecto individual de cada uno de los factores de interés sobre la presencia o ausencia de discrepancia diagnóstica en el estudio sobre los 359 estudios programados seleccionados, construyendo finalmente un modelo multivariante que nos permita identificar aquellos factores que influyen de forma independiente en la probabilidad de discrepancia diagnóstica. En el Anexo 2 puede observarse una tabla resumen de los resultados de todos los análisis univariantes y multivariantes que se han llevado a cabo.

4.4.1.1. Análisis univariante de factores significativos en los estudios programados.

Desarrollado en el Anexo 1. Son significativos, desde un punto de vista univariante, la modalidad del estudio (mayor asociación con discrepancia diagnóstica en los estudios de TC respecto a los de RM), la calidad técnica de las imágenes adecuada respecto a la ausencia de la misma, la presencia de imágenes previas respecto a la ausencia de las mismas, la presencia de un monitor respecto a disponer de más de uno, la ausencia de reconocimiento de voz respecto a la presencia del sistema, la edad del radiólogo, la experiencia del radiólogo, el género del radiólogo femenino respecto al género masculino, la especialidad del estudio MSK frente a columna o la concordancia entre la especialidad del radiólogo y el estudio, siendo además todos ellos factores precursores de discrepancia diagnóstica, a excepción de la edad y la experiencia del radiólogo que son factores protectores.

4.4.1.2. Análisis multivariante de factores significativos en los estudios programados.

En la Tabla 11 se aprecia el resultado del ajuste multivariante para modelizar la probabilidad de discrepancia diagnóstica en los estudios telerradiología programados. Se observa que la modalidad del estudio continúa siendo un factor latente muy relacionado con la discrepancia diagnóstica (p-valor = 0,012), con un incremento de la probabilidad de discrepancia diagnóstica de 7,6 veces para un estudio de TC respecto a uno de RM. Existe también un incremento significativo (p-valor = 0,012) del riesgo de discrepancia diagnóstica en los estudios con imágenes con una calidad técnica inadecuada respecto a los estudios con imágenes adecuadas. Del mismo modo, existe un incremento significativo (p-valor = 0,035) del riesgo de discrepancia diagnóstica en los estudios con imágenes previas respecto de los estudios sin imágenes previas. Otro factor relacionado con la discrepancia diagnóstica es el número de monitores utilizados en la valoración inicial (p-valor = 0,002), con un incremento de la probabilidad de discrepancia

diagnóstica de 3,0 veces para un estudio valorado con un único monitor respecto a otro valorado con dos o más monitores. También influye de forma significativa sobre la probabilidad de discrepancia diagnóstica el uso de un sistema de reconocimiento de voz en la valoración inicial (p -valor = 0,010), con un incremento de la probabilidad de discrepancia diagnóstica de 3,7 veces para un estudio valorado sin este sistema respecto a otro valorado con este sistema. La concordancia entre la especialidad del radiólogo y la del estudio también representa un factor significativo en la valoración de las discrepancias diagnósticas (p -valor < 0,001), con un incremento de la probabilidad de discrepancia diagnóstica de 3,9 veces para un estudio con concordancia media/baja respecto a otro con concordancia alta. Se observa también que el horario en que se ha realizado la valoración inicial es también un factor significativo en el modelo multivariante (p -valor = 0,031) al explicar una parte significativa de la variabilidad de la presencia de discrepancia diagnóstica que no es explicada por otros factores, con un incremento significativo de la probabilidad de discrepancia diagnóstica en los estudios realizados en horario no laboral o en día festivo respecto a los realizados en horario laboral.

El error diagnóstico en Radiología. Análisis de variables explicativas.

	Modelo univariante			Modelo multivariante		
	OR	IC95%	p-valor	OR	IC95%	p-valor
Complejidad del estudio			0,282			0,296
Normal	(línea base)					
Alta	1,872	0,598 5,866				
Modalidad del estudio			0,020			0,012
Resonancia Magnética	(línea base)			(línea base)		
Tomografía Computarizada	5,718	1,310 24,958		7,629	1,555 37,429	
Calidad técnica de las imágenes			0,035			0,012
Buena	(línea base)			(línea base)		
Regular / Mala	8,930	1,166 68,407		15,005	1,800 125,045	
Imágenes de estudios previos			0,020			0,035
No	(línea base)			(línea base)		
Sí	5,718	1,310 24,958		5,694	1,126 28,787	
Información clínica de la solicitud			0,077			0,294
Adecuada	(línea base)					
Insuficiente	1,615	0,950 2,745				
Tipo de visor utilizado			0,153			0,702
Clear Canvas®	(línea base)					
OsiriX®	1,745	0,814 3,741				
Número de monitores			<0,001			0,002
Dos o más monitores	(línea base)			(línea base)		
Un monitor	3,024	1,738 5,261		2,996	1,517 5,915	
Uso sistema reconocimiento voz			0,022			0,010
Sí	(línea base)			(línea base)		
No	2,678	1,154 6,218		3,706	1,374 9,993	
Edad del radiólogo	0,940	0,913 0,968	<0,001			0,585
Experiencia del radiólogo	0,945	0,921 0,969	<0,001			0,324
Género del radiólogo			0,036			0,413
Masculino	(línea base)					
Femenino	1,602	1,031 2,489				
Horario valoración inicial			0,120			0,031
Laboral 8:00-17:59h	(línea base)			(línea base)		
Laboral Otro horario	1,593	0,973 2,608	0,064	1,928	1,066 3,484	0,030
Festivo	1,703	0,912 3,179	0,095	2,441	1,167 5,104	0,018
Especialidad del estudio			0,023			0,709
Columna	(línea base)					
MSK	2,200	1,247 3,883	0,007			0,510
Neuro / Body	2,138	0,906 5,043	0,083			0,920
Concordancia especialidades			<0,001			<0,001
Alta	(línea base)			(línea base)		
Media/Baja	3,467	2,209 5,441		3,920	2,309 6,655	

Tabla 11. Modelo multivariante del efecto de los factores latentes considerados sobre la discrepancia diagnóstica en los estudios programados.

En la Tabla 11 se observa que en el modelo multivariante final no tienen efecto significativo sobre la discrepancia diagnóstica la complejidad del estudio, la información clínica de la solicitud, la especialidad del estudio, el tipo de visor utilizado, la edad del radiólogo, los años de experiencia del radiólogo o el género del radiólogo.

En ninguno de los análisis multivariantes de nuestro estudio se ha introducido como variable independiente la especialidad del radiólogo, pues se han analizado la especialidad del estudio y la concordancia del estudio con la especialidad del radiólogo, por lo que si también se añade la especialidad del radiólogo existirían tres variables con problemas de colinealidad (ya que la concordancia se obtiene de las otras dos) y aparecerían problemas de convergencia de los ajustes. En este punto hay que elegir dos, y se han elegido la especialidad del estudio y la concordancia del mismo con la especialidad del radiólogo.

Aunque el objetivo principal del análisis multivariante realizado mediante regresión logística es evaluar la fuerza de asociación de los factores de riesgo considerados sobre la probabilidad de cometer una discrepancia diagnóstica de una manera independiente, se tratará de aplicar este mismo modelo con fines predictivos a partir de las variables incluidas en el modelo final. A partir del modelo multivariante de la Tabla 11 puede obtenerse, para cada uno de los estudios programados, una estimación de la probabilidad de discrepancia diagnóstica en la segunda valoración en función de los factores significativos incluidos en el estudio.

En la Figura 33 puede observarse la curva ROC asociada a la predicción realizada con este modelo multivariante, con un área bajo la curva ROC estimada de 0,765 (IC_{95%} 0,714-0,815) (p-valor < 0,001). Si utilizamos como punto de corte una probabilidad de 0,66 (esto es, si predecimos una discrepancia en los estudios programados con una probabilidad estimada mayor a 0,66, y no discrepancia en los estudios con una probabilidad estimada menor o igual), que es el valor que maximiza el índice de Youden, (la suma de especificidad + sensibilidad - 1) obtenemos una prueba diagnóstica de discrepancia con una sensibilidad del 71,3% y una especificidad del 67,4%.

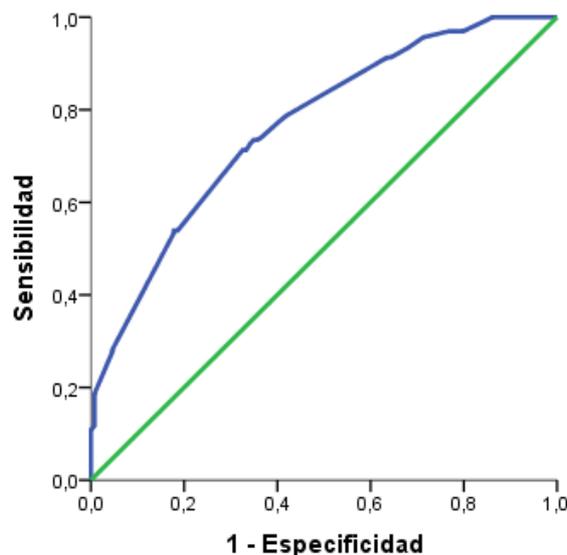


Figura 33. Curva ROC asociada al modelo predictivo obtenido para los estudios programados.

4.4.2. Evaluación de los factores relacionados con la discrepancia diagnóstica en los estudios de telerradiología urgentes.

Como puede observarse en la Figura 24, se han incluido en este análisis los 21.882 estudios de telerradiología de carácter urgente realizados por radiólogos en la empresa IBERORAD SL durante un periodo de dos años, habiendo recibido todos ellos una revisión y por tanto una evaluación de la discrepancia diagnóstica.

Para el análisis del efecto de los diferentes factores de interés considerados sobre la presencia de discrepancia diagnóstica en los estudios de telerradiología urgentes se han seleccionado los 307 estudios urgentes que presentaron discrepancias diagnósticas (casos) y una muestra de 408 estudios de los 21.575 estudios urgentes que no presentaron discrepancias diagnósticas (controles).

4.4.2.1. Análisis univariante de factores significativos en los estudios urgentes.

Desarrollado en el Anexo 1. Se observa que el efecto de una complejidad alta del estudio respecto a una complejidad normal es un factor precursor significativo de discrepancia diagnóstica. También son factores precursores significativos de discrepancia diagnóstica, desde un punto de vista univariante, la presencia de imágenes de estudios previos respecto a la ausencia de las mismas, una información clínica de la solicitud insuficiente respecto a una información adecuada, el uso del visor *OsiriX*® respecto al visor *Clear*

Canvas®, la presencia de un monitor respecto a disponer de más de uno, la ausencia de reconocimiento de voz respecto a la presencia del sistema, el género del radiólogo femenino respecto al género masculino, la especialidad del estudio Body respecto a Neurorradiología, la concordancia media entre la especialidad del radiólogo y el estudio respecto a una concordancia alta o la concordancia baja entre la especialidad del radiólogo y el estudio respecto a una concordancia alta, mientras que la edad del radiólogo y los años de experiencia del radiólogo son factores protectores de la discrepancia diagnóstica. En cambio no son significativos la modalidad del estudio, la calidad técnica de las imágenes o el horario de la valoración inicial.

4.4.2.2. Análisis multivariante de factores significativos en los estudios urgentes.

El resultado se aprecia en la Tabla 12. Se observa que la complejidad del estudio continúa siendo un factor latente muy relacionado con la discrepancia diagnóstica (p-valor = 0,026), con un incremento del riesgo de discrepancia diagnóstica de 2,1 veces para un estudio de alta complejidad respecto a un estudio con complejidad normal. Del mismo modo, existe un incremento significativo (p-valor = 0,001) del riesgo de discrepancia diagnóstica en los estudios con imágenes previas respecto de los estudios sin imágenes previas. También se observa un incremento significativo (p-valor < 0,001) del riesgo de discrepancia diagnóstica en los estudios con información clínica insuficiente en la solicitud respecto a los que la tienen adecuada. También son factores relevantes la presencia de un único monitor (p-valor = 0,011) y la ausencia de un sistema de reconocimiento de voz (p-valor = 0,015). Otro factor latente relacionado con la discrepancia diagnóstica es la especialidad del estudio (p-valor = 0,006), siendo importante el incremento de riesgo de discrepancia de los estudios de Body respecto a los de Neuro. El grado de concordancia entre la especialidad del radiólogo y del estudio a evaluar es también un factor relacionado con la discrepancia diagnóstica independiente (p-valor < 0,001), observándose una mayor discrepancia en los estudios con concordancia media que en los estudios con concordancia baja, que a su vez son peores que los de concordancia alta. Se observa en la Tabla 12 que en el modelo multivariante final no tienen efecto significativo sobre la presencia de discrepancia diagnóstica de los estudios urgentes la modalidad del estudio, la calidad técnica de las imágenes, el horario de la valoración inicial, el tipo de visor utilizado, la edad del radiólogo, los años de experiencia del radiólogo o el género del radiólogo. No obstante hay que indicar que la modalidad

del estudio, la calidad técnica de las imágenes y el horario rozan la significación estadística.

	Modelo univariante			Modelo multivariante		
	OR	IC95%	p-valor	OR	IC95%	p-valor
Complejidad del estudio			0,040			0,026
Normal	(línea base)			(línea base)		
Alta	1,871	1,028	3,403	2,097	1,090	4,031
Modalidad del estudio			0,630			0,076
Resonancia Magnética	(línea base)			(línea base)		
Tomografía Computarizada	0,895	0,569	1,407			
Calidad técnica de las imágenes			0,114			0,074
Buena	(línea base)			(línea base)		
Regular / Mala	2,434	0,808	7,338			
Imágenes de estudios previos			0,001			0,001
No	(línea base)			(línea base)		
Sí	6,291	2,107	18,783	8,170	2,460	27,141
Información clínica de la solicitud			<0,001			<0,001
Adecuada	(línea base)			(línea base)		
Insuficiente	3,155	1,681	5,922	3,664	1,873	7,168
Tipo de visor utilizado			0,005			0,507
Clear Canvas®	(línea base)			(línea base)		
OsiriX®	1,652	1,166	2,342			
Número de monitores			0,027			0,011
Dos o más monitores	(línea base)			(línea base)		
Un monitor	1,471	1,046	2,071	1,608	1,115	2,318
Uso sistema reconocimiento voz			0,006			0,015
Sí	(línea base)			(línea base)		
No	3,943	1,492	10,421	3,563	1,278	9,936
Edad del radiólogo	0,980	0,966	0,994			0,268
Experiencia del radiólogo	0,969	0,955	0,984			0,161
Género del radiólogo			<0,001			0,406
Masculino	(línea base)			(línea base)		
Femenino	1,911	1,410	2,590			
Horario valoración inicial			0,165			0,098
Laboral 8:00-17:59h	(línea base)			(línea base)		
Laboral Otro horario	1,232	0,849	2,788			0,273
Festivo	0,850	0,601	1,204			0,361
Especialidad del estudio			0,010			0,006
Neuro	(línea base)			(línea base)		
Body	1,659	1,196	2,301	1,729	1,220	2,450
MSK / Columna	1,262	0,823	1,936	1,538	0,967	2,447
Concordancia especialidades			<0,001			<0,001
Alta	(línea base)			(línea base)		
Media	3,591	2,170	5,944	4,035	2,347	6,937
Baja	2,085	1,154	3,767	2,500	1,336	4,681

Tabla 12. Modelo multivariante del efecto de los factores latentes considerados sobre la discrepancia diagnóstica en los estudios urgentes.

Aunque el objetivo principal del análisis multivariante realizado mediante regresión logística es evaluar la fuerza de asociación de los factores de riesgo considerados sobre la probabilidad de cometer una discrepancia diagnóstica de una manera independiente, se

tratará de aplicar este mismo modelo con fines predictivos a partir de las variables incluidas en el modelo final. A partir del modelo multivariante de la Tabla 12 se obtiene, para cada uno de los estudios urgentes, una estimación de la probabilidad de discrepancia diagnóstica en la segunda valoración en función de los factores significativos incluidos en el estudio.

En la Figura 34 se observa la curva ROC asociada a la predicción realizada con este modelo multivariante, con un área encerrada bajo la curva ROC estimada de 0,684 (IC95% 0,645-0,723) (p -valor < 0,001). Si utilizamos como punto de corte una probabilidad de 0,42 (esto es, si predecimos una discrepancia en los estudios programados con una probabilidad estimada mayor a 0,42, y no discrepancia en los estudios con una probabilidad estimada menor o igual), que es el valor que maximiza el índice de Youden, obtenemos una prueba diagnóstica de discrepancia con una sensibilidad del 56,7% y una especificidad del 69,1%.

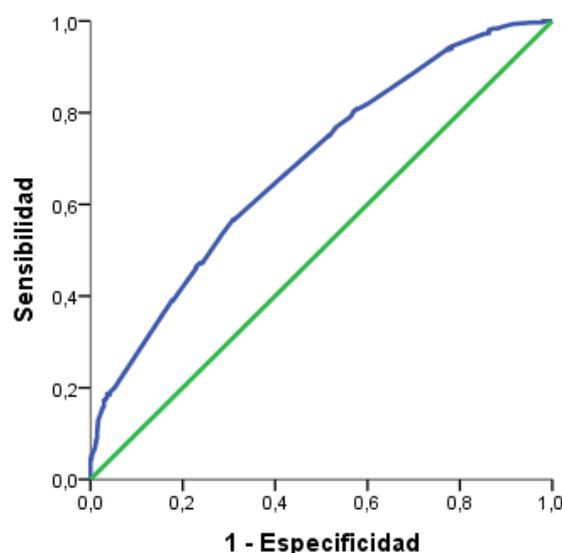


Figura 34. Curva ROC asociada al modelo predictivo obtenido para los estudios urgentes.

4.4.3. Comparación de los modelos ajustados para ambos estratos.

La Tabla 13 muestra los ajustes multivariantes ajustados para ambos estratos. La presencia de imágenes de estudios previos, el número de monitores, el uso de reconocimiento de voz y la concordancia entre las especialidades del radiólogo y del estudio son factores relevantes para los estudios de ambos estratos. En cambio la modalidad de estudio, la calidad técnica de las imágenes y el horario son relevantes en los estudios programados, mientras que la complejidad del estudio, la información

clínica de la solicitud y la especialidad del estudio son significativas para los estudios urgentes.

		Programados			Urgentes				
		OR	IC (95%)		p-valor	OR	IC (95%)		p-valor
Complejidad del estudio	Alta vs Normal				0,296	2,097	1,090	4,031	0,026
Modalidad del estudio	TC vs RM	7,629	1,555	37,429	0,012				0,076
Calidad técnica	Inadecuada vs Adecuada	15,005	1,800	125,045	0,012				0,074
Imágenes estud. previos	Sí vs No	5,694	1,126	28,787	0,035	8,170	2,460	27,141	0,001
Inf. clínica solicitud	Insuf vs Suf				0,294	3,664	1,873	7,168	<0,001
Tipo de visor utilizado	OsiriX® vs Clear Canvas®				0,702				0,507
Número de monitores	Uno vs 2 o más	2,996	1,517	5,915	0,002	1,608	1,115	2,318	0,011
Uso sistema rec. voz	No vs Sí	3,706	1,375	9,993	0,010	3,563	1,278	9,936	0,015
Edad radiólogo					0,585				0,268
Experiencia radiólogo					0,324				0,161
Género radiólogo	Fem vs Masc				0,413				0,406
Horario					0,031				0,098
	No Lab vs Lab	1,928	1,066	3,484	0,030				0,050
	Festivo vs Lab	2,441	1,167	5,104	0,018				0,085
Especialidad estudio					0,709				0,006
	MSK vs Colum				0,510				
	Neuro/Body vs Colum				0,920				
	Body vs Neuro					1,729	1,220	2,450	0,002
	MSK/Colum vs Neuro					1,538	0,967	2,447	0,069
Concordancia de especialidades					<0,001				<0,001
	Media/Baja vs Alta	3,467	2,209	5,441					
	Media vs Alta					4,035	2,347	6,937	<0,001
	Baja vs Alta					2,500	1,336	4,681	0,004

Tabla 13. Modelos multivariantes para los estudios programados y urgentes.

4.4.4. Evaluación de los factores relacionados con la discrepancia diagnóstica sobre todos los estudios de telerradiología.

En este apartado se considerarán todos los estudios, tanto programados como urgentes, seleccionados para el análisis del efecto de los diferentes factores de interés considerados sobre la presencia de discrepancia diagnóstica en los estudios de telerradiología. En concreto se habían seleccionado, tal y cómo se muestra en la Figura 24, los 537 estudios que presentaron discrepancias diagnósticas (casos) y una muestra de 537 estudios que no presentaron discrepancias diagnósticas (controles).

Para llevar a cabo inferencias causales hay que poder asumir que los controles proceden y son representativos de la población de la que han sido seleccionados los casos. En este caso esta población no sería la población total, sino la población accesible formada por todos los estudios urgentes y únicamente un 10% de los estudios programados. De esta

población (población accesible) es de donde se han extraído los casos, y de donde se ha obtenido una muestra de controles. Por tanto, los OR univariantes asociados a cada factor latente estimados sobre esta muestra de casos–controles será representativa de la población accesible, aunque podrían no ser representativos de la población total para aquellos factores de riesgo relacionados con el carácter del estudio, ya que en la población accesible (y por tanto en la muestra seleccionada) el porcentaje de estudios urgentes es mucho mayor que el que se ha observado en la población total. En cualquier caso se ha llevado a cabo un estudio multivariante de los factores significativos desde el punto de vista univariante incluyendo el carácter del estudio para controlar el efecto del carácter del estudio sobre cada factor.

En primer lugar se ha analizado el efecto individual de cada uno de los factores de interés sobre la presencia o ausencia de discrepancia diagnóstica en el estudio, construyendo finalmente un modelo multivariante que nos permita identificar aquellos factores que influyen de forma independiente en la probabilidad de discrepancia diagnóstica. Como ya se ha expuesto anteriormente, hay que ser especialmente cautelosos con las interpretaciones de los resultados de este apartado al existir un claro sesgo de verificación con una mayor presencia de estudios urgentes en la muestra respecto a la población total.

4.4.4.1. Análisis univariante de factores significativos. Desarrollado en el Anexo 1. Se aprecia que todos los factores son factores significativos respecto a la presencia o ausencia de discrepancia diagnóstica desde el punto de vista univariante.

4.4.4.2. Análisis multivariante de factores significativos.

El resultado se observa en la Tabla 14. Se evidencia que la complejidad del estudio continúa siendo un factor latente muy relacionado con la discrepancia diagnóstica (p -valor = 0,014), con un incremento de la probabilidad de discrepancia diagnóstica de 2,1 veces para un estudio de alta complejidad respecto a un estudio con complejidad normal. Otro de los factores muy relacionados con la discrepancia diagnóstica en la segunda valoración es el tipo del estudio, con un incremento significativo del riesgo de discrepancia diagnóstica en los estudios programados respecto de los estudios urgentes (p -valor < 0,001). Del mismo modo, existe un incremento significativo del riesgo de discrepancia diagnóstica en los estudios con imágenes previas respecto de los estudios

sin imágenes previas (p -valor $< 0,001$), en los estudios con información insuficiente en la solicitud con los que tienen una información adecuada (p -valor $< 0,001$) o en los estudios con una calidad técnica de las imágenes regular o mala respecto a los que tienen una buena calidad (p -valor = $0,009$). En el modelo multivariante también se aprecia el efecto precursor significativo de discrepancia diagnóstica de la presencia de un único monitor (p -valor $< 0,001$) y de la ausencia de un sistema de reconocimiento de voz (p -valor $< 0,001$). También existen diferencias relevantes respecto del horario en que se realizó la valoración inicial (p -valor = $0,042$), con un incremento del riesgo de discrepancia diagnóstica en los estudios valorados en día laborable fuera del horario normal de trabajo respecto de los estudios valorados en día y horario laborable. Otro factor latente relacionado con la discrepancia diagnóstica es la especialidad del estudio (p -valor = $0,007$), siendo importante el incremento de riesgo de discrepancia de los estudios de Body respecto a los de Neuro, mientras que el incremento de riesgo de los MSK o de Columna respecto de los de Neuro no es estadísticamente significativo. Por último, también es muy relevante la concordancia de las especialidades del estudio con la del radiólogo (p -valor $< 0,001$), con un incremento relevante en la probabilidad de discrepancia diagnóstica de los de concordancia baja o media respecto a los de concordancia alta. En el modelo multivariante final no tienen efecto significativo sobre la presencia o ausencia de discrepancia diagnóstica la modalidad del estudio, el tipo de visor utilizado, la edad, los años de experiencia o el género del radiólogo que realiza la valoración inicial.

	Modelo univariante			Modelo multivariante		
	OR	IC95%	p-valor	OR	IC95%	p-valor
Complejidad del estudio			0,041			0,014
Normal	(línea base)			(línea base)		
Alta	1,720	1,022 2,896		2,112 1,167 3,822		
Modalidad del estudio			<0,001			0,619
Resonancia Magnética	(línea base)					
Tomografía Computarizada	0,550	0,429 0,705				
Tipo de estudio			<0,001			<0,001
Urgente	(línea base)			(línea base)		
Programado	2,370	1,824 3,078		2,651 1,677 4,191		
Calidad técnica de las imágenes			0,002			0,009
Buena	(línea base)			(línea base)		
Regular/Mala	4,140	1,679 10,212		3,800 1,401 10,310		
Imágenes de estudios previos			<0,001			<0,001
No	(línea base)			(línea base)		
Sí	6,549	2,740 15,651		7,768 2,960 20,386		
Información clínica de la solicitud			<0,001			<0,001
Adecuada	(línea base)			(línea base)		
Insuficiente	2,745	1,851 4,070		2,337 1,499 3,644		
Tipo de visor utilizado			<0,001			0,715
Clear Canvas®	(línea base)					
OsiriX®	1,996	1,465 2,718				
Número de monitores			<0,001			<0,001
Dos o más monitores	(línea base)			(línea base)		
Un monitor	1,932	1,448 2,578		1,909 1,379 2,643		
Uso sistema reconocimiento voz			0,001			<0,001
Sí	(línea base)			(línea base)		
No	2,725	1,484 5,006		3,418 1,725 6,774		
Edad del radiólogo	0,967	0,955 0,980	<0,001			0,406
Experiencia del radiólogo	0,963	0,951 0,975	<0,001			0,233
Género del radiólogo			0,001			0,288
Masculino	(línea base)					
Femenino	1,828	1,430 2,337				
Horario valoración inicial			0,002			0,042
Laboral 8:00-17:59h	(línea base)			(línea base)		
Laboral Otro horario	1,542	1,156 2,057	0,003	1,484 1,067 2,064		0,019
Festivo	0,963	0,716 1,296	0,804	1,055 0,757 1,470		0,754
Especialidad del estudio			<0,001			0,007
Neuro	(línea base)			(línea base)		
MSK	2,571	1,880 3,515	<0,001	1,341 0,831 2,164		0,230
Body	1,727	1,257 2,371	0,001	1,778 1,266 2,495		0,001
Columna	1,596	1,048 2,431	0,029	1,639 0,985 2,728		0,057
Concordancia especialidades			<0,001			<0,001
Alta	(línea base)					
Media	2,277	1,744 3,063	<0,001	3,881 2,694 5,591		<0,001
Baja	1,033	0,671 1,591	0,881	2,528 1,513 4,222		<0,001

Tabla 14. Modelo multivariante del efecto de los factores latentes considerados sobre la discrepancia diagnóstica.

4.4.5. Evaluación sobre una submuestra de los estudios de telerradiología muestreados.

Se selecciona aleatoriamente una muestra del 10% de estudios urgentes, estratificando por la presencia de discrepancia diagnóstica, obteniendo un subconjunto de 41 estudios urgentes sin discrepancia diagnóstica y 31 estudios urgentes con discrepancia diagnóstica. De esta forma, el conjunto de 261 estudios con discrepancia diagnóstica (230 programados y 31 urgentes) son una muestra aleatoria de los 2.607 estudios estimados con discrepancia diagnóstica (2.300 programados y 307 urgentes) con la misma distribución según carácter del estudio (programado o urgente). Del mismo modo, el conjunto de 170 estudios sin discrepancia diagnóstica (129 programados y 41 urgentes) son una muestra aleatoria de los 83.055 estudios estimados sin discrepancia diagnóstica (61.480 programados y 21.575 urgentes) con la misma distribución según carácter del estudio (programado o urgente). Por tanto, se obtiene una muestra de casos – controles representativa de los casos y controles del total de estudios.

En primer lugar se ha analizado el efecto individual de cada uno de los factores de interés sobre la presencia o ausencia de discrepancia diagnóstica en el estudio, construyendo finalmente un modelo multivariante que nos permita identificar aquellos factores que influyen de forma independiente en la probabilidad de discrepancia diagnóstica.

4.4.5.1. Análisis univariante de factores significativos. Desarrollado en el Anexo 1. Se observa asociación estadísticamente significativa con la discrepancia diagnóstica de todas las variables excepto la complejidad del estudio, la modalidad del estudio (TC/RM), el tipo de visor utilizado y el uso de sistemas de reconocimiento de voz. También se evidencia asociación estadísticamente significativa con la discrepancia diagnóstica de la nueva variable introducida en la submuestra: el tipo de estudio (urgente/programado).

4.4.5.2. Análisis multivariante de factores significativos.

El resultado puede apreciarse en la Tabla 15. Se observa que la calidad técnica de las imágenes del estudio continua siendo un factor relacionado con la discrepancia diagnóstica (p -valor = 0,018), con un incremento del riesgo de discrepancia diagnóstica de 6,6 veces para un estudio de calidad insuficiente respecto a un estudio con

complejidad aceptable. El tipo de estudio también es un factor significativo (p-valor = 0,001), con un incremento significativo de la probabilidad de discrepancia diagnóstica de los estudios programados respecto a los urgentes. También se observa un incremento significativo (p-valor = 0,015) del riesgo de discrepancia diagnóstica en los estudios con imágenes previas respecto de los estudios sin imágenes previas, y en los estudios con información clínica en la solicitud insuficiente respecto a los que la tienen adecuada (p-valor = 0,029). También son factores relevantes la presencia de un único monitor (p-valor = 0,003) y la especialidad del estudio (p-valor = 0,015) (siendo significativo el incremento de riesgo de los estudios de Body respecto a los de Neuro). El grado de concordancia entre la especialidad del radiólogo y del estudio a evaluar es también un factor relacionado con la discrepancia diagnóstica independiente (p-valor < 0,001), observándose una mayor discrepancia en los estudios con concordancia media que en los estudios con concordancia alta, mientras que los de concordancia baja se sitúan en un nivel intermedio. En la Tabla 15 se evidencia que en el modelo multivariante final no tienen efecto significativo sobre la presencia de discrepancia diagnóstica la complejidad del estudio, la modalidad del estudio, el tipo de visor utilizado, el uso de sistema de reconocimiento de voz, el horario de la valoración inicial, la edad del radiólogo, los años de experiencia del radiólogo o el género del radiólogo.

El error diagnóstico en Radiología. Análisis de variables explicativas.

	Modelo univariante			Modelo multivariante				
	OR	IC95%	p-valor	OR	IC95%	p-valor		
Complejidad del estudio			0,185			0,265		
Normal	(línea base)			(línea base)				
Alta	1,904	0,735	4,932					
Modalidad del estudio			0,197			0,088		
Resonancia Magnética	(línea base)			(línea base)				
Tomografía Computarizada	0,732	0,456	1,175					
Calidad técnica de las imágenes			0,024			0,018		
Buena	(línea base)			(línea base)				
Regular / Mala	5,486	1,245	24,171	6,575	1,378	31,362		
Tipo de estudio			0,001			0,001		
Urgente	(línea base)			(línea base)				
Programado	2,358	1,411	3,942	4,286	1,753	10,475		
Imágenes de estudios previos			0,011			0,015		
No	(línea base)			(línea base)				
Sí	4,871	1,430	16,593	5,604	1,407	22,318		
Información clínica de la solicitud			0,005			0,029		
Adecuada	(línea base)			(línea base)				
Insuficiente	2,101	1,257	3,510	1,904	1,067	3,397		
Tipo de visor utilizado			0,133			0,203		
Clear Canvas®	(línea base)			(línea base)				
OsiriX®	1,615	0,864	3,018					
Número de monitores			<0,001			0,003		
Dos o más monitores	(línea base)			(línea base)				
Un monitor	3,180	1,932	5,233	2,447	1,351	4,433		
Uso sistema reconocimiento voz			0,055			0,051		
Sí	(línea base)			(línea base)				
No	2,199	0,985	4,912					
Edad del radiólogo	0,957	0,933	0,980	<0,001		0,335		
Experiencia del radiólogo	0,961	0,940	0,982	<0,001		0,974		
Género del radiólogo			0,040			0,607		
Masculino	(línea base)			(línea base)				
Femenino	1,514	1,020	2,247					
Horario valoración inicial			0,091			0,221		
Laboral 8:00-17:59h	(línea base)			(línea base)				
Laboral Otro horario	1,639	1,050	2,558	0,030		0,083		
Festivo	1,398	0,822	2,377	0,216		0,312		
Especialidad del estudio			<0,001			0,015		
Neuro	(línea base)			(línea base)				
MSK	3,255	1,761	6,015	<0,001	1,225	0,523	2,869	0,640
Body	3,300	1,457	7,475	0,004	3,655	1,478	9,036	0,005
Columna	1,365	0,647	2,880	0,414	0,775	0,293	2,049	0,608
Concordancia especialidades			<0,001			<0,001		
Alta	(línea base)			(línea base)				
Media	2,442	1,127	3,664	<0,001	2,926	1,844	4,642	<0,001
Baja	1,090	0,337	3,520	0,886	2,603	0,674	10,054	0,165

Tabla 15. Modelo multivariante del efecto de los factores latentes considerados sobre la discrepancia diagnóstica.

A partir del modelo multivariante de la Tabla 15 se obtiene, para cada uno de los estudios, una estimación de la probabilidad de discrepancia diagnóstica en la segunda valoración en función de los factores significativos incluidos en el estudio.

En la Figura 35 se observa la curva ROC asociada a la predicción realizada con este modelo multivariante, con un área encerrada bajo la curva ROC estimada de 0,749 (IC95% 0,703-0,796) (p -valor < 0,001). Si utilizamos como punto de corte una probabilidad de 0,67 (esto es, si predecimos una discrepancia en los estudios programados con una probabilidad estimada mayor a 0,67, y no discrepancia en los estudios con una probabilidad estimada menor o igual), que es el valor que maximiza el índice de Youden, obtenemos una prueba diagnóstica de discrepancia con una sensibilidad del 63,2% y una especificidad del 74,7%.

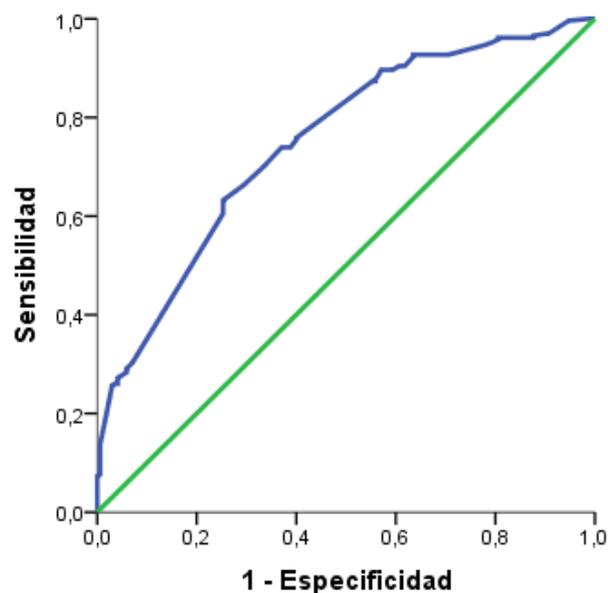


Figura 35. Curva ROC asociada al modelo predictivo obtenido en la submuestra.

5. DISCUSIÓN.

5.1. ASPECTOS GENERALES.

Para el total de estudios existe una tasa de discrepancia diagnóstica del 3,0%, que desciende al 1,4% en los estudios urgentes y se sitúa en el 3,6% en los estudios programados. La tasa global de discrepancia diagnóstica observada en nuestro estudio es por tanto coherente con las cifras publicadas en la bibliografía y expuestas en la introducción.

Para explicar por qué la tasa de discrepancia observada es superior en los estudios programados que en los urgentes es fundamental tener en cuenta que en los estudios programados no podemos analizar si la revisión fue por selección estrictamente aleatoria en el contexto del control de calidad ordinario de la empresa o si fue generada por una solicitud expresa del médico receptor tras leer el informe. Esta última situación introduce un sesgo en la distribución porcentual del tipo de errores detectados (por ejemplo, puede generar un aumento en el porcentaje de errores de comunicación en detrimento de los errores de infralectura) (Brigham et al., 2015). Además, con cierta frecuencia, los médicos receptores del estudio solicitan revisiones de los mismos tras conocer información adicional posterior al estudio radiológico (ej: endoscopias, pruebas de laboratorio, estudios de imagen adicionales...) de la que el radiólogo autor del informe inicial no disponía cuando interpretó las imágenes. Puede ocurrir además que entre la fecha del estudio radiológico y la fecha de los hallazgos descubiertos a posteriori (incluyendo evidencia quirúrgica o de Anatomía Patológica) haya habido progresión de la enfermedad, con lo que dicha evidencia posterior no sería aplicable al estudio radiológico previo. Podría argumentarse en base a ello que la tasa de error de los estudios programados encontrada en nuestro análisis está falsamente magnificada debido a un sesgo de selección. La selección aleatorizada de los casos para las dobles lecturas conduce a menores tasas de discrepancia, como se ilustra en la tasa de discrepancia del 0,69% del estudio de O'Keeffe (O'Keeffe et al., 2016). Esto puede explicarse por la inclusión de grandes cantidades de casos poco complejos, poco susceptibles a la discrepancia en las dobles lecturas, que conducirían a un mayor número de categorías 1 y A. Son por tanto necesarios estudios adicionales para clarificar estas

posibles diferencias entre la selección aleatoria y no aleatoria de los casos sometidos a doble lectura.

También se comprueba que el porcentaje de cada uno de los tipos de errores se ajusta a los resultados de trabajos previos, siendo el tipo de error más frecuente el de infralectura (habitualmente correspondiente a falsos negativos o errores de percepción), el segundo error más frecuente el de sobrelectura (en su mayoría correspondientes a falsos positivos o errores cognitivos) y en tercer lugar se situarían los errores de comunicación. Se confirma por tanto el hallazgo ampliamente descrito en la bibliografía de que el número de errores de infralectura (falsos negativos) supera en más de cinco veces al de sobrelectura (falsos positivos) (Birkelo & Chamberlain, 1947) (Yerushalmy et al., 1951) (Newell et al., 1954) (Garland, 1959) (Renfrew et al., 1992).

Es importante tener en cuenta que dentro de los errores de comunicación, en nuestro estudio sólo medimos los errores en la transmisión del resultado al médico solicitante. Sin embargo se ha observado que la mayoría de los errores de comunicación ocurren en otros puntos del estudio de imagen, como la solicitud, la programación, la realización o la interpretación (Siewert B. et al., 2015)

La tasa media de estudios con discrepancias clínicamente significativas se sitúa en el 23,8%. Este resultado refleja de forma coherente la gravedad clínica general de las patologías, siendo una constante ampliamente reconocida en Medicina que las patologías leves o con menor significación clínica predominan ampliamente sobre aquellas de mayor gravedad; es lógico por tanto que esta distribución se traslade a los errores diagnósticos. Aunque los estudios programados presentan un mayor porcentaje de discrepancias con repercusiones significativas (27,4% del total de las discrepancias respecto al 21,2% en los urgentes), estas diferencias no son estadísticamente significativas (p-valor = 0,102).

Se evidencia que la asociación de las distintas variables analizadas con el error diagnóstico varía en función de si el estudio es urgente o programado. En el análisis univariante del total de estudios de telerradiología (urgentes y programados) todas y cada una de las variables analizadas tienen relación estadísticamente significativa con la presencia de discrepancia diagnóstica. Probablemente sea porque el tamaño muestral es mucho mayor, lo que permite ajustar el modelo con mayor precisión y detectar si el OR

asociado es distinto de 1. Sin embargo, los resultados del análisis multivariante conjunto de todos los estudios (programados y urgentes) han de valorarse con mucha cautela, pues ya se ha explicado que existe un sesgo de selección consistente en una sobrerrepresentación de los estudios urgentes respecto a la población real de nuestro estudio. Lo ideal sería haber analizado el mismo número de estudios programados que urgentes, pero la forma de selección de los estudios de la muestra ha sido inevitablemente determinada por el propio funcionamiento diario de la revisión de casos en la empresa; se considera prioritario revisar el 100% de los estudios urgentes debido a la potencial gravedad de un error diagnóstico en los mismos, y sólo el 10% de los estudios programados debido a la ingente cantidad de recursos que consumiría realizar una revisión de la totalidad de estos últimos. Sí podemos evaluar una aproximación teórica a este escenario ideal a partir de los resultados del análisis de la submuestra.

5.2. COMPLEJIDAD DEL ESTUDIO.

La complejidad del estudio (en áreas anatómicas, cantidad de imágenes, reconstrucciones...) presenta asociación estadísticamente significativa con la discrepancia diagnóstica en los estudios urgentes y también en el análisis conjunto de estudios urgentes y programados (en este último caso probablemente debido a la sobrerrepresentación de estudios urgentes en la selección de casos). Hay que tener en cuenta que el plazo límite de realización de un informe urgente en la empresa IBERORAD S.L. es de 1 hora en contraste con el par de días de plazo de entrega de un informe programado. Es más fácil por tanto que la complejidad del estudio genere un impacto en términos de discrepancia diagnóstica cuando el radiólogo tiene menos margen de tiempo para analizar el estudio, algo aún más evidente si se tiene en cuenta que algunas características que definen la complejidad del estudio son un alto número de imágenes, áreas anatómicas y/o reconstrucciones.

Se sabe que los errores diagnósticos pueden estar relacionados simplemente con la dificultad de los casos en cuestión (Renfrew et al., 1992). En nuestro análisis sólo hemos podido probar esta asociación de la complejidad de la prueba con las discrepancias diagnósticas en los estudios urgentes. Es posible que exista por tanto un conflicto entre el intenso ritmo de trabajo de la Radiología de Urgencias y la especial dedicación que requieren aquellos estudios que o bien constan de un elevado número de imágenes o bien presentan una especial complejidad anatómica. Los errores diagnósticos en estudios complejos podrían ser por otra parte especialmente susceptibles de mejora con entrenamiento específico. Buchanan et al. observaron que los radiólogos especializados tenían un rendimiento diagnóstico significativamente superior al de los radiólogos inexpertos en la clasificación prequirúrgica mediante RM de las fistulas anales, pero también descubrieron que los radiólogos inexpertos podían alcanzar una precisión diagnóstica aceptable tras un periodo de educación específica (Buchanan et al., 2004).

5.3. MODALIDAD DEL ESTUDIO (TC/RM).

Dentro de los estudios programados se observa mayor tasa de discrepancia diagnóstica en los estudios de TC respecto a los estudios de RM, tanto en el análisis univariante como en el multivariante. Para explicar esta asociación deberemos analizar las características propias de las TC programadas. Con gran frecuencia se trata de estudios oncológicos, una especialidad transversal de la Radiología que implica a múltiples órganos-sistemas y que presenta unas características propias. En estos estudios de TC generalmente se evalúan múltiples sistemas simultáneamente, con adquisiciones multifásicas, por lo que no es infrecuente que el radiólogo deba evaluar más de 100 imágenes en un solo estudio. Son estudios con gran número de imágenes, no tantas en términos generales como los estudios de cuerpo completo de los politraumatizados (no han sido incluidos por tanto en la categoría de estudios complejos) pero sí suficientemente significativas como para considerarlas al explicar esta asociación con la discrepancia diagnóstica. El rango de posibles patologías y la lista de diagnósticos diferenciales con frecuencia es mucho más amplia que cuando se estudia una sola región anatómica como ocurre habitualmente en los estudios de RM (Leslie et al., 2000).

También suelen ser estudios que requieren evaluación comparativa con otros anteriores, y en este sentido son susceptibles de sufrir los sesgos propios de la existencia de otro estudio obviamente informado con anterioridad (aspecto que desarrollaremos más tarde al explicar la asociación de la disponibilidad de estudios de imagen previos con mayor riesgo de discrepancia diagnóstica). Estas TC oncológicas además, por la necesidad de comparación evolutiva y también por la minuciosidad de su informe (no limitado únicamente al diagnóstico sino que incluye también la estadificación y, cada vez más, información altamente estructurada y detallada en relación a las características de la patología que puede incluir biomarcadores, sistemas diagnósticos específicos para varios tumores – por ejemplo, PI-RADS, LI-RADS, Lung-RADS...) generalmente consumen más tiempo de informe radiológico que las TC urgentes. Supone un 40% más de tiempo informar una TC abdominopélvica de pacientes oncológicos respecto a informar una TC de pacientes que acuden con dolor abdominal (Muchantef & Forman, 2004). Por otra parte la información clínica disponible para los radiólogos externos al propio centro, es decir, aquellos que como en el caso de nuestro estudio trabajan mediante telerradiología, no es tan amplia como aquella con la que cuentan los radiólogos presenciales sin

restricciones de acceso a toda la historia clínica del paciente (fechas de intervenciones quirúrgicas, informes de la técnica quirúrgica realizada, datos de laboratorio, fechas de inicio y de final de regímenes de quimioterapia/radioterapia, fármacos usados en el tratamiento que con frecuencia afectan directamente al aspecto radiológico de las lesiones tumorales...etc).

5.4. CALIDAD TÉCNICA DEL ESTUDIO.

Se observa asociación estadísticamente significativa entre una deficiente calidad técnica de las imágenes y las discrepancias diagnósticas en los estudios programados. También parece ser relevante en los estudios urgentes, pero el reducido número de estudios urgentes con calidad técnica inadecuada dificulta la estimación del riesgo asociado con la suficiente precisión.

La pobre calidad de imagen y la técnica subóptima pueden influenciar la habilidad del radiólogo para identificar alteraciones o para interpretarlas correctamente (Siewert B. Et al., 2008). Brenner et al observaron que una pobre calidad de imagen era un factor contribuyente en el 17% de los errores diagnósticos en mamografía (Brenner et al., 1998). Yee et al observaron también que la distensión del colon en la colonografía por TC mejoraba significativamente al usar de forma combinada las posiciones en decúbito y en prono, y que esta mejora en la técnica del estudio se correlacionaba directamente con mejoras de sensibilidad en la detección de pólipos (Yee et al., 2003). Y también se ha visto en otro estudio que dos causas principales de error en la clasificación por TC de las masas renales son la variabilidad interobservador y una técnica radiológica incorrecta (Bosniak, 1997) (Siegel C L. et al., 1997), siendo muy difícil implementar la primera de ambas.

La generación de la imagen radiológica es fuente potencial de múltiples errores, por ejemplo, aquellos relacionados con el número y tipo de imágenes radiológicas (proyecciones, series, secuencias... según la técnica radiológica en cuestión), el posicionamiento del paciente, la selección de los parámetros, el uso inapropiado de los colimadores y las rejillas, los errores de procesamiento y los artefactos de la imagen (Alexander, 2010). Y si la realización de una técnica radiológica correcta es de por sí importante como práctica general, lo es más cuando deben solventarse características especiales del paciente (por ejemplo talla, complexión corporal, incapacidad para permanecer inmóvil...) que puedan degradar la calidad del estudio (Brown et al., 1999). Las causas de estos errores pueden residir en insuficiente formación de los técnicos, baja calidad de los equipos, o condiciones de trabajo subóptimas como cuando un técnico de Radiodiagnóstico está sobrepasado por el trabajo y sin apoyo (Bhimani & Bruno, 2018).

Reducir la explicación de las imágenes deficientes a un contexto laboral o económico susceptible de control sería una visión muy simplista. Las posibles razones subyacentes a un estudio de imagen deficiente son múltiples, desde los artefactos de movimiento del paciente, la dificultad del paciente para mantener apneas prolongadas, el diferente grado de experiencia de los técnicos...etc. Sin embargo, sí debemos ser conscientes de que debido a un contexto de precarización y presión asistencial creciente, los técnicos pueden intentar acortar los estudios en detrimento de su calidad técnica, y por tanto los radiólogos deben controlar de forma exhaustiva y permanente la calidad de las imágenes de los estudios radiológicos. La telerradiología no debe sustituir la contratación de radiólogos que gestionen localmente los servicios radiológicos. El servicio de telerradiología debe instaurarse para mejorar la atención al paciente y no por la falta de radiólogos o la reducción de los costes que comprometen siempre los estándares de calidad en la asistencia sanitaria. Además el contrato de telerradiología debería incluir los protocolos de actuación en el entorno clínico y la descripción de los aspectos técnicos de la instalación que aseguren la calidad y la seguridad de los procedimientos, garantizando el cumplimiento de los estándares de calidad de las imágenes en cualquier momento (Martí-Bonmatí et al., 2011).

Independientemente del origen del error, si un radiólogo identifica un estudio de pobre calidad o no diagnóstico, debe reseñar claramente la limitación técnica y solicitar que se repita el estudio o se realice un procedimiento alternativo y, en la mayoría de los casos, diferir la interpretación hasta que la calidad del estudio sea suficientemente buena. El estudio debe también etiquetarse para revisión interna para que pueda llevarse a cabo una adecuada gestión de riesgos, y abordar el origen de la calidad de imagen deficiente (Ryan et al., 2018).

5.5. DISPONIBILIDAD DE ESTUDIOS PREVIOS.

La presencia de estudios de imagen previos es uno de los factores que presenta asociación estadísticamente significativa con la discrepancia diagnóstica en todos los análisis del estudio, tanto en los estudios programados como en los urgentes. Este dato resulta muy sorprendente, pues siempre se ha pensado que la situación era exactamente la contraria, es decir, que la precisión diagnóstica de los radiólogos podría verse reducida por el acceso limitado a los estudios de imagen o informes previos (Robinson, 1997), (Berlin, 1999a). Y este razonamiento es totalmente lógico, ya que usar un estudio previo como referencia para valorar la progresión temporal de una lesión puede ayudar a catalogar la agresividad de un hallazgo. Del mismo modo, al evaluar la progresión de una lesión puede estrecharse el diagnóstico diferencial o valorar si los hallazgos de imagen están mejorando o empeorando. Para los radiólogos, esto es por tanto similar a la historia clínica y es coherente pensar que cuando no estén disponibles estudios previos aumentará el riesgo de omitir o interpretar erróneamente una lesión (Ryan et al., 2018).

Para poder justificar la asociación, aparentemente paradójica, que observamos en nuestro trabajo entre presencia de estudios de imagen previos y discrepancias diagnósticas debemos considerar varios elementos.

En primer lugar el sesgo aliterativo, por el que el radiólogo tiende a reproducir un diagnóstico previo erróneo en el informe que en ese momento está realizando (Berlin, 2000a). No sólo pueden transferirse de un informe previo al actual los errores diagnósticos; palabras o números mal escritos, información imprecisa, u otros datos en el registro informático de los datos también pueden reproducirse con “copia y pega” en los informes sucesivos.

La prevalencia de error aliterativo podría estar influenciada por el contexto laboral actual. En entornos de trabajo como el de la telerradiología, en que la retribución depende directamente del número de informes realizados, conforme la presión asistencial aumenta y la remuneración del trabajo médico sigue una tendencia descendente, los profesionales sanitarios con frecuencia intentan compensar la situación aumentando su productividad individual y colectiva (Reiner & Krupinski, 2012). Esto implica que con frecuencia asuman como válido el contenido de los informes previos sin

reinterpretar las imágenes por el sobreesfuerzo que esto último implica. A veces esta sobrecarga de tiempo y de esfuerzo es inevitable, como ocurre en los controles evolutivos de imagen (por ejemplo en estudios oncológicos, enfermedades desmielinizantes, fracturas...), y en estos casos sería el propio incremento de carga laboral de la evaluación comparativa entre ambos estudios el que podría generar un aumento en la tasa de error médico. Es probable que esta constituya la explicación de que como se explicó anteriormente, la modalidad del estudio, en concreto los estudios de TC, se asocian a mayor discrepancia diagnóstica únicamente cuando son programados pero no cuando son estudios urgentes.

Habría que evaluar si los estudios de imagen con exploraciones previas se asocian o no a mayor discrepancia diagnóstica en aquellos entornos de trabajo, por ejemplo la red sanitaria pública de nuestro país, donde la retribución económica de los radiólogos no está ligada tan directamente a la productividad.

5.6. ADECUACIÓN DE LA INFORMACIÓN CLÍNICA.

Otro factor asociado a la presencia de discrepancia diagnóstica en nuestro estudio es la información clínica ausente o deficiente en la solicitud del estudio de imagen. Las deficiencias en la información clínica aportada se asocian a discrepancia diagnóstica en la mayor parte de los análisis del estudio, un resultado concordante con la mayoría de la bibliografía médica publicada hasta el momento. La única falta de asociación estadísticamente significativa con la discrepancia diagnóstica en nuestro trabajo se observa en el análisis uni y multivariante de los estudios programados, lo cual podría deberse al tamaño muestral porque en ellos roza la significación estadística; sería interesante evaluar el comportamiento de esta variable en un grupo más amplio de sujetos, en el cual posiblemente sí alcanzase mayor significación. Otra posible explicación de que las deficiencias en la información clínica se asocien con la discrepancia diagnóstica sólo en los estudios urgentes y no en los estudios programados es que la sobrecarga y falta de estandarización de los tiempos de asistencia médicos en los servicios de Urgencias hace que con frecuencia los radiólogos reciban información clínica incompleta o irrelevante, con el consiguiente riesgo de error diagnóstico. Incluso en muchos casos, en el momento en que se solicita el estudio de imagen la recogida de datos clínicos todavía está en curso. Por ello es de enorme importancia que los radiólogos en Urgencias mantengan una actitud proactiva de comunicación con el médico solicitante (Ryan et al., 2018). Además se ha visto que el impacto de una información clínica deficitaria sobre la tasa de error diagnóstico aumenta con la complejidad del estudio (Leslie et al., 2000) y en este sentido los estudios urgentes son principalmente de TC, con numerosas imágenes, muchas estructuras anatómicas y varias patologías de diagnóstico diferencial, y en los cuales la información clínica puede ayudar mucho a centrar el área de interés y la sospecha diagnóstica.

La información clínica rigurosa es un componente esencial de la solicitud de los estudios de imagen, que aumenta el rendimiento diagnóstico si se usa adecuadamente. Leslie et al observaron que cuando la información clínica aportada era correcta aumentaba la precisión diagnóstica del 83% de los informes (Leslie et al., 2000). El trabajo de Dalla Palma et al. mostró que la comunicación clinicoradiológica producía cambio del diagnóstico clínico en el 50% de los casos y cambio de tratamiento en el 60% de los casos discutidos (Dalla Palma et al., 2000). Se ha objetivado además que el conocimiento de los

datos clínicos pertinentes aumenta la precisión en la interpretación de estudios de radiología convencional del 16 al 72% en residentes y del 38 al 84% en radiólogos titulados (Berlin & Berlin, 1995).

Idealmente, la solicitud de los estudios de imagen debería incluir información clínica pertinente que ayude al radiólogo a centrarse en el área de interés y a responder preguntas específicas. Por ejemplo, el conocer que un paciente con disnea presenta un estado de hipercoagulabilidad es fundamental para que el radiólogo considere una búsqueda dirigida de tromboembolismos pulmonares.

Desafortunadamente muchos clínicos consideran que el informe radiológico es definitivo; creen que la detección de alteraciones y la interpretación de los hallazgos es un proceso objetivo y que no debería variar entre un radiólogo y otro. Es más, algunos clínicos piensan que ocultar información clínica al radiólogo tendrá como consecuencia un informe más objetivo (Leslie et al., 2000).

En muchos casos las anotaciones cortas que los radiólogos reciben de los médicos solicitantes son insuficientes para proporcionar la historia clínica más apropiada para que los primeros emitan una impresión diagnóstica precisa y trasladable a medidas terapéuticas concretas (Tudor et al, 1997) (Leslie et al., 2000). Además de la historia clínica, la integración con un registro médico electrónico también permite al radiólogo acceder a los parámetros de laboratorio, anotaciones quirúrgicas e informes de anatomía patológica, todos los cuales pueden jugar un papel significativo en la mejora de la precisión de la interpretación, evitando así el error cognitivo (Kabadi & Krishnaraj, 2017). Y la telerradiología tiene una potencial debilidad en este aspecto, pues con frecuencia los radiólogos informan los estudios en base a documentos aislados y anotaciones cortas anexadas a la solicitud, sin tener acceso directo e inmediato al conjunto de la historia clínica del paciente. La menor fluidez que existe en telerradiología en el acceso a la historia clínica electrónica del paciente no está en relación, como aparentemente podría pensarse, con la protección legal de los datos del paciente (pues la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal habilita al radiólogo a acceder a la historia clínica electrónica del paciente en tanto que constituye parte integrante del equipo asistencial) sino con motivos de índole operativa. Las empresas de telerradiología suelen operar con múltiples centros sanitarios, cada uno de ellos con un Sistema de Información Radiológica (RIS) distinto, y por lo tanto es mucho más práctico

que a los radiólogos se les indique directamente la información de la solicitud y, en caso de que considerasen necesario algún dato en particular, puedan solicitarlo directamente al centro. Sin embargo esta manera de proceder inevitablemente condiciona barreras en la comunicación y dificultad para el trabajo multidisciplinario en equipo que podrían eventualmente afectar a la calidad del proceso sanitario.

Por otra parte, la información clínica inadecuada o las expectativas erróneas sobre las capacidades de una técnica radiológica pueden conducir a una incorrecta comprensión o comunicación entre el médico solicitante y el radiólogo conduciendo, en vez de a una evaluación más precisa, a errores diagnósticos (Brady A., Laoide, McCarthy, & McDermott, 2012). Eldevik et al. consideraron el efecto del sesgo clínico en la interpretación de mielografías y TC espinales; observaron que la historia clínica aumentaba el número de diagnósticos falsos positivos y reducía el número de diagnósticos falsos negativos (Eldevik et al., 1982). Esto sugiere una tendencia en los observadores a interpretar los hallazgos indeterminados como patológicos cuando se correlacionan con clínica positiva. Y no se trata de un asunto de escasa relevancia ya que, en los últimos años, se ha observado un progresivo crecimiento de las demandas por diagnósticos falsos positivos (por ejemplo, de patología meniscal o ligamentosa) que conducen a procedimientos quirúrgicos innecesarios (Berlin & Berlin, 1995) (Halpin, 2009) (van Breest Smallenburg et al., 2012).

También tienen una relación directa con el rendimiento diagnóstico el énfasis o la gravedad con que el médico solicitante describe los datos clínicos. De acuerdo con un estudio de Newmann-Toker, el 9% de todos los eventos cerebrovasculares pasan inadvertidos inicialmente, y las probabilidades de error diagnóstico se multiplican al menos por 5 cuando los síntomas son leves o transitorios (Newman-Toker, Robinson, & Edlow, 2008).

Por otra parte se ha comprobado que la información clínica confusa puede sesgar al radiólogo, conduciendo a un aumento de los diagnósticos falsos positivos (Doubilet & Herman, 1981) (Fitzgerald, 2001). En un estudio prospectivo ciego en el que los radiólogos interpretaban los estudios antes y después de tener acceso a la información clínica, el conocer la historia clínica produjo un informe más preciso en el 67% de los casos; sin embargo, en el 33% de los casos el informe tras conocer la información clínica fue menos preciso (Leslie et al., 2000).

Un posible mecanismo de estos errores asociados a la información clínica es el sesgo de composición (“framing bias”). Es importante evitarlo revisando primero las imágenes antes de leer la información clínica (Hare, 2007), para asegurar que la impresión radiológica no esté excesivamente influenciada por aquella. Además el radiólogo debe usar su juicio clínico para determinar qué partes de la historia clínica son relevantes y factibles y cuáles no.

5.7. TIPO DE VISOR.

El tipo de visor sólo se asocia a discrepancia diagnóstica en el análisis univariante de los estudios urgentes, y en el análisis univariante de todos los estudios (con sobrerrepresentación de estudios urgentes frente a programados). Hay que considerar que el modelo multivariante evalúa el efecto de cada una de las variables en presencia de las otras, esto es, si explican una parte importante de la variabilidad en la discrepancia diagnóstica que no es explicada por otras variables. El tipo de visor es significativo en el modelo univariante de estudios urgentes para explicar la discrepancia diagnóstica (es decir, que aisladamente y al igual que otras variables explica una parte importante de las diferencias en la discrepancia diagnóstica de los estudios analizados), en cambio en el modelo multivariante hay otras variables que continúan siendo significativas y el tipo de visor ha dejado de serlo.

5.8. NÚMERO DE MONITORES.

El número de monitores usados para visualizar los estudios de imagen es una variable que en todos los análisis presenta asociación estadísticamente significativa con las discrepancias diagnósticas. Un trabajo ya publicado muestra, de forma similar a lo observado en nuestro estudio, que la situación óptima de trabajo es una estación de dos monitores para visualización simultánea tanto del estudio actual como de los previos (Siegel E L. & Reiner, 2002). No cabe duda de que, a diferencia de la mayoría de los clínicos cuyo trabajo está fundamentalmente centrado en el paciente, el trabajo de los radiólogos está casi exclusivamente ligado al ordenador (con la excepción de los procedimientos de intervencionismo, mamografía y fluoroscopia) (Vertinsky & Forster, 2005). Se sabe que la disminución de estrés cognitivo y de la fatiga causado por una estación de trabajo óptima no sólo mejora el confort del operador sino que disminuye la probabilidad de error individual (Wickens et al., 2004).

Aquellos radiólogos que trabajan con un solo monitor, al informar estudios comparativos con previos deben, o bien desplegar alternativamente las imágenes de uno y de otro en la pantalla, lo cual podría constituir un factor de interrupción en el análisis visual, o bien ocupar la pantalla del monitor de diagnóstico con ambos estudios simultáneamente, con lo que el tamaño de las imágenes se reduce a la mitad, creando también una fuente potencial de error.

5.9. USO DE SISTEMAS DE RECONOCIMIENTO DE VOZ.

La no utilización de sistemas de reconocimiento de voz se asocia a discrepancia diagnóstica durante todo el análisis inicial (tanto en programados como en urgentes, tanto en el análisis uni como multivariante).

El hecho de que no haya asociación estadísticamente significativa con discrepancia diagnóstica en la submuestra de 431 estudios analizados (261 casos –con discrepancia diagnóstica y 170 controles –sin discrepancia diagnóstica-), aunque el p-valor=0,055 roza dicha significación), se debe a que el sistema de reconocimiento de voz es una característica con una reducida presencia en la población analizada (9,8%), a que en este análisis el tamaño muestral es más reducido y a que en esta submuestra tienen un peso mucho mayor los estudios programados (donde el uso de sistemas de reconocimiento de voz es menos significativo) que los urgentes (donde esta característica tiene más influencia). Podemos desarrollar este argumento comprobando, en la Tabla 23, que de los n=359 estudios programados, existen únicamente 24 realizados con sistema de reconocimiento de voz, con p-valor = 0,022, ya que el porcentaje de pacientes con sistema de reconocimiento de voz es mayor en los que no presentan discrepancias respecto a los que presentan discrepancias (10,9% vs 4,3%). En la Tabla 40 observamos por otra parte que en los n=715 estudios urgentes, existen 30 realizados con sistema de reconocimiento de voz, con p-valor = 0,006, ya que el porcentaje de pacientes con sistema de reconocimiento de voz es mayor en los que no presentan discrepancias respecto a los que presentan discrepancias (6,1% vs 1,6%). Por tanto, en una mezcla de pacientes de ambos tipos, la significación será mayor cuanto mayor sea el número de urgentes y menor el de programados, siempre teniendo en cuenta el número de sujetos incluidos, ya que como el uso de sistema de reconocimiento de voz es tan reducido, con tamaños muestrales pequeños no existe suficiente potencia para detectar diferencias significativas. En la submuestra analizada en el último punto (Tabla 53) hay 72 estudios urgentes y 359 programados, por lo que la proporción de urgentes es reducida (como ocurre en la realidad clínica) y por tanto hay menos efecto del sistema de reconocimiento de voz. Existen 26 estudios con sistema de reconocimiento de voz, con p-valor = 0,055, ya que el porcentaje de pacientes con sistema de reconocimiento de voz es mayor en los que no presentan discrepancias, 8,8% (que está entre 10,9% de los programados y 6,1% de los urgentes) respecto a los que presentan discrepancias, 4,2%

(que está entre 4,3% de los programados y 1,6% de los urgentes) pero las diferencias no son lo suficientemente grandes con este tamaño muestral más reducido como para que salgan estadísticamente significativas (aunque rozan la significación estadística). Hay que considerar también que en este análisis se ha obtenido una muestra, por lo que se incluye el azar en la selección; tal vez podrían haberse extraído muchas muestras de este tamaño y calcular los valores promedio (por ejemplo, obtener 100 submuestras con este procedimiento y obtener el p-valor medio).

El resultado de nuestro estudio en relación a esta variable aparentemente contradice la bibliografía, en la que se refleja que el uso de sistemas de reconocimiento automático de voz puede conducir a una tasa significativamente mayor de error en relación con la transcripción manual (Rana et al., 2005) (McGurk et al., 2008) (Quint et al., 2008) (Strahan & Schneider-Kolsky, 2010) (Basma et al., 2011).

Cuando se comparan los informes realizados mediante sistema de reconocimiento de voz con los realizados por transcripción convencional se observa una tasa de error superior en los primeros (Rana et al., 2005) (McGurk et al., 2008) (Quint et al., 2008) (Strahan & Schneider-Kolsky, 2010) (Basma et al., 2011) (Hawkins et al., 2012) (Minn, Zandieh, & Filice, 2015).

Pero las condiciones de los estudios anteriores no son extrapolables a las de nuestro trabajo, en el que, independientemente de usar o no sistemas de reconocimiento de voz, existen plantillas de texto preformadas para cada uno de los tipos de informe y accesibles para todos los radiólogos. Por ello con frecuencia la redacción de los informes no ocurre únicamente por transcripción directa o por dictado, sino por combinación de uno de los dos métodos con plantillas de texto. Los estudios publicados sobre la tasa de error diagnóstico en sistemas de reconocimiento de voz establecen como criterio de exclusión el uso de plantillas de texto; no hemos encontrado por tanto en la bibliografía científica ningún trabajo que evalúe la asociación del uso del sistema automático de reconocimiento de voz con las discrepancias diagnósticas en condiciones comparables a las de nuestro estudio. En nuestro trabajo el uso de sistema de reconocimiento de voz parece ejercer un efecto protector frente a la discrepancia diagnóstica. Se plantean varias explicaciones posibles para este hecho. Por un lado, las plantillas predefinidas de texto eliminarían la mayor parte de los errores tipográficos y de transcripción tradicionalmente asociados a los sistemas de reconocimiento de voz. Por otra parte, el

uso de los sistemas de reconocimiento de voz permitiría al radiólogo redistribuir más tiempo de trabajo hacia la visualización directa de las imágenes en vez de orientar la mirada hacia el teclado o hacia el documento de texto. Oestmann et al. comprobaron que la detección de lesiones en radiografías de tórax disminuía conforme se reducía el tiempo de visualización (Oestmann et al., 1988), y es probable que este efecto sea extrapolable a las técnicas radiológicas más complejas como la TC y la RM.

5.10. EDAD Y EXPERIENCIA DEL RADIÓLOGO.

La edad y la experiencia del radiólogo sólo muestran asociación estadísticamente significativa con la discrepancia diagnóstica en los análisis univariantes pero no en los multivariantes. Para comprender por qué ocurre esto, hay que considerar la propia naturaleza de los análisis uni y multivariantes, como ya se hizo con anterioridad al analizar la asociación del tipo de visor con discrepancia diagnóstica en los estudios urgentes. Supongamos por ejemplo que tenemos tres variables V_1 , V_2 y V_3 , que explican, desde el punto de vista univariante, el 70%, el 55% y el 65% de las diferencias en la discrepancia diagnóstica respectivamente. Puede ocurrir que en el análisis multivariante la V_1 ya no esté en el modelo final, ya que una parte de lo que explicaba (por ejemplo el 40%) ya lo explica V_2 , y otra parte de lo que explicaba (por ejemplo el 25%) ya lo explica V_3 , por lo que únicamente explica un 5% que no explican V_2 y V_3 , por lo que esto no es significativo y el modelo no la introduce. Es exactamente lo que ocurre con estas dos variables.

Como puede observarse en la Tabla 11, los años de experiencia del radiólogo están relacionados con la presencia de discrepancia diagnóstica para los estudios programados. Se observa un OR estimado de 0.945 (IC95% 0.921-0.969), esto es, disminuye de media un 5,5% la probabilidad de discrepancia diagnóstica por cada año de experiencia de más que tiene el radiólogo. Múltiples estudios apoyan este hallazgo de que una mayor experiencia se asocia a una menor tasa de error diagnóstico. Eng et al. compararon las lecturas de radiografías simples realizadas por médicos adjuntos y médicos residentes de Radiodiagnóstico y de Medicina de Urgencias. Encontraron que el rendimiento diagnóstico de los adjuntos de Radiodiagnóstico era superior al de los residentes de Radiodiagnóstico, el de los residentes de Radiodiagnóstico era superior al de los adjuntos de Medicina de Urgencias y el de estos últimos superior al de los residentes de Medicina de Urgencias (Eng et al., 2000). En otro estudio similar de pruebas radiológicas interpretadas por médicos del Servicio de Urgencias porque no había disponibilidad de radiólogos, hasta el 16% de las radiografías y el 35% de las TC craneales estaban incorrectamente diagnosticados (Neale et al., 2001). Alfaro et al. observaron tasas de discrepancia incluso superiores, del 39%, entre las interpretaciones que realizaban de estudios de TC craneal los neurorradiólogos respecto a las que hacían los médicos del Servicio de Urgencias (Alfaro et al., 1995).

No obstante, en el modelo multivariante la experiencia del radiólogo no permanece como una variable significativa, debido a que una gran parte de la variabilidad de la discrepancia diagnóstica explicada por esta variable es explicada por algunas de las otras variables incluidas. En concreto, existe una relación muy grande entre la concordancia estudio-especialidad radiólogo y los años de experiencia del radiólogo, con una media de 16,7 años de experiencia para aquellos estudios con una concordancia alta y de sólo 2,5 años para los estudios de concordancia media / baja. Al introducir esta variable de concordancia en el modelo, la significatividad de los años de experiencia desaparece en el modelo multivariante.

5.11. GÉNERO DEL RADIÓLOGO.

Al igual que ocurre con la edad y la experiencia del radiólogo, el género del radiólogo sólo se asocia a discrepancia diagnóstica en el análisis univariante y no en el multivariante. Esto es porque las diferencias en el género del radiólogo realmente son debidas a diferencias en otras variables más relevantes. Por ejemplo, para los estudios programados se observa que al entrar en el modelo el número de monitores, pierde toda significación matemática el sexo. Se debe a que existe una relación importante entre el sexo y el número de monitores, ya que mientras en el 23,6% de los estudios programados realizados por radiólogos varones se han utilizado dos o más monitores, únicamente ocurre esto en el 10,6% de los estudios realizados por radiólogas mujeres, por lo que los radiólogos varones utilizan más de un monitor en más ocasiones. Existen otras variables sexo-dependientes que también podrían capturar parte de la información explicada por el sexo y por tanto reducir su impacto en el modelo multivariante. Hay que tener en cuenta que para las características relacionadas con el radiólogo existe muy poca variabilidad, ya que únicamente hay estudios asignados a 29 radiólogos.

5.12. HORARIO DE LA PRIMERA LECTURA. FATIGA LABORAL.

El horario de la primera lectura presenta asociación estadísticamente significativa con la discrepancia diagnóstica en el análisis multivariante de los estudios programados (mayor riesgo de discrepancia diagnóstica en aquellos estudios informados fuera de horario laboral 8:00-17:59 h), aunque no en el univariante de estos mismos estudios. Esto es porque el horario de la primera lectura explica una parte reducida de la discrepancia diagnóstica (no significativa en el univariante) pero esa pequeña parte no interseca con las otras variables en el estudio multivariante; es decir, mantiene estable esa pequeña parte de asociación con la discrepancia diagnóstica en el estudio multivariante. Dicho de otro modo, es una variable que explica algo genuino que no consiguen las otras variables del estudio multivariante. Este resultado de nuestro estudio refleja un paralelismo entre la fatiga del radiólogo y el error diagnóstico, es decir, aquellos estudios informados fuera del horario laboral ordinario (lunes a viernes de 8:00 a 17:59 h) muestran un riesgo aumentado de error, posiblemente debido a un mayor cansancio o menor nivel de alerta del radiólogo. Nuestro trabajo refuerza por tanto la evidencia científica de que el esfuerzo y la fatiga visual de los radiólogos afectan negativamente a la productividad y la precisión diagnóstica al favorecer los errores de percepción y cognitivos (Krupinski, 2003). El aumento del tiempo que los radiólogos dedican a ver imágenes produce un mayor esfuerzo en el sistema óculomotor, conduciendo a fatiga visual (astenopia), que se manifiesta como dificultad para enfocar y visión borrosa (Ebenholtz, 2001) (MacKenzie, 1843). Estos síntomas son más acentuados al final de una larga jornada de trabajo con una consiguiente menor precisión para detectar hallazgos (Krupinski & Berbaum, 2009) (Brogdon et al., 1978).

Por el contrario no se observa asociación estadísticamente significativa entre el horario de la primera lectura y la discrepancia diagnóstica en los estudios urgentes. Una posible explicación sería que los turnos de guardia en que trabajan los radiólogos en la empresa IBERORAD S.L. suelen ser reducidos, prácticamente siempre en horario diurno, con un flujo de trabajo ininterrumpido y un nivel de alerta elevado, a diferencia de los turnos de guardia habituales en el Sistema Nacional de Salud que suelen oscilar entre 18 y 24 horas, potencialmente con mayor propensión a fatiga laboral o a un menor nivel de alerta, especialmente durante la noche.

Existen limitaciones en el diseño de nuestro estudio para evaluar la asociación de la fatiga del radiólogo con la discrepancia diagnóstica. Se sabe por ejemplo que las tasas de error al informar TC de abdomen son más del doble cuando un radiólogo informa más de 20 estudios al día (Bechtold et al., 1997) (Fitzgerald, 2001). Pero en nuestro estudio no se ha podido realizar una medida directa de la asociación entre la carga de informes diaria y las discrepancias diagnósticas, ya que varios radiólogos trabajan a tiempo parcial en la empresa, desempeñando también actividad en otros centros que no podemos cuantificar. Por ello la única variable en base a la cual podemos estimar dicha asociación es el horario en que se realizaron las primeras lecturas de los estudios, teniendo en cuenta además que la densidad de estudios por hora tiene un pico máximo en torno a las 18:00 h conforme se muestra en la Figuras 26 y 27. Otro elemento que no analizamos en nuestro estudio son los efectos de la privación de sueño por trabajo en horario nocturno, una situación frecuente en muchos entornos de trabajo, especialmente en nuestro Sistema Nacional de Salud. Por ejemplo, se ha llegado a cuantificar una reducción en el rendimiento de los especialistas en cirugía tras una jornada de guardia (Grantcharov et al., 2001). Pero esta no es una situación que pueda evaluarse en la empresa IBERORAD S.L. debido a que, por la dispersión geográfica de los radiólogos, todos asumen la actividad tanto programada como urgente mayoritariamente en horario diurno (Figura 26).

En cualquier caso, hay que considerar que las técnicas de imagen más actuales (TC y RM) adquieren cada vez un mayor número de imágenes que deben ser interpretadas cada vez en menos tiempo (Bhargavan & Sunshine, 2005) (Meghea & Sunshine, 2007) (Lu Y. et al., 2008) (Nakajima et al., 2008) (Sunshine & Maynard, 2008). Es preocupante que la carga de trabajo de los radiólogos pueda ser tan exigente que la fatiga y el reducido tiempo disponible para la interpretación estén impactando negativamente en la precisión diagnóstica (MacKenzie, 1843) (Ebenholtz, 2001) (Goo et al., 2004) (Vertinsky & Forster, 2005) (Krupinski & Kallergi, 2007). Este aumento en la carga de trabajo y en las expectativas en torno a la calidad de su trabajo conducen con frecuencia a errores diagnósticos (Reiner & Krupinski, 2012). Un estudio nacional sobre la carga de trabajo de radiólogos titulados en Irlanda ha demostrado que, en 2009, los radiólogos irlandeses de media asumían un 128% de la carga de trabajo considerada como apropiada en Australia (Brady, 2011). Sorprendentemente, no existen límites legales fijados para la actividad diaria que cada radiólogo puede asumir con garantías razonables de seguridad, al contrario que en otras profesiones con riesgo potencial de ocasionar eventos adversos.

Los controladores aéreos del Reino Unido tienen una carga laboral semanal de 32 horas, descansan durante 30 minutos cada 90 minutos de trabajo, trabajan por parejas, y tienen indicaciones estrictas sobre el número de aviones que permiten en su sector [J. Rhodes (controlador de tráfico aéreo, Manchester), comunicación personal]. En Radiodiagnóstico tan sólo se han publicado guías orientativas sobre el volumen de cada técnica de diagnóstico que puede asumir un radiólogo en su jornada laboral para conservar parámetros de seguridad (*Workload and Manpower in Clinical Radiology*, 1999). También se han descrito las implicaciones medicolegales de informar a un ritmo muy acelerado. Al menos en una ocasión, un radiólogo en los EE.UU. fue demandado por mala praxis debido a un cáncer de mama no diagnosticado en una mamografía, porque dicho radiólogo “leyó demasiadas radiografías el día en cuestión, demostrando una inaceptable falta de respeto al bienestar del paciente al sacrificar la calidad de la asistencia por el volumen de trabajo a fin de maximizar los ingresos” (Berlin, 2000b).

Los servicios de Radiología deben por tanto tener mecanismos para responder de forma dinámica al aumento de la carga de trabajo, incluyendo cómo distribuir adecuadamente los recursos y el personal donde sea necesario. Los volúmenes de trabajo deben también evaluarse de forma periódica para que las tendencias longitudinales puedan identificarse y el aumento en la carga de trabajo pueda ser anticipado (Ryan et al., 2018).

5.13. ESPECIALIDAD DEL ESTUDIO. ESPECIALIDAD DEL RADIÓLOGO. CONCORDANCIA ENTRE AMBAS.

La especialidad del estudio muestra una asociación con la discrepancia diagnóstica variable en función de si los estudios son urgentes o programados. En el caso de los estudios urgentes (y también en el análisis conjunto de la totalidad de los estudios, con sobrerrepresentación de los estudios urgentes) se observa un aumento del riesgo de discrepancia diagnóstica en los estudios de Body y, en menor medida, de MSK/columna respecto a los de Neuro. Esto se entiende considerando que los estudios urgentes de Neuro son básicamente TC craneales, con un campo de visión relativamente reducido, pocas imágenes, a menudo sin reconstrucciones multiplanares y con un abanico relativamente reducido de alteraciones a identificar (fracturas, isquemia, hemorragia, lesiones ocupantes de espacio-efecto masa, hidrocefalia) en comparación con las posibles alteraciones en un estudio de Body, que incluyen campos de visión mayores, con muchas más imágenes, que implican a múltiples áreas de conocimiento (patología torácica, cardiovascular, gastrointestinal, urológica, ginecológica e incluso de columna y de musculoesquelético al hallarse estas últimas estructuras incluidas en el campo del estudio) y que con cierta frecuencia incluyen evaluación de pacientes postquirúrgicos, con gran complejidad diagnóstica. Por su parte, los estudios urgentes de MSK/columna, principalmente encuadrados en pacientes con traumatismos, conllevan un riesgo apreciable de discrepancias diagnósticas en comparación con los estudios de Neuro debido al mayor número de imágenes, reconstrucciones multiplanares... y a la presencia de estructuras anatómicas de menor tamaño (por ejemplo, huesos del carpo, apófisis vertebrales...) cuya lesión puede pasar fácilmente inadvertida.

Por el contrario, en los estudios programados la especialidad del estudio se asocia a discrepancia diagnóstica únicamente en el análisis univariante, no en el multivariante. Si se atiende al resultado del análisis univariante podrá apreciarse que los estudios con menor asociación a discrepancia diagnóstica son los de columna (habitualmente RM de columna por espondiloartrosis, con poca variabilidad en los diagnósticos y un nivel de complejidad cognitiva bajo). Respecto a los estudios de columna, existe un incremento de la probabilidad de discrepancia diagnóstica de 2,2 veces en un estudio de MSK y un incremento de 5,0 veces de un estudio de Body. Las posibles razones de este aumento del riesgo de discrepancia son múltiples. Por una parte la mayor variabilidad y complejidad

en términos generales de la patología en los estudios programados de MSK y Body respecto a los de Columna, y por otros factores específicos de cada una de las especialidades de estudios. En el caso de los estudios de MSK interviene el sesgo de selección de casos con discrepancia diagnóstica ya comentado con anterioridad; en este tipo de estudio no es infrecuente que los traumatólogos soliciten a demanda revisiones de los diagnósticos iniciales, para hacerlos más coherentes con información clínica de la que disponen pero que no fue aportada en la solicitud del estudio e incluso información adicional posterior al estudio radiológico (por ejemplo endoscopias, pruebas de laboratorio, estudios de imagen adicionales...). También se solicitan con frecuencia revisiones de los estudios de MSK debido a que una parte importante de los mismos son informes de mutuas sanitarias, que revisten complejidad por el frecuente fraude y simulación por parte de los pacientes; los inspectores de trabajo consideran que los estudios radiológicos son pruebas objetivas y en base a los informes de los mismos deciden las compensaciones económicas y el tipo de incapacidad (accidente de trabajo-enfermedad profesional o accidente no laboral-enfermedad común), por lo que es importante evitar el sobrediagnóstico. Algunos ejemplos de rectificaciones en los informes que se solicitan con cierta frecuencia son: eliminar el diagnóstico de “contusión” o “fractura” en un paciente con edema óseo pero sin traumatismo (es un diagnóstico incorrecto pues el término contusión significa que existe un golpe o impacto), eliminar el diagnóstico de “fractura de estrés” en un paciente atropellado por un coche o que recibió un martillazo (las fracturas de estrés son debidas a sobreuso o traumatismo repetitivo, por ejemplo en una maratón), eliminar de los informes el término “secuela” (que se considera como daño residual al traumatismo, y que no debe usarse para describir una normal evolución de la lesión), y eliminar los términos “agudo” o “crónico” cuando no existen datos concluyentes que orienten hacia uno u otro. Todas estas rectificaciones a demanda podrían aumentar falsamente la asociación de los estudios programados de MSK con la discrepancia diagnóstica. Los estudios programados de Body tienen sus características propias que los predisponen a discrepancia diagnóstica y que ya se han mencionado con anterioridad en la discusión: campos de estudio amplios con múltiples estructuras anatómicas, gran cantidad de imágenes y de series/fases de contraste, un alto porcentaje de estudios oncológicos muy exigentes en términos de tiempo de informe...etc.

La explicación de que no se observe asociación con la discrepancia diagnóstica en el análisis multivariante de los estudios programados es que de estos la gran mayoría son

de MSK (73,3%), con muchos menos de Columna (16,7%) y muy pocos de Neuro (5,3%) o Body (4,7%). Esto a su vez condiciona que la precisión de las estimaciones de los OR asociados a cada especialidad respecto de la de referencia (la de Columna) sea muy baja, y por tanto más difícil obtener significación estadística de una de ellas respecto de otra. En cambio la distribución por especialidades de los estudios urgentes muestra mucha menor concentración de estudios en cada especialidad. En el modelo multivariante de los estudios programados, la concordancia entre la especialidad del radiólogo y la especialidad del estudio captura la significación de esta segunda variable, y por ello la especialidad del estudio no acaba mostrando asociación estadísticamente significativa con la discrepancia diagnóstica.

En cuanto a la especialidad del radiólogo, nuestro estudio constata que tanto al analizarse de forma independiente como a través de la concordancia con la especialidad del estudio, un factor de riesgo para discrepancia diagnóstica es que el radiólogo sea un “radiólogo general”, sin dedicación preferente a ningún aparato-sistema. Este resultado se suma a la abundante evidencia científica de que los radiólogos expertos en un área de conocimiento tienen mejor rendimiento diagnóstico que los radiólogos generales (Robinson, 1997) (Halligan, 2002). Por poner algunos ejemplos, en la población adulta, los informes de estudios de imagen realizados en centros especializados de referencia muestran una tasa de discrepancia que oscila entre el 12% y el 41% respecto a los informes de esos mismos estudios realizados en centros de atención primaria (Kalbhen et al., 1998) (Gollub et al., 1999) (Loevner et al., 2002) (Tilleman et al., 2003) (Sung et al., 2009). Un análisis que comparaba segundas lecturas de estudios de TC y RM neurológicos realizadas por neuroradiólogos subespecialistas respecto a los informes iniciales reveló unas tasas de error significativo del 13% y no significativo del 21% (Briggs et al. 2008). En estudios que evalúan revisiones por radiólogos oncológicos de TC y RM informados por radiólogos generales, se cambió la estadificación o se observaron desacuerdos importantes en el diagnóstico en el 12-19% de los casos, con necesidad de modificar el tratamiento en el 3-9% (Gollub et al., 1999) (Loughrey et al., 1999) (Tilleman et al., 2003). Por último, los estudios de RM endorrectal de próstata informados por radiólogos expertos presentan menor variabilidad interobservador y son más precisos que las evaluaciones realizadas por radiólogos no experimentados (Tsuda et al., 1999).

Los radiólogos más especializados cometen menos errores tanto cognitivos como de percepción. La menor tasa de errores cognitivos se explica por tener una mejor

comprensión del proceso patológico, que ayuda al radiólogo a formarse una impresión más precisa en lugar de tener que dar un diagnóstico ambiguo (Kabadi & Krishnaraj, 2017). La menor tasa de errores de percepción se debe por una parte a que en términos de búsqueda visual los radiólogos más experimentados tienden a detectar lesiones en menor tiempo y necesitan también menos tiempo para decidir si un hallazgo constituye o no una lesión (Nodine & Mello-Thomas, 2010), y por otra al conocimiento de asociaciones lesionales dentro de un mismo proceso patológico (por ejemplo, fracturas de calcáneo con fracturas de columna vertebral, pancreatitis aguda y trombosis de la vena esplénica, tumores pineales y metástasis meníngeas...etc) que deben investigarse de forma activa (Nodine & Mello-Thomas, 2010) (Kabadi & Krishnaraj, 2017).

Podrían sin embargo existir algunos sesgos al considerar que existe una menor tasa de discrepancia diagnóstica en aquellos estudios informados por radiólogos especialistas. Se ha establecido que la mayor corrección diagnóstica de los oncorradiólogos especialistas respecto a los radiólogos generales podría deberse a información clínica adicional o a informar en un contexto diferente (por ejemplo conocimiento de un diagnóstico ya establecido de cáncer) (FitzGerald & Mehra, 2000). En este sentido sí que es más probable que a los radiólogos especialistas se les asignen estudios mejor orientados desde el punto de vista clínico, que con cierta frecuencia tienen patologías de base ya diagnosticadas. Y aunque no es el caso de nuestro trabajo donde los estudios se informan por vía telemática, también ha de considerarse que los radiólogos especialistas forman parte con mucha más frecuencia de equipos de trabajo multidisciplinarios, que les aportan un feedback clínico y anatomopatológico constante. Esta forma de trabajar tiene un reconocimiento creciente como el estándar oro para cualquier disciplina médica (Briggs et al., 2008).

La abundancia y distribución de la patología en la muestra de casos analizada también puede sesgar las tasas de error/discrepancia. Erly et al. encontraron que la tasa de discrepancia significativa entre los radiólogos generales y los neurorradiólogos en estudios de TC craneal urgentes era sólo del 2% en una gran serie de casos. Sin embargo, de los 716 casos sólo había enfermedad significativa/activa en 47, y en este último subgrupo la tasa de discrepancia fue muy superior, del 23% (Erly et al., 2003).

Otro factor a considerar al analizar la magnitud de las discrepancias diagnósticas es la especialidad del estudio. Mientras que la tasa de discrepancia oscila entre el 1,3% y el

34% cuando se comparan radiólogos generales y neurorradiólogos informando estudios de Neurorradiología (Jordan, Jordan, Lightfoote, & Ragland, 2012) (Zan et al., 2010) (Briggs et al., 2008), en Radiología Pediátrica la tasa de discrepancia entre radiólogos generales y radiólogos especialistas alcanza el 42%, siendo estas discrepancias significativas en el 22% de los casos (Eakins et al., 2012).

En nuestro estudio también se observa una tasa de discrepancia diagnóstica superior en los radiólogos generales que en los radiólogos especialistas que informan estudios fuera de sus áreas de conocimiento específico. Una posible explicación sería que el contacto más frecuente con estudios de todas las especialidades diagnósticas hace a los radiólogos generales tener una mayor confianza en su diagnóstico y por tanto un menor nivel de alerta. El exceso de confianza se ha descrito como un contribuyente importante al error diagnóstico (Berner & Graber, 2008). Aun así debe mencionarse que en los estudios programados hay pocos con concordancia baja, debido a la práctica en IBERORAD S.L. de asignar siempre dentro de lo posible cada tipo de estudio a radiólogos con experiencia en cada una de las áreas de conocimiento.

5.14. TIPO DE ESTUDIO (URGENTE/PROGRAMADO).

En el análisis de la submuestra se detecta una asociación estadísticamente significativa entre el tipo de estudio (urgente/programado) y la discrepancia diagnóstica, siendo la probabilidad de discrepancia mayor en los estudios programados. La posible explicación ya se ha planteado al inicio de esta discusión: en los estudios programados no podemos analizar si la revisión fue por selección estrictamente aleatoria en el contexto del control de calidad ordinario de la empresa o si fue causada por una solicitud expresa del médico receptor tras leer el informe. Ello introduce un sesgo en los porcentajes del tipo de errores detectados (como ya se comentó con anterioridad, puede generar un aumento en el porcentaje de errores de comunicación en detrimento de los errores de infralectura) (Brigham et al., 2015). Es por ello probable que la tasa de error de los estudios programados encontrada en nuestro análisis esté falsamente magnificada debido a un sesgo de selección.

Podemos vernos tentados, tras todo nuestro análisis de variables explicativas del error, a pensar que puede hacerse una distinción entre el error humano y el fallo técnico, y que un error debe ser causado por uno o por otro, sin darnos cuenta de que esto es una construcción teórica sobresimplificada. Pero como indicó Dekker en su libro *“The Field Guide to Understanding ‘Human Error’”*, cuando se asume la complejidad de un error, uno se da cuenta de que la distinción entre los contribuyentes técnicos y humanos es muy imprecisa. Pensamos que hay algo que constituye “la causa” de un error, y que si investigamos en suficiente profundidad la encontraremos. Pero en realidad lo que se denomina “causa raíz” es simplemente el lugar en el que dejamos de buscar en mayor profundidad. Es más acertado por tanto hablar de “explicaciones” que de “causas”. (Dekker, 2017).

6. LIMITACIONES DEL ESTUDIO.

La referencia para considerar las discrepancias diagnósticas no es el seguimiento del paciente ni el resultado quirúrgico o de Anatomía Patológica, sino una segunda lectura realizada por un radiólogo experto y por lo tanto puede no ser 100% exacta.

Identificar los errores en tiempo real con frecuencia es complicado. A veces tienen que transcurrir semanas o meses de estudios de seguimiento para que se haga evidente el diagnóstico correcto (Shojania et al., 2003).

En los estudios programados no podemos analizar si la revisión fue por selección estrictamente aleatoria en el contexto del control de calidad ordinario de la empresa o si fue causada por una solicitud expresa del médico receptor tras leer el informe.

No existe un “ciego” real para los radiólogos revisores de los casos. El radiólogo revisor conoce el nombre del autor del diagnóstico inicial y el contenido del informe.

Con frecuencia la revisión de estudios de imagen previos sólo se registra electrónicamente cuando existen discrepancias clínicamente significativas, lo cual conduce al infrarregistro de las categorías 1 (no discrepancia) y A (clínicamente no significativas).

Existe un sesgo cuando una institución (pública o privada) audita su propio trabajo, ya que su prestigio depende de una baja tasa de discrepancia. Quizá la revisión de casos debería ser realizada de forma anónima por varios radiólogos externos a la empresa, los cuales a su vez evaluarían una mezcla anonimizada de casos procedentes de varias instituciones (Strickland, 2015).

7. CONCLUSIONES.

A partir de los resultados obtenidos y como respuesta a los objetivos planteados en este trabajo establecemos las siguientes conclusiones:

1. Para el total de estudios existe una tasa de discrepancia diagnóstica del 3,0%, que desciende al 1,4% en los estudios urgentes y se sitúa en el 3,6% en los estudios programados.
2. La tasa media de estudios con discrepancias clínicamente significativas se sitúa en el 23,8% del total de discrepancias. Los estudios programados presentan un mayor porcentaje de discrepancias con repercusiones significativas pero estas diferencias no son estadísticamente significativas.
3. En el análisis multivariante de los estudios programados se observa asociación estadísticamente significativa con la discrepancia diagnóstica, en orden decreciente de magnitud, de las siguientes variables: deficiente calidad técnica del estudio, estudio de TC frente a estudio de RM, disponibilidad de estudios de imagen previos, concordancia entre la especialidad del radiólogo y la especialidad del estudio media o baja, no utilización de sistemas de reconocimiento de voz, uso de un solo monitor de visualización de las imágenes y estudio informado fuera del horario laboral de un día laborable.
4. El análisis multivariante de los estudios urgentes muestra asociación estadísticamente significativa con la discrepancia diagnóstica, en orden decreciente de magnitud, de las siguientes variables: disponibilidad de estudios de imagen previos, concordancia entre la especialidad del radiólogo y del estudio media, información clínica insuficiente, no utilización de sistemas de reconocimiento de voz, concordancia entre la especialidad del radiólogo y del estudio baja, estudio de complejidad alta, estudio de Body, uso de un solo monitor de visualización de las imágenes y estudio de MSK/Columna.
5. En el análisis multivariante de la submuestra de todos los estudios analizados existe asociación estadísticamente significativa con la discrepancia diagnóstica,

en orden decreciente de magnitud, de las siguientes variables: deficiente calidad técnica del estudio, disponibilidad de estudios de imagen previos, estudio de Body, estudio programado, concordancia entre la especialidad del radiólogo y del estudio media o baja, uso de un solo monitor de visualización de las imágenes, información clínica insuficiente, estudio de MSK y estudio de Neuro.

Finalmente y como respuesta a la hipótesis inicialmente planteada, concluimos que existen múltiples factores asociados al error diagnóstico en Radiología, dependientes tanto de los estudios de imagen como del propio radiólogo, y que la intensidad en que cada uno de estos factores se asocia al error diagnóstico es distinta en cada factor y variable en función del resto de características presentes en el proceso diagnóstico.

8. BIBLIOGRAFÍA.

Abujudeh HH, Boland GW, Kaewlai R, Rabiner P, Halpern EF, Gazelle GS, et al. Abdominal and pelvic computed tomography (CT) interpretation: discrepancy rates among experienced radiologists. *Eur Radiol* 2010 August 01;20(8):1952-1957.

Abujudeh HH, Kaewlai R, Shaqdan K, Bruno MA. Key Principles in Quality and Safety in Radiology. *AJR Am J Roentgenol* 2017 Mar;208(3):W109.

Alexander K. Reducing error in radiographic interpretation. *Can Vet J* 2010 May;51(5):533-536.

Alfaro D, Levitt MA, English DK, Williams V, Eisenberg R. Accuracy of interpretation of cranial computed tomography scans in an emergency medicine residency program. *Ann Emerg Med* 1995 February 01;25(2):169-174.

Allen B, Jr. The value of radiologists in reducing diagnostic error in health care. *J Am Coll Radiol* 2014 Dec;11(12 Pt A):1101-1102.

Alpert HR, Hillman BJ. Quality and variability in diagnostic radiology. *J Am Coll Radiol* 2004 Feb;1(2):127-132.

Anderson RE, Hill RB, Key CR. The sensitivity and specificity of clinical diagnostics during five decades. Toward an understanding of necessary fallibility. *JAMA* 1989 March 17;261(11):1610-1617.

Armato SG, Roberts RY, Kocherginsky M, Aberle DR, Kazerooni EA, Macmahon H, et al. Assessment of radiologist performance in the detection of lung nodules: dependence on the definition of "truth". *Acad Radiol* 2009 January 01;16(1):28-38.

Ashman CJ, Yu JS, Wolfman D. Satisfaction of search in osteoradiology. *AJR Am J Roentgenol* 2000 August 01;175(2):541-544.

Baines CJ, McFarlane DV, Miller AB. The role of the reference radiologist. Estimates of inter-observer agreement and potential delay in cancer detection in the national breast screening study. *Invest Radiol* 1990 September 01;25(9):971-976.

Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *CMAJ* 2004 May 25;170(11):1678-1686.

Ballard-Barbash R, Taplin SH, Yankaskas BC, Ernster VL, Rosenberg RD, Carney PA, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *AJR Am J Roentgenol* 1997 October 01;169(4):1001-1008.

Balsamo RR, Brown MD. Risk management. In: Sanbar SS, Gibofsky A, Firestone MH, LeBlang TR, editors. *Legal Medicine* St Louis: Mosby; 1998. p. 223-244.

Basma S, Lord B, Jacks LM, Rizk M, Scaranelo AM. Error rates in breast imaging reports: comparison of automatic speech recognition and dictation transcription. *AJR Am J Roentgenol* 2011 October 01;197(4):923-927.

Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists. Findings from a national sample. *Arch Intern Med* 1996 January 22;156(2):209-213.

Bechtold RE, Chen MY, Ott DJ, Zagoria RJ, Scharling ES, Wolfman NT, et al. Interpretation of abdominal CT: analysis of errors and their causes. *J Comput Assist Tomogr* 1997 October 01;21(5):681-685.

Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR Am J Roentgenol* 2000 June 01;174(6):1769-1777.

Berlanstein B. Patient Safety. In: Abujudeh H, Bruno MA, editors. *Radiology Noninterpretive Skills: The Requisites* Pennsylvania: Elsevier; 2018. p. 25-32.

Berlin L. Does the "missed" radiographic diagnosis constitute malpractice? *Radiology* 1977 May 01;123(2):523-527.

Berlin L. Malpractice and radiologists, update 1986: an 11.5-year perspective. *AJR Am J Roentgenol* 1986 December 01;147(6):1291-1298.

Berlin L. Reporting the "missed" radiologic diagnosis: medicolegal and ethical considerations. *Radiology* 1994 July 01;192(1):183-187.

Berlin L, Berlin JW. Malpractice and radiologists in Cook County, IL: trends in 20 years of litigation. *AJR Am J Roentgenol* 1995 October 01;165(4):781-788.

Berlin L. Errors in judgment. *AJR Am J Roentgenol* 1996a June 01;166(6):1259-1261.

Berlin L. Perceptual errors. *AJR* 1996b;167:587-590.

Berlin L. Comparing new radiographs with those obtained previously. *AJR Am J Roentgenol* 1999a Jan;172(1):3-6.

Berlin L. The missed breast cancer: perceptions and realities. *AJR Am J Roentgenol* 1999b November 01;173(5):1161-1167.

Berlin L. Malpractice issues in radiology. Alliterative errors. *AJR Am J Roentgenol* 2000a April 01;174(4):925-931.

Berlin L. Liability of interpreting too many radiographs. *AJR Am J Roentgenol* 2000b Jul;175(1):17-22.

Berlin L. Hindsight bias. *AJR Am J Roentgenol* 2000c September 01;175(3):597-601.

Berlin L. Defending the "missed" radiographic diagnosis. *AJR Am J Roentgenol* 2001 February 01;176(2):317-322.

Berlin L. Fear of cancer. *AJR Am J Roentgenol* 2004a August 01;183(2):267-272.

Berlin L. Outcome bias. *AJR. American journal of roentgenology* 2004b Sep;183(3):557.

Berlin L. Accuracy of diagnostic procedures: has it improved over the past five decades? *AJR Am J Roentgenol* 2007a May;188(5):1173-1178.

Berlin L. Radiologic errors and malpractice: a blurry distinction. *AJR Am J Roentgenol* 2007b Sep;189(3):517-522.

Berlin L. Radiologic errors, past, present and future. *Diagnosis* 2014 Jan 1,;1(1):79-84.

Berlin L. Malpractice and Radiology : A Hapless Relationship. In: Abujudeh H, Bruno MA, editors. *Radiology Noninterpretive Skills: The Requisites* Pennsylvania: Elsevier; 2018. p. 256-266.

Berner ES, Miller RA, Graber ML. Missed and delayed diagnoses in the ambulatory setting. *Ann Intern Med* 2007 March 20;146(6):1.

- Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008 May;121(5 Suppl):2.
- Bhargavan M, Sunshine JH. Utilization of radiology services in the United States: levels and trends in modalities, regions, and populations. *Radiology* 2005 March 01;234(3):824-832.
- Bhimani C, Bruno MA. Error in Radiology. In: Abujudeh H, Bruno MA, editors. *Radiology Noninterpretive Skills The Requisites* Pennsylvania: Elsevier; 2018. p. 105-109.
- Birkelo CC, Chamberlain WE. Tuberculosis case finding; a comparison of the effectiveness of various roentgenographic and photofluorographic methods. *J Am Med Assoc* 1947 February 08;133(6):359-366.
- Blendon RJ, DesRoches CM, Brodie M, Benson JM, Rosen AB, Schneider E, et al. Views of practicing physicians and the public on medical errors. *N Engl J Med* 2002 December 12;347(24):1933-1940.
- Borgstede JP, Lewis RS, Bhargavan M, Sunshine JH. RADPEER quality assurance program: a multifacility study of interpretive disagreement rates. *J Am Coll Radiol* 2004 Jan;1(1):59-65.
- Bosniak MA. Diagnosis and management of patients with complicated cystic lesions of the kidney. *AJR Am J Roentgenol* 1997 September 01;169(3):819-821.
- Brady A. Measuring Consultant Radiologist workload: method and results from a national survey. *Insights Imaging* 2011 Jun;2(3):247-260.
- Brady A, Laoide RO, McCarthy P, McDermott R. Discrepancy and error in radiology: concepts, causes and consequences. *Ulster Med J* 2012 Jan;81(1):3-9.
- Brady AP. Error and discrepancy in radiology: inevitable or avoidable? *Insights Imaging* 2017 Feb;8(1):171-182.
- Brenner RJ, Lucey LL, Smith JJ, Saunders R. Radiology and medical malpractice claims: a report on the practice standards claims survey of the Physician Insurers Association of America and the American College of Radiology. *AJR Am J Roentgenol* 1998 July 01;171(1):19-22.
- Briggs GM, Flynn PA, Worthington M, Rennie I, McKinstry CS. The role of specialist neuroradiology second opinion reporting: is there added value? *Clin Radiol* 2008 Jul;63(7):791-795.
- Brigham LR, Mansouri M, Abujudeh HH. JOURNAL CLUB: Radiology Report Addenda: A Self-Report Approach to Error Identification, Quantification, and Classification. *AJR Am J Roentgenol* 2015 Dec;205(6):1230-1239.
- Brogdon BG, Kelsey CA, Moseley RD. Effect of fatigue and alcohol on observer perception. *AJR Am J Roentgenol* 1978 May 01;130(5):971-974.
- Brook OR, O'Connell AM, Thornton E, Eisenberg RL, Mendiratta-Lala M, Kruskal JB. Quality initiatives: anatomy and pathophysiology of errors occurring in clinical radiology practice. *Radiographics* 2010 Sep;30(5):1401-1410.
- Brown JH, Lustrin ES, Lev MH, Ogilvy CS, Taveras JM. Reduction of aneurysm clip artifacts on CT angiograms: a technical note. *AJNR Am J Neuroradiol* 1999 Apr;20(4):694-696.
- Bruno MA, Walker EA, Abujudeh HH. Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. *Radiographics* 2015 Oct;35(6):1668-1676.

Buchanan GN, Halligan S, Taylor S, Williams A, Cohen R, Bartram C. MRI of fistula in ano: inter- and intraobserver agreement and effects of directed education. *AJR Am J Roentgenol* 2004 Jul;183(1):135-140.

Burroughs TE, Waterman AD, Gallagher TH, Waterman B, Adams D, Jeffe DB, et al. Patient concerns about medical errors in emergency departments. *Acad Emerg Med* 2005 January 01;12(1):57-64.

Cascade PN. Quality improvement in diagnostic radiology. *AJR Am J Roentgenol* 1990 May 01;154(5):1117-1120.

Chang CA, Strahan R, Jolley D. Non-clinical errors using voice recognition dictation software for radiology reports: a retrospective audit. *J Digit Imaging* 2011 August 01;24(4):724-728.

Chellis M, Olson J, Augustine J, Hamilton G. Evaluation of missed diagnoses for patients admitted from the emergency department. *Acad Emerg Med* 2001 February 01;8(2):125-130.

Committee on Diagnostic Error in Health Care. *Improving Diagnosis in Health Care*. Washington, D.C: National Academies Press; 2015.

Conklin TA. *A story of failure. Pre-accident investigations*. Farnham, Surrey, England: Ashgate; 2012. p. 1-16.

Cooper JB, Newbower RS, Long CD, McPeck B. Preventable anesthesia mishaps: a study of human factors. *Anesthesiology* 1978 December 01;49(6):399-406.

Cosby KS, Croskerry P. Profiles in patient safety: authority gradients in medical error. *Academic emergency medicine: official journal of the Society for Academic Emergency Medicine* 2004 Dec;11(12):1341-1345.

Croskerry P. Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Acad Emerg Med* 2002 November 01;9(11):1184-1204.

Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 2003 August 01;78(8):775-780.

Croskerry P. Perspectives on diagnostic failure and patient safety. *Healthc Q* 2012;15 Spec No:50-56.

Dalla Palma L, Stacul F, Meduri S, Geitung JT. Relationships between radiologists and clinicians: results from three surveys. *Clin Radiol* 2000 August 01;55(8):602-605.

Davis P, Lay-Yee R, Briant R, Ali W, Scott A, Schug S. Adverse events in New Zealand public hospitals II: preventability and clinical context. *N Z Med J* 2003 October 10;116(1183):U624.

Dekker S. *You can't count errors. The Field Guide To Understanding Human Errors* Surrey, England: Ashgate; 2006. p. 65-72.

Dekker S. *The Field Guide to Understanding 'Human Error'*. 3rd ed. Farnham: CRC Press; 2017.

Donald JJ, Barnard SA. Common patterns in 558 diagnostic radiology errors. *J Med Imaging Radiat Oncol* 2012 Apr;56(2):173-178.

Donnelly LF, Strife JL. Performance-based assessment of radiology faculty: a practical plan to promote improvement and meet JCAHO standards. *AJR Am J Roentgenol* 2005 May 01;184(5):1398-1401.

- Doubilet P, Herman PG. Interpretation of radiographs: effect of clinical history. *AJR Am J Roentgenol* 1981 November 01;137(5):1055-1058.
- Drozda J, Messer JV, Spertus J, Abramowitz B, Alexander K, Beam CT, et al. ACCF/AHA/AMA-PCPI 2011 performance measures for adults with coronary artery disease and hypertension: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures and the American Medical Association-Physician Consortium for Performance Improvement. *Circulation* 2011 July 12;124(2):248-270.
- Eakins C, Ellis WD, Pruthi S, Johnson DP, Hernanz-Schulman M, Yu C, et al. Second opinion interpretations by specialty radiologists at a pediatric hospital: rate of disagreement and clinical implications. *AJR Am J Roentgenol* 2012 Oct;199(4):916-920.
- Ebenholtz SM. *Oculomotor systems and perception*. 1. publ. ed. Cambridge [u.a.]: Cambridge Univ. Press; 2001.
- Eldevik OP, Dugstad G, Orrison WW, Houghton VM. The effect of clinical bias on the interpretation of myelography and spinal computed tomography. *Radiology* 1982 October 01;145(1):85-89.
- Elliot DL. Evaluation of physical examination skills. Reliability of faculty observers and patient instructors. *JAMA: The Journal of the American Medical Association* 1987 Dec 18;258(23):3405-3408.
- Erly WK, Ashdown BC, Lucio RW, 2nd, Carmody RF, Seeger JF, Alcalá JN. Evaluation of emergency CT scans of the head: is there a community standard? *AJR Am J Roentgenol* 2003 Jun;180(6):1727-1730.
- Failure to diagnose. Misdiagnosis of conditions and diseases. *Medical Malpractice Lawyers and Attorneys Online*, 2006.
- Filatova IS, Bruno MA. Error Management and Reduction. In: Abujudeh H, Bruno Michael A, editors. *Radiology Noninterpretive Skills: The Requisites* Pennsylvania: Elsevier; 2018. p. 110-119.
- FitzGerald R, Mehra R. How accurate is cancer scan reporting? *Hosp Med* 2000 September 01;61(9):637-642.
- Fitzgerald R. Error in radiology. *Clin Radiol* 2001 December 01;56(12):938-946.
- FitzGerald R. Radiological error: analysis, standard setting, targeted instruction and teamworking. *Eur Radiol* 2005 Aug;15(8):1760-1767.
- Fotinos A, Nagy P. What are your goals for peer review? A framework for understanding differing methods. *J Am Coll Radiol* 2012 Dec;9(12):929-930.
- Garland LH. On the scientific evaluation of diagnostic procedures. *Radiology* 1949 March 01;52(3):309-328.
- Garland LH. Studies on the accuracy of diagnostic procedures. *Am J Roentgenol Radium Ther Nucl Med* 1959 July 01;82(1):25-38.
- Geyer LL, Korner M, Linsenmaier U, Huber-Wagner S, Kanz KG, Reiser MF, et al. Incidence of delayed and missed diagnoses in whole-body multidetector CT in patients with multiple injuries after trauma. *Acta Radiol* 2013 Jun;54(5):592-598.
- Gladwell M. The picture problem: mammography, air power, and the limits of looking. *The New Yorker* 2004 December 13;74-81.

Goddard P, Leslie A, Jones A, Wakeley C, Kabala J. Error in radiology. *Br J Radiol* 2001 October 01;74(886):949-951.

Goldberg-Stein S, Frigini LA, Long S, Metwalli Z, Nguyen XV, Parker M, et al. ACR RADPEER Committee White Paper with 2016 Updates: Revised Scoring System, New Classifications, Self-Review, and Subspecialized Reports. *J Am Coll Radiol* 2017 Aug;14(8):1080-1086.

Gollub MJ, Panicek DM, Bach AM, Penalver A, Castellino RA. Clinical importance of reinterpretation of body CT scans obtained elsewhere in patients referred for care at a tertiary cancer center. *Radiology* 1999 January 01;210(1):109-112.

Goo JM, Choi JY, Im JG, Lee HJ, Chung MJ, Han D, et al. Effect of monitor luminance and ambient light on observer performance in soft-copy reading of digital chest radiographs. *Radiology* 2004 September 01;232(3):762-766.

Graber M. Diagnostic errors in medicine: a case of neglect. *Jt Comm J Qual Patient Saf* 2005 February 01;31(2):106-113.

Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005 July 11;165(13):1493-1499.

Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Laparoscopic performance after one night on call in a surgical department: prospective study. *BMJ* 2001 November 24;323(7323):1222-1223.

Grover SA, Barkun AN, Sackett DL. The rational clinical examination. Does this patient have splenomegaly? *JAMA* 1993 November 10;270(18):2218-2221.

Gunderman RB. Biases in Radiologic Reasoning. *American Journal of Roentgenology* 2009 Mar 1;192(3):561-564.

Halligan S. Subspecialist radiology. *Clin Radiol* 2002 November 01;57(11):982-98.

Halpin SF. Medico-legal claims against English radiologists: 1995-2006. *Br J Radiol* 2009 December 01;82(984):982-988.

Halsted MJ, Kumar H, Paquin JJ, Poe SA, Bean JA, Racadio JM, et al. Diagnostic errors by radiology residents in interpreting pediatric radiographs in an emergency setting. *Pediatr Radiol* 2004 April 01;34(4):331-336.

Halsted MJ. Radiology peer review as an opportunity to reduce errors and improve patient care. *J Am Coll Radiol* 2004 Dec;1(12):984-987.

Hansson K, Haggstrom J, Kvarn C, Lord P. Reader performance in radiographic diagnosis of signs of mitral regurgitation in cavalier King Charles spaniels. *J Small Anim Pract* 2009 September 01;50 Suppl 1:44-53.

Hare W. *Medicolegal Radiology*. Sydney: Elsevier Australia; 2007. p. 37-43.

Harrigan CL, Ely WK. On-call radiology: community standards and current trends. *Semin Ultrasound CT MR* 2007 April 01;28(2):85-93.

Harris M, Hartley AL, Blair V, Birch JM, Banerjee SS, Freemont AJ, et al. Sarcomas in north west England: I. Histopathological peer review. *Br J Cancer* 1991 August 01;64(2):315-320.

Harvey JA, Fajardo LL, Innis CA. Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blinded interpretation. 1993 ARRS President's Award. *AJR Am J Roentgenol* 1993 December 01;161(6):1167-1172.

- Hawkins CM, Hall S, Hardin J, Salisbury S, Towbin AJ. Prepopulated radiology report templates: a prospective analysis of error rate and turnaround time. *J Digit Imaging* 2012 August 01;25(4):504-511.
- Hawkins CM, Hall S, Zhang B, Towbin AJ. Creation and implementation of department-wide structured reports: an analysis of the impact on error rate in radiology reports. *J Digit Imaging* 2014 October 01;27(5):581-587.
- Hayward R. VOMIT (victims of modern imaging technology)—an acronym for our times. *BMJ* 2003;326(7401):1273.
- Heinrich HW. *Industrial accident prevention: a scientific approach*. New York [u.a.]: McGraw-Hill; 1931.
- Henson RM, Wyatt SW, Lee NC. The National Breast and Cervical Cancer Early Detection Program: a comprehensive public health response to two major health issues for women. *J Public Health Manag Pract* 1996 January 01;2(2):36-47.
- Herman PG, Gerson DE, Hessel SJ, Mayer BS, Watnick M, Blesser B, et al. Disagreements in chest roentgen interpretation. *Chest* 1975 September 01;68(3):278-282.
- Hobby JL, Tom BD, Todd C, Bearcroft PW, Dixon AK. Communication of doubt and certainty in radiological reports. *Br J Radiol* 2000 September 01;73(873):999-1001.
- Hudson PTW. Process indicators: Managing safety by the numbers. *Safety Science* 2009;47(4):483-485.
- Huntley JH, Carone M, Yousem DM, Babiarz LS. Opportunities for Targeted Education: Critical Neuroradiologic Findings Missed or Misinterpreted by Residents and Fellows. *AJR Am J Roentgenol* 2015 Dec;205(6):1155-1159.
- Irish J. Mala praxis: a study of malpractice claims and litigation in the field of radiology. *Radiol Manage* 2008 August 01;30(4):3.
- Jackson VP, Cushing T, Abujudeh HH, Borgstede JP, Chin KW, Grimes CK, et al. RADPEER scoring white paper. *J Am Coll Radiol* 2009 Jan;6(1):21-25.
- James JT. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf* 2013 September 01;9(3):122-128.
- Janjua KJ, Sugrue M, Deane SA. Prospective evaluation of early missed injuries and the role of tertiary trauma survey. *J Trauma* 1998 June 01;44(6):7.
- Johnson CD, Krecke KN, Miranda R, Roberts CC, Denham C. Quality initiatives: developing a radiology quality and safety program: a primer. *Radiographics* 2009;29(4):951-959.
- Johnson MR, Good CD, Penny WD, Barnes PR, Scadding JW. Lesson of the week: Playing the odds in clinical decision making: lessons from berry aneurysms undetected by magnetic resonance angiography. *BMJ* 2001 June 02;322(7298):1347-1349.
- Jones DN, Benveniste KA, Schultz TJ, Mandel CJ, Runciman WB. Establishing national medical imaging incident reporting systems: issues and challenges. *J Am Coll Radiol* 2010 August 01;7(8):582-592.
- Jordan YJ, Jordan JE, Lightfoote JB, Ragland KD. Quality outcomes of reinterpretation of brain CT studies by subspecialty experts in stroke imaging. *AJR Am J Roentgenol* 2012 Dec;199(6):1365-1370.

Kabadi SJ, Krishnaraj A. Strategies for Improving the Value of the Radiology Report: A Retrospective Analysis of Errors in Formally Over-read Studies. *J Am Coll Radiol* 2017 Apr;14(4):459-466.

Kaewlai R, Abujudeh H. Peer review in clinical radiology practice. *AJR Am J Roentgenol* 2012 Aug;199(2):158.

Kalbhen CL, Yetter EM, Love L, Moncada R, Lawson TL, Albain KA. Outside film reviews of thoracic imaging studies in oncology patients (abstract). *AJR* 1998;170(suppl):77.

Kerlikowske K, Grady D, Barclay J, Frankel SD, Ominsky SH, Sickles EA, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J Natl Cancer Inst* 1998 December 02;90(23):1801-1809.

Khullar D, Jha AK, Jena AB. Reducing Diagnostic Errors--Why Now? *N Engl J Med* 2015 Dec 24;373(26):2491-2493.

Kim H, Zelman RJ, Fox MA, Bennett JM, Berard CW, Butler JJ, et al. Pathology Panel for Lymphoma Clinical Studies: a comprehensive analysis of cases accumulated since its inception. *J Natl Cancer Inst* 1982 January 01;68(1):43-67.

Kim YW, Mansfield LT. Fool me twice: delayed diagnoses in radiology with emphasis on perpetuated errors. *AJR Am J Roentgenol* 2014 Mar;202(3):465-470.

Kline TJ, Kline TS. Radiologists, communication, and Resolution 5: a medicolegal issue. *Radiology* 1992 July 01;184(1):131-134.

Kohn LT, Corrigan JM, Donaldson MS. Errors in health care: a leading cause of death and injury. *To err is human: building a safer health system*. Washington, DC: National Academy Press.; 2000. p. 26-48.

Kronz JD, Westra WH, Epstein JI. Mandatory second opinion surgical pathology at a large referral hospital. *Cancer* 1999 December 01;86(11):2426-2435.

Krupinski EA. Medical image perception issues for PACS deployment. *Semin Roentgenol* 2003 July 01;38(3):231-243.

Krupinski EA, Kallergi M. Choosing a radiology workstation: technical and clinical considerations. *Radiology* 2007 Mar;242(3):671-682.

Krupinski EA, Berbaum KS. Measurement of visual strain in radiologists. *Acad Radiol* 2009 August 01;16(8):947-950.

Krupinski EA, Berbaum KS, Caldwell RT, Scharzt KM, Kim J. Long radiology workdays reduce detection and accommodation accuracy. *J Am Coll Radiol* 2010 Sep;7(9):698-704.

Krupinski EA, Samei E. Medical image perception. *The handbook of medical image perception and techniques*. 1. publ. ed. Cambridge [u.a.]: Cambridge Univ. Press; 2010.

Kruskal JB, Anderson S, Yam CS, Sosna J. Strategies for establishing a comprehensive quality and performance improvement program in a radiology department. *Radiographics* 2009 April 01;29(2):315-329.

Kundel HL. History of research in medical image perception. *J Am Coll Radiol* 2006 June 01;3(6):402-408.

Kuzminski SJ. Sticks and Stones Can Break Your Bones, Words Can Also Hurt You. *J Am Coll Radiol* 2016 Jan;13(1):7.

- Kwaan MR, Studdert DM, Zinner MJ, Gawande AA. Incidence, patterns, and prevention of wrong-site surgery. *Arch Surg* 2006 April 01;141(4):8.
- Larson DB, Towbin AJ, Pryor RM, Donnelly LF. Improving consistency in radiology reporting through the use of department-wide standardized structured reporting. *Radiology* 2013 April 01;267(1):240-250.
- Larson DB, Kruskal JB, Krecke KN, Donnelly LF. Key Concepts of Patient Safety in Radiology. *Radiographics* 2015 Oct;35(6):1677-1693.
- Laura Landro. The Informed Patient: Hospitals Move to Cut Dangerous Lab Errors; Improved Specimen Collection And Efficiency Help Increase Accuracy of Medical Testing. *Wall Street Journal* 2006 Jun 14.
- Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med* 1991 February 07;324(6):377-384.
- Leape LL. Error in medicine. *JAMA* 1994 December 21;272(23):1851-1857.
- Lee CS, Nagy PG, Weaver SJ, Newman-Toker DE. Cognitive and system factors contributing to diagnostic errors in radiology. *AJR Am J Roentgenol* 2013 Sep;201(3):611-617.
- Lehr JL, Lodwick GS, Farrell C, Braaten MO, Virtama P, Kolvisto EL. Direct measurement of the effect of film miniaturization on diagnostic accuracy. *Radiology* 1976 February 01;118(2):257-263.
- Leslie A, Jones AJ, Goddard PR. The influence of clinical information on the reporting of CT by radiologists. *Br J Radiol* 2000 Oct;73(874):1052-1055.
- Levine D, Bankier AA, Halpern EF. Submissions to Radiology: Our Top 10 List of Statistical Errors. *Radiology* 2009 Nov;253(2):288-290.
- Loevner LA, Sonners AI, Schulman BJ, Slawek K, Weber RS, Rosenthal DI, et al. Reinterpretation of cross-sectional images in patients with head and neck cancer in the setting of a multidisciplinary cancer center. *AJNR Am J Neuroradiol* 2002 December 01;23(10):1622-1626.
- Loughrey GJ, Carrington BM, Anderson H, Dobson MJ, Lo Ying Ping F. The value of specialist oncological radiology review of cross-sectional imaging. *Clin Radiol* 1999 March 01;54(3):5.
- Lu Y, Zhao S, Chu PW, Arenson RL. An Update Survey of Academic Radiologists' Clinical Productivity. *Journal of the American College of Radiology* 2008;5(7):817-826.
- Luetmer MT, Hunt CH, McDonald RJ, Bartholmai BJ, Kallmes DF. Laterality errors in radiology reports generated with and without voice recognition software: frequency and clinical significance. *J Am Coll Radiol* 2013 July 01;10(7):538-543.
- MacKenzie W. On asthenopia or weak-sightedness.. *Edinburgh J Med & Surg* 1843;60:73-103.
- Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ* 2016 May 03;353:i2139.
- Markus JB, Somers S, Franic SE, Moola C, Stevenson GW. Interobserver variation in the interpretation of abdominal radiographs. *Radiology* 1989 Apr 1,;171(1):69-71.

Martí-Bonmatí L, Morales A, Donoso Bach L. Hacia un uso adecuado de la telerradiología. *Radiologia* 2011 Sep.

May DS, Lee NC, Nadel MR, Henson RM, Miller DS. The National Breast and Cervical Cancer Early Detection Program: report on the first 4 years of mammography provided to medically underserved women. *AJR Am J Roentgenol* 1998 January 01;170(1):97-104.

McGinnis KS, Lessin SR, Elder DE, Guerry D, Schuchter L, Ming M, et al. Pathology review of cases presenting to a multidisciplinary pigmented lesion clinic. *Arch Dermatol* 2002 May 01;138(5):617-621.

McGurk S, Brauer K, Macfarlane TV, Duncan KA. The effect of voice recognition software on comparative error rates in radiology reports. *Br J Radiol* 2008 October 01;81(970):767-770.

Meghea C, Sunshine JH. Determinants of radiologists' desired workloads. *J Am Coll Radiol* 2007 March 01;4(3):166-170.

Miller CS. Skin-deep diagnosis: affective bias and zebra retreat complicating the diagnosis of systemic sclerosis. *The American journal of the medical sciences* 2013 Jan;345(1):53-56.

Minn MJ, Zandieh AR, Filice RW. Improving Radiology Report Quality by Rapidly Notifying Radiologist of Report Errors. *J Digit Imaging* 2015 August 01;28(4):492-498.

Moray N. Error reduction as a systems problem. In: Bogner MS, editor. *Human Error in Medicine* Hillsdale, NJ: Lawrence Erlbaum Associates; 1994. p. 67-91.

Moskop JC, Geiderman JM, Hobgood CD, Larkin GL. Emergency physicians and disclosure of medical errors. *Ann Emerg Med* 2006 November 01;48(5):523-531

Muchantef K, Forman HP. Professional resource cost of body CT examinations: analysis of interpretation costs in different patient populations. *J Am Coll Radiol* 2004 September 01;1(9):652-658.

Muhm JR, Miller WE, Fontana RS, Sanderson DR, Uhlenhopp MA. Lung cancer detected during a screening program using four-month chest radiographs. *Radiology* 1983 September 01;148(3):609-615.

Nakajima Y, Yamada K, Imamura K, Kobayashi K. Radiologist supply and workload: international comparison--Working Group of Japanese College of Radiology. *Radiat Med* 2008 October 01;26(8):455-465.

Naylor CD. Physical Examination of the Liver. *JAMA* 1994 Jun 15;271(23):1859-1865.

Neale G, Woloshynowych M, Vincent C. Exploring the causes of adverse events in NHS hospital practice. *J R Soc Med* 2001 July 01;94(7):322-330.

Newell RR, Chamberlain WE, Rigler L. Descriptive classification of pulmonary shadows; a revelation of unreliability in the roentgenographic diagnosis of tuberculosis. *Am Rev Tuberc* 1954 April 01;69(4):566-584.

Newman-Toker DE, Robinson KA, Edlow JA. Frontline misdiagnosis of cerebrovascular events in the era of modern neuroimaging: a systematic review. *Ann Neurol* 2008;64(suppl 12):S18.

Newman-Toker DE, Pronovost PJ. Diagnostic errors--the next frontier for patient safety. *JAMA* 2009 Mar 11;301(10):1060-1062.

Nodine C, Mello-Thomas C. The role of expertise in radiologic image interpretation. In: Samei E, Krupinski EA, editors. The handbook of medical image perception and techniques New York: Cambridge University Press; 2010. p. 139-159.

Nolan TW. System changes to improve patient safety. *BMJ* 2000 March 18;320(7237):771-773.

Normal accidents: living with high-risk technologies / Charles Perrow. New York: Basic Books; 1984.

O'Connor PM, Doney KE, Bell PM, Irwin ST, Dearden CH. Unnecessary delays in accident and emergency departments: do medical and surgical senior house officers need to vet admissions? *J Accid Emerg Med* 1995 December 01;12(4):251-254.

Oestmann JW, Greene R, Kushner DC, Bourgouin PM, Linetsky L, Llewellyn HJ. Lung lesions: correlation between viewing time and detection. *Radiology* 1988 Feb;166(2):451-453.

O'Keefe MM, Davis TM, Siminoski K. Performance results for a workstation-integrated radiology peer review quality assurance program. *Int J Qual Health Care* 2016 June 01;28(3):294-298.

Physician Insurers Association of America and the American College of Radiology. Practice Standards Claims Survey. 1997.

Physician Insurers Association of America, Rockville. General and Family Practice Claim Summary. 2002.

Pichichero ME, Poole MD. Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. *Arch Pediatr Adolesc Med* 2001 October 01;155(10):1137-1142.

Pinto A, Brunese L. Spectrum of diagnostic errors in radiology. *World J Radiol* 2010 Oct 28;2(10):377-383.

Plebani M. Errors in clinical laboratories or errors in laboratory medicine? *Clin Chem Lab Med* 2006;44(6):750-759.

Potchen EJ. Measuring observer performance in chest radiology: some experiences. *J Am Coll Radiol* 2006 June 01;3(6):423-432.

Quekel LG, Kessels AG, Goei R, van Engelshoven JM. Miss rate of lung cancer on the chest radiograph in clinical practice. *Chest* 1999 March 01;115(3):720-724.

Quint LE, Quint DJ, Myles JD. Frequency and spectrum of errors in final radiology reports generated with automatic speech recognition technology. *J Am Coll Radiol* 2008 Dec;5(12):1196-1199.

Rabol LI, Andersen ML, Ostergaard D, Bjorn B, Lilja B, Mogensen T. Descriptions of verbal communication errors between staff. An analysis of 84 root cause analysis-reports from Danish hospitals. *BMJ Qual Saf* 2011 March 01;20(3):268-274.

Rana DS, Hurst G, Shepstone L, Pilling J, Cockburn J, Crawford M. Voice recognition for radiology reporting: Is it good enough? *Clinical Radiology* 2005;60(11):1205-1212.

Raskin MM. Survival Strategies for Radiology: Some Practical Tips on How to Reduce the Risk of Being Sued and Losing. *Journal of the American College of Radiology* 2006;3(9):689-693.

Reason J. The contribution of latent human failures to the breakdown of complex systems. *Philos Trans R Soc Lond B Biol Sci* 1990 Apr 12;327(1241):475-484.

Reason J. Understanding adverse events: human factors. *Qual Health Care* 1995 June 01;4(2):80-89.

Reason J. Human error: models and management. *West J Med* 2000 Jun;172(6):393-396.

Reason J. Perceptions of unsafe acts. The human contribution: unsafe acts, accidents and heroic recoveries Burlington, VT: Ashgate; 2008. p. 69-106.

Reiner BI, Krupinski E. The insidious problem of fatigue in medical imaging practice. *J Digit Imaging* 2012 Feb;25(1):3-6.

Renfrew DL, Franken EA, Jr, Berbaum KS, Weigelt FH, Abu-Yousef MM. Error in radiology: classification and lessons in 182 cases presented at a problem case conference. *Radiology* 1992 Apr;183(1):145-150.

Robinson PJ. Radiology's Achilles' heel: error and variation in the interpretation of the Rontgen image. *Br J Radiol* 1997 November 01;70(839):1085-1098.

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 April 01;72(856):323-330.

Rogers LF. Keep looking: satisfaction of search. *AJR Am J Roentgenol* 2000 August 01;175(2):287.

Roosen J, Frans E, Wilmer A, Knockaert DC, Bobbaers H. Comparison of premortem clinical diagnoses in critically ill patients and subsequent autopsy findings. *Mayo Clin Proc* 2000 June 01;75(6):562-567.

Ryan B, O'Malley, Tarun P, Heller MT, Chen LE, McGrath AL, et al. Emergency Radiology. In: Abujudeh H, Bruno MA, editors. *Radiology Noninterpretive Skills: The Requisites* Pennsylvania: Elsevier; 2018. p. 229-255.

Saber Tehrani AS, Lee H, Mathews SC, Shore A, Makary MA, Pronovost PJ, et al. 25-Year summary of US malpractice claims for diagnostic errors 1986-2010: an analysis from the National Practitioner Data Bank. *BMJ Qual Saf* 2013 August 01;22(8):672-680.

Sackett DL, Haynes RB, Tugwell P. The interpretation of diagnostic data. In: Sackett DL, Haynes RB, Tugwell P, eds. *Clinical epidemiology*. Boston: Little, Brown, 1985:59-138.

Sackett DL. A Primer on the Precision and Accuracy of the Clinical Examination. *JAMA* 1992 May 20;267(19):2638-2644.

Sadigh G, Loehfelm T, Applegate KE, Tridandapani S. JOURNAL CLUB: Evaluation of Near-Miss Wrong-Patient Events in Radiology Reports. *AJR Am J Roentgenol* 2015 Aug;205(2):337-343.

Sangwaiya MJ, Saini S, Blake MA, Dreyer KJ, Kalra MK. Errare humanum est: frequency of laterality errors in radiology reports. *AJR Am J Roentgenol* 2009 May 01;192(5):239.

Seiden SC, Barach P. Wrong-side/wrong-site, wrong-procedure, and wrong-patient adverse events: Are they preventable? *Arch Surg* 2006 September 01;141(9):931-939.

Shojania KG, Wald H, Gross R. Understanding medical error and improving patient safety in the inpatient setting. *Med Clin North Am* 2002 July 01;86(4):847-867.

Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. *JAMA* 2003 June 04;289(21):2849-2856.

- Sickles EA. Breast imaging: from 1965 to the present. *Radiology* 2000 April 01;215(1):1-16.
- Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology* 2002 September 01;224(3):861-869.
- Siegel CL, McFarland EG, Brink JA, Fisher AJ, Humphrey P, Heiken JP. CT of cystic renal masses: analysis of diagnostic performance and interobserver variation. *American Journal of Roentgenology* 1997 Sep 1;169(3):813.
- Siegel EL, Reiner B. Image workflow. In: Dreyer KJ, Mehta A, Thrall JH, editors. *PACS: a guide to the digital revolution* New York: Springer-Verlag; 2002. p. 161-190.
- Siegle RL, Baram EM, Reuter SR, Clarke EA, Lancaster JL, McMahan CA. Rates of disagreement in imaging interpretation in a group of community hospitals. *Acad Radiol* 1998 March 01;5(3):148-154.
- Siewert B, Sosna J, McNamara A, Raptopoulos V, Kruskal JB. Missed lesions at abdominal oncologic CT: lessons learned from quality assurance. *Radiographics* 2008 June 01;28(3):623-638.
- Siewert B, Hochman MG. Improving Safety through Human Factors Engineering. *Radiographics* 2015 Oct;35(6):1694-1705.
- Siewert B, Brook OR, Hochman M, Eisenberg RL. Impact of Communication Errors in Radiology on Patient Care, Customer Satisfaction, and Work-Flow Efficiency. *AJR Am J Roentgenol* 2016 Mar;206(3):573-579.
- Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Saf* 2014 September 01;23(9):727-731.
- Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, et al. Comparison of screening mammography in the United States and the United Kingdom. *JAMA* 2003 October 22;290(16):2129-2137.
- Soffa DJ, Lewis RS, Sunshine JH, Bhargavan M. Disagreement in interpretation: a method for the development of benchmarks for quality assurance in imaging. *J Am Coll Radiol* 2004 March 01;1(3):212-217.
- Stevenson CA. Accuracy of the x-ray report. *JAMA* 1969 February 10;207(6):1140-1141.
- Strack F, Mussweiler T. Explaining the Enigmatic Anchoring Effect. *Journal of Personality and Social Psychology* 1997 Sep;73(3):437-446.
- Strahan RH, Schneider-Kolsky ME. Voice recognition versus transcriptionist: error rates and productivity in MRI reporting. *J Med Imaging Radiat Oncol* 2010 October 01;54(5):411-414.
- Strickland NH. Quality assurance in radiology: peer review and peer feedback. *Clinical radiology* 2015 Nov;70(11):1158-1164.
- Studdert DM, Mello MM, Sage WM, DesRoches CM, Peugh J, Zapert K, et al. Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *JAMA* 2005 June 01;293(21):2609-2617.
- Sung JC, Sodickson A, Ledbetter S. Outside CT imaging among emergency department transfer patients. *J Am Coll Radiol* 2009 September 01;6(9):626-632.

Sunshine JH, Maynard CD. Update on the Diagnostic Radiology Employment Market: Findings Through 2007-2008. *Journal of the American College of Radiology* 2008;5(7):827-833.

The Big Picture: Radiology Quality Improvement in the Hospital QA Setting: Communication Is (almost) Everything. *Radiological Society of North America Scientific Assembly and Annual Meeting Program*; December 1 - December 6, 2013 ,Chicago IL.

Thomas EJ, Studdert DM, Burstin HR, Orav EJ, Zeena T, Williams EJ, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000 March 01;38(3):261-271.

Tilleman EH, Phoa SS, Van Delden OM, Rauws EA, van Gulik TM, Lameris JS, et al. Reinterpretation of radiological imaging in patients referred to a tertiary referral centre with a suspected pancreatic or hepatobiliary malignancy: impact on treatment strategy. *Eur Radiol* 2003 May 01;13(5):1095-1099.

Tomaszewski JE, Bear HD, Connally JA, Epstein JI, Feldman M, Foucar K, et al. Consensus conference on second opinions in diagnostic anatomic pathology. Who, What, and When. *Am J Clin Pathol* 2000 September 01;114(3):329-335.

Tsuda K, Yu KK, Coakley FV, Srivastav SK, Scheidler JE, Hricak H. Detection of extracapsular extension of prostate cancer: role of fat suppression endorectal MRI. *J Comput Assist Tomogr* 1999 February 01;23(1):74-78.

Tuddenham WJ. Roentgen image perception--a personal survey of the problem. *Radiol Clin North Am* 1969 December 01;7(3):499-501.

Tudor GR, Finlay D, Taub N. An assessment of inter-observer agreement and accuracy when reporting plain radiographs. *Clin Radiol* 1997 March 01;52(3):235-238.

UK National Health Service Breast Screening Web site. The NHS Breast Screening Programme. Available at: <http://www.cancerscreening.nhs.uk/breastscreen>. Accessibility verified September 10, 2003.

van Beuzekom M, Akerboom SP, Boer F. Assessing system failures in operating rooms and intensive care units. *Qual Saf Health Care* 2007 Feb;16(1):45-50.

van Beuzekom M, Boer F, Akerboom S, Hudson P. Patient safety: latent risk factors. *Br J Anaesth* 2010 Jul;105(1):52-59.

van Breest Smalenburg V, Setz-Pels W, Groenewoud JH, Voogd AC, Jansen FH, Louwman MW, et al. Malpractice claims following screening mammography in The Netherlands. *Int J Cancer* 2012 September 15;131(6):1360-1366.

Van Such M, Lohr R, Beckman T, Naessens JM. Extent of diagnostic agreement among medical referrals. *J Eval Clin Pract* 2017 Aug;23(4):870-874.

Vehmas T. Hawthorne effect: shortening of fluoroscopy times during radiation measurement studies. *Br J Radiol* 1997 October 01;70(838):1053-1055.

Vertinsky T, Forster B. Prevalence of eye strain among radiologists: influence of viewing variables on symptoms. *AJR Am J Roentgenol* 2005 Feb;184(2):681-686.

Wachter RM. Why diagnostic errors don't get any respect--and what can be done about them. *Health Aff (Millwood)* 2010 Sep;29(9):1605-1610.

Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive Error in Radiology. *American Journal of Roentgenology* 2016 Dec 27;:1-11.

- Weeks WB, Foster T, Wallace AE, Stalhandske E. Tort claims analysis in the Veterans Health Administration for quality improvement. *J Law Med Ethics* 2001 January 01;29(3-4):335-345.
- Weingart SN, Ship AN, Aronson MD. Confidential clinician-reported surveillance of adverse events among medical inpatients. *J Gen Intern Med* 2000 July 01;15(7):470-477.
- West CP, Huschka MM, Novotny PJ, Sloan JA, Kolars JC, Habermann TM, et al. Association of perceived medical errors with resident distress and empathy: a prospective longitudinal study. *JAMA* 2006 September 06;296(9):1071-1078.
- Whang JS, Baker SR, Patel R, Luk L, Castro A. The causes of medical malpractice suits against radiologists in the United States. *Radiology* 2013 February 01;266(2):548-554.
- White CS, Romney BM, Mason AC, Austin JH, Miller BH, Protopapas Z. Primary carcinoma of the lung overlooked at CT: analysis of findings in 14 patients. *Radiology* 1996 April 01;199(1):109-115.
- Wickens CD, Lee JD, Liu Y, Becker S. Engineering anthropometry and work space design. *An introduction to human factors engineering*: Upper Saddle River, NJ: Pearson Prentice Hall; 2004. p. 243-268.
- Williams JW, Simel DL. Does This Patient Have Ascites?: How to Divine Fluid in the Abdomen. *JAMA* 1992 May 20;267(19):2645-2648.
- Woolf SH, Kuzel AJ, Dovey SM, Phillips RL, Jr. A string of mistakes: the importance of cascade analysis in describing, counting, and preventing medical errors. *Ann Fam Med* 2004;2(4):317-326.
- Working Group chaired by Professor Sir Patrick Forrest. *Breast Cancer Screening: Report to Health Ministers of England, Wales, Scotland and Northern Ireland*. Her Majesty's Stationery Office; 1986.
- Workload and manpower in clinical radiology. Iain McCall. Royal College of Radiologists. Faculty of Clinical Radiology, 1999.
- Wu AW. Medical error: the second victim. The doctor who makes the mistake needs help too. *BMJ* 2000 March 18;320(7237):726-727.
- Wu AW, Folkman S, McPhee SJ, Lo B. Do house officers learn from their mistakes? *Qual Saf Health Care* 2003 June 01;12(3):8.
- (68) Wu MZ, McInnes MDF, Macdonald DB, Kielar AZ, Duigenan S. CT in adults: systematic review and meta-analysis of interpretation discrepancy rates. *Radiology* 2014 Mar;270(3):717-735.
- Yee J, Kumar NN, Hung RK, Akerkar GA, Kumar PR, Wall SD. Comparison of supine and prone scanning separately and in combination at CT colonography. *Radiology* 2003 March 01;226(3):653-661.
- Yerushalmy J, Garland LH, Harkness JT, Hinshanw HC, Miller ER, Shipmen SJ, et al. An evaluation of the role of serial chest roentgenograms in estimating the progress of disease in patients with pulmonary tuberculosis. *Am Rev Tuberc* 1951 September 01;64(3):225-248.
- YouGov survey of medical misdiagnosis. Isabel Healthcare-Clinical Decision Support System. 2005.
- Zan E, Yousem DM, Carone M, Lewin JS. Second-opinion consultations in neuroradiology. *Radiology* 2010 April 01;255(1):135-141.

9. ANEXOS.

9.1. ANEXO 1. ANÁLISIS UNIVARIANTE DE FACTORES RELACIONADOS CON DISCREPANCIA DIAGNÓSTICA EN RADIOLOGÍA.

9.1.1 Análisis univariante de factores significativos en los estudios programados.

9.1.1.1 Efecto de la complejidad del estudio.

La complejidad del estudio (en áreas anatómicas y reconstrucciones de imagen) se ha clasificado en normal o alta. En la Tabla 16 se evidencia un mayor porcentaje de estudios programados con alta complejidad en los que presentan discrepancia diagnóstica (5,7%) que en los que no la presentan (3,1%), aunque el número de estudios con complejidad alta es muy reducido en ambos casos. También se muestra el ajuste de un modelo de regresión logística univariante para modelizar la probabilidad de la presencia de discrepancia diagnóstica en los estudios de telerradiología programados en función de su complejidad, donde no se observan diferencias significativas (p-valor = 0,282) a partir de los datos recogidos en el presente estudio.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Complejidad del estudio					0,282
Normal	125 (96,9%)	217 (94,3%)	(línea base)		
Alta	4 (3,1%)	13 (5,7%)	1,872	0,598-5,866	

Tabla 16. Relación entre la complejidad y la discrepancia diagnóstica en los estudios programados.

9.1.1.2. Efecto de la modalidad del estudio.

En la Tabla 17 se aprecia un porcentaje similar de RM en los estudios programados con discrepancias diagnósticas (8,3%) que en los que no las presentan (1,6%). También se muestra el resultado del ajuste de un modelo de regresión logística univariante para la probabilidad de discrepancia diagnóstica en un estudio de telerradiología programado en función de la modalidad del estudio, donde se observa que, a pesar del reducido número de TC incluidas, la modalidad está relacionada significativamente con la presencia de discrepancia diagnóstica (p-valor = 0,020), con un incremento estimado de 5,7 veces de la probabilidad de discrepancia diagnóstica en los estudios de TC respecto a los estudios de RM.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Modalidad del estudio					0,020
Resonancia Magnética.	127 (98,4%)	211 (91,7%)	(línea base)		
Tomografía Comp.	2 (1,6%)	19 (8,3%)	5,718	1,310-24,958	

Tabla 17. Relación entre la modalidad y la discrepancia diagnóstica en los estudios programados.

9.1.1.3. Efecto de la calidad técnica de las imágenes.

La calidad técnica de las imágenes se puede clasificar en adecuada o inadecuada. En la Tabla 18 se aprecia que en los estudios programados sin discrepancia tenemos únicamente un 0,8% de estudios con una calidad técnica inadecuada, mientras que un 6,5% de los estudios programados con discrepancia presenta una calidad técnica deficiente. Se trata por tanto de un factor importante para la determinación de la discrepancia diagnóstica en un estudio de telerradiología programado (p-valor = 0,035), aunque el reducido número de estudios con esta clasificación da lugar a una estimación del riesgo asociado con poca precisión.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Calidad técnica de las imágenes					0,035
Buena	128 (99,2%)	215 (93,5%)	(línea base)		
Regular / Mala	1 (0,8%)	15 (6,5%)	8,930	1,166-68,407	

Tabla 18. Relación entre calidad técnica de las imágenes y discrepancia diagnóstica en los estudios programados.

9.1.1.4. Efecto de la disponibilidad de imágenes de estudios previos.

En la Tabla 19 se observa que en los estudios programados con discrepancias diagnósticas hay un mayor porcentaje de estudios con imágenes de estudios previos (8,3%) que en los estudios sin discrepancias diagnósticas (1,6%). Se evidencia que en los estudios programados la presencia de imágenes previas está relacionada significativamente con la presencia de discrepancia diagnóstica (p-valor = 0,020), con un incremento estimado de 5,7 veces en la probabilidad de discrepancia diagnóstica en un estudio con imágenes previas respecto a otro sin ellas.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Imágenes de estudios previos					0,020
No	127 (98,4%)	211 (91,7%)	(línea base)		
Sí	2 (1,6%)	19 (8,3%)	5,718	1,310-24,958	

Tabla 19. Relación entre disponibilidad de imágenes previas y discrepancia diagnóstica en los estudios programados.

9.1.1.5. Efecto de la adecuación de la información clínica de la solicitud del estudio.

En la Tabla 20 se aprecia que en los estudios programados sin discrepancias diagnósticas hay un mayor porcentaje de estudios con información clínica adecuada (81,4%) que en los estudios con discrepancias diagnósticas (73,0%). Se observa que la presencia de información clínica está relacionada con la presencia de discrepancia diagnóstica, aunque las diferencias no son estadísticamente significativas (p-valor = 0,077).

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Información clínica de la solicitud					0,077
Adecuada	105 (81,4%)	168 (73,0%)	(línea base)		
Insuficiente	24 (18,6%)	62 (27,0%)	1,615	0,950-2,745	

Tabla 20. Relación entre información clínica de la solicitud y discrepancia diagnóstica en los estudios programados.

9.1.1.6. Efecto del tipo de visor radiológico utilizado en la primera lectura.

En la Tabla 21 se evidencia que en los estudios programados sin discrepancias diagnósticas hay un mayor porcentaje de estudios en los que se ha utilizado *Clear Canvas*[®] (10,9%) que en los estudios con discrepancias diagnósticas (6,5%). También se muestra el resultado del ajuste de un modelo de regresión logística univariante para la probabilidad de discrepancia diagnóstica en un estudio de telerradiología programado en función del tipo de visor utilizado en la primera lectura, donde se aprecia que el tipo de visor utilizado en la primera lectura no está relacionado significativamente con la presencia de discrepancia diagnóstica (p-valor = 0,153), aunque se observa que el uso del visor *Clear Canvas* es muy residual en las valoraciones de estudios de carácter programado.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Tipo de visor utilizado					0,153
<i>Clear Canvas</i> [®]	14 (10,9%)	15 (6,5%)	(línea base)		
<i>OsiriX</i> [®]	115 (89,1%)	215 (93,5%)	1,745	0,814-3,741	

Tabla 21. Relación entre tipo de visor utilizado y discrepancia diagnóstica en los estudios programados.

9.1.1.7. Efecto del número de monitores utilizados en la primera lectura.

Como se describe anteriormente, hay radiólogos que utilizan un monitor o más de uno (dos monitores o incluso tres monitores). En la Tabla 22 se aprecia que en los estudios programados sin discrepancias diagnósticas hay un mayor porcentaje de estudios en los que se han utilizado dos o más monitores (28,7%) que en los estudios con discrepancias diagnósticas (11,7%). Se observa que el número de monitores utilizados en la primera lectura está relacionado muy significativamente con la presencia de discrepancia diagnóstica (p-valor < 0,001), con un incremento estimado de 3,0 veces en la probabilidad de discrepancia diagnóstica en un estudio en el que se ha utilizado un único monitor respecto a uno en que se han utilizado dos o más monitores.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Número de monitores					<0,001
Dos o más monitores	37 (28,7%)	27 (11,7%)	(línea base)		
Un monitor	92 (71,3%)	203 (88,3%)	3,024	1,738-5,261	

Tabla 22. Relación entre número de monitores utilizados y discrepancia diagnóstica en los estudios programados.

9.1.1.8. Efecto del uso de sistemas de reconocimiento de voz en la primera lectura.

En la Tabla 23 se evidencia que en los estudios programados sin discrepancias diagnósticas hay un mayor porcentaje de estudios en los que se han utilizado sistemas de reconocimiento de voz (10,9%) que en los estudios con discrepancias diagnósticas (4,3%). El uso de este tipo de sistema en la primera lectura está relacionado significativamente con la presencia de discrepancia diagnóstica (p-valor = 0,022), con un incremento estimado de 2,7 veces en la probabilidad de discrepancia diagnóstica en un estudio en el que no se ha utilizado un sistema de reconocimiento de voz respecto a uno en que sí se ha utilizado.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Uso sistema reconocimiento voz					0,022
Sí	14 (10,9%)	10 (4,3%)	(línea base)		
No	115 (89,1%)	220 (95,7%)	2,678	1,154- 6,218	

Tabla 23. Relación entre el uso de un sistema de reconocimiento de voz y la discrepancia diagnóstica en los estudios programados.

9.1.1.9. Efecto de la edad del radiólogo de la primera lectura.

En la Tabla 24 se aprecia que en los estudios programados sin discrepancias diagnósticas hay un mayor porcentaje de estudios realizados por radiólogos de 50 y más años de edad (58,7%) que en los estudios con discrepancias diagnósticas (33,3%). También se muestra el resultado del ajuste de un modelo de regresión logística univariante para la probabilidad de discrepancia diagnóstica en un estudio de telerradiología programado en función de la edad del radiólogo que ha realizado la primera lectura, donde se evidencia que la edad es un factor muy significativo (p-valor < 0,001), con un descenso estimado de la probabilidad de discrepancia diagnóstica de un 6% por cada año de edad de más que tiene el radiólogo que ha realizado la valoración inicial.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Edad radiólogo valoración inicial			0,940	0,913- 0,968	<0,001
>= 50 años	34 (26,4%)	24 (10,4%)			
40-49 años	52 (40,3%)	71 (30,9%)			
< 40 años	43 (33,3%)	135 (58,7%)			

Tabla 24. Relación entre edad del radiólogo y discrepancia diagnóstica en los estudios programados.

9.1.1.10. Efecto de la experiencia del radiólogo de la primera lectura.

La Tabla 25 muestra que en los estudios programados sin discrepancias diagnósticas hay un mayor porcentaje de estudios realizados por radiólogos con más de 15 años de experiencia (38,0%) que en los estudios con discrepancias diagnósticas (22,2%). Se evidencia que la experiencia del radiólogo es un factor muy significativo desde el punto

de vista univariante (p -valor $< 0,001$), con un descenso estimado en la probabilidad de discrepancia diagnóstica de un 5,5% por cada año de experiencia de más que tiene el radiólogo que ha realizado la valoración inicial.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Experiencia radiólogo valoración inicial			0,945	0,921- 0,969	<0,001
>15 años	49 (38,0%)	51 (22,2%)			
3-15 años	41 (31,8%)	64 (27,8%)			
≤ 2 años	39 (30,2%)	115 (50,0%)			

Tabla 25. Relación entre experiencia radiólogo y discrepancia diagnóstica en los estudios programados.

9.1.1.11. Efecto del género del radiólogo que realiza la primera lectura.

En la Tabla 26 se observa que en los estudios programados sin discrepancias diagnósticas hay un mayor porcentaje de estudios que se han valorado inicialmente por radiólogos del género masculino (62,8%) que en los estudios con discrepancias diagnósticas (51,3%). Se aprecia también que el género del radiólogo está relacionada significativamente con la presencia de discrepancia diagnóstica (p -valor = 0,036), con un incremento estimado en la probabilidad de discrepancia diagnóstica del 60,2% de un estudio valorado por un radiólogo de género femenino respecto a otro valorado en la primera lectura por un radiólogo de género masculino.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Género del radiólogo					0,036
Masculino	81 (62,8%)	118 (51,3%)	(línea base)		
Femenino	48 (37,2%)	112 (48,7%)	1,602	1,031- 2,489	

Tabla 26. Relación entre género radiólogo y discrepancia diagnóstica en los estudios programados.

9.1.1.12. Efecto del horario en que se realiza la primera lectura.

En la Tabla 27 se evidencia que en los estudios programados sin discrepancias diagnósticas hay un mayor porcentaje de estudios realizados en día y horario laborable (36,4%) que en los estudios con discrepancias diagnósticas (26,1%), mientras que en los estudios con discrepancias diagnósticas predominan los otros horarios. No pueden establecerse diferencias estadísticamente significativas (p -valor = 0,120) a partir de los datos observados, aunque para algunos tramos horarios las diferencias con las valoraciones realizadas en día y horario laboral rozan la significación estadística.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Horario valoración inicial					0,120
Laboral 8:00-17:59h	47 (36,4%)	60 (26,1%)	(línea base)		
Laboral 18:00-7:59h.	59 (45,7%)	120 (52,2%)	1,593	0,973-2,608	0,064
Festivo	23 (17,8%)	50 (21,7%)	1,703	0,912-3,179	0,095

Tabla 27. Relación entre horario valoración inicial y discrepancia diagnóstica en los estudios programados.

9.1.1.13. Efecto de la especialidad del estudio.

En la Tabla 28 se evidencia que en los estudios sin discrepancias diagnósticas hay un mayor porcentaje de estudios de Columna (24,0%) que en los estudios con discrepancias diagnósticas (12,6%), mientras que hay un menor porcentaje de estudios de MSK (66,7% frente a 77,0%), siendo similar el porcentaje de estudios de Neuro o Body en ambas muestras. Puede apreciarse también el ajuste de un modelo de regresión logística univariante para modelizar la probabilidad de la presencia de una discrepancia diagnóstica en un estudio de telerradiología en función de la especialidad del estudio, que muestra una relación significativa (p-valor = 0,013), con un incremento de la probabilidad de discrepancia diagnóstica de 2,2 veces de un estudio de MSK respecto a uno de Columna, y un incremento de 5,0 veces de un estudio de Body respecto de uno de Columna.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Especialidad del estudio					0,013
Columna	31 (24,0%)	29 (12,6%)	(línea base)		
Neuro	9 (7,0%)	10 (4,3%)	1,188	0,423- 3,337	0,744
MSK	86 (66,7%)	177 (77,0%)	2,200	1,247- 3,883	0,007
Body	3 (2,3%)	14 (6,1%)	4,989	1,299- 19,163	0,019

Tabla 28. Relación entre la especialidad del estudio y discrepancia diagnóstica en los estudios programados.

En la Tabla 28 se observa que en ambas muestras (estudios con discrepancia y sin discrepancia), los estudios de MSK son los más frecuentes, seguidos a mucha distancia de los estudios de Columna y con muy pocos casos analizados de estudios Neurológicos y de Body que agruparemos en una única categoría para poder estimar el OR con mayor precisión. En la Tabla 29 se observan los resultados obtenidos al utilizar esta recodificación.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Especialidad del estudio					0,023
Columna	31 (24,0%)	29 (12,6%)	(línea base)		
MSK	86 (66,7%)	177 (77,0%)	2,200	1,247- 3,883	0,007
Neuro / Body	12 (9,3%)	24 (10,4%)	2,138	0,906- 5,043	0,083

Tabla 29. Relación entre la especialidad del estudio recodificada y la discrepancia diagnóstica en los estudios programados.

9.1.1.14. Efecto de la especialidad del radiólogo que valora inicialmente el estudio.

En la Tabla 30 se evidencia que en los estudios sin discrepancias diagnósticas hay un mayor porcentaje de estudios valorados por radiólogos especialistas de Neuro (19,4%) o

MSK (40,3%) que en los estudios con discrepancias diagnósticas (7,7% y 22,2%, respectivamente), mientras que hay un menor porcentaje de estudios valorados por radiólogos generales (38,0% frente a 67,0%). Se aprecia que existe una relación significativa (p -valor $< 0,001$), con un incremento medio de la probabilidad de discrepancia diagnóstica de 4,4 veces de un estudio valorado por un radiólogo general respecto a uno de Neuro.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Especialidad del radiólogo					<0,001
Neuro	25 (19,4%)	18 (7,8%)	(línea base)	0,664-	
MSK	52 (40,3%)	51 (22,2%)	1,362	2,794	0,120
Body	3 (2,3%)	7 (3,0%)	3,241	0,736-	0,399
General	49 (38,0%)	154 (67,0%)	4,365	14,265	<0,001
				2,198-	
				8,667	

Tabla 30. Relación entre la especialidad del radiólogo y la discrepancia diagnóstica en los estudios programados.

En la Tabla 31 se observa el cruce de ambas características (especialidad del radiólogo y especialidad del estudio a valorar) para las dos muestras de estudios programados consideradas en este análisis. Se aprecia que para los estudios programados, muchos estudios son evaluados por radiólogos especialistas, aunque también hay un gran porcentaje de estudios valorados por radiólogos generales, particularmente en los estudios con discrepancias.

	Total	Especialidad del radiólogo			
		Neuro	MSK	Body	General
Estudios sin discrepancias					
Especialidad del estudio					
Neuro	9	7 (77,8%)	0 (0%)	0 (0%)	2 (22,2%)
MSK	86	0 (0%)	49 (57,0%)	0 (0%)	37 (43,0%)
Body	3	0 (0%)	0 (0%)	2 (66,7%)	1 (33,3%)
Columna	31	18 (58,1%)	3 (9,7%)	1 (3,2%)	9 (29,0%)
Estudios con discrepancias					
Especialidad del estudio					
Neuro	10	4 (40%)	1 (10%)	0 (0%)	5 (50,0%)
MSK	177	1 (0,6%)	49 (27,7%)	2 (1,1%)	125 (70,6%)
Body	14	0 (0%)	0 (0%)	5 (35,7%)	9 (64,3%)
Columna	29	13 (44,8%)	1 (3,4%)	0 (0%)	15 (51,7%)

Tabla 31. Relación entre la especialidad del radiólogo y la especialidad de los estudios programados.

Hay que tener en cuenta que el número de observaciones para algunas de las combinaciones posibles es muy reducido, lo que dará lugar a estimaciones poco precisas del efecto de dichos valores sobre la presencia o ausencia de discrepancia diagnóstica. Por tanto se agruparán los estudios según el tipo de estudio y el tipo de radiólogo que lo

ha valorado, estableciendo el grado de concordancia existente entre estudio-evaluador en base al criterio siguiente:

- Grado de concordancia alta: 151 estudios
 - o 11 estudios de Neuro informados por un neurorradiólogo.
 - o 98 estudios de MSK informados por un radiólogo musculoesquelético.
 - o 7 estudios de Body informados por un radiólogo de Body.
 - o 31 estudios de Columna informados por un neurorradiólogo.
 - o 4 estudios de Columna informados por un radiólogo musculoesquelético.
- Grado de concordancia media: 203 estudios
 - o 7 estudios de Neuro informados por un radiólogo general.
 - o 162 estudios de MSK informados por un radiólogo general.
 - o 10 estudios de Body informados por un radiólogo general.
 - o 24 estudios de Columna informados por un radiólogo general.
- Grado de concordancia baja: 5 estudios
 - o 1 estudios de Neuro informados por un radiólogo musculoesquelético.
 - o 1 estudios de MSK informados por un neurorradiólogo.
 - o 2 estudios de MSK informados por un radiólogo de Body.
 - o 1 estudios de Columna informados por un radiólogo de Body.

Por tanto, con esta recodificación se han clasificado los estudios en estudios valorados por un radiólogo especialista de la misma especialidad (concordancia alta), estudios valorados por un radiólogo generalista (concordancia media) y estudios valorados por un radiólogo especialista de diferente especialidad a la del estudio (concordancia baja). En los estudios programados hay muy pocos estudios con concordancia baja, por lo que los agruparemos con los de concordancia media para el análisis. A partir de la recodificación descrita anteriormente se han clasificado los estudios, según la relación de la especialidad del estudio a valorar y del radiólogo que lo valora, en estudios con concordancia alta y media/baja. En la Tabla 32 se evidencia que en los estudios sin discrepancias diagnósticas hay un mayor porcentaje de estudios con concordancia alta (61,2%) respecto a los estudios con discrepancias diagnósticas (31,3%). Se aprecia que existe una relación muy significativa (p -valor $< 0,001$), con un incremento medio de la probabilidad de discrepancia diagnóstica de 3,5 veces de un estudio programado con concordancia media/baja respecto a uno con concordancia alta.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Concordancia de especialidades					<0,001
Alta	79 (61,2%)	72 (31,3%)	(línea base)		
Media / Baja	50 (38,8%)	158 (68,7%)	3,467	2,209- 5,441	

Tabla 32. Relación de la concordancia de especialidades estudio-radiólogo y la discrepancia diagnóstica en los estudios programados.

En la Figura 36 se muestran los OR estimados, junto con sus intervalos de confianza, para los diferentes factores de riesgo de discrepancia diagnóstica considerados en el estudio. Son factores significativos desde el punto de vista univariante aquellos en los que el intervalo de confianza no contiene el valor 1, siendo factores protectores de la discrepancia diagnóstica aquellos con valores menores de la unidad y factores precursoros de la discrepancia diagnóstica aquellos con valores mayores de la unidad. Se observa en la Figura 36 que, por ejemplo, que el efecto de una complejidad alta del estudio respecto a una complejidad normal es un factor precursor de discrepancia diagnóstica, pero no es significativo al dar lugar a un intervalo de confianza muy amplio. Son significativos, desde un punto de vista univariante, la modalidad del estudio (mayor asociación con discrepancia diagnóstica en los estudios de TC respecto a los de RM), la calidad técnica de las imágenes adecuada respecto a la ausencia de la misma, la presencia de imágenes previas respecto a la ausencia de las mismas, la presencia de un monitor respecto a disponer de más de uno, la ausencia de reconocimiento de voz respecto a la presencia del sistema, la edad del radiólogo, la experiencia del radiólogo, el género del radiólogo femenino respecto al género masculino, la especialidad del estudio MSK frente a columna o la concordancia entre la especialidad del radiólogo y el estudio, siendo además todos ellos factores precursoros de discrepancia diagnóstica, a excepción de la edad y la experiencia del radiólogo que son factores protectores. No obstante, hay que tener en cuenta que todos ellos son análisis univariantes, por lo que no tienen en cuenta la posible relación existente entre factores. Es necesario llevar a cabo un análisis multivariante que permita establecer el efecto que tienen los diferentes factores en la discrepancia diagnóstica de forma independiente, controlando el efecto del resto de factores significativos.

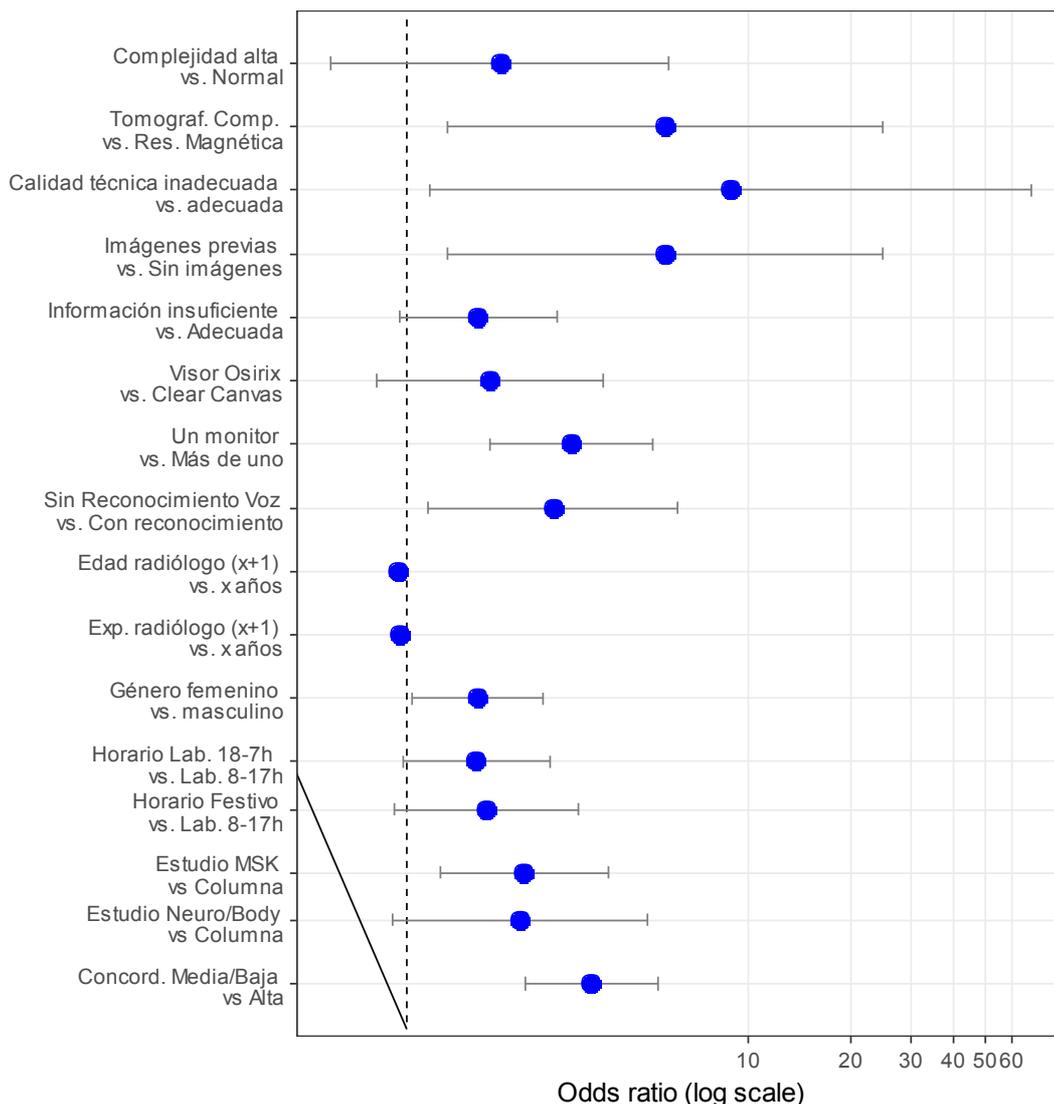


Figura 36. OR estimado e IC95% univariante de los factores de riesgo considerados para los estudios programados.

9.1.2. Análisis univariante de factores significativos en los estudios urgentes.

9.1.2.1. Efecto de la complejidad del estudio.

En la Tabla 33 se observa un mayor porcentaje de estudios con alta complejidad en los que presentan discrepancia diagnóstica (8,8%) que en los que no la presentan (4,9%). También se muestra el ajuste de un modelo de regresión logística univariante para modelizar la probabilidad de la presencia de discrepancia diagnóstica en los estudios de telerradiología urgentes en función de su complejidad, donde se aprecia que la complejidad del estudio está relacionada significativamente con la presencia de discrepancia diagnóstica (p-valor = 0,040), con un incremento estimado de la probabilidad de discrepancia diagnóstica del 87,1% en un estudio de alta complejidad respecto a otro con complejidad normal.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Complejidad del estudio					0,040
Normal	388 (95,1%)	280 (91,2%)	(línea base)		
Alta	20 (4,9%)	27 (8,8%)	1,871	1,028-3,403	

Tabla 33. Relación entre la complejidad y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.2. Efecto de la modalidad del estudio.

En la Tabla 34 se aprecia un porcentaje similar de RM en los estudios urgentes con discrepancias diagnósticas (12,7%) que en los que no las presentan (11,5%). Se evidencia que en los estudios urgentes la modalidad no está relacionada significativamente con la presencia de discrepancia diagnóstica (p-valor = 0,630).

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Modalidad del estudio					0,630
Resonancia Magnética	47 (11,5%)	39 (12,7%)	(línea base)		
Tomografía Comp.	361 (88,5%)	268 (87,3%)	0,895	0,569-1,407	

Tabla 34. Relación entre la modalidad y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.3. Efecto de la calidad técnica de las imágenes.

La calidad técnica de las imágenes se puede clasificar en adecuada o inadecuada. En la Tabla 35 se observa un mayor porcentaje de estudios con una calidad técnica de las imágenes inadecuada en los que presentan discrepancia diagnóstica (2,9%) que en los que no presentan una calidad adecuada (1,2%). También se muestra el ajuste de un modelo de regresión logística univariante para modelizar la probabilidad de la presencia de discrepancia diagnóstica en los estudios de telerradiología urgentes en función de la calidad técnica de las imágenes, aunque el reducido número de casos con una calidad inadecuada dificultan el ajuste del modelo asociado, dando lugar a un modelo no significativo (p-valor = 0,114).

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Calidad técnica de las imágenes					0,114
Buena	403 (98,8%)	298 (97,1%)	(línea base)		
Regular / Mala	5 (1,2%)	9 (2,9%)	2,434	0,808-7,338	

Tabla 35. Relación entre calidad técnica de las imágenes y discrepancia diagnóstica en los estudios urgentes.

9.1.2.4. Efecto de la disponibilidad de imágenes de estudios previos.

En la Tabla 36 se observa que en los estudios urgentes con discrepancias diagnósticas hay un mayor porcentaje de estudios con imágenes de estudios previos (5,9%) que en los estudios sin discrepancias diagnósticas (1,0%). Se evidencia que en los estudios urgentes la presencia de imágenes previas está relacionada muy significativamente con la presencia de discrepancia diagnóstica (p-valor = 0,001), con un incremento estimado de

6,3 veces en la probabilidad de discrepancia diagnóstica en un estudio con imágenes previas respecto a otro sin ellas.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Imágenes de estudios previos					0,001
No	404 (99,0%)	289 (94,1%)	(línea base)	2,107-	
Si	4 (1,0%)	18 (5,9%)	6,291	18,783	

Tabla 36. Relación entre la disponibilidad de imágenes de estudios previos y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.5. Efecto de la adecuación de la información clínica de la solicitud del estudio.

En la Tabla 37 se aprecia que en los estudios urgentes sin discrepancias diagnósticas hay un mayor porcentaje de estudios con información clínica adecuada (96,3%) que en los estudios con discrepancias diagnósticas (89,3%). Se observa que la presencia de información clínica está relacionada muy significativamente con la presencia de discrepancia diagnóstica (p-valor < 0,001), con un incremento estimado de 3,2 veces en la probabilidad de discrepancia diagnóstica en un estudio con información clínica insuficiente respecto a uno con información clínica adecuada.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Información clínica de la solicitud					<0,001
Adecuada	393 (96,3%)	274 (89,3%)	(línea base)	1,681-	
Insuficiente	15 (3,7%)	33 (10,7%)	3,155	5,922	

Tabla 37. Relación entre la suficiencia de la información clínica de la solicitud y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.6. Efecto del tipo de visor radiológico utilizado en la primera lectura.

En la Tabla 38 se evidencia que en los estudios urgentes sin discrepancias diagnósticas hay un mayor porcentaje de estudios en los que se ha utilizado *Clear Canvas*® (29,9%) que en los estudios con discrepancias diagnósticas (20,5%). Se aprecia que el tipo de visor utilizado en la primera lectura está relacionada muy significativamente con la presencia de discrepancia diagnóstica (p-valor = 0,005), con un incremento estimado de 1,7 veces en la probabilidad de discrepancia diagnóstica en un estudio en el que se ha utilizado un visor *OsiriX*® respecto a uno en que se ha utilizado *Clear Canvas*®.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Tipo de visor utilizado					0,005
Clear Canvas®	122 (29,9%)	63 (20,5%)	(línea base)	1,166-	
OsiriX®	286 (70,1%)	244 (79,5%)	1,652	2,342	

Tabla 38. Relación entre el tipo de visor utilizado en la primera lectura y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.7. *Efecto del número de monitores utilizados en la primera lectura.*

En la Tabla 39 se observa que en los estudios urgentes sin discrepancias diagnósticas hay un mayor porcentaje de estudios en los que se han utilizado dos o más monitores (29,9%) que en los estudios con discrepancias diagnósticas (22,5%). Se evidencia que el número de monitores utilizados en la primera lectura está relacionado muy significativamente con la presencia de discrepancia diagnóstica (p -valor = 0,027), con un incremento estimado de 1,5 veces en la probabilidad de discrepancia diagnóstica en un estudio en el que se ha utilizado un único monitor respecto a uno en que se han utilizado dos o más monitores.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Número de monitores					0,027
Dos o más monitores	122 (29,9%)	69 (22,5%)	(línea base)	1,046-	
Un monitor	286 (70,1%)	238 (77,5%)	1,471	2,071	

Tabla 39. Relación entre el número de monitores utilizados en la primera lectura y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.8. *Efecto del uso de sistemas de reconocimiento de voz en la primera lectura.*

En la Tabla 40 se observa que en los estudios programados sin discrepancias diagnósticas hay un mayor porcentaje de estudios en los que se han utilizado sistemas de reconocimiento de voz (6,1%) que en los estudios con discrepancias diagnósticas (1,6%). Se aprecia que el uso de este tipo de sistema en la primera lectura está relacionado muy significativamente con la presencia de discrepancia diagnóstica (p -valor = 0,006), con un incremento estimado de 3,9 veces en la probabilidad de discrepancia diagnóstica en un estudio en el que no se ha utilizado un sistema de reconocimiento de voz respecto a uno en que sí se ha utilizado.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Uso sistema reconocimiento voz					0,006
Sí	25 (6,1%)	5 (1,6%)	(línea base)	1,492-	
No	383 (93,9%)	302 (98,4%)	3,943	10,421	

Tabla 40. Relación entre el uso de un sistema de reconocimiento de voz y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.9. *Efecto de la edad del radiólogo de la primera lectura.*

En la Tabla 41 se evidencia que en los estudios urgentes sin discrepancias diagnósticas hay un mayor porcentaje de estudios realizados por radiólogos de 50 y más años de edad (39,0%) que en los estudios con discrepancias diagnósticas (29,0%). También se muestra el resultado del ajuste de un modelo de regresión logística univariante para la probabilidad de discrepancia diagnóstica en un estudio de telerradiología urgente en

función de la edad del radiólogo que ha realizado la primera lectura, donde se observa que la edad es un factor muy significativo (p -valor = 0,005), con un descenso estimado de la probabilidad de discrepancia diagnóstica del 2,0% por cada año de más que tiene el radiólogo que ha realizado la valoración inicial.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Edad radiólogo valoración inicial			0,980	0,966-0,994	0,005
>= 50 años	159 (39,0%)	89 (29,0%)			
40-49 años	98 (24,0%)	67 (21,8%)			
< 40 años	151 (37,0%)	151 (49,2%)			

Tabla 41. Relación entre edad del radiólogo y discrepancia diagnóstica en los estudios urgentes.

9.1.2.10. Efecto de la experiencia del radiólogo de la primera lectura.

En la Tabla 42 se aprecia que en los estudios urgentes sin discrepancias diagnósticas hay un mayor porcentaje de estudios realizados por radiólogos con más de 15 años de experiencia (37,7%) que en los estudios con discrepancias diagnósticas (21,5%). Se evidencia que la experiencia del radiólogo es un factor muy significativo desde el punto de vista univariante (p -valor < 0,001), con un descenso estimado de la probabilidad de discrepancia diagnóstica del 3,1% por cada año de más de experiencia del radiólogo que ha realizado la valoración inicial.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Experiencia radiólogo valoración inicial			0,969	0,955-0,984	<0,001
>15 años	154 (37,7%)	66 (21,5%)			
3-15 años	89 (21,8%)	83 (27,0%)			
<=2 años	165 (40,4%)	158 (51,5%)			

Tabla 42. Relación entre la experiencia del radiólogo y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.11. Efecto del género del radiólogo que realiza la primera lectura.

En la Tabla 43 se observa que en los estudios urgentes sin discrepancias diagnósticas hay un mayor porcentaje de estudios que se han valorado inicialmente por radiólogos del género masculino (66,7%) que en los estudios con discrepancias diagnósticas (51,1%). Se observa también que el género del radiólogo está relacionado significativamente con la presencia de discrepancia diagnóstica (p -valor < 0,001), con un incremento estimado en la probabilidad de discrepancia diagnóstica del 91,1% de un estudio valorado por un radiólogo de género femenino respecto a otro valorado en la primera lectura por un radiólogo de género masculino.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Género del radiólogo					<0,001
Masculino	136 (33,3%)	150 (48,9%)	(línea base)		
Femenino	272 (66,7%)	157 (51,1%)	1,911	1,410-2,590	

Tabla 43. Relación entre el género del radiólogo que ha valorado la primera lectura y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.12. Efecto del horario en que se realiza la primera lectura.

En la Tabla 44 se aprecia que en los estudios urgentes sin discrepancias diagnósticas hay un mayor porcentaje de estudios realizados en día festivo (37,5%) que en los estudios con discrepancias diagnósticas (31,9%). No obstante estas diferencias no son estadísticamente significativas (p-valor = 0,165).

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Horario valoración inicial					0,165
Laboral 8:00-17:59h	158 (38,7%)	119 (38,8%)	(línea base)		
Laboral Otro horario	97 (23,8%)	90 (29,3%)	1,232	0,849-2,788	0,273
Festivo	153 (37,5%)	98 (31,9%)	0,850	0,601-1,204	0,361

Tabla 44. Relación entre el horario de la valoración inicial y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.13. Efecto de la especialidad del estudio.

En la Tabla 45 se evidencia que en los estudios sin discrepancias diagnósticas hay un mayor porcentaje de estudios de Neurorradiología (49,5%) que en los estudios con discrepancias diagnósticas (39,1%), mientras que hay un menor porcentaje de estudios de Body (33,8% frente a 44,3%), siendo similar el porcentaje de estudios de MSK o Columna en ambas muestras. Se observa también el ajuste de un modelo de regresión logística univariante para modelizar la probabilidad de la presencia de una discrepancia diagnóstica en un estudio de telerradiología en función de la especialidad del estudio, que muestra una relación muy significativa (p-valor = 0,007), con un incremento de la probabilidad de discrepancia diagnóstica del 65,9% de un estudio de Body respecto a uno de Neurorradiología, y un incremento del 74,1% de un estudio de Columna respecto a uno de Neurorradiología.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Especialidad del estudio					0,007
Neuro	202 (49,5%)	120 (39,1%)	(línea base)		
MSK	39 (9,6%)	21 (6,8%)	0,906	0,509-1,613	0,738
Body	138 (33,8%)	136 (44,3%)	1,659	1,196-2,301	0,002
Columna	29 (7,1%)	30 (9,8%)	1,741	0,997-3,043	0,051

Tabla 45. Relación entre la especialidad del estudio y la discrepancia diagnóstica en los estudios urgentes.

En la Tabla 45 se aprecia que en ambas muestras (estudios con discrepancia y sin discrepancia), los estudios de Neurología y Body son los más frecuentes, seguidos a mucha distancia de los estudios de MSK y Columna, que agruparemos en una única categoría para poder estimar el OR con mayor precisión. En la Tabla 46 se observan los resultados obtenidos al utilizar esta recodificación, que permite un mejor ajuste del OR asociado.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Especialidad del estudio					0,010
Neuro	202 (49,5%)	120 (39,1%)	(línea base)		
Body	138 (33,8%)	136 (44,3%)	1,659	1,196-2,301	0,002
MSK / Columna	68 (16,7%)	51 (16,6%)	1,262	0,823-1,936	0,285

Tabla 46. Relación entre la especialidad del estudio recodificada y la discrepancia diagnóstica en los estudios urgentes.

9.1.2.14. Efecto de la especialidad del radiólogo que valora inicialmente el estudio.

En la Tabla 47 se observa que en los estudios sin discrepancias diagnósticas hay un mayor porcentaje de estudios valorados por radiólogos especialistas de Neuro (15,4%), MSK (7,4%) o Body (17,4%) que en los estudios con discrepancias diagnósticas (7,1%, 2,9% y 12,4%, respectivamente), mientras que hay un menor porcentaje de estudios valorados por radiólogos generales (59,8% frente a 77,5%). Se aprecia que existe una relación significativa (p-valor < 0,001), con un incremento estimado de la probabilidad de discrepancia diagnóstica de 2,8 veces de un estudio valorado por un radiólogo general respecto a uno de Neuro. También se observa que los estudios valorados por un radiólogo de MSK son los de menor riesgo de discrepancia diagnóstica, aunque las diferencias con el resto de especialidades no son estadísticamente significativas.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Especialidad del radiólogo					<0,001
Neuro	63 (15,4%)	22 (7,2%)	(línea base)		
MSK	30 (7,4%)	9 (2,9%)	0,859	0,353-2,090	0,738
Body	71 (17,4%)	38 (12,4%)	1,533	0,820-2,864	0,181
General	244 (59,8%)	238 (77,5%)	2,793	1,665-4,685	<0,001

Tabla 47. Relación entre especialidad del radiólogo y discrepancia diagnóstica en los estudios urgentes.

En la Tabla 48 se observa el cruce de ambas características (especialidad del radiólogo y especialidad del estudio a valorar) para las dos muestras de estudios urgentes consideradas en este análisis. Puede apreciarse que en ambas muestras consideradas, los estudios de todas las especialidades son valorados mayoritariamente por radiólogos generalistas.

	Total	Especialidad del radiólogo			
		Neuro	MSK	Body	General
Estudios sin discrepancias					
Especialidad del estudio					
Neuro	202	29 (14,4%)	10 (5,0%)	28 (13,9%)	135 (66,8%)
MSK	39	7 (17,9%)	11 (28,2%)	8 (20,5%)	13 (33,3%)
Body	138	23 (16,7%)	3 (2,2%)	31 (22,5%)	81 (58,7%)
Columna	29	4 (13,8%)	6 (20,7%)	4 (13,8%)	15 (51,7%)
Estudios con discrepancias					
Especialidad del estudio					
Neuro	210	5 (2,4%)	3 (1,4%)	18 (8,6%)	94 (44,8%)
MSK	21	2 (9,5%)	1 (4,8%)	1 (4,8%)	17 (81,0%)
Body	136	14 (10,3%)	3 (2,2%)	13 (9,6%)	106 (77,9%)
Columna	30	1 (3,3%)	2 (6,7%)	6 (20,0%)	21 (70,0%)

Tabla 48. Relación entre la especialidad del radiólogo y la de los estudios urgentes.

	N	OR	IC 95%		p-valor
Combinación de especialidades					
Estudio Neuro + Radiólogo Neuro	34	(línea base)			<0,001
Estudio Columna + Radiólogo MSK	8	1,933	0,301	12,426	0,487
Estudio MSK + Radiólogo Body	9	0,725	0,074	7,125	0,783
Estudio Neuro + Radiólogo MSK	13	1,740	0,351	8,633	0,498
Estudio MSK + Radiólogo Neuro	9	1,657	0,264	10,391	0,590
Estudio Body + Radiólogo Body	44	2,432	0,771	7,674	0,129
Estudio Body + Radiólogo Neuro	37	3,530	1,108	11,245	0,033
Estudio Columna + Radiólogo Neuro	5	1,450	0,133	15,793	0,760
Estudio Neuro + Radiólogo Body	46	3,729	1,218	11,412	0,021
Estudio Neuro + Radiólogo General	229	4,039	1,508	10,814	0,005
Estudio MSK + Radiólogo MSK	12	0,527	0,055	5,035	0,578
Estudio Body + Radiólogo MSK	6	5,800	0,902	37,279	0,064
Estudio Columna + Radiólogo Body	10	8,700	1,789	42,304	0,007
Estudio Body + Radiólogo General	187	7,590	2,814	20,471	<0,001
Estudio Columna + Radiólogo General	36	8,120	2,552	25,837	<0,001
Estudio MSK + Radiólogo General	30	7,585	2,301	24,995	0,001

Tabla 49. Efecto de la relación de especialidades sobre la discrepancia diagnóstica en los estudios urgentes.

En la Tabla 49 se observa el ajuste de un modelo de regresión logística univariante para modelizar la probabilidad de la presencia de una discrepancia diagnóstica en un estudio de telerradiología en función de la relación de especialidades del estudio y del radiólogo que lo valora, donde se estima el OR asociado a cada una de las posibles combinaciones respecto a la de menor tasa de discrepancia. Hay que tener en cuenta que el número de observaciones para algunas de las combinaciones posibles es muy reducido, lo que dará lugar a estimaciones poco precisas del efecto de dichos valores sobre la presencia o ausencia de discrepancia diagnóstica.

A partir del tipo de estudio y del tipo de radiólogo que lo ha valorado se establecerá el grado de concordancia existente entre estudio-evaluador, en base al criterio siguiente:

- Grado de concordancia alta: 103 estudios
 - o 34 estudios de Neurorradiología informados por un neurorradiólogo.
 - o 12 estudios de Musculoesquelético informados por un radiólogo musculoesquelético.
 - o 44 estudios de Body informados por un radiólogo de Body.
 - o 5 estudios de Columna informados por un neurorradiólogo.
 - o 8 estudios de Columna informados por un radiólogo musculoesquelético.
- Grado de concordancia media: 482 estudios
 - o 229 estudios de Neurorradiología informados por un radiólogo general.
 - o 30 estudios de Musculoesquelético informados por un radiólogo general.
 - o 187 estudios de Body informados por un radiólogo general.
 - o 36 estudios de Columna informados por un radiólogo general.
- Grado de concordancia baja: 130 estudios
 - o 46 estudios de Neuro informados por un radiólogo de Body.
 - o 13 estudios de Neuro informados por un radiólogo musculoesquelético.
 - o 9 estudios de Musculoesquelético informados por un neurorradiólogo.
 - o 9 estudios de Musculoesquelético informados por un radiólogo de Body.
 - o 37 estudios de Body informados por un neurorradiólogo.
 - o 6 estudios de Body informados por un radiólogo musculoesquelético.
 - o 10 estudios de Columna informados por un radiólogo de Body.

Por tanto, con esta recodificación se han clasificado los estudios en estudios valorados por un radiólogo especialista de la misma especialidad (concordancia alta), estudios valorados por un radiólogo generalista (concordancia media) y estudios valorados por un radiólogo especialista de diferente especialidad a la del estudio (concordancia baja). En la Tabla 50 se evidencia que en los estudios sin discrepancias diagnósticas hay un mayor porcentaje de estudios con concordancia alta (19,9%) o baja (20,3%) respecto a los estudios con discrepancias diagnósticas (7,2% y 15,3%, respectivamente), mientras que hay un menor porcentaje de estudios con concordancia media (59,8% frente a 77,5%) que han sido valorados por radiólogos generalistas. Se aprecia que existe una relación

significativa (p -valor $< 0,001$), con un incremento estimado de la probabilidad de discrepancia diagnóstica de 3,6 veces de un estudio urgente con concordancia media respecto a uno con concordancia alta, y un incremento de la probabilidad de discrepancia diagnóstica de 2,1 veces de un estudio urgente con concordancia media respecto a uno con concordancia baja.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Concordancia de especialidades					<0,001
Alta	81 (19,9%)	22 (7,2%)	(línea base)		
Media	244 (59,8%)	238 (77,5%)	3,591	2,170-5,944	<0,001
Baja	83 (20,3%)	47 (15,3%)	2,085	1,154-3,767	0,015

Tabla 50. Relación de la concordancia de especialidades estudio-radiólogo y la discrepancia diagnóstica en los estudios urgentes.

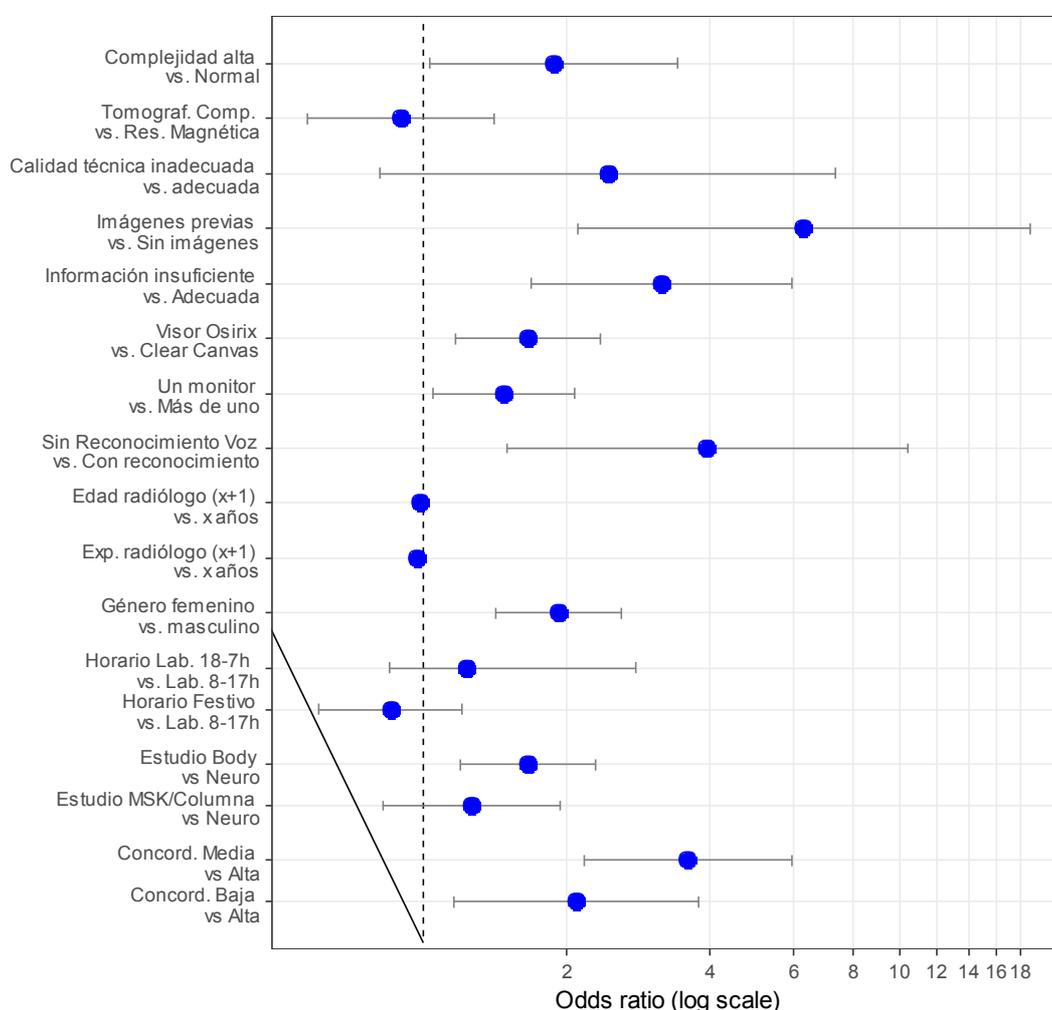


Figura 37. OR estimado e IC95% en el análisis univariante de los factores de riesgo considerados para los estudios urgentes.

En la Figura 37 se muestran los OR estimados, junto con sus intervalos de confianza, para los diferentes factores de riesgo de discrepancia diagnóstica considerados en el

estudio. Se observa que, por ejemplo, el efecto de una complejidad alta del estudio respecto a una complejidad normal es un factor precursor significativo de discrepancia diagnóstica. También son factores precursores significativos de discrepancia diagnóstica, desde un punto de vista univariante, la presencia de imágenes de estudios previos respecto a la ausencia de las mismas, una información clínica de la solicitud insuficiente respecto a una información adecuada, el uso del visor *OsiriX*® respecto al visor *Clear Canvas*®, la presencia de un monitor respecto a disponer de más de uno, la ausencia de reconocimiento de voz respecto a la presencia del sistema, el género del radiólogo femenino respecto al género masculino, la especialidad del estudio Body respecto a Neurorradiología, la concordancia media entre la especialidad del radiólogo y el estudio respecto a una concordancia alta o la concordancia baja entre la especialidad del radiólogo y el estudio respecto a una concordancia alta, mientras que la edad del radiólogo y los años de experiencia del radiólogo son factores protectores de la discrepancia diagnóstica. En cambio no son significativos la modalidad del estudio, la calidad técnica de las imágenes o el horario de la valoración inicial. Hay que tener en cuenta que todos ellos son análisis univariantes, por lo que no tienen en cuenta la posible relación existente entre factores. Es necesario llevar a cabo un análisis multivariante que permita establecer el efecto que tienen los diferentes factores en la discrepancia diagnóstica de forma independiente, controlando el efecto del resto de factores significativos.

En las secciones 4.4.1 y 4.4.2 se ha expuesto el efecto de los diferentes factores considerados en este estudio en los estudios programados y urgentes. En la Tabla 51 se aprecia el efecto univariante de cada uno de los factores estimado para cada uno de los dos estratos de estudios definidos. Se evidencia que la mayoría de los factores se relacionan con la discrepancia diagnóstica de forma similar en los estudios programados y en los estudios urgentes. En cambio se observan diferencias en el efecto de la modalidad de estudio, con una presencia mayor de discrepancias en las TC en los estudios programados que no se observan en los estudios urgentes. También hay diferencias en los efectos del tipo de estudio sobre la discrepancia diagnóstica, ya que en los programados se observa que los estudios de columna presentan mejores resultados que el resto de estudios, con un comportamiento similar, mientras que en los urgentes los estudios neurológicos son los que mejor comportamiento muestran, mientras que el resto son similares respecto a la presencia de discrepancia diagnóstica. Respecto a la calidad técnica de las imágenes, parece que en ambos tipos de estudios es un factor

relevante, pero el reducido número de estudios urgentes con calidad inadecuada dificulta la estimación del riesgo asociado con la suficiente precisión.

Estudios programados				Estudios urgentes			
	OR	IC (95%)	p-valor		OR	IC (95%)	p-valor
Complejidad del estudio			0,282	Complejidad del estudio			0,040
Normal (línea base)				Normal (línea base)			
Alta	1,872	0,598-5,866		Alta	1,871	1,028-3,403	
Modalidad del estudio			0,020	Modalidad del estudio			0,630
Reson. Mag. (línea base)				Reson. Mag. (línea base)			
Tomog. Comp.	5,718	1,310-24,958		Tomog. Comp.	0,895	0,569-1,407	
Calidad técnica de las imágenes			0,035	Calidad técnica de las imágenes			0,114
Buena (línea base)				Buena (línea base)			
Regular / Mala	8,930	1,166-68,407		Regular / Mala	2,434	0,808-7,338	
Imágenes de estudios previos			0,020	Imágenes de estudios previos			0,001
No (línea base)				No (línea base)			
Sí	5,718	1,310-24,958		Sí	6,291	2,107-18,783	
Información clínica de la solicitud			0,077	Información clínica de la solicitud			<0,001
Adecuada (línea base)				Adecuada (línea base)			
Insuficiente	1,615	0,950-2,745		Insuficiente	3,155	1,681-5,922	
Tipo de visor utilizado			0,153	Tipo de visor utilizado			0,005
Clear Canvas® (línea base)				Clear Canvas® (línea base)			
OsiriX®	1,745	0,814-3,741		OsiriX®	1,652	1,166-2,342	
Número de monitores			<0,001	Número de monitores			0,027
>=2 monitores (línea base)				>=2 monitores (línea base)			
Un monitor	3,024	1,738-5,261		Un monitor	1,471	1,046-2,071	
Uso sistema reconocimiento voz			0,022	Uso sistema reconocimiento voz			0,006
Sí (línea base)				Sí (línea base)			
No	2,678	1,154-6,218		No	3,943	1,492-10,421	
Edad radiólogo	0,940	0,913-0,968	<0,001	Edad radiólogo	0,980	0,966-0,994	0,005
Experiencia radiólogo	0,945	0,921-0,969	<0,001	Experiencia radiólogo	0,969	0,955-0,984	<0,001
Género del radiólogo			0,036	Género del radiólogo			<0,001
Masculino (línea base)				Masculino (línea base)			
Femenino	1,602	1,031-2,489		Femenino	1,911	1,410-2,590	
Horario valoración inicial			0,120	Horario valoración inicial			0,165
Laboral 8:00-17:59h (línea base)				Laboral 8:00-17:59h (línea base)			
Laboral Otro horario	1,593	0,973-2,608	0,064	Laboral Otro horario	1,232	0,849-2,788	0,273
Festivo	1,703	0,912-3,179	0,095	Festivo	0,850	0,601-1,204	0,361
Especialidad del estudio			0,023	Especialidad del estudio			0,010
Columna (línea base)				Neuro (línea base)			
MSK	2,200	1,247-3,883	0,007	Body	1,659	1,196-2,301	0,002
Neuro / Body	2,138	0,906-5,043	0,083	MSK / Columna	1,262	0,823-1,936	0,285
Concordancia de especialidades			<0,001	Concordancia de especialidades			<0,001
Alta (línea base)				Alta (línea base)			
Media / Baja	3,467	2,209-5,441		Media	3,591	2,170-5,944	<0,001
				Baja	2,085	1,154-3,767	0,015

Tabla 51. Modelos univariantes ajustados para los estudios programados y urgentes.

9.1.3. Análisis univariante de factores significativos en la totalidad de los estudios (programados y urgentes).

En la Tabla 52 se muestra la presencia de cada uno de los factores de interés sobre la muestra de 537 elementos con discrepancia diagnóstica y sobre la de 537 elementos sin discrepancia diagnóstica. También aparecen los resultados del ajuste univariante mediante un modelo de regresión logística de la probabilidad de discrepancia diagnóstica para los diferentes valores del factor de interés. Además de las especialidades el estudio y del radiólogo, se ha añadido el grado de concordancia entre ambas especialidades calculado mediante el procedimiento descrito en secciones anteriores. Se aprecia que todos los factores son factores significativos respecto a la presencia o ausencia de discrepancia diagnóstica desde el punto de vista univariante.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Complejidad del estudio					0,041
Normal	513 (95,5%)	497 (92,6%)	(línea base)		
Alta	24 (4,5%)	40 (7,4%)	1,720	1,022-2,896	
Modalidad del estudio					<0,001
Resonancia Magnética	174 (32,4%)	250 (46,6%)	(línea base)		
Tomografía Comp.	363 (67,6%)	287 (53,4%)	0,550	0,429-0,705	
Tipo de estudio					<0,001
Urgente	408 (76,0%)	307 (57,2%)	(línea base)		
Programado	129 (24,0%)	230 (42,8%)	2,370	1,824-3,078	
Calidad técnica de las imágenes					0,002
Buena	531 (98,9%)	527 (95,5%)	(línea base)		
Regular/Mala	6 (1,1%)	24 (4,5%)	4,140	1,679-10,212	
Imágenes de estudios previos					<0,001
No	531 (98,9%)	500 (93,1%)	(línea base)		
Sí	6 (1,1%)	37 (6,9%)	6,549	2,740-15,651	
Información clínica de la solicitud					<0,001
Adecuada	498 (92,7%)	442 (82,3%)	(línea base)		
Insuficiente	39 (7,3%)	95 (17,7%)	2,745	1,851-4,070	
Tipo de visor utilizado					<0,001
Clear Canvas®	136 (25,3%)	78 (14,5%)	(línea base)		
OsiriX®	401 (74,7%)	459 (85,5%)	1,996	1,465-2,718	
Número de monitores					<0,001
Dos o más monitores	159 (29,6%)	96 (17,9%)	(línea base)		
Un monitor	378 (70,4%)	441 (82,1%)	1,932	1,448-2,578	
Uso sistema reconocimiento voz					0,001
Sí	39 (7,3%)	15 (2,8%)	(línea base)		
No	498 (92,7%)	522 (97,2%)	2,725	1,484-5,006	
Edad radiólogo valoración inicial			0,967	0,955-0,980	<0,001
Experiencia radiólogo valoración inicial			0,963	0,951-0,975	<0,001
Género del radiólogo					<0,001
Masculino	353 (65,7%)	275 (51,2%)	(línea base)		
Femenino	184 (34,3%)	262 (48,8%)	1,828	1,430-2,337	
Horario valoración inicial					0,002
Laboral 8:00-17:59h	205 (38,2%)	179 (33,3%)	(línea base)		
Laboral Otro horario	156 (29,1%)	210 (39,1%)	1,542	1,156-2,057	0,003
Festivo	176 (32,8%)	148 (27,6%)	0,963	0,716-1,296	0,804
Especialidad del estudio					<0,001
Neuro	211 (39,3%)	130 (24,2%)	(línea base)		
MSK	125 (23,3%)	198 (36,9%)	2,571	1,880-3,515	<0,001
Body	141 (26,3%)	150 (27,9%)	1,727	1,257-2,371	0,001
Columna	60 (11,2%)	59 (11,0%)	1,596	1,048-2,431	0,029
Especialidad del radiólogo					<0,001
Neuro	88 (16,4%)	40 (7,4%)	(línea base)		
MSK	82 (15,3%)	60 (11,2%)	1,610	0,976-2,656	0,062
Body	74 (13,8%)	45 (8,4%)	1,338	0,790-2,264	0,278
General	293 (54,6%)	392 (73,0%)	2,943	1,967-4,405	<0,001
Concordancia de especialidades					<0,001
Alta	160 (29,8%)	94 (17,5%)	(línea base)		
Media	293 (54,6%)	392 (73,0%)	2,277	1,693-3,063	<0,001
Baja	84 (15,6%)	51 (9,5%)	1,033	0,671-1,591	0,881

Tabla 52. Relación de los factores latentes considerados con la discrepancia diagnóstica.

9.1.4. Análisis univariante de factores significativos en la submuestra de los estudios de telerradiología muestreados.

En la Tabla 53 se muestra la presencia de cada uno de los factores de interés sobre las muestras de elementos con discrepancia diagnóstica y sobre la muestra de elementos sin discrepancia diagnóstica. También muestra los resultados del ajuste univariante mediante un modelo de regresión logística de la probabilidad de discrepancia diagnóstica para los diferentes valores del factor de interés. Se observan los siguientes aspectos:

- Un mayor porcentaje de estudios con alta complejidad en los estudios con discrepancias diagnósticas (6,5%) que en los estudios sin discrepancias diagnósticas (3,5%), aunque esta diferencia no es estadísticamente significativa (p-valor = 0,185).
- Un mayor porcentaje de RM en los estudios con discrepancias diagnósticas (81,6%) que en los estudios sin discrepancias diagnósticas (76,5%), aunque esta diferencia no es estadísticamente significativa (p-valor = 0,197).
- Un mayor porcentaje de estudios con imágenes con una calidad técnica inadecuada en los estudios con discrepancias diagnósticas (6,1%) que en los estudios sin discrepancias diagnósticas (1,2%), con un incremento de la probabilidad de discrepancia diagnóstica de 5,5 veces de un estudio programado respecto a otro urgente (p-valor = 0,024).
- Un mayor porcentaje de estudios programados en los estudios con discrepancias diagnósticas (88,1%) que en los estudios sin discrepancias diagnósticas (76,3%), con un incremento de la probabilidad de discrepancia diagnóstica de 2,4 veces de un estudio programado respecto a otro urgente (p-valor = 0,001).
- Un mayor porcentaje de estudios con imágenes de estudios previos en los estudios con discrepancias diagnósticas (8,0%) que en los estudios sin discrepancias diagnósticas (1,8%), con un incremento de la probabilidad de discrepancia diagnóstica de 4,9 veces de un estudio con imágenes previas respecto a otro sin imágenes (p-valor = 0,011).
- Un mayor porcentaje de estudios con información clínica insuficiente en los estudios con discrepancias (25,7%) que en los estudios sin discrepancias (14,1%), con un incremento de la probabilidad de discrepancia diagnóstica de 2,1 veces de un estudio con información clínica insuficiente respecto a otro con información clínica adecuada (p-valor = 0,005).

- Un mayor porcentaje de estudios en los que se ha utilizado *OsiriX*[®] en los estudios con discrepancias diagnósticas (91,6%) que en los estudios con discrepancias diagnósticas (87,1%), aunque esta diferencia no es estadísticamente significativa (p-valor = 0,133).
- Un mayor porcentaje de estudios en los que se ha utilizado un único monitor en los estudios con discrepancias diagnósticas (88,1%) que en los estudios sin discrepancias diagnósticas (70,0%), con un incremento de la probabilidad de discrepancia diagnóstica de 3,2 veces de un estudio en el que se ha utilizado un único monitor respecto a otro en el que se han utilizado más de un monitor en la primera lectura (p-valor < 0,001).
- Un mayor porcentaje de estudios en los que no se han utilizado sistemas de reconocimiento de voz en los estudios con discrepancias diagnósticas (95,8%) que en los estudios sin discrepancias diagnósticas (91,2%), con una diferencia que roza la significación estadística (p-valor = 0,055).
- Un descenso medio estimado de la probabilidad de discrepancia diagnóstica estimado del 4,3% por cada año de edad de más que tiene el radiólogo (p-valor < 0,001).
- Un descenso estimado de la probabilidad de discrepancia diagnóstica estimado del 3,9% por cada año de experiencia de más que tiene el radiólogo (p-valor < 0,001).
- Un mayor porcentaje de estudios valorados inicialmente por radiólogos de género femenino en los estudios con discrepancias diagnósticas (47,1%) que en los estudios sin discrepancias diagnósticas (37,1%). Se aprecia que el género del radiólogo está relacionado significativamente con la presencia de discrepancia diagnóstica en el mismo (p-valor = 0,040), con un incremento estimado de la probabilidad de discrepancia diagnóstica de 1,5 veces de un estudio valorado por un radiólogo de género femenino respecto a otro valorado en la primera lectura por un radiólogo de género masculino.
- Un mayor porcentaje de estudios realizados en horario laboral en los estudios sin discrepancias diagnósticas que en los estudios con discrepancias diagnósticas (37,6% vs 28,0%), mientras que en los estudios con discrepancias diagnósticas predominan los estudios en horario no laboral (49,4% vs 40,6%), y con un porcentaje similar de estudios en días festivos. No obstante, las diferencias globales no son estadísticamente significativas (p-valor = 0,091), aunque sí se

detectan diferencias importantes entre los estudios en días laborales según si el horario es o no laboral.

- Un mayor porcentaje de estudios de Neurorradiología y Columna en los estudios sin discrepancias diagnósticas (18,8% y 20,0%, respectivamente) que en los estudios con discrepancias diagnósticas (7,7% y 11,1%, respectivamente), mientras que hay un menor porcentaje de estudios de MSK y Body en los estudios sin discrepancias diagnósticas (51,8% y 9,4%, respectivamente) que en los estudios con discrepancias diagnósticas (68,6% y 12,6%, respectivamente). Se observa que existe una relación significativa entre ambas características (p-valor < 0,001), con un incremento medio de la probabilidad de discrepancia diagnóstica de 3,3 veces de un estudio de MSK o de un estudio de Body respecto de uno de Neurorradiología.
- Un mayor porcentaje de estudios valorados por radiólogos especialistas de Neurorradiología (15,9%), musculoesquelético (32,4%) o Body (5,3%) en los estudios sin discrepancias diagnósticas que en los estudios con discrepancias diagnósticas (7,7%, 19,5% y 5,0%, respectivamente), mientras que hay un menor porcentaje de estudios valorados por radiólogos generales (46,5% frente a 67,8%). Se evidencia una relación significativa entre ambas características (p-valor < 0,001), con un incremento medio de la probabilidad de discrepancia diagnóstica de 3,0 veces de un estudio valorado por un radiólogo general respecto a uno valorado por un neurorradiólogo.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Complejidad del estudio					0,185
Normal	164 (96,5%)	244 (93,5%)	(línea base)		
Alta	6 (3,5%)	17 (6,5%)	1,904	0,735-4,932	
Modalidad del estudio					0,197
Resonancia Magnética	130 (76,5%)	213 (81,6%)	(línea base)		
Tomografía Comp.	40 (23,5%)	48 (18,4%)	0,732	0,456-1,175	
Calidad técnica de las imágenes					0,024
Buena	168 (98,9%)	245 (93,9%)	(línea base)		
Regular / Mala	2 (1,2%)	16 (6,1%)	5,486	1,245-24,171	
Tipo de estudio					0,001
Urgente	41 (24,1%)	31 (11,9%)	(línea base)		
Programado	129 (75,9%)	230 (88,1%)	2,358	1,411-3,942	
Imágenes de estudios previos					0,011
No	167 (98,2%)	240 (92,0%)	(línea base)		
Sí	3 (1,8%)	21 (8,0%)	4,871	1,430-16,593	
Información clínica de la solicitud					0,005
Adecuada	146 (85,9%)	194 (74,3%)	(línea base)		
Insuficiente	24 (14,1%)	67 (25,7%)	2,101	1,257-3,510	
Tipo de visor utilizado					0,133
Clear Canvas®	22 (12,9%)	22 (8,4%)	(línea base)		
OsiriX®	148 (87,1%)	239 (91,6%)	1,615	0,864-3,018	
Número de monitores					<0,001
Dos o más monitores	51 (30,0%)	31 (11,9%)	(línea base)		
Un monitor	119 (70,0%)	230 (88,1%)	3,180	1,932-5,233	
Uso sistema reconocimiento voz					0,055
Sí	15 (8,8%)	11 (4,2%)	(línea base)		
No	155 (91,2%)	250 (95,8%)	2,199	0,985-4,912	
Edad radiólogo valoración inicial			0,957	0,933-0,980	<0,001
Experiencia radiólogo valoración inicial			0,961	0,940-0,982	<0,001
Género del radiólogo					0,040
Masculino	107 (62,9%)	138 (52,9%)	(línea base)		
Femenino	63 (37,1%)	123 (47,1%)	1,514	1,020-2,247	
Horario valoración inicial					0,091
Laboral 8:00-17:59h	64 (37,6%)	73 (28,0%)	(línea base)		
Laboral Otro horario	69 (40,6%)	129 (49,4%)	1,639	1,050-2,558	0,030
Festivo	37 (21,8%)	59 (22,6%)	1,398	0,822-2,377	0,216
Especialidad del estudio					<0,001
Neuro	32 (18,8%)	20 (7,7%)	(línea base)		
MSK	88 (51,8%)	179 (68,6%)	3,255	1,761-6,015	<0,001
Body	16 (9,4%)	33 (12,6%)	3,300	1,457-7,475	0,004
Columna	34 (20,0%)	29 (11,1%)	1,365	0,647-2,880	0,414
Especialidad del radiólogo					<0,001
Neuro	27 (15,9%)	20 (7,7%)	(línea base)		
MSK	55 (32,4%)	51 (19,5%)	1,252	0,626-2,502	0,525
Body	9 (5,3%)	13 (5,0%)	1,950	0,698-5,451	0,203
General	79 (46,5%)	177 (67,8%)	3,025	1,601-5,714	0,001

Tabla 53. Relación de los factores latentes considerados con la discrepancia diagnóstica en la submuestra.

Como se observa en la Tabla 53, se han registrado la especialidad del radiólogo que ha valorado inicialmente el estudio y la especialidad del estudio. En la Tabla 54 se muestra el cruce de la especialidad del radiólogo y la especialidad del estudio a valorar para las dos muestras consideradas en este análisis. Se observa que en ambas muestras consideradas, los estudios de todas las especialidades son valorados mayoritariamente por radiólogos generalistas.

	Total	Especialidad del radiólogo			
		Neuro	MSK	Body	General
Estudios sin discrepancias					
Especialidad del estudio					
Neuro	32	8 (25%)	0 (0%)	2 (6,3%)	22 (68,8%)
MSK	88	0 (0%)	49 (55,7%)	1 (1,1%)	38 (43,2%)
Body	16	1 (6,3%)	1 (6,3%)	5 (31,3%)	9 (56,3%)
Columna	34	18 (52,9%)	5 (14,7%)	1 (2,9%)	10 (29,4%)
Estudios con discrepancias					
Especialidad del estudio					
Neuro	20	5 (25%)	1 (5%)	1 (5%)	13 (65%)
MSK	179	2 (1,1%)	49 (27,4%)	2 (1,1%)	126 (70,4%)
Body	33	0 (0%)	0 (0%)	10 (30,3%)	23 (69,7%)
Columna	29	13 (44,8%)	1 (3,4%)	0 (0%)	15 (51,7%)

Tabla 54. Relación entre la especialidad del radiólogo y la especialidad del estudio valorado en la submuestra.

En la Tabla 55 se aprecia el ajuste de un modelo de regresión logística univariante para modelizar la probabilidad de la presencia de discrepancia diagnóstica en un estudio de telerradiología en función de la relación de especialidades del estudio y del radiólogo que lo valora, donde se estima el OR asociado a cada una de las posibles combinaciones respecto a la de menor tasa de discrepancia. Hay que tener en cuenta que el número de observaciones para algunas de las combinaciones posibles es muy reducido, lo que dará lugar a estimaciones poco precisas del efecto de dichos valores sobre la presencia o ausencia de discrepancia diagnóstica.

	N	OR	IC 95%		p-valor
Combinación de especialidades					<0,001
Estudio Neuro + Radiólogo Neuro	13	(línea base)			
Estudio Columna + Radiólogo MSK	6	0,320	0,028	3,600	0,356
Estudio MSK + Radiólogo Body	3	-	-	-	-
Estudio Neuro + Radiólogo MSK	1	-	-	-	-
Estudio MSK + Radiólogo Neuro	2	-	-	-	-
Estudio Body + Radiólogo Body	15	3,200	0,680	15,069	0,141
Estudio Body + Radiólogo Neuro	1	-	-	-	-
Estudio Columna + Radiólogo Neuro	31	1,156	0,307	4,350	0,831
Estudio Neuro + Radiólogo Body	3	-	-	-	-
Estudio Neuro + Radiólogo General	35	0,945	0,255	3,507	0,933
Estudio MSK + Radiólogo MSK	98	1,600	0,489	5,235	0,437
Estudio Body + Radiólogo MSK	1	-	-	-	-
Estudio Columna + Radiólogo Body	1	-	-	-	-
Estudio Body + Radiólogo General	32	4,089	1,052	15,888	0,042
Estudio Columna + Radiólogo General	25	2,400	0,607	9,486	0,212
Estudio MSK + Radiólogo General	164	5,305	1,639	17,175	0,005

Tabla 55. Efecto de la relación de especialidades sobre la discrepancia diagnóstica en la submuestra.

A partir del tipo de estudio y del tipo de radiólogo que lo ha valorado se establecerá el grado de concordancia existente entre estudio-evaluador, en base al criterio siguiente:

- Grado de concordancia alta: 163 estudios
 - o 13 estudios de Neurorradiología informados por un neurorradiólogo.
 - o 98 estudios de MSK informados por un radiólogo musculoesquelético.
 - o 15 estudios de Body informados por un radiólogo de Body.
 - o 31 estudios de Columna informados por un neurorradiólogo.
 - o 6 estudios de columna informados por un radiólogo musculoesquelético.
- Grado de concordancia media: 256 estudios
 - o 35 estudios de Neurorradiología informados por un radiólogo general.
 - o 164 estudios de MSK informados por un radiólogo general.
 - o 32 estudios de Body informados por un radiólogo general.
 - o 25 estudios de Columna informados por un radiólogo general.
- Grado de concordancia baja: 12 estudios
 - o 3 estudios de Neurorradiología informados por un radiólogo de Body.
 - o 1 estudio de Neurorradiología informado por un radiólogo musculoesquelético.
 - o 2 estudios de MSK informados por un neurorradiólogo.
 - o 3 estudios de MSK informados por un radiólogo de Body.
 - o 1 estudio de Body informado por un neurorradiólogo.
 - o 1 estudio de Body informado por un radiólogo musculoesquelético.

- 1 estudio de Columna informado por un radiólogo de Body.

A partir de la recodificación descrita anteriormente se han clasificado los estudios, según la relación de la especialidad del estudio a valorar y del radiólogo que lo valora, en estudios con concordancia alta, media y baja. En la Tabla 56 se observa que en los estudios sin discrepancias diagnósticas hay un mayor porcentaje de estudios con concordancia alta (50,0%) o baja (3,5%) respecto a los estudios con discrepancias diagnósticas (29,9% y 2,3%, respectivamente), mientras que hay un menor porcentaje de estudios con concordancia media (46,5% frente a 67,8%) que han sido valorados por radiólogos generalistas. Se evidencia una relación significativa entre ambas características (p-valor < 0,001), con un incremento estimado de la probabilidad de discrepancia diagnóstica de 2,4 veces de un estudio con concordancia media respecto a uno con concordancia alta. En cambio no hay diferencias significativas entre los estudios con concordancia alta y baja, aunque el número de estudios de discrepancia alta analizados es muy reducido.

	Sin discrepancia	Con discrepancia	OR	IC (95%)	p-valor
Concordancia especialidades					<0,001
Alta	85 (50,0%)	78 (29,9%)	(línea base)		
Media	79 (46,5%)	177 (67,8%)	2,442	1,127-3,664	<0,001
Baja	6 (3,5%)	6 (2,3%)	1,090	0,337-3,520	0,886

Tabla 56. Relación entre la concordancia de especialidades estudio-radiólogo y la discrepancia diagnóstica en la submuestra.

9.2. ANEXO 2. RESUMEN DE RESULTADOS.

	PROGRAMADOS		URGENTES		TOTAL		SUBMUESTRA	
	UNIVARIANTE	MULTIVARIANTE	UNIVARIANTE	MULTIVARIANTE	UNIVARIANTE	MULTIVARIANTE	UNIVARIANTE	MULTIVARIANTE
Complejidad del estudio	X	X	✓	✓	✓	✓	X	X
Modalidad del estudio (TC/RM)	✓	✓	X	X	✓	X	X	X
Calidad técnica	✓	✓	X	X	✓	✓	✓	✓
Estudios previos	✓	✓	✓	✓	✓	✓	✓	✓
Información clínica	X	X	✓	✓	✓	✓	✓	✓
Tipo de visor	X	X	✓	X	✓	X	X	X
Número de monitores	✓	✓	✓	✓	✓	✓	✓	✓
Sistemas de reconocimiento de voz	✓	✓	✓	✓	✓	✓	X	X
Edad del radiólogo	✓	X	✓	X	✓	X	✓	X
Experiencia del radiólogo	✓	X	✓	X	✓	X	✓	X
Género del radiólogo	✓	X	✓	X	✓	X	✓	X
Horario de la primera lectura	X	✓	X	X	✓	✓	✓	X
Especialidad del estudio	✓	X	✓	✓	✓	✓	✓	✓
Especialidad del radiólogo	✓	*	✓	*	✓	*	✓	*
Concordancia entre especialidad del estudio y del radiólogo	✓	✓	✓	✓	✓	✓	✓	✓
Tipo de estudio (Urgente/Programado)							✓	✓

✓ Existe asociación estadísticamente significativa con discrepancia diagnóstica.

X No existe asociación estadísticamente significativa con discrepancia diagnóstica.

* Variable no incluida en el análisis multivariante por problemas de colinealidad.

9.3. ANEXO 3. DOCUMENTACIÓN DEL COMITÉ ÉTICO DE INVESTIGACIÓN EN HUMANOS DE LA COMISIÓN DE ÉTICA EN INVESTIGACIÓN EXPERIMENTAL DE LA UNIVERSITAT DE VALÈNCIA.



D. José María Montiel Company, Profesor Contratado Doctor Interino del departamento de Estomatología, y Secretario del Comité Ético de Investigación en Humanos de la Comisión de Ética en Investigación Experimental de la Universitat de València,

CERTIFICA:

Que el Comité Ético de Investigación en Humanos, en la reunión celebrada el día 4 de julio de 2016, una vez estudiado el proyecto de tesis doctoral titulado:

"Error diagnóstico en Radiología, análisis de variables explicativas", número de procedimiento H1460322382143,

cuyo responsable es D. Fernando Facal de Castro, dirigido por D. Estanislao Arana Fernández de Moya y Dña. Delfina Dualde Beltrán,

ha acordado informar favorablemente el mismo dado que se respetan los principios fundamentales establecidos en la Declaración de Helsinki, en el Convenio del Consejo de Europa relativo a los derechos humanos y cumple los requisitos establecidos en la legislación española en el ámbito de la investigación biomédica, la protección de datos de carácter personal y la bioética.

Y para que conste, se firma el presente certificado en Valencia, a cinco de julio de dos mil dieciséis.

