

MODELOS CLINICO-DOSIMETRICOS
PREDICTIVOS DEL CONTROL BIOQUIMICO
DEL CANCER DE PROSTATA
ORGANO-CONFINADO TRATADO MEDIANTE
IMPLANTES RADIATIVOS PERMANENTES

Alfredo Polo Rubio
Universidad de Barcelona
Departamento de Ciencias Clínicas
Año 2007

6. PACIENTES, MATERIAL Y METODOS

6.1 DESCRIPCION DE LAS SERIES DE PACIENTES USADAS EN ESTA TESIS Y DE LA TECNICA DE IMPLANTE

6.1.1 Serie 1: monoterapia con ^{125}I

Esta serie constituye el cuerpo de datos sobre los que se ha realizado la inferencia estadística y la construcción de los modelos clínico-dosimétricos predictivos. De los 200 pacientes tratados en el IEO entre 1999 y 2001 se han incluido en esta serie 113 pacientes tratados consecutivamente con braquiterapia exclusiva con semillas de ^{125}I para los cuales la dosimetría post-implante ha sido realizada por un mismo equipo constituido por un observador que ha diseñado los volúmenes de interés y un operador a identificar las semillas y calcular las dosis y los HDV correspondientes. No se han incluido los pacientes iniciales, tratados con la ayuda de un tutor, ni un grupo de pacientes cuya dosimetría post-implante fue realizada fuera del IEO. No se han incluido los pacientes tratados con ^{103}Pd . Por último, no se incluyeron los pacientes tratados con una combinación de radioterapia externa e implante permanente.

La técnica de implante será descrita a continuación separadamente, ya que es común para las tres series. El protocolo de dosimetría post-implante consistía en la adquisición de una TC inmediatamente tras el implante (en el mismo día, normalmente en las primeras dos horas post-implante) y una segunda TC 50 días post-implante. La razón para tomar un intervalo de 50 días fue puramente administrativa y tenía relación con el sistema de cálculo de DRG (“Diagnostic Related Groups”) y su facturación. Con el protocolo de doble dosimetría se pretendió desde el principio estudiar el efecto del edema sobre la dosimetría y estudiar la dinámica del mismo. Después de revisar la bibliografía relevante al tema, hemos comprobado que nuestra serie es la más amplia que recoge este tipo de información dosimétrica. Asimismo, aunque muchas series recogen datos dosimétricos del día del implante, y otras, datos a 30 días post-implante, no hemos encontrado ninguna serie que informe de resultados a día 0 y día 50.

Los datos recogidos para cada una de las series incluyen un conjunto de variables clínicas y un conjunto de variables dosimétricas. Entre las variables clínicas destacan las correspondientes al paciente, a la biología tumoral y al seguimiento post-implante. Los pacientes se agruparon en grupos de riesgo siguiendo la estratificación del MSKCC. Entre las variables dosimétricas se incluyen las recomendadas por el AAPM TG64 y por la ABS, que han sido revisadas anteriormente, para cada uno de los estudios dosimétricos post-implante. El fracaso bioquímico se ha definido siguiendo el criterio del consenso ASTRO (68). Hay que recordar que el criterio ASTRO-Phoenix (72) no estaba descrito en el momento en que acometimos el análisis de los datos de nuestros pacientes. No se ha recogido la recidiva local clínica, ni la aparición de metástasis ni la supervivencia global, ya que la hipótesis planteada y los objetivos concretos hacen referencia exclusivamente al control bioquímico de la enfermedad.

6.1.2 Serie 2: fusión de imágenes

Esta serie se construyó para analizar el efecto de la modalidad de imagen sobre los resultados dosimétricos. Para ello se incluyeron 23 pacientes tratados consecutivamente indistintamente con ^{125}I y ^{103}Pd . El protocolo de dosimetría post-implante incluía la realización de una TC post-implante (normalmente en la primera hora post-implante) y una RMN (normalmente dentro de las dos primeras horas post-implante). Posteriormente se realizó una fusión de imágenes en una estación de trabajo dedicada, ubicada en el Departamento de Radiología del IEO, y se transfirieron los datos en formato DICOM vía red hasta el Departamento de Oncología radioterápica, donde se realizó la delimitación de volúmenes y el cálculo dosimétrico. Para cada paciente se realizó un estudio dosimétrico sobre las imágenes procedentes de TC y otro sobre las imágenes procedentes de fusión TC-RMN, y se compararon entre sí. El equipo que llevó a cabo el estudio post-implante fue el mismo que para la serie 1. El método empleado está descrito con detalle en el artículo que nuestro grupo publicó en diciembre de 2004 (242).

6.1.3 Serie 3: estudio de CT seriados

A partir de la revisión de la bibliografía y del análisis preliminar de los resultados de los pacientes que se iban incluyendo en la serie 1 se hizo evidente que los volúmenes determinados mediante TC para el cálculo de la dosimetría post-implante presentaban diferencias evidentes respecto a los volúmenes pre-implante. Para estudiar este fenómeno e intentar caracterizar el fenómeno del edema y elaborar un modelo predictivo se decidió iniciar un pequeño estudio de TC seriados para estudiar el efecto del edema en la dosimetría post-implante. De los datos derivados de este estudio se intentará derivar un modelo predictivo de la dosimetría del día 50 a partir de los datos de la dosimetría del día 0 y se evaluará su utilidad práctica en el análisis de la supervivencia.

Se han incluido 5 pacientes en la serie 3. Los criterios de inclusión fueron: pacientes tratados con ^{125}I como tratamiento exclusivo (monoterapia) que no hubieran recibido tratamiento hormonal, cercanía a nuestro centro y firma de consentimiento informado. Se realizó la primera TC inmediatamente tras el implante, como era norma general. Posteriormente se realizó TC los días 7, 14, 21 y 28 post-implante. Se recogió también TC a día 50 tal como establecía el protocolo general del IEO. Ningún paciente presentó desviación respecto a los días de toma de TC. El número de pacientes es reducido pero dado que con este estudio no se seguía el criterio ALARA (“As Low As Reasonably Achievable”) se consideró oportuno evaluar la serie con sólo 5 pacientes. De la revisión de la (escasa) literatura relevante también se observa que las series publicadas incluyen un número reducido de pacientes por lo que se consideró suficiente a los efectos que se pretendía.

La dosimetría post-implante se realizó por el mismo equipo. Se delinearon los volúmenes de interés de manera ciega, de manera que el observador no sabía en ningún momento qué estudio estaba siendo contorneado. Se recogieron los datos generales y además los datos de volumen prostático, D90 y V100 para cada estudio TC (día 0, 7, 14, 21 y 28). Posteriormente se realizó un ajuste a un modelo exponencial, siguiendo los ejemplos presentes en la literatura revisada y se obtuvo el valor para el tiempo de semi-desintegración del edema prostático. Dicho valor será de gran utilidad para la caracterización del edema sobre la serie 1.

Este estudio presenta una gran utilidad para el propósito de caracterizar el edema post-implante. Sin embargo presenta una gran limitación que hay que explicar para estar en condiciones de valorar adecuadamente los resultados obtenidos. La medida del volumen pre-implante se obtuvo a partir de TRUS mientras que los volúmenes post-implante se midieron a partir de TC, lo que necesariamente conduce a una discrepancia que se debe tener en cuenta. Se profundizará en este aspecto en el epígrafe 7.3 correspondiente a la caracterización del edema mediante un modelo de regresión múltiple. Sin embargo, de nuevo, la revisión de la bibliografía nos indicó que la metodología del resto de grupos que publicaban resultados similares era la misma que habíamos aplicado nosotros.

6.1.4 Técnicas de braquiterapia empleadas en esta Tesis

En el epígrafe correspondiente se han revisado las diferentes técnicas de implante y se ha insistido por una parte en la necesidad de optimizar los flujos de trabajo sobre nuestros pacientes, y por otra parte la necesidad de evaluar dichas técnicas y su utilidad para aumentar el índice terapéutico de la Braquiterapia. Tal y como se ha mencionado, nuestra técnica de implante sufrió una serie de modificaciones durante el periodo de tiempo en que se reclutaron los pacientes que han conformado esta Tesis. Estas modificaciones se fueron solapando entre si, híbridándose, de forma que la técnica resultante es un “continuum” de pequeñas variaciones (software de planificación, abordaje intraoperatorio, navegación en tiempo real) acumuladas durante cinco años (figura 6.1).

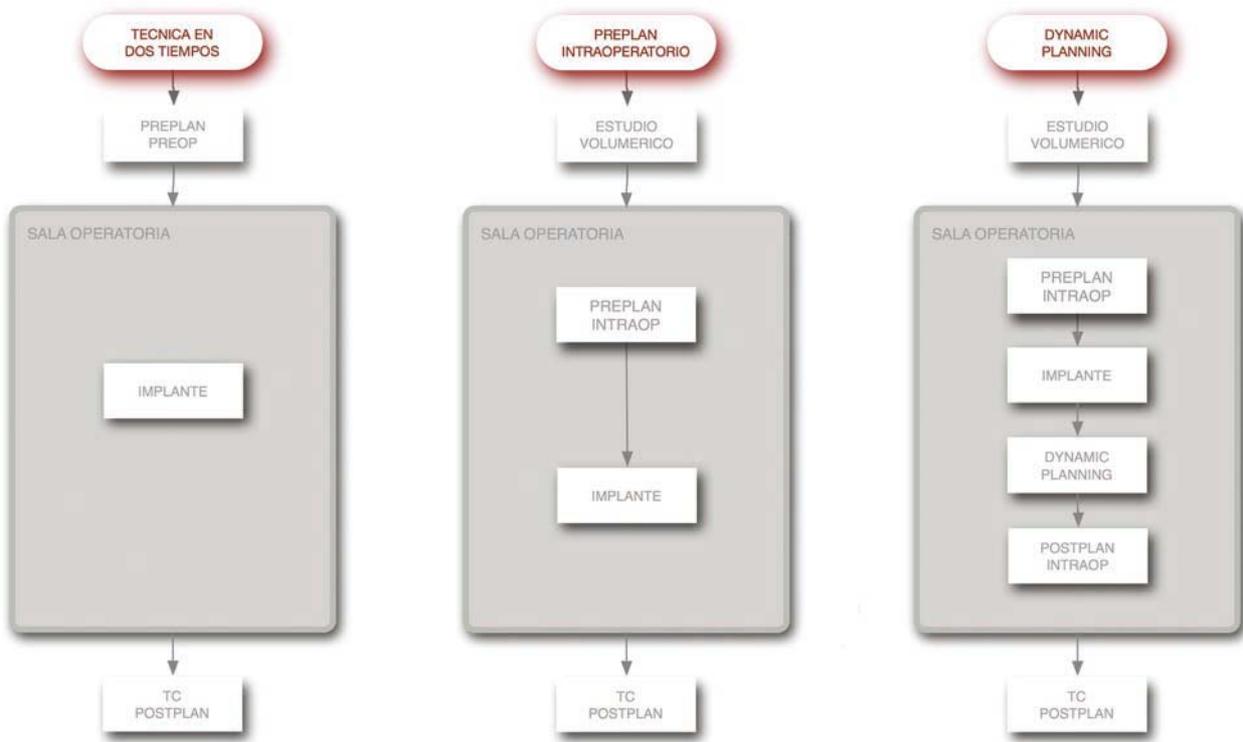


Figura 6.1 Se observa la evolución de la técnica, de izquierda a derecha: técnica en dos tiempos, técnica de preplanificación intraoperatoria y técnica de planificación dinámica intraoperatoria.

La dosimetría se ha realizado en todos los casos siguiendo la formulación propuesta por la AAPM en su informe del grupo de trabajo n° 43 (216, 288). La dosis prescrita ha sido de 145 Gy para el ^{125}I y de 115 Gy (Pre-NIST 99) para el ^{103}Pd . La evaluación y aceptación de la dosimetría previsual y de la dosimetría post-implante se ha realizado siguiendo las recomendaciones de la AAPM (TG 64) y de la ABS (215, 140). Los aspectos dosimétricos han sido revisados anteriormente en el epígrafe 4.3. Todos los volúmenes de interés han sido delimitados por el mismo operador (el autor de esta Tesis) y todos los cálculos de dosis, incluyendo la identificación de las semillas han sido realizados por un único operador (Dra. Cattani). De esta forma se asegura que al no existir más que un operador, las incertidumbres quedaban reducidas a la variabilidad intra-observador, que hemos supuesto que son mínimas. Las definiciones del volumen blanco se hicieron siguiendo el informe ICRU 58. El CTV (“clinical target volume”) se definió como el volumen prostático delimitado por ecografía transrectal pre-implante. No se realizó expansión de CTV por lo que [PTV = CTV]. El ETV (“evaluation target volume”) se definió como el volumen prostático delimitado por TC post-implante. Se añadió un sufijo al ETV (v.g ETV_0, ETV_50) para indicar el día en que se tomaron las imágenes TC. El ETV es un concepto *ad hoc* que no se encuentra en el informe ICRU 58.

∴ La primera técnica aplicada fue la conocida como técnica con pre-planificación (septiembre 1999 – mayo 2000): en primer lugar se realizaba un estudio volumétrico del paciente que constaba de una TC para la evaluación del arco púbico y una ecografía transrectal para planificación. El paciente entonces era dado de alta. Una vez obtenidas las imágenes ecográficas se digitalizaban y transferían al sistema de planificación (Plato®). Posteriormente se delimitaban los volúmenes de interés (próstata sin márgenes, uretra y pared anterior del recto) y se realizaba una dosimetría previsual, con la que se realizaba la orden de las semillas. Al cabo de varias semanas (normalmente eran cuatro allá por 1999) el paciente ingresaba de nuevo para realizar el implante siguiendo la dosimetría previsual. Para ello se habían anotado cuidadosamente las características de posicionamiento del paciente (en la mayoría de los casos se incluían fotografías Polaroid® del mismo) y se trataba de reproducir fielmente dicha posición. Posteriormente se realizaba el implante usando agujas que habían sido precargadas con las semillas correspondientes a la dosimetría previsual (semillas de ^{103}Pd de Theraseed® o semillas de ^{125}I en forma de RAPIDstrand®). El implante era controlado en tiempo real mediante fluoroscopia y ecografía transrectal (imagen transversa). Tras el implante se realizaba una cistoscopia de verificación y se extraían semillas alojadas en vejiga, si procedía.

Sobre la técnica de pre-planificación se realizaron tres modificaciones simples: por un lado se incorporó un software que permitía transferir las imágenes ecográficas directamente al sistema de planificación sin necesidad de un digitalizador (lo que suponía comodidad, ahorro de tiempo y rapidez de ejecución de la dosimetría previsual). Por otro lado comenzamos a usar con más profusión las imágenes ecográficas procedentes del transductor longitudinal, lo que nos daba una información de profundidad que complementaba la de la fluoroscopia, y nos permitió iniciar el concepto de “navegación” en tiempo real, y por último decidimos suprimir la cistoscopia del procedimiento tras comprobar que su rendimiento había sido escasísimo.

∴ La segunda técnica aplicada fue la conocida como pre-planificación intraoperatoria (mayo 2000 – septiembre 2001): el estudio volumétrico preoperatorio dejó de realizarse.

En su lugar se tomaba una ecografía transrectal simple (en nuestro Departamento o por parte del Centro de referencia). Sobre dicha base y con ayuda de un nomograma que recogía nuestra experiencia pasada se realizaba la orden de las semillas. Posteriormente, en el acto operatorio se realizaba una pre-planificación con ejecución inmediata de la misma, gracias al nuevo sistema de planificación portátil (Variseed®). No era necesario un re-posicionamiento del paciente por lo que el procedimiento resultaba más ágil. Se utilizaban agujas precargadas con semillas de ^{125}I (se abandonó el ^{103}Pd por distintas razones, entre ellas las dificultades en la logística de su distribución en virtud de su vida media más corta) que debían ser preparadas simultáneamente a la ejecución del implante. La guía del procedimiento se realizaba igualmente con la ayuda de la fluoroscopia y la ecografía transrectal.

•• La tercera técnica aplicada fue la conocida como planificación dinámica que se simultaneó con la técnica anterior desde septiembre de 2001. Dicha técnica se basa en la utilización de un nuevo sistema integrado de implante conocido como FIRST®, que integra un sistema de planificación (SPOT®), un sistema de navegación sobre imagen ecográfica tridimensional reconstruida en tiempo real y un sistema de carga diferida de semillas libres (SeedSelectron®). La secuencia es la siguiente (todo en un solo acto operatorio de 2h de duración): realización de pre-planificación y dosimetría previsual, preparación de las agujas precargadas, implante y planificación dinámica (captura de la posición de las agujas y re-cálculo de dosis). La planificación dinámica permite comprobar la progresión del implante y la necesidad de realizar modificaciones sobre el plan previsto para ajustar defectos de cobertura, presencia de zonas frías o sobredosis en la proximidad de los órganos críticos. Mediante la planificación dinámica se tiene otra perspectiva sobre el implante y la experiencia del operador es más satisfactoria. Gracias al sistema de navegación la utilización de la fluoroscopia se hizo innecesaria y también desaparecieron determinados problemas de radioprotección que su uso intensivo podría plantear.

En el DVD acompañante a esta Tesis se presenta la técnica de implante permanente que se utiliza actualmente en el Instituto Catalán de Oncología. Dicha técnica es la culminación de todos los esfuerzos combinados de un equipo constituido por anesestesiólogos, urólogos, físicos, enfermeras, técnicos de radioterapia y braquiterapeutas y representa una forma de realizar dichos implantes. Es imposible describir “la” técnica que se ha empleado al realizar nuestra tesis, por que como hemos dicho, es la suma de múltiples aportaciones humanas y tecnológicas.

La sucesión de técnicas aplicadas durante el periodo de reclutamiento de los pacientes que han servido para elaborar esta Tesis no introduce sesgos ni resta validez a nuestros resultados. Y ello se debe a que, independientemente de la metodología utilizada para realizar el implante, la metodología para la obtención de imágenes, delimitación de volúmenes de interés y cálculo de la dosimetría post-implante se ha mantenido constante a lo largo del tiempo que ha durado la elaboración de esta Tesis. Esto añade un valor a sus resultados y evita que se pueda apelar a diferencias de técnica para explicar los resultados clínicos, ya que las diferentes técnicas se analizaron siguiendo un protocolo común.

6.2 PLATAFORMA DE ANALISIS DE DATOS

Para la elaboración de nuestra tesis hemos realizado un esfuerzo para la creación de una plataforma de análisis de datos multi-propósito, reutilizable y escalable a partir de un conjunto de aplicaciones informáticas registradas y con licencia de utilización. Aunque no se trata del objetivo principal de nuestro trabajo, esta plataforma de análisis es uno de los aspectos más trabajados y de los que estamos más satisfechos. Y ello por dos motivos: en primer lugar por que es independiente de los datos que se quieran analizar. La plataforma se comportaría igual si fuera alimentada con datos de un estudio de cáncer de próstata, o mama o cualquier otra localización. Esto permite su aplicación en multitud de estudios distintos dentro de nuestros servicios asistenciales. Permite llevar simultáneamente el seguimiento de varias tesis doctorales y permite simulación de alto nivel sobre datos clínicos reales actualizados en tiempo real. Por otra parte, una plataforma de este tipo demuestra que se puede obtener un grado de autonomía importante en el análisis de datos clínicos que permite un ciclo de producción científica (recogida de datos – análisis – generación de hipótesis – publicación - recogida de datos...) más corto y compacto. Por todo ello nuestra plataforma de análisis de datos es un producto sobre el cual estamos trabajando para dotarlo de autonomía más allá de los límites temporales impuestos por este trabajo.

En esta sección analizaremos los módulos de que se compone nuestra plataforma de análisis y los flujos de trabajo que hemos establecido con ella (figura 6.2). Por último plantearemos los desarrollos futuros que pretendemos alcanzar con su aplicación.

∴ El primer módulo es el repositorio de datos: en nuestro caso se usó FileMaker8®. Se trata de una base de datos relacional con una interface de usuario muy intuitiva y capacidades de programación y “script” avanzadas. Las bases de datos relacionales creadas pueden ser alimentadas directamente o bien sincronizarse con otros programas como Microsoft Excel® o Microsoft Access® a través de AppleScript® o Automator®. En nuestro caso, la construcción de la base de datos fue distribuida. Cada parte de la misma fue compilada por una persona distinta, realizándose un corte temporal en diciembre de 2005 para el análisis de la supervivencia. En su versión más avanzada (y que no tenemos implementada todavía) los módulos de análisis de datos o de simulación podrían interrogar a la base de datos en tiempo real, y el ciclo de análisis se acortaría sensiblemente.

∴ La simulación se realizó usando Wolfram Matemática 5.2®. Se trata de un paquete de software científico multi-función que integra un motor de cálculo numérico y simbólico, un lenguaje de programación, un sistema de gestión de gráficos y una amplia conectividad con otros programas, en particular a través del protocolo JDBC con Filemaker8® (figura 6.3). La comunidad científica utiliza ampliamente este paquete (figura 6.4), aunque dentro de la comunidad de la Radioterapia no es muy utilizado, siendo las soluciones basadas en Matlab® mucho más populares y existiendo un repositorio de aplicaciones de libre disposición^(289, 290, 291, 292). En nuestro caso se ha utilizado este paquete para calcular los modelos que caracterizan el edema, la obtención del periodo de semidesintegración del edema y las simulaciones correspondientes.

∴ El análisis estadístico se realizó usando JMP6®. Se trata de un producto muy avanzado, con una interface muy intuitiva, que permite la realización del análisis estadístico clásico (integrable con distintos repositorios de datos como Microsoft Access® o Filemaker8® a través de protocolos ODBC/JDBC) y la implementación



MATERIAL
Y METODOS

Figura 6.2 Plataforma de análisis de datos y ciclo de producción de nuestra tesis. El núcleo central lo constituye el repositorio de datos (Filemaker Pro ®) y las aplicaciones que acceden a él. Por un lado Mathematica 5.2 ® es un software de simulación para el cálculo de modelos y JMP6 ® es un paquete estadístico de alto nivel para simulación y construcción de modelos estadísticos. Los tres módulos se comunican a través de protocolos ODBC/JDBC. El resto de flujos confluyen en la generación de un documento científico final.

de los modelos predictivos utilizados en esta Tesis. Asimismo dispone de un lenguaje propio de guiones (“scripting”) que le permite automatizar parte de las tareas o integrarlas en una forma de análisis en tiempo real.

•• La gestión de la bibliografía se realizó con Endnote7®. La confección de la memoria se realizó con Microsoft Word®.

•• La integración entre los distintos módulos se realizó mediante AppleScript® y Automator®, que son lenguajes propios del sistema operativo Mac OS X.

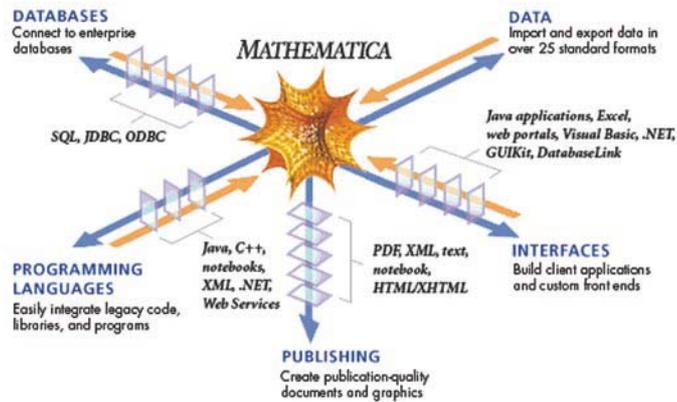


Figura 6.3 El entorno de Mathematica ® permite realizar una gran variedad de operaciones sobre bases de datos, crear aplicaciones a medida, automatizar tareas, realizar análisis gráfico y científico y ayudar a confección del producto final.

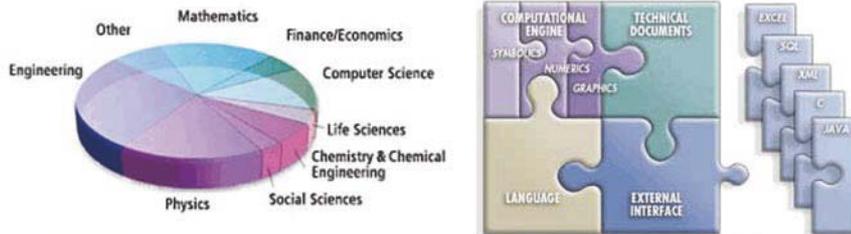


Figura 6.4 A la izquierda se puede ver la penetración de Mathematica ® en las diferentes disciplinas del conocimiento. A la derecha se puede ver la estructura del paquete y las posibilidades de conectividad con el mundo exterior.

6.3 METODOS ESTADISTICOS

Bajo esta sección describiremos los métodos estadísticos utilizados en nuestro trabajo. Los hemos dividido en dos grandes grupos: métodos tradicionales de análisis y métodos de inteligencia artificial. Tanto la clasificación como la nomenclatura utilizada son arbitrarios pero nos ayuda a diferenciar los dos grandes grupos de herramientas que hemos utilizado.

6.3.1 Métodos tradicionales

Análisis descriptivo de las variables:

Para el análisis descriptivo se ha utilizado la plataforma distribución (“Distribution platform”) de JMP6. Se ha realizado un análisis descriptivo de las variables, tanto cualitativas como cuantitativas, en los grupos de pacientes estudiados. Las variables cuantitativas se han expresado según la media y la desviación estándar (DE) en el caso de que se ajustaran a una distribución normal (se ha usado el test de Shapiro-Wilk para calcular la bondad y significación del ajuste de las poblaciones a una distribución normal), o en su defecto por la media, mediana, intervalo de confianza del 95% (IC95%) y rango (valor máximo y mínimo). Para las variables cualitativas se presentarán las frecuencias de las diferentes categorías.

Inferencia estadística:

•. Comparación de variables: para examinar la distribución de una variable de respuesta categórica Y en función de los valores de una variable categórica X se ha empleado la plataforma de Análisis de la Contingencia (“Contingency platform”) de JMP6. La tabla de frecuencias se obtiene mediante el gráfico de mosaico (“mosaic plot”) y los valores numéricos se presentan en la tabla de contingencia (“Contingency table”). El análisis de independencia se realiza mediante el test de χ^2 (“chi-cuadrado”) para dos tests de hipótesis (“Likelihood ratio” y test de Pearson). Se calculó asimismo el valor de R^2 , que indica que porción de la incertidumbre estadística es explicada por el modelo ($R^2 = 1$) y que parte por el azar ($R^2 = 0$). Para tablas de 2 x 2 el análisis de independencia se realizó mediante el test exacto de Fisher y se reportaron los valores de riesgo relativo y “odds ratio”.

El análisis de la distribución de una variable continua Y entre distintos grupos de población definidos por una variable categórica X se realizó mediante la plataforma “Oneway” de JMP6, que calcula el análisis de la varianza (ANOVA) para determinar si hay diferencias entre las medias. El análisis de independencia se realizó con test paramétricos (t-Student, Tukey-Kramer-HSD y Hsu-MCB) y no paramétricos (test de Wilcoxon para dos grupos y test de Kruskal-Wallis para más de dos grupos). La bondad de ajuste se midió igualmente con el valor de R^2 , que indica que porción de la incertidumbre estadística es explicada por el modelo ($R^2 = 1$) y que parte por el azar ($R^2 = 0$). La efectividad final del modelo se calculó con el F-ratio y su probabilidad asociada (si es pequeña, el modelo explica mejor la realidad que la media de la población).

El análisis de la distribución de una variable continua Y respecto a una variable continua X se realizó mediante la plataforma bivariada (“Bivariate platform”) de JMP6. Se obtuvo un diagrama de puntos (“scatterplot”) para cada par de variables estudiado y a partir de él se calcularon los coeficientes de regresión (una variable dependiente y una variable independiente) y de correlación de Pearson (dos variables dependientes).

La bondad de ajuste del modelo de regresión se calculó con el valor de R^2 , con la misma interpretación dada anteriormente. La efectividad final del modelo de regresión se calculó con un F-ratio y su probabilidad asociada, con la misma interpretación dada anteriormente. Para el modelo de correlación, se calculó la probabilidad asociada al coeficiente de Pearson (probabilidad de obtener por azar una correlación mayor a la dada por el coeficiente de Pearson sin que exista efectivamente una relación lineal entre las variables X e Y).

Por último, se utilizó el análisis de regresión logística simple contenido en la plataforma logística (“Logistic platform”) de JMP6 para analizar la relación entre una variable Y categórica (nominal u ordinal) y una variable X continua. La regresión logística nominal estima la probabilidad de obtener uno de los niveles de respuesta en Y como función suavizada de un factor X. Se obtuvo una gráfica con las probabilidades representadas en el eje-y y los valores en el eje-x. Para cada valor de x, la escala de probabilidad en la dirección del eje-y se divide en varias probabilidades, una para cada categoría de Y, con una curva para cada una de las categorías. Esta gráfica da una idea global del modelo. La bondad de ajuste del modelo se calculó con el valor de R^2 , con la misma interpretación dada anteriormente. La efectividad final del modelo se calculó con un χ^2 (“chi-cuadrado”) y su probabilidad asociada (llamada valor p del χ^2), que es la probabilidad de encontrar por azar, valores de χ^2 mayores del computado. Para obtener valores de corte (umbral para el valor de X a partir del cual una condición Y existe o no) se obtuvo la curva ROC (“Receiver Operating Characteristic”).

•• Estudio de la supervivencia: Se utilizó la plataforma de análisis de supervivencia de JMP6 (“Survival and Reliability Platform”). La función de supervivencia para el grupo general y para los distintos subgrupos analizados se calculó usando el método de Kaplan-Meier (“product-limit method”). Las comparaciones entre subgrupos y su homogeneidad se calcularon mediante los estadísticos χ^2 para el Log-Rank y el test de Wilcoxon (Log-Rank tiene más valor para tiempos de supervivencia largos mientras que el test de Wilcoxon estima mejor para tiempos de supervivencia cortos). El análisis univariado se completó con el análisis de causas (“estimation of competing causes”). Los cálculos de supervivencia se realizaron a partir del día del implante.

El análisis de supervivencia se completó con el estudio multivariado. Para ello se ajustó el modelo de riesgos proporcionales de Cox (“Cox’s Proportional Hazards Model”) para examinar el efecto de determinadas variables explicatorias en los tiempos de supervivencia.

6.3.2 Métodos de Inteligencia Artificial

Recursive Partitioning Analysis (RPA)

•• Concepto: el RPA es una técnica de clasificación que puede aplicarse al análisis de grandes bases de datos para descubrir relaciones ocultas entre las variables y para definir subgrupos con diferencias estadísticamente significativas entre ellos. En términos generales, RPA es un método de análisis de datos para relacionar una variable “dependiente” (Y) con una serie de variables “independientes” (X) también llamadas “predictores”. El RPA puede ser utilizado para establecer relaciones de significación o bien para realizar una exploración de los datos previa a la modelización con otras herramientas estadísticas. Está especialmente indicado en escenarios donde las técnicas estadísticas clásicas (v.g. regresión) no pueden establecer predicciones

adecuadas, si existen datos ausentes o la relación entre las variables es compleja y difícil de interpretar (293, 294, 295).

•. RPA en Medicina y Oncología: el RPA es una metodología estadística bien establecida y se aplica a problemas de clasificación en oncología desde hace 20 años. El RPA se ha utilizado de forma amplia para el análisis de la supervivencia en los estudios de el RTOG de glioma maligno y de metástasis cerebrales (296, 297, 298, 299). Además, el RTOG ha utilizado el RPA para definir subgrupos de pacientes con pronóstico similar en cáncer de cabeza y cuello y cáncer de pulmón no-célula pequeña (300, 301, 302). También se ha usado el RPA para la estimación de supervivencia en casos de carcinoma metastásico de origen desconocido (303), esófago (304) y para el análisis de la recidiva en cáncer de mama tras el tratamiento combinado de radioterapia y quimioterapia (305, 306). Numerosos trabajos recientes han reportado la utilización del RPA en el análisis de “microarrays” de ADN en cáncer de colon (307, 308).

•. RPA en cáncer de próstata: en el caso del cáncer de próstata el RPA se ha utilizado en varios frentes. Determinados autores la han usado como herramienta de predicción del hallazgo de cáncer en biopsia prostática (309, 310). Se ha usado para explorar la cinética del PSA tras la administración de RTE (311). Otros autores han usado RPA para explorar el efecto dosis (312, 313), y para establecer grupos de riesgo (311, 313, 314, 315, 316, 317, 318). El RPA también se ha usado en casos de enfermedad hormono-resistente para establecer grupos de supervivencia (319).

•. Algoritmo de partición: el algoritmo usado por JMP6 es una variedad del CART (“classification and regression tree analysis”). El resultado del RPA es un diagrama de árbol (también conocido como “árbol de decisión”) en el cual los datos se organizan mediante divisiones (particiones o “splits”) en nodos a lo largo de las ramas. Los datos similares se localizan en el mismo nodo, mientras que los datos sin relación ocupan nodos diferentes. El algoritmo de partición es simple e intuitivo. En cada paso se determina para cada variable X un valor de corte que divide de forma óptima el grupo en dos subgrupos y selecciona la variable que mejor ajuste el valor del criterio de partición (G2). Todas las variables independientes son evaluadas analizando cada valor de corte posible. En el paso sucesivo se toman los subgrupos y se repite la operación para cada uno de ellos hasta que no se puede realizar una partición sucesiva, bien por que el subgrupo contiene una sola clase de individuos o porque es demasiado pequeño para ofrecer una subdivisión con valor estadístico.

•. Evaluación del árbol: una vez generado el árbol de decisión se puede analizar la bondad de ajuste. Para variables dependientes categóricas (recidiva vs. no-recidiva) se usa el ROC (“Receiver Operating Characteristic”), que es la representación gráfica de la sensibilidad (eje y) por (1-especificidad) en el eje x. Se calcula el área bajo la curva obtenida que oscila entre 0.5, que indica falta de valor predictivo y 1, que indicaría una predicción perfecta para el árbol de decisión propuesto. La clasificación propuesta puede ser asimismo evaluada mediante un método de validación cruzada (“cross-validation”) para evitar el efecto del fenómeno de sobre-ajuste (“overfitting”) y la pérdida de utilidad del modelo en la población real. La validación puede realizarse por dos métodos: “k-fold-cross-validation” y “random subsampling” (usamos los términos ingleses por que son de uso muy común y es difícil una traducción concisa. Sus equivalentes en castellano serían, para el primero: validación cruzada usando k-subgrupos y para el segundo: validación cruzada usando subgrupos aleatorios. El

primero de los métodos asigna aleatoriamente todas las filas no excluidas a uno de k grupos y calcula el error para cada punto usando medias o tasas estimadas para todos los grupos excepto aquél. El resultado es un coeficiente de correlación R^2 entre el grupo general y el grupo de validación. En el segundo método se emplea una parte de los datos para computar el modelo (subgrupo de entrenamiento) y el resto de datos para validarlo (subgrupo de validación). El resultado es un coeficiente de correlación R^2 entre el grupo de entrenamiento y el de validación.

∴ Generación de reglas: un subproducto interesante de los árboles de decisión es la generación de reglas del tipo “si...entonces” que cubran todo el espectro de casos. Una regla de clasificación, $r = [a,c]$ consta de un antecedente (si...) y de un consecuente (entonces...). Cada antecedente contiene un predicado que puede ser evaluado como verdadero o falso en cada elemento de la base de datos. Estas reglas están relacionadas directamente con el árbol de decisión creado. Un árbol de decisión siempre puede ser usado para generar un grupo de reglas pero no son equivalentes. El árbol contiene información acerca del orden en que se ha realizado la partición (las reglas no contienen orden). Por otra parte el árbol se crea valorando todas las variables predictoras a la vez, mientras que las reglas examinan una clase cada vez.

Redes Neuronales Artificiales (“Artificial Neural Networks”):

∴ Concepto: Utilizaremos indistintamente el acrónimo “ANN” y los términos “redes neuronales”, “redes neuronales artificiales” o “modelos neuronales”. ANN son una técnica de clasificación, al igual que RPA que puede aplicarse al análisis de bases de datos para descubrir relaciones ocultas entre las variables y para definir subgrupos con diferencias estadísticamente significativas entre ellos. En términos generales, ANN se usan para predecir una o más variables respuesta o “dependientes” (Y) a partir de una red flexible de funciones de variables de entrada o “independientes” (X). ANN puede ser utilizadas para establecer relaciones de significación o bien para realizar una exploración de los datos previa a la modelización con otras herramientas estadísticas. Está especialmente indicado en escenarios donde no es necesario conocer la forma de la superficie de respuesta o la relación entre las variables es compleja y difícil de interpretar (294, 295).

∴ ANN en Medicina y Oncología: la aplicación de redes neuronales en Medicina data prácticamente desde su invención en 1943. Una revisión entre 1990 y 1997 cifró en 2000 artículos aproximadamente haciendo mención a la aplicación de redes neuronales en medicina (320). Prácticamente todas las áreas de especialidad han sido objeto de evaluación mediante redes neuronales. Particularmente interesante es su aplicación en el ámbito de la radiología para el reconocimiento de patrones, filtrado de ruidos y detección de bordes en tumores cerebrales, de mama y próstata (321, 322, 323). En Oncología prácticamente todas las localizaciones han sido objeto de estudio mediante redes neuronales. ANN se han utilizado de forma amplia para el análisis de la supervivencia en cáncer de mama, cáncer de cabeza y cuello y otras localizaciones (324, 325, 326, 327, 328). También se han utilizado para predecir el riesgo de toxicidad asociado al tratamiento con radioterapia (329).

∴ ANN en cáncer de próstata: el cáncer de próstata parece un terreno propicio para la aplicación de las redes neuronales (330). En una revisión multicéntrica comparando 28 estudios, los modelos neuronales se comportaban tan bien como los modelos logísticos en el 50% de los casos y en 7 de los 8 estudios mayores, con una muestra

de más de 5000 pacientes (331). ANN se han aplicado a la detección precoz del cáncer de próstata a partir del valor de PSA y de su fracción libre y de otros factores clínicos (332, 333). Una segunda aplicación de las redes neuronales ha sido la predicción del estadio patológico post-quirúrgico a partir de variables pre-quirúrgicas (334). Por último, también se han usado como herramientas predictivas de recidiva bioquímica tras el tratamiento del cáncer de próstata (335, 336, 337, 338), incluyendo la braquiterapia (151).

•. Algoritmo de clasificación: al igual que para el RPA, el resultado de la ANN es un modelo gráfico para representar el modelo, que puede aplicarse a distintos grupos de datos. Este gráfico consta de unos nodos de entrada, nodos de salida y nodos intermedios, que son los que realizan los cálculos. Cada nodo tiene asociada una función que determina como se comporta dicho nodo ante la información entrante. Las funciones más comunes son: umbral, sigmoidea y de campana. Cada conexión entre nodos aporta un peso a dicha relación. Tanto las funciones como los pesos se pueden modificar de forma supervisada o no supervisada. Para realizar la predicción un elemento de la base de datos entra a través de los nodos de entrada y tras su procesamiento, sale por un nodo de salida correspondiente a la predicción. A diferencia del RPA que posee un solo punto de entrada, la ANN posee tantos nodos de entrada como variables “independientes” (X) se quieran introducir en el modelo. A diferencia del RPA, la ANN puede cambiar su motor de clasificación tras cada interacción (aunque no cambie la forma del diagrama) modificando las funciones de activación de cada nodo, permitiendo así un proceso de aprendizaje de la ANN.

•. Evaluación de la ANN: los modelos neurales por lo general están sobre-ajustados. Esto quiere decir que ajustan perfectamente a los datos empleados en su construcción pero a veces fallan al realizar la predicción sobre otro grupo de datos en la población real. Por ello no resulta conveniente elegir los modelos por su bondad de ajuste (ROC) sino que deben evaluados mediante un método de validación cruzada (“cross-validation”) para evitar dicho fenómeno de sobre-ajuste. La validación puede realizarse por dos métodos: “k-fold-cross-validation” y “random holdback” (usamos los términos ingleses por que son de uso muy común y es difícil una traducción concisa. Sus equivalentes en castellano serían, para el primero: validación cruzada usando k subgrupos y para el segundo: validación cruzada usando subgrupos para training y validación. El primero de los métodos funcional igual que el algoritmo del mismo nombre de la plataforma de RPA y es ideal en situaciones donde la cantidad de datos no es muy grande para permitir una partición de los mismos en dos grupos independientes. El resultado es un coeficiente de correlación R^2 entre el grupo general y el grupo de validación. En el segundo método se emplea una parte de los datos (por defecto en JMP6, el 66.67%) para computar el modelo (subgrupo de entrenamiento) y el resto de datos para validarlo (subgrupo de validación). Este método es muy útil en el caso de bases de datos muy grandes. El resultado es un coeficiente de correlación R^2 entre el grupo de entrenamiento y el de validación.