



TRABAJO FIN DE MÁSTER EN BIOESTADÍSTICA

PROYECTO KILISUN

Junio 2018

Autores: Miguel Ángel Calvo Walling y Onintze Zaballa Larumbe

Tutores: José Luis Brita-Paja Segoviano

Pedro Girón Daviña

AGRADECIMIENTOS

Este Trabajo de Fin de Máster ha sido posible gracias a la ayuda y apoyo de varias personas a las que nos gustaría mostrar todo nuestro agradecimiento.

De manera especial, a Pepa García de Polavieja, por su capacidad para guiarnos, no solamente en el desarrollo de este proyecto, sino también en nuestra formación como bioestadísticos. Una pieza clave del buen trabajo grupal realizado y sin duda, una persona referente para nosotros.

A Mafalda Soto, por darnos la oportunidad de formar parte en un proyecto tan fascinante y benéfico como Kilisun.

Y finalmente a nuestros tutores, Pedro Girón y José Luis Brita-Paja, por su disponibilidad y confianza en nosotros a lo largo de todo este trabajo.

INDICE DE CONTENIDOS

| | |
|--|-----|
| AGRADECIMIENTOS..... | ii |
| INDICE DE CONTENIDOS..... | iii |
| INDICE DE FIGURAS | v |
| INDICE DE TABLAS | vii |
| RESUMEN | 1 |
| 1. INTRODUCCIÓN Y OBJETIVOS..... | 2 |
| 1.1 ALBINISMO EN ÁFRICA | 3 |
| 1.2 PROYECTO KILISUN..... | 4 |
| 1.3 OBJETIVOS | 6 |
| 2. FUENTES DE DATOS..... | 8 |
| 2.1 CUESTIONARIO DE KILISUN CARE..... | 8 |
| 2.2 BASES DE DATOS | 9 |
| 3. EVALUACIÓN DE LA CALIDAD DE LAS BASES DE DATOS..... | 12 |
| 3.1 DEPURACIÓN DE LA BASE DE DATOS | 12 |
| 3.1.1 OBTENCIÓN DE DATOS TÉCNICAMENTE CORRECTOS..... | 13 |
| 3.1.2 OBTENCIÓN DE DATOS CONSISTENTES..... | 15 |
| 3.2 EVALUACIÓN DE LAS INCONSISTENCIAS | 16 |
| 3.3 ANÁLISIS DE LA REALIDAD | 20 |
| 4. METODOLOGÍA..... | 22 |
| 4.1 ANÁLISIS ESTADÍSTICO | 22 |
| 4.1.1 ANÁLISIS DE SUPERVIVENCIA | 22 |
| 4.1.2 EVALUACIÓN DEL SEGUIMIENTO | 26 |
| 4.2 PLATAFORMA DE RECOGIDA DE DATOS | 26 |
| 4.2.1 REDISEÑO DE LA BASE DE DATOS..... | 26 |
| 4.2.2 EXPORTACIÓN DE LAS BASES DE DATOS | 27 |
| 4.2.3 DIGITALIZACIÓN DEL CUESTIONARIO..... | 27 |
| 5. RESULTADOS | 29 |
| 5.1 ANÁLISIS DE LAS VARIABLES DEMOGRÁFICAS | 29 |
| 5.1.1 CURVAS DE SUPERVIVENCIA DE KAPLAN-MEIER | 30 |
| 5.1.2 MODELO DE REGRESIÓN DE COX..... | 33 |
| 5.2 EVALUACIÓN DEL TIEMPO MEDIO ENTRE VISITAS..... | 33 |
| 5.3 PLATAFORMA DE RECOGIDA DE DATOS | 35 |
| 5.3.1 CODIFICACIÓN DEL CUESTIONARIO PILOTO | 35 |
| 5.3.2 MODIFICACIONES DEL CUESTIONARIO | 39 |

| | |
|---|----|
| 5.3.3 DESCRIPCIÓN DE LOS CUESTIONARIOS DE EPI INFO | 40 |
| 5.3.4 PARTICIPANTES..... | 42 |
| 5.3.5 ANÁLISIS DESCRIPTIVO | 42 |
| 5.3.6 GUÍA PARA EL USUARIO | 43 |
| 5.3.7 APLICACIÓN PARA DISPOSITIVOS MÓVILES | 43 |
| 6. DISCUSIÓN..... | 45 |
| 6.1 EVALUACIÓN DE LAS BASES DE DATOS | 45 |
| 6.2 CUESTIONARIO EN EPI INFO..... | 46 |
| 6.2.1 SOFTWARE PARA EL CUESTIONARIO..... | 46 |
| 6.2.2 BASE DE DATOS | 48 |
| 6.3 ANÁLISIS DE SUPERVIVENCIA..... | 48 |
| 6.4 VISITAS..... | 49 |
| 6.5 EL FUTURO DEL PROYECTO KILISUN:..... | 50 |
| 6.5.1 NECESIDAD DE APOYO ESTADÍSTICO | 50 |
| 6.5.2 PERSPECTIVAS DE FUTURO | 50 |
| 7. CONCLUSIONES | 52 |
| 8. BIBLIOGRAFÍA | 54 |
| ANEXOS | A |
| A. CUESTIONARIO PILOTO DE KILISUN | A |
| B. PAQUETES Y FUNCIONES DE R | F |
| C. DATOS NO DISPONIBLES (CUESTIONARIO PILOTO) | G |
| D. MODELOS DE COX Y DIAGNÓSTICO | J |
| E. ESTADÍSTICOS DESCRIPTIVOS DEL TIEMPO ENTRE VISITAS | M |
| F. MANUAL DE USUARIO DE EPI INFO..... | M |

INDICE DE FIGURAS

Figura 1. Esquema del estudio prospectivo. En el momento que se inscriben al programa, los pacientes comienzan a estar bajo tratamiento. *(pág., 5)*

Figura 2. Esquema del cuestionario que ha de realizarse en función de la visita del paciente. *(pág. 9)*

Figura 3. Comparación de las dos bases de datos proporcionadas por Kilisun. En negrita se muestran las nuevas variables incluidas en la base de datos del cuestionario piloto. *(pág. 10)*

Figura 4. Mapa de las regiones de Tanzania. Se muestra en una escala de verdes el número de participantes que se han inscrito al proyecto Kilisun. Aquellas regiones sin pacientes se muestran en blanco. *(pág. 11)*

Figura 5. Esquema de la depuración de bases de datos. *(pág. 13)*

Figura 6. Diagrama de flujo de los pacientes registrados en las bases de datos de Kilisun. Se indica la distribución a lo largo del seguimiento. *(pág. 17)*

Figura 7. Razón de abandono del proyecto Kilisun. Los valores se muestran en frecuencia absoluta. *(pág. 18)*

Figura 8. Distribución de los pacientes en función de las visitas que han realizado (marrón) y visitas teóricas (verde). La visita 1 corresponde con la visita basal de inscripción al programa. *(pág. 20)*

Figura 9. Curvas de supervivencia teórica (A) y real al utilizar datos (B). *(pág. 23)*

Figura 10. Diagrama de flujo de los pacientes registrados en las bases de datos integrada para el análisis de supervivencia. *(pág. 29)*

Figura 11. Histograma de la variable edad en los sujetos de la base de datos para el análisis de supervivencia. Se superpone la función de densidad de la variable edad. *(pág. 30)*

Figura 12. Curva de supervivencia para las visitas de los sujetos realizada en R. Se indica en la parte inferior el número de pacientes en riesgo. *(pág. 31)*

Figura 13. Curva de supervivencia para las visitas de los sujetos por sexo (M = hombre, F = mujer) realizada en R. Se indica en la parte inferior el número de pacientes en riesgo. Se indica el p-valor del test de Log-Rank. (pág. 31)

Figura 14. Curva de supervivencia para las visitas de los participantes por nivel de edad realizada en R. Se indica en la parte inferior el número de pacientes en riesgo. (pág. 32)

Figura 15. Histograma del tiempo medio transcurrido entre visitas en los sujetos del proyecto Kilisun. Se superpone la función de densidad de dicha variable. (pág. 34)

Figura 16. Diagrama de cajas para el tiempo medio transcurrido entre visitas de los participantes de las distintas regiones de Tanzania. Se indica con un punto sin rellenar el valor de la media en cada caso. (pág. 35)

Figura 17. Cuestionario *DEMOG* del proyecto de Epi Info (pág. 37)

Figura 18. Cuestionario *QUESTIONNAIRE* del proyecto de Epi Info compuesto por: *SHABITS* (A), *DISTRIBUTION* (B) y *EDUCATION* (C). (pág. 37)

Figura 19. Cuestionario *SATISFACTION* del proyecto de Epi Info. (pág. 38)

Figura 20. Cuestionario *DISCONTINUATION* del proyecto de Epi Info. (pág. 38)

Figura 21. Búsqueda automática de registros del paciente K-01. (pág. 40)

Figura 22. Cuestionario de Epi Info en un dispositivo móvil. En (A) se muestra la pantalla de elección de cuestionario. Las imágenes (B) y (C) corresponden a *QUESTIONNAIRE*: la de la izquierda es el inicio de *SHABITS*; la de la derecha forma parte del apartado de *EDUCATION*. (pág. 44)

Figura 23. Ejemplo de distribución de los datos a lo largo de las visitas. El valor “B” indica que es la visita basal, el valor “X” indica cualquier otra visita. (pág. 46)

INDICE DE TABLAS

Tabla 1. Frecuencias y porcentajes de respuestas a los cuestionarios en la base de datos piloto. (pág. 15)

Tabla 2. Frecuencias y porcentajes de datos obtenidos y perdidos de los registros de seguimiento por región de Tanzania. (pág. 19)

Tabla 3. Análisis de riesgos proporcionales de Cox univariante para eventos recurrentes estratificado por el número de visita. (pág. 33)

Tabla 4. Reglas de asignación del número de visita en función del cuestionario y tipo de la visita. (pág. 39)

Tabla 5. Distribución de los pacientes, en función de la visita y cuestionario incluidos en Epi Info. En la visita basal se toman en cuenta todos los participantes, y en las de seguimiento solamente aquellos que han realizado la visita. (pág. 42)

RESUMEN

Introducción: *Kilisun Care* surge en Tanzania con el objetivo de solventar, por un lado, la discriminación a la que las personas con albinismo están sometidas cada día y, por otro lado, la carencia de educación en los hábitos de sus cuidados para así reducir los casos de cáncer de piel.

Objetivos: Tras realizar un análisis de la realidad del proyecto se decide analizar el efecto de variables demográficas sobre el beneficio de continuar en el proyecto Kilisun; evaluar el tiempo entre visitas de los pacientes desde el comienzo del proyecto; y establecer una plataforma electrónica de recogida y gestión de datos.

Métodos: Estudio prospectivo de intervención de brazo único en pacientes con albinismo en Tanzania. Los datos registrados fueron proporcionados por Kilisun, y se toman a partir de dos bases de datos; la primera, desde el 28 de noviembre de 2012 hasta el 24 de marzo de 2017; la segunda, desde el 27 de abril de 2017 hasta el 8 de febrero de 2018. Para analizar cuáles son los beneficios de continuar en el proyecto se realizó un modelo de Cox. Por otro lado, se realizó un análisis descriptivo del tiempo entre visitas para evaluar el seguimiento de los pacientes. Finalmente, se utilizó Epi Info como plataforma electrónica para elaborar un cuestionario digital.

Resultados y conclusiones: La predisposición de realizar una nueva visita se veía incrementada un 20% en aquellos participantes con menos de 20 años respecto al resto de la población. En cuanto al seguimiento, el tiempo mediano transcurrido entre dos visitas fue de 189 días, aproximadamente 6 meses, que es el tiempo establecido por Kilisun. Por último, se creó un cuestionario electrónico que implanta un nuevo protocolo de recogida de datos para los proyectos que Kilisun llevará a cabo próximamente.

Palabras clave: Albinismo, Kilisun Care, Epi Info, Depuración de bases de datos, Análisis de supervivencia, Gestión de datos.

1. INTRODUCCIÓN Y OBJETIVOS

La melanina es un pigmento natural presente en la mayoría de los organismos. Este pigmento es producido por un grupo especializado de células conocidas como melanocitos. En humanos, la biosíntesis de melanina es iniciada por la exposición a la radiación ultravioleta (UV), lo que provoca que la piel se oscurezca. La principal característica de esta biomolécula es su capacidad fotoprotectora, siendo capaz de disipar más del 99.9% de la radiación UV absorbida (1). Debido a esta propiedad, la melanina protege las células de la piel del daño de la radiación UV, reduciendo el riesgo de cáncer de piel. Se ha demostrado que, en individuos con mayor acumulación de melanina en la piel, es decir, con un tono de piel más oscuro, se reducía significativamente la incidencia de melanomas. A pesar de esto, aún no está clara la relación entre la pigmentación de la piel y la fotoprotección (2).

El albinismo es un conjunto de desórdenes congénitos caracterizado por la reducción completa o parcial de la producción de melanina en la piel, el pelo o los ojos. En concreto, el albinismo oculocutáneo (OCA) engloba un total de cuatro defectos genéticos autosómicos recesivos que afectan a la biosíntesis de melanina por parte de los melanocitos, resultando en la hipopigmentación del pelo, la piel y los ojos conjuntamente (3). El déficit de melanina en los ojos produce discapacidad visual, alterando el desarrollo de la retina y las conexiones nerviosas oculares. Esta discapacidad no es degenerativa y no empeora con el tiempo, y puede mejorarse utilizando gafas o dispositivos de baja visión. Por otro lado, el riesgo de padecer cáncer de piel también se puede controlar reduciendo la exposición a la luz solar, aplicando protector solar y usando ropa de manga larga y sombreros (4).

A nivel mundial, 1 de cada 18.000 personas tiene albinismo, sin embargo, en zonas del África subsahariana se observa una alta prevalencia de OCA tipo II, llegando a estimarse en 1 de cada 1.400 personas (4), siendo más de diez veces superior que la prevalencia global.

1.1 ALBINISMO EN ÁFRICA

El albinismo es una condición poco conocida en muchas sociedades africanas. Las personas con esta enfermedad son estigmatizadas debido a su apariencia pálida, incluso en la cultura popular africana existen mitos y supersticiones acerca de los albinos, con conceptos erróneos que conducen a la burla, el abandono social y la violencia.

En el África subsahariana, el albinismo a menudo se entiende como una maldición o una enfermedad contagiosa. Algunos creen que las personas con albinismo no son humanos y no mueren, sino que son demonios. Se dice que los cuerpos de personas con albinismo poseen propiedades mágicas, capaces de curar enfermedades o entregar fortuna. Esta patología, debido a la falta de conocimiento, se considera con frecuencia como un error exclusivo de la madre, conllevando la ruptura de las familias y dejando a las madres condenadas al aislamiento y al abuso.

La discriminación impide el acceso a los servicios de salud y aísla a las personas con albinismo de sus familias, comunidades y cuidadores. En Tanzania, donde el estigma abunda, se estima que menos del 10% de las personas con albinismo alcanzan la edad de 30 años debido al cáncer de piel. Solo la mitad de todos los niños con albinismo en Tanzania completan la escuela primaria, y aún menos asisten a secundaria. Además, muchos padres y profesores carecen del conocimiento para satisfacer las necesidades de estos estudiantes. Aquellos que se gradúan a menudo abandonan la escuela excluidos y no calificados, con sólo el 7% encontrando empleo formal. Muchos se ven obligados a trabajar al sol, donde están en constante riesgo de contraer cáncer de piel. Mientras tanto, las lesiones visibles causadas por el daño solar contribuyen aún más a las percepciones de personas con albinismo como contagiosas o inferiores.

Las personas con albinismo en África se enfrentan además a la amenaza de la violencia física desde la infancia. Algunos incluso son blanco de sus partes del cuerpo, se venden por miles de dólares y se usan en hechizos de brujería que se cree que aportan riqueza y fortuna. Tanzania es un caso particularmente grave, con más de 76 asesinatos a albinos desde 2006. Se cree que muchos asesinatos no han sido denunciados, y es imposible saber cuántos bebés han muerto al nacer. En Tanzania, el gobierno ha colocado a cientos de niños con albinismo en centros especializados en su protección y seguridad.

En resumen, la falta de educación en hábitos saludables y el estigma crean un ciclo de violencia que coloca a las personas con albinismo en África entre los más pobres y vulnerables del planeta (4).

1.2 PROYECTO KILISUN

Las provisiones de fotoprotectores en África son muy escasas, no se encuentran al alcance de cualquiera y en general, son donaciones procedentes de otros países (5). En particular, Tanzania es un país donde el 80% de la población vive con menos de 1.50\$ diarios, y el coste de la crema solar ronda los 10-15\$ (6). Estas cremas solares contienen químicos que producen irritación en la piel de las personas con albinismo y es habitual que estén caducadas o que su protección no sea la adecuada. El suministro es esporádico y caro, sin ningún tipo de programa de distribución. En ocasiones, ni siquiera son protectores solares si no *After Sun* o crema hidratante. Muchos de estos productos terminan en el mercado negro, siendo su precio extremadamente alto e inaccesible para la gran mayoría. Todo ello, junto con una educación muy pobre en los hábitos de protección solar y otras medidas preventivas como se ha comentado anteriormente, crea un problema alarmante en la población albina de África (5).

Por esta razón y por la exclusión social que sufren por parte de la comunidad tanzana nace *Kilimanjaro Sunscreen Production Unit* (KSPU) en 2012 (7), organización pionera en producir y distribuir crema solar diseñada para la prevención del cáncer de piel en personas con albinismo (5). El fotoprotector *Kilimanjaro Suncare*, o más conocido como Kilisun, es una emulsión de agua con aceite con una protección de nivel 30+ (Factor de Protección Solar alto) (4) que actúa como la melanina y contribuye a salvar la vida de personas con albinismo. La KSPU cuenta con el apoyo de los laboratorios dermocosméticos Frida-Dorsh y la empresa química líder mundial BASF. Ellos son los creadores de la fórmula de Kilisun y se encargan de los controles periódicos de calidad. Actualmente, la producción del fotoprotector se realiza en una unidad del Centro Regional de Dermatología del Kilimanjaro Christian Center (KMCM), referencia en el cuidado de enfermedades de la piel (5). Para fomentar la inclusión, el 30% de los empleados de producción de Kilisun son personas con albinismo (Datos propios de Kilisun).

LABOR DE *KILISUN CARE* EN TANZANIA

No sirve de nada fabricar un buen producto si no se aplica de la manera adecuada y en el momento oportuno. Por ello, desde la KSPU se regula la distribución gratuita de Kilisun hasta las áreas más remotas de Tanzania por medio de clínicas móviles, colegios y otras plataformas locales. Se hace especial hincapié en la sensibilización y educación sobre fotoprotección tanto a los beneficiarios como a sus respectivos familiares, cuidadores, profesores, dermatólogos comunitarios, personal sanitario y a la comunidad en general (5). Uno de los propósitos es concienciar a los participantes de que una aplicación correcta y consistente de Kilisun acompañada de gafas de sol, sombreros y ropa que les proteja, reduce considerablemente el riesgo de padecer cáncer de piel (7).

Desde el nacimiento de KSPU en 2012 se recogen los datos básicos de los pacientes a los que se realiza el seguimiento. Sin embargo, no fue hasta abril de 2017 cuando se introdujo de manera piloto en las regiones de Kilimanjaro y Arusha un cuestionario en el que se recogen datos demográficos, de eficacia, así como información relacionada con el uso del producto y la satisfacción con la labor de la organización.

DISEÑO DEL ESTUDIO

Se diseñó un estudio prospectivo de intervención de brazo único en pacientes externos con albinismo de las distintas regiones geográficas de Tanzania. No hay ningún detonante específico para inscribir a los sujetos en el proyecto, lo que significa que iniciarán el programa en diferentes edades o etapas de su albinismo. Una vez estén registrados, realizarán visitas cada cuatro o seis meses (en función de la región), donde se proporcionará un suministro renovado de Kilisun. (Figura 1).

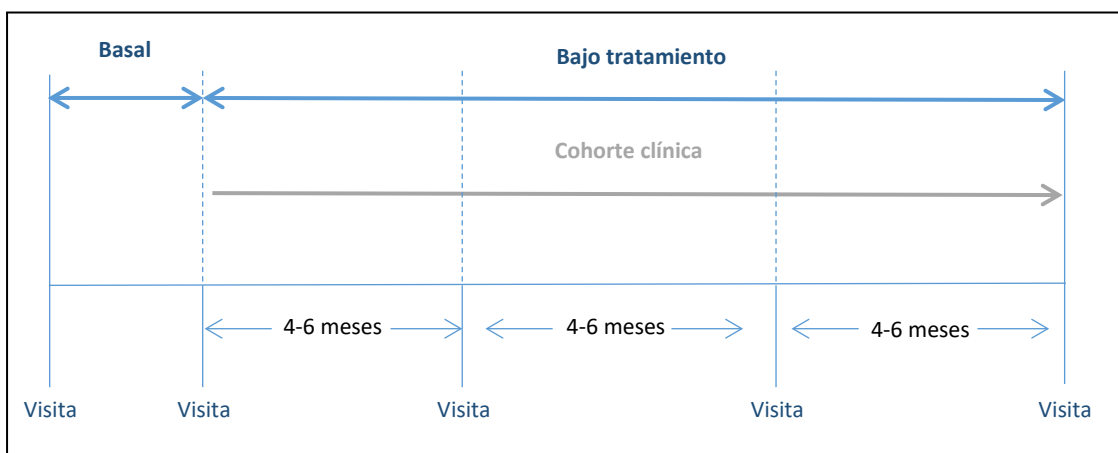


Figura 1. Esquema del estudio prospectivo. En el momento que se inscriben al programa, los pacientes comienzan a estar bajo tratamiento.

IMPACTO SOCIAL Y LOGROS

Si bien *Kilisun Care* comenzó siendo un pequeño proyecto con apenas 25 niños de una escuela de Kilimanjaro (5), no ha dejado de dar pasos agigantados desde entonces, llegando a la mayoría de regiones de Tanzania. En 2017, más de 3500 personas tanzanas con albinismo recibieron botes Kilisun, siendo el 54% de ellos niños. Además, se redujo en un 80% las quemaduras solares en los nuevos beneficiarios registrados, y el 97% de ellos, recomiendan los servicios de Kilisun (Datos propios de Kilisun). De hecho, es ganador de varios premios y se ha presentado en numerosos congresos científicos internacionales como el Congreso Mundial de Cáncer de Piel (Edimburgo 2014) y la Conferencia Internacional de Fotoprotección (Londres 2015) (5).

PROYECTOS PARA EL FUTURO

De cara al futuro, el objetivo es claro: abrir nuevas fronteras a países vecinos como Uganda, Kenia y Malawi al mismo tiempo que se refuerzan las campañas educativas (5), llegando a conseguir ser 10.000 participantes con albinismo en el proyecto (Datos propios de Kilisun). Conseguir así un mundo donde las personas con albinismo en África tengan las mismas oportunidades, acceso al cuidado de su salud y expectativas de vida, así como gozar de los mismos derechos y la libertad que les permita valerse por sí mismos (7).

Para cumplir con los propósitos que se plantean de aquí a 2020, se necesita, entre otros, apoyo estadístico. Es fundamental una metodología adecuada y consolidada para la recogida de los datos si después se quiere realizar un análisis estadístico que permita corroborar la calidad del proyecto. Esto es lo que este Trabajo de Fin de Máster aporta a la organización *Kilisun Care*, resultados en los que apoyarse y la iniciativa de un cuestionario electrónico para la optimización de los recursos que se disponen.

1.3 OBJETIVOS

Los objetivos de partida del presente trabajo de acuerdo con las demandas de Kilisun fueron:

- Valoración del contenido y estructura del cuestionario piloto antes de su lanzamiento a gran escala.
- Establecer una plataforma de recogida electrónica y gestión de datos.

- Elaborar un plan de análisis en el que se defina la mejor estrategia para tratar los datos perdidos
- Definir los indicadores claves para obtener respuestas sobre:
 - Capacidad de los beneficiarios para adoptar medidas de prevención de cuidados
 - Evolución en la prevalencia de condiciones médicas asociadas con el desarrollo de cáncer de piel
 - Nivel de satisfacción con el programa.

2. FUENTES DE DATOS

2.1 CUESTIONARIO DE KILISUN CARE

Kilisun desarrolló un primer cuestionario donde se incluían los datos generales de los participantes, la fecha de visita, la clínica a la que acudían, los botes de crema solar que se entregaban/devolvían y el personal que les atendía. Esta información era insuficiente para las preguntas que la organización se planteaba. Por esta razón, en abril de 2017 se hizo una nueva propuesta de cuestionario piloto para el seguimiento de los participantes que completaba la anterior. Este es el formulario que se utiliza actualmente y consta de 3 secciones (seguimiento, satisfacción y discontinuación del participante):

1. **Seguimiento:** se rellena en cada visita a la que acude el participante. Recoge los datos generales del individuo, información acerca de sus hábitos de protección solar, la educación que ha recibido, los botes de crema solar que se le proporcionan cuando asiste al centro, etc. En función de la visita en la que se encuentre el paciente, se realiza un formulario u otro (Figura 2):
 - a. *Screening visit o enrollment visit:* es el cuestionario inicial y se completa cuando el participante se inscribe en el proyecto. Primero se introducen los datos acerca de su procedencia, nombre, género, familiares con albinismo, diagnóstico previo de cáncer de piel, y otros datos de interés. Además, incluye aspectos relacionados con los hábitos de protección solar, los botes que recibe el paciente y la educación adquirida mediante el programa Kilisun. Está formado por un total de 50 preguntas.
 - b. *Follow-up:* se trata de un cuestionario con formato similar al inicial y se utiliza cuando el paciente ya forma parte del programa. Una vez almacenados los datos demográficos en el registro de inscripción (*screening visit*), basta con un seguimiento de su protección solar, educación y distribución de los botes en las posteriores visitas. Este formulario recoge toda esta información y está establecido por 40 preguntas.

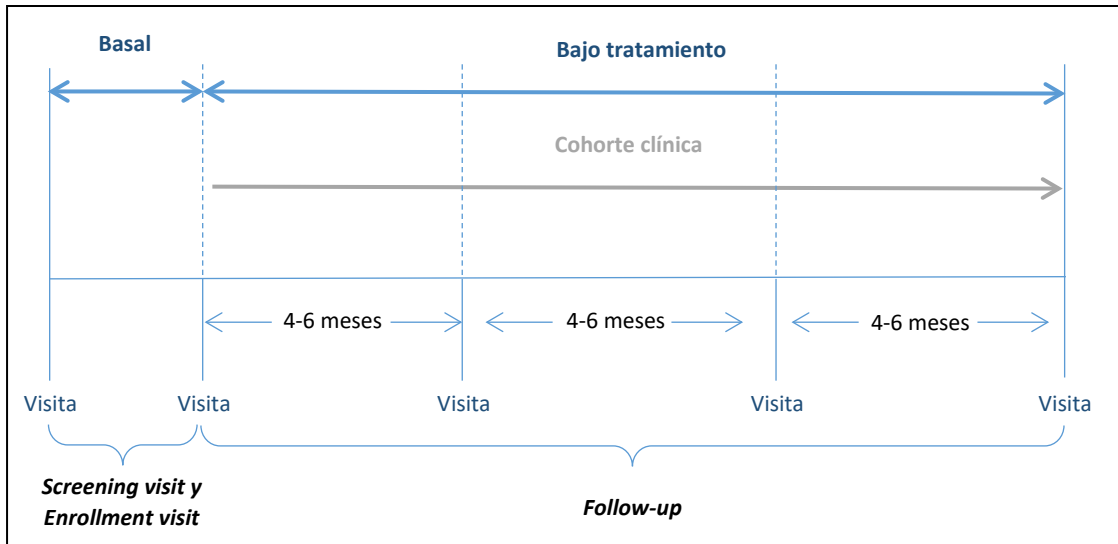


Figura 2. Esquema del cuestionario que ha de realizarse en función de la visita del paciente.

2. **Satisfacción:** tiene como objetivo conocer la satisfacción de los participantes con la labor de Kilisun. Se realizan un total de 9 de preguntas, entre ellas, si notan mejoría en la piel, si el tratamiento limita sus actividades diarias, y si recomiendan el producto. Este cuestionario es opcional, aunque se recomienda realizarlo una vez al año.
3. **Discontinuación:** este último cuestionario se lleva a cabo únicamente por los participantes que deciden no continuar con el tratamiento y escogen entre 8 opciones la razón de su abandono.

Los cuestionarios se rellenan en formato papel y son transportados periódicamente a la sede central de Kilisun ubicada en Kilimanjaro, donde una persona es la encargada de introducir manualmente los datos en una tabla de Microsoft Excel. Así es como se crean las bases de datos que se describen a continuación. El cuestionario completo puede ser visto en el apartado de Anexos (A. *Cuestionario piloto de Kilisun*).

2.2 BASES DE DATOS

Todos los datos recogidos por Kilisun se encuentran recabados en dos bases de datos distintas en Microsoft Excel. Por un lado, una base de datos inicial que recoge los datos de los adscritos al proyecto desde su creación en 2012 hasta la implantación del nuevo cuestionario (Base de datos básica de aquí en adelante) y, por otro lado, una base de datos que recoge las respuestas al nuevo cuestionario (Base de datos piloto de aquí en adelante) (Figura 3):

- **Base de datos básica:** la base de datos se abrió el 28 de noviembre de 2012 y fue cerrada el 24 de marzo de 2017 para este estudio. En ese período se recogió información de un total de 3856 pacientes distribuidos por 19 de las 30 regiones administrativas que posee Tanzania. El archivo Excel disponía de una hoja diferente para cada una de las regiones y mostraba la información en formato ancho para cada visita de seguimiento que realizó el paciente.
- **Base de datos piloto:** recoge las respuestas al cuestionario piloto. Se abrió el 27 de abril de 2017 y fue cerrada el 8 de febrero de 2018 para este estudio. Esta base de datos está formada por pacientes únicamente de las regiones de Kilimanjaro y Arusha. Mantiene todos los participantes de dichas regiones que se encuentran en la base de datos básica, lo que se traduce en un total 790 individuos, de los cuales, sólo 62 son nuevos pacientes. El archivo Excel disponía de una hoja diferente para cada región y mostraba la información en formato ancho para cada una de las encuestas de seguimiento que realizó cada individuo.

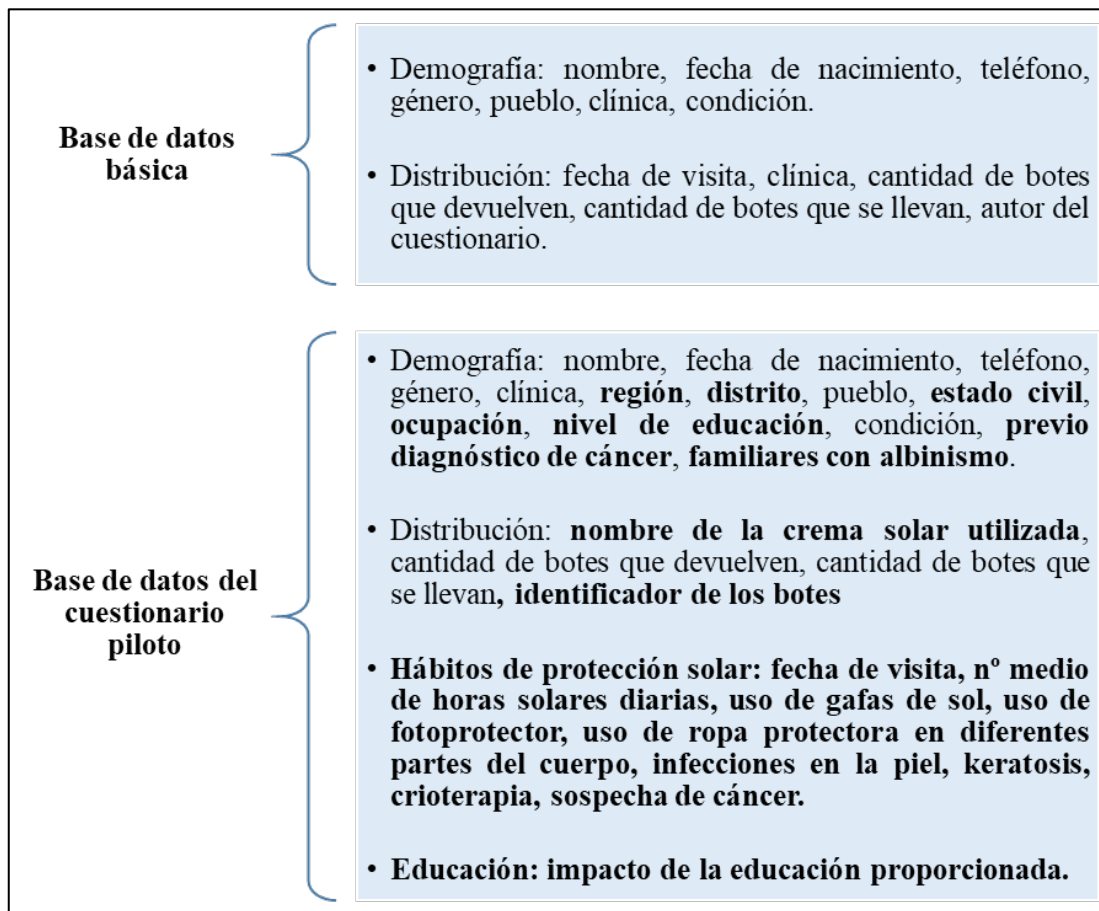


Figura 3. Comparación de las dos bases de datos proporcionadas por Kilisun. En negrita se muestran las nuevas variables incluidas en la base de datos del cuestionario piloto.

Adicionalmente en este trabajo se utilizó una tercera base de datos que integra los datos recogidos en la base de datos básica y piloto para aquellas variables que son comunes en ambas (**Base de datos integrada** de aquí en adelante) de tal forma que se minimizase la pérdida de información entre ambas bases de datos.

La distribución de los pacientes adscritos al programa Kilisun a lo largo de las divisiones administrativas o regiones de Tanzania se puede observar en la Figura 4.

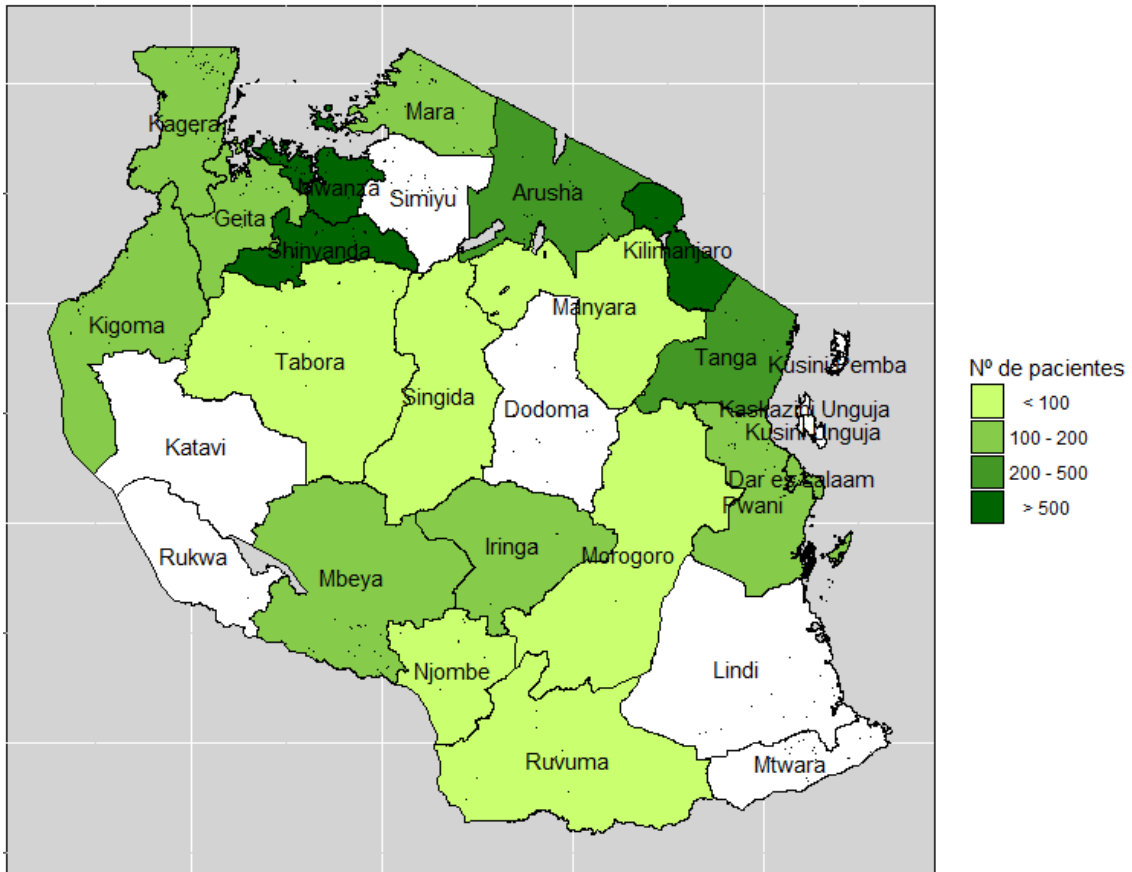


Figura 4. Mapa de las regiones de Tanzania. Se muestra en una escala de verdes el número de participantes que se han inscrito al proyecto Kilisun. Aquellas regiones sin pacientes se muestran en blanco.

3. EVALUACIÓN DE LA CALIDAD DE LAS BASES DE DATOS

Al recibir una nueva base de datos es habitual que los datos brutos que dispone no estén en el formato correcto, contengan errores, estén incompletos y no tengan todas las etiquetas y códigos correctos que se necesitan para el análisis. La depuración de las bases de datos de Kilisun y la preparación de estas es una parte esencial y necesaria antes de realizar cualquier análisis estadístico. Además, permite localizar inconsistencias y errores que pueden aparecer. La presencia de datos incorrectos o inconsistentes puede comprometer los resultados de los análisis, llegando incluso a imposibilitar la realización de estos al reducir el volumen de datos utilizable (8).

Esta depuración previa de los datos, en la práctica, a menudo consume mucho más tiempo que el análisis estadístico en sí mismo. Todas las acciones realizadas durante la depuración influyen en los resultados de los análisis estadísticos posteriores. Por tanto, la depuración de datos se debe considerar una operación estadística, que se realizará de una manera reproducible (9). Se utilizó R (v 3.4.4) como software estadístico para llevar a cabo la depuración y análisis de la base de datos al ser un buen entorno para la limpieza reproducible de datos, ya que todas las acciones de limpieza se pueden programar y, por lo tanto, reproducir. El listado de funciones de R no incluidos en el paquete base se pueden localizar en Anexos (*B. Paquetes y funciones de R*).

3.1 DEPURACIÓN DE LA BASE DE DATOS

Antes de comenzar la depuración de la base de datos es necesario comprender cuál o cuáles pueden ser las fuentes de los errores que pueda contener. Un correcto entendimiento de la base de datos y el origen de sus errores es útil para la elaboración de técnicas de limpieza dirigidas que permitan detectar y manejar los errores con mayor facilidad (8).

Debido a la naturaleza de la recopilación de datos que se realizó (toma de datos en un cuestionario en papel y posterior introducción manual en Excel), se espera que la mayoría de los errores a encontrar en las bases de datos de Kilisun sean en la entrada de datos.

El proceso de depuración de bases de datos se puede simplificar en dos etapas (Figura 5). Los datos brutos deben tratarse para que cada variable sea nombrada y codificada de manera uniforme y se eliminen los datos que contienen valores imposibles o no preestablecidos. Además, al exportarse la base de datos a R cada variable deberá ser reconocida de acuerdo con su tipo (numérico, factor, fecha, etc). Una vez se ha realizado este proceso, los datos serían técnicamente correctos. Pero que los datos sean técnicamente correctos no quiere decir que estén exentos de errores. Un ejemplo de esto sería una persona que tuviese 3 años y su profesión fuese “Profesor”, ya que, las dos variables por separado son correctas, pero conjuntamente es un dato erróneo. Es por esto que es necesario realizar un segundo filtro para obtener unos datos consistentes que permitan elaborar un correcto análisis estadístico.

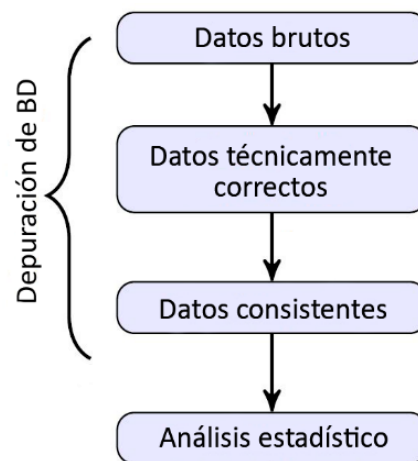


Figura 5. Esquema de la depuración de bases de datos. Adaptación de *Introduction to data cleaning with R* (9).

3.1.1 OBTENCIÓN DE DATOS TÉCNICAMENTE CORRECTOS

La obtención de datos técnicamente correctos implica que estén en un formato adecuado para R y que cada variable sea catalogada de acuerdo con el tipo que debe ser, por ejemplo, una variable que recoja la edad de un paciente debe ser reconocida como numérica.

El primer paso fue la exportación de las bases de datos de Microsoft Excel a R. Una vez incorporada la base de datos a Excel se siguen las siguientes estrategias en R para la obtención de datos técnicamente correctos:

- **Conversión del tipo de variable:**

Para la conversión de los tipos de las variables se utilizan las funciones de coerción que incluye el paquete base de R. Estas funciones son conocidas como “as.” y por defecto, aquellos valores que no se puedan convertir al tipo especificado se convertirán en valores NA. En este apartado se excluyen las variables tipo fecha, las cuales se tratarán de manera individual y se explicarán en uno de los siguientes apartados.

- **Recodificación de variables tipo factor**

Las variables categóricas en R son guardadas con el tipo factor. Un factor es un vector que especifica qué valor corresponde a cada nivel de la variable. Únicamente para aquellas variables que no mostraban ningún valor categórico extraño se realizó una recodificación con el fin de unificar el formato de la base de datos. Ejemplo de esto sería codificar todas las variables en mayúsculas o utilizar siglas preespecificadas para las variables con valores complejos.

- **Manipulación de cadenas**

Las variables que recogen caracteres son tratadas como cadenas en R. Una variable de este tipo puede poseer una gran variedad pese a no tener un excesivo número de valores posibles, por ejemplo, al recoger el nombre del profesional que realiza la entrevista puede referirse de numerosas formas a la misma persona: uso de mayúsculas y minúsculas, incluir el nombre completo o nombre sin apellidos, uso de abreviaturas, etc

La manipulación de las cadenas busca reducir el número posible de valores que posee dicha cadena para poder manipularla con mayor facilidad en un futuro. El paquete base de R viene con una amplia funcionalidad de manipulación de cadenas, en este caso se utilizó la familia de funciones “grep” que es capaz de localizar coincidencias de texto en un vector, lo que seguido de un reemplazamiento de los valores permitirá unificar los valores de las variables.

La manipulación de cadenas también permite subsanar parte de los errores que pudiesen darse en variables tipo factor que no tenían estandarizado el formato, y poder así realizar la recodificación. Adicionalmente, aquellos valores que no coincidían con ningún valor posible de las variables fueron convertidos en NA.

- **Conversión de fechas**

La fecha es una de las variables más complejas dentro de una base de datos ya que puede ser almacenada de muchas formas distintas. Es por esto que el procesamiento de este tipo de datos requiere una dedicación adicional.

La base de datos sin depurar de Kilisun contiene numerosas inconsistencias dentro de estas variables ya que para cada región existe un formato de fecha predominante específico, incluso dentro de una misma región pueden existir distintos formatos. Entre ellos, se pueden encontrar *dd/mm/aaaa*, *mm/dd/aaaa*, *aaaa/mm/dd* y otros formatos como el número de días desde una fecha de referencia.

Es fundamental reconocer los distintos formatos y unificarlos, en este caso a un formato de *dd/mm/aaaa*, antes de su exportación a R. Hay que remarcar además que las variables de tipo fecha son muy susceptibles de errores al ser introducidas manualmente en una base de datos.

Una vez la variable está unificada, el entorno de trabajo de R permite la utilización de distintos formatos en función del análisis que se quiera realizar. R posee distintos tipos que puede adoptar una variable para almacenar los datos de fechas; por defecto, el formato fecha de Excel al importarse a R es de tipo *POSIXt*, un formato que permite almacenar fechas y horas.

3.1.2 OBTENCIÓN DE DATOS CONSISTENTES

Los datos técnicamente correctos pueden contener errores, para obtener datos consistentes es necesario eliminar y corregir valores perdidos, errores obvios, valores atípicos... Para poder transformar los datos técnicamente correctos a datos consistentes no basta con conocer el tipo de variable, sino que es necesario un conocimiento real de la base de datos y el tema a tratar en cuestión.

El proceso de obtención de datos consistentes en las bases de datos de Kilisun implica tres pasos (9). Primero, la detección de inconsistencias, por ejemplo, la variable fecha de realización del cuestionario no puede tomar un valor posterior al cierre de la base de datos. En segundo lugar, establecer reglas que puedan causar inconsistencias, un ejemplo sería una variable que, de ser respondida de una manera concreta, abra la puerta a la contestación de una segunda pregunta relacionada. La presencia de un valor en esta segunda variable cuando la primera no lo permite sería una inconsistencia de los datos.

Por último, se debe establecer un método de corrección de los campos que se consideren erróneos.

El principal tipo de inconsistencia presente en las bases de datos de Kilisun son los valores perdidos o valores *missing*. Se realizó un conteo de estos para evaluar la posibilidad de realizar análisis estadísticos consistentes. También se establecieron una serie de reglas entre variables para detectar inconsistencias, que fueron las siguientes:

- Acotación de los valores posibles que pueden presentar las variables numéricas.
- Aquellas variables que dependan de la respuesta previa de otra debían estar en concordancia.
- No puede existir un registro de visita sin la existencia previa de una visita basal en la que se recogen los datos del paciente.

Aquellas inconsistencias detectadas fueron convertidas en valores *missing*. Otro aspecto a tener en cuenta es que cuando los datos son introducidos por humanos en lugar de automáticamente, es probable que se produzcan algunos errores típicos generados por la acción humana (9). Estos errores a veces se pueden detectar a partir de datos sin procesar con certeza, un ejemplo de esto sería la utilización de la función arrastrar de Microsoft Excel cuando se quiere mantener un mismo dato en un conjunto de celdas. Como resultado de esto, se puede observar un conjunto de datos ordenados en lugar del mismo dato repetido.

La depuración de una base de datos, pese a ser un proceso muy laborioso no permite asegurar que los datos sean correctos. De hecho, algunas inconsistencias como los errores tipográficos o en las fechas son difíciles de detectar incluso realizando un plan exhaustivo de depuración.

3.2 EVALUACIÓN DE LAS INCONSISTENCIAS

La base de datos integrada permite analizar la distribución de los registros dentro de ambas bases de datos (básica y piloto) a lo largo de las visitas (Figura 6). Del total de 3918 sujetos, en 3794 se dispone de algún tipo de información basal, sin embargo, el porcentaje de pacientes con alguna visita se reduce considerablemente. Un 72,6% de los pacientes realizaron al menos una visita, el 49,2% dos visitas, el 31% tres visitas y sólo un 21% de los participantes tendrá más de tres visitas de seguimientos. Este porcentaje

continúa descendiendo hasta 0,1% correspondiente a 12 visitas, siendo el valor máximo de visitas observado en la base de datos integrada.

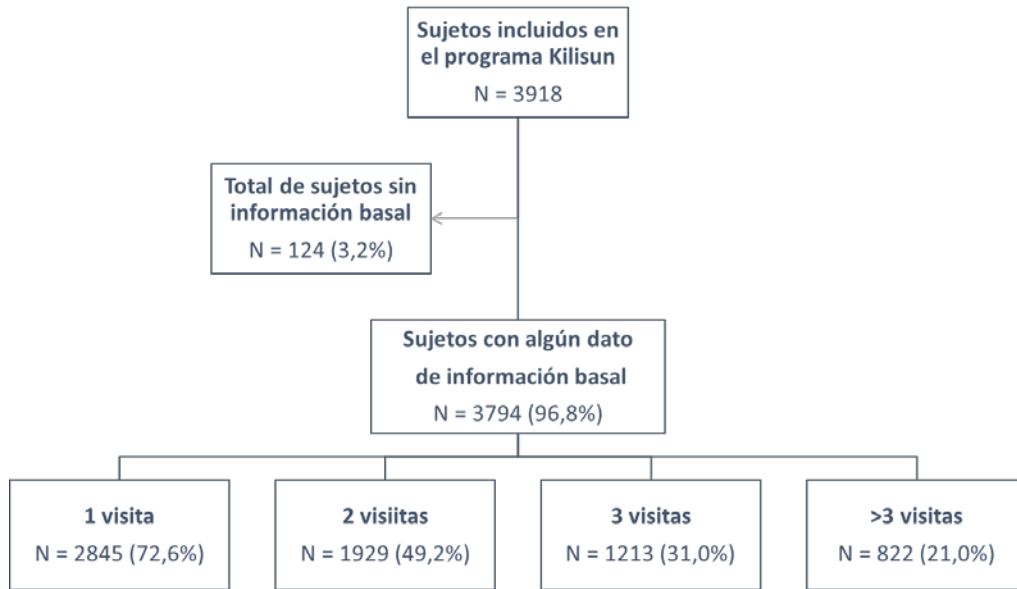


Figura 6. Diagrama de flujo de los pacientes registrados en las bases de datos de Kilisun. Se indica la distribución de los pacientes a lo largo del seguimiento.

Para el análisis de indicadores claves que puedan responder a los objetivos iniciales propuestos fue necesario comprobar la calidad de la base de datos piloto. En el periodo de recogida de esta base de datos se almacenan datos del cuestionario inicial de inscripción (*enrollment visit*), dando tiempo, en el mejor de los casos, a realizar dos visitas *follow-up* de los pacientes. Dentro del registro de pacientes se observa un elevado número de datos no disponibles (Tabla 1), llegando a tener únicamente un 4,3% de los datos correspondientes a las respuestas del cuestionario de inscripción, un 40% de las respuestas del cuestionario correspondiente a la primera visita y un 31,8% de las respuestas de la segunda visita. Se puede consultar los datos disponibles para cada variable en Anexos (*C. Datos no disponibles (cuestionario piloto)*). El bajo número de datos del cuestionario de inscripción es debido a que la mayoría de los pacientes, como se indica en el apartado 2.2 provenían de la base de datos básica.

| | N | % datos disponibles | % datos no disponibles |
|-----------------------------------|-----|---------------------|------------------------|
| Visita basal (inscripción) | 790 | 4.3% | 95.7% |
| Visita 1 | 790 | 40.0% | 60.0% |
| Visita 2 | 790 | 31.8% | 68.2% |

Tabla 1. Frecuencias y porcentajes de respuestas a los cuestionarios en la base de datos piloto.

El cuestionario piloto, además de las respuestas proporcionadas en las visitas, disponía de dos secciones correspondientes por un lado a la satisfacción del paciente con el proyecto y por otro a los motivos por los que un paciente decidía abandonar el tratamiento. Sin embargo, Kilisun no dispone de un registro de respuestas para la encuesta de satisfacción dentro de su base de datos piloto. En cuanto a los datos de discontinuación, se registraron las respuestas de un total de 26 pacientes (Figura 7), donde cada uno podía indicar más de un motivo. Para los pacientes adscritos al proyecto antes de la implantación del cuestionario piloto o en regiones en las que no se realizó no se tiene ningún dato de discontinuación ni del motivo de esta.

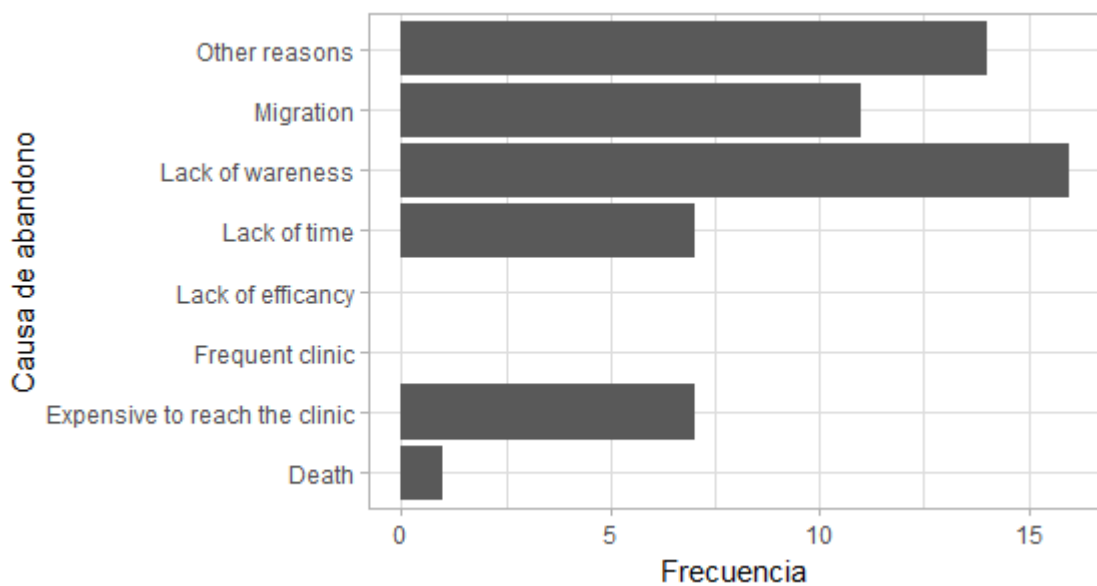


Figura 7. Razón de abandono del proyecto Kilisun. Los valores se muestran en frecuencia absoluta. Los datos se obtienen de las respuestas de 26 pacientes al cuestionario de discontinuación.

La base de datos básica también muestra un porcentaje elevado de visitas perdidas (Tabla 2). Disponiendo de un 40% del global de los datos de seguimiento. El alto porcentaje de datos perdidos puede ser debido a la estructura física de la base de datos y errores a la hora de introducir nuevas visitas de pacientes, este punto se trata con mayor profundidad en el apartado de Discusión.

A la vista de los datos expuestos existen numerosas incoherencias dentro de las bases de datos. Algunas inconsistencias no habrán sido detectadas en la depuración, y otras, sin embargo, son detectadas sin necesidad de serlo.

| Región | n (sujetos) | Nº de visitas | % Datos | % Missing |
|--|--------------------|----------------------|----------------|------------------|
| Global (Todas las regiones) | 3856 | 28370 | 11385 (40 %) | 16985 (60%) |
| Kilimanjaro | 503 | 5040 | 2031 (40 %) | 3009 (60 %) |
| Arusha | 223 | 2007 | 648 (32 %) | 1359 (68 %) |
| Geita | 182 | 1456 | 456 (31%) | 1000 (69 %) |
| Mwanza | 560 | 5058 | 2161 (43 %) | 2897 (57 %) |
| Shinyanga | 543 | 4905 | 1811 (37 %) | 3094 (63 %) |
| Kagera | 136 | 408 | 307 (75 %) | 101 (25 %) |
| Mara | 175 | 1225 | 471 (38 %) | 754 (62 %) |
| Manyara | 54 | 378 | 64 (17 %) | 314 (83 %) |
| Mbeya | 131 | 923 | 345 (37 %) | 578 (63 %) |
| Njombe | 37 | 259 | 113 (44 %) | 146 (56 %) |
| Dar-Es-Salaam | 179 | 1281 | 420 (33 %) | 861 (67 %) |
| Iringa | 199 | 1038 | 581 (56 %) | 457 (44 %) |
| Tabora | 80 | 270 | 152 (56 %) | 118 (44 %) |
| Kigoma | 133 | 399 | 335 (84 %) | 64 (16 %) |
| Pwani | 104 | 208 | 166 (80 %) | 42 (20 %) |
| Ruvuma | 95 | 190 | 181 (95 %) | 9 (5 %) |
| Tanga | 446 | 3122 | 1056 (34 %) | 2066 (66 %) |
| Morogoro | 25 | 50 | 41 (82 %) | 9 (18 %) |
| Singida | 51 | 153 | 46 (30 %) | 107 (70 %) |

Tabla 2. Frecuencias y porcentajes de datos obtenidos y perdidos de los registros de seguimiento por región de Tanzania

Para tener una idea aproximada del valor real de visitas perdidas en la base de datos integrada se compara el número de visitas que realizaron los pacientes frente al valor teórico de visitas que deberían haber realizado de acuerdo con su fecha de ingreso en el programa Kilisun tomando como tiempo entre visitas 6 meses (Figura 8).

El número total de visitas de la base de datos integrada es de 11804 visitas, al realizar la estimación teórica se obtiene que la base de datos debería tener 23617 registros de visitas para los pacientes adscritos, lo que equivale a un 50% de visitas perdidas, siendo este valor inferior al 60% de datos *missing* de la base de datos básica (Tabla 2) o a los porcentajes de datos no disponibles de la base de datos piloto (Tabla 1).

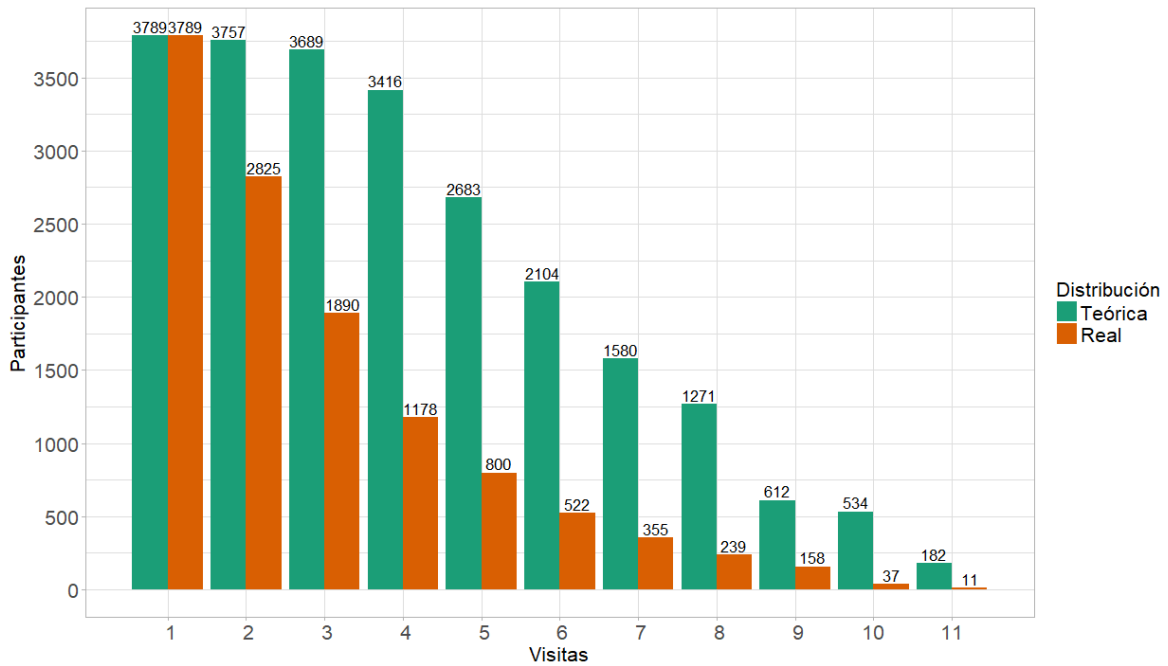


Figura 8. Distribución de los pacientes en función de las visitas que han realizado (marrón) y visitas teóricas (verde). La visita 1 corresponde con la visita basal de inscripción al programa.

3.3 ANÁLISIS DE LA REALIDAD

La falta de supervisión durante la recogida y almacenamiento de los registros ha conllevado numerosos errores en las bases de datos y a no disponer de la mayoría de los registros del cuestionario y el seguimiento en las tres bases de datos. Llegando incluso a no tener ningún dato como es en el caso de la encuesta de satisfacción. Las razones por las que ha podido suceder esto se analizarán en el apartado de Discusión de este trabajo.

La gran cantidad de datos perdidos y datos no disponibles imposibilita la realización de un análisis estadístico que cubra las expectativas iniciales que se plantean en el epígrafe 1.3. Además, no cabe la posibilidad de imputarlos, ya que la base de datos debe estar compuesta por una cantidad coherente de datos presentes en proporción a los ausentes, y a su vez, tener un número razonable de observaciones que permitan una buena predicción.

La base de datos integrada no dispone de variables de interés presentes en el cuestionario piloto, y la base de datos piloto en el momento en que se cierra para este estudio no dispone de un número consistente de visitas.

Por tanto, en este punto se tuvo que replantear el rumbo del proyecto y priorizar nuevos objetivos que permitieran, en la manera de lo posible, obtener resultados acerca de la calidad del seguimiento de los pacientes en el programa.

Los nuevos objetivos planteados fueron:

- Analizar el efecto de variables demográficas sobre el riesgo de continuar el proyecto.
- Evaluar el seguimiento de los pacientes desde el comienzo del proyecto de Kilisun.
- Establecer una plataforma electrónica de recogida y gestión de datos.

4. METODOLOGÍA

4.1 ANÁLISIS ESTADÍSTICO

Este apartado recogerá la metodología estadística a seguir para el desarrollo de los objetivos planteados tras la evaluación de la calidad de los datos de partida.

4.1.1 ANÁLISIS DE SUPERVIVENCIA

El análisis de supervivencia es una colección de procedimientos estadísticos para el análisis de datos en los que la variable resultado de interés es el tiempo hasta que ocurre un evento (10). A partir de la base de datos integrada se realizó un análisis de supervivencia con el objetivo obtener una medida del riesgo de un paciente de realizar una nueva visita respecto a sus datos demográficos básicos (edad y sexo), presentes en dicha base de datos. En este caso el evento se define como la realización de una nueva visita por parte de los participantes.

La mayoría de los análisis de supervivencia consideran un problema de los datos llamado censura. La censura se produce cuando tenemos información sobre el tiempo de supervivencia de un individuo, pero no se conoce exactamente el tiempo de supervivencia. Al no poseer datos de discontinuación, se consideró como censura para la base de datos integrada la observación de los participantes en el momento en que se cierra la base de datos para este estudio.

La función de supervivencia, denotada por $S(t)$, da la probabilidad de que la variable T exceda el tiempo t especificado. Teóricamente, como t varía desde 0 hasta infinito, la función de supervivencia se representa gráficamente como una curva continua decreciente, que comienza en $S(t) = 1$ y $t = 0$ y se dirige de manera decreciente hacia cero a medida que t tiende a infinito. En la práctica, al usar los datos reales de Kilisun, se obtuvieron curvas de supervivencia estimadas que son funciones escalonadas, como se muestra en la Figura 9, en lugar de curvas continuas (10).

La función de riesgo $h(t)$, proporciona el riesgo instantáneo por unidad de tiempo para que el evento ocurra dado que el individuo ha sobrevivido al tiempo t . La función de riesgo se enfoca en que se produzca el evento, por tanto, cuanto mayor es el riesgo

promedio, peor es el impacto sobre la supervivencia. El riesgo es una tasa, no una probabilidad. Por lo tanto, los valores de $h(t)$ oscilan entre cero e infinito.

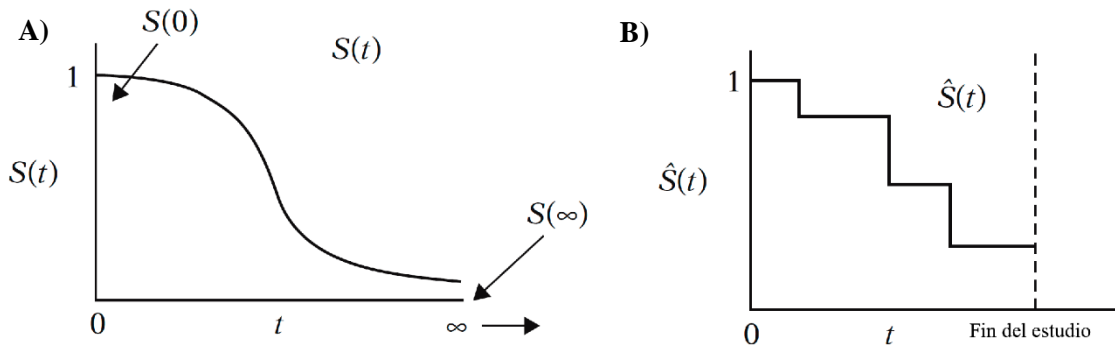


Figura 9. Curvas de supervivencia teórica (A) y real al utilizar datos (B). Adaptación de (10).

ESTIMACIÓN DE LA FUNCIÓN DE SUPERVIVENCIA

El estimador de Kaplan-Meier para la función de supervivencia (11) es el estimador utilizado por defecto en la mayoría de los softwares estadísticos. Asumiendo que se parte de una muestra de n observaciones independientes denotadas como (t_i, c_i) , $i=1,2,\dots,n$ de la variable de supervivencia T y una variable indicadora de la censura C . Se asume que entre las n observaciones, hay $m \leq n$ tiempos de evento y $n - m$ valores censurados. Se denotan los tiempos ordenados de supervivencia como $t_1 < t_2 < \dots < t_m$. Definiendo el número de sujetos en riesgo de morir en t como h_i y el número de sujetos en los que se ha producido el evento como d_i . El estimador de Kaplan-Meier de la función de supervivencia se calcula como (12):

$$\hat{S}(t) = \prod_{t_i \leq t} \left(\frac{h_i - d_i}{h_i} \right) \quad i = 1, 2, \dots, n$$

Este estimador permitió visualizar las funciones de supervivencia globales de los sujetos de la base de datos integrada, así como la realización de dichos gráficos en función de las variables edad y el sexo.

MODELO DE REGRESIÓN DE COX CON EVENTOS RECURRENTES

El modelo de riesgos proporcionales (PH) de Cox (13) es un modelo matemático utilizado para analizar datos de supervivencia. El modelo Cox generalmente se escribe en términos de la función de riesgo. Este modelo da una expresión del riesgo en el tiempo t para un

individuo con unas características de un conjunto de variables explicativas denotadas por \mathbf{X} (vector de las variables explicativas del modelo).

$$h(t, \mathbf{X}) = h_0(t) e^{\sum_{i=1}^p \beta_i X_i} \quad \mathbf{X} = (X_1, X_2, \dots, X_p)$$

La fórmula del modelo de Cox muestra que el riesgo en el tiempo t es el producto de $h_0(t)$ (función de riesgo de referencia) y una expresión exponencial que depende de las variables explicativas. Una característica importante de esta fórmula es la suposición de riesgos proporcionales (PH), es decir, que el riesgo de referencia es una función de t , pero no dependa las variables \mathbf{X} . La relación entre los riesgos o *hazard ratio* (HR) de los distintos grupos debe mantenerse constante en el tiempo.

$$HR = \frac{h_i(t/X_i)}{h_0(t)} = \exp(\beta X_i) \quad i = 1, 2, \dots, n$$

En el caso de los datos de Kilisun, de acuerdo con la definición de evento establecida puede darse el caso de que el evento ocurra varias veces a lo largo del seguimiento del paciente, es decir, un mismo paciente puede realizar más de una visita. Dichos eventos para este caso son idénticos entre sí y se denominan "eventos recurrentes", caracterizándose por poseer datos no independientes y que cada sujeto permanece en riesgo hasta que el estudio finaliza.

El modelado de este tipo de datos se puede llevar a cabo utilizando un modelo de Cox con el diseño de los datos construido de modo que cada sujeto tenga una línea de datos correspondiente a cada evento recurrente.

Este formato que han de tener los datos se denomina enfoque *Counting Process* (CP) (14). Una característica distintiva del diseño de la base de datos para el enfoque del *Counting Process* es que cada línea de datos enumera el tiempo de inicio y el tiempo de finalización para cada intervalo de seguimiento. Si el último intervalo de tiempo para el sujeto termina en censura, entonces el número de eventos recurrentes para este sujeto es el número de filas que tenga menos una.

Por otro lado, además de modelar los eventos de Kilisun como recurrentes, se consideró que el orden en que se producían dichos eventos era relevante. Para abordar la relevancia del orden se utilizó un modelo de Cox estratificado de eventos recurrentes condicional 1, en el que se estratifica en el orden en que ocurren los eventos recurrentes mediante una variable indicadora. La interpretación de los *hazard ratio* resultantes de la estratificación

es equivalente salvo que ha de aplicarse para cada estrato de manera independiente (12). El enfoque condicional 1 usa el mismo formato utilizado para el enfoque de CP (tiempo de inicio, tiempo de finalización), diferenciándose de este por la estratificación.

Dado que los coeficientes del modelo se encuentran correlados entre eventos, se ajustó la varianza de los mismos a partir de estimaciones conocidas como "estimaciones de varianza robustas" (10).

BASE DE DATOS EN FORMATO *COUNTING PROCESS*

La base de datos integrada no está en el formato adecuado para la realización del análisis planteado. El primer paso realizado fue la conversión de los datos de formato ancho a formato largo. Las variables seleccionadas para el análisis fueron:

- ID: identificador del paciente
- DOB: Año de nacimiento
- SEX: sexo del paciente (M = Hombre, F = mujer)
- DOV: Fecha en la que se realizó la visita

Una vez obtenido el formato largo fue necesario preparar las variables específicas para la realización del análisis de supervivencia. Se realizaron las siguientes modificaciones de los datos seleccionados para obtener un formato de *Counting Process*:

- Conversión de la variable DOB en Edad.
- Creación de variables:
 - START: tiempo de inicio de cada paciente. Toma el valor 0 para todas las observaciones.
 - EVENT: variable que recoge la presencia de evento en cada observación (1=Evento, 0=Censurado).
 - STOP: tiempo en días entre el comienzo del seguimiento del paciente y la visita realizada si EVENT = 1 o el cierre de la base de datos si EVENT = 0.
- Eliminación de los pacientes sin dato para la fecha en que se realizó la visita inicial (Pacientes sin referencia temporal de inicio).

- Eliminación de las visitas con datos faltantes para la fecha (Falta de referencia temporal).
- Eliminación de los pacientes que no tengan al menos una visita inicial y una visita (Falta de observaciones).

4.1.2 EVALUACIÓN DEL SEGUIMIENTO

La evaluación del seguimiento tiene como objetivo determinar la frecuencia de los pacientes a la hora de realizar una nueva visita médica y si esta visita se realiza con los estándares propuestos por Kilisun (visita cada 4-6 meses en función de la región).

Para ello, se realizará un análisis descriptivo del tiempo entre visitas para el conjunto de todos los pacientes registrados y por región (diagramas de cajas). Se utilizó de partida la misma base de datos que fue creada para el análisis de supervivencia con modificaciones. Las modificaciones de la base de datos fueron:

- Creación de una nueva variable correspondiente al tiempo entre visitas a partir de los datos de la variable STOP.
- Eliminación de las filas correspondientes a las censuras (selección de los eventos)

4.2 PLATAFORMA DE RECOGIDA DE DATOS

4.2.1 REDISEÑO DE LA BASE DE DATOS

Hasta ahora se ha visto que la estructura y organización de las dos bases de datos que dispone Kilisun (básica y piloto) no proporciona la calidad de datos necesaria para el análisis estadístico. Por esta razón y dado que la información que se maneja tiende a crecer, es conveniente optar por desarrollar una nueva base de datos que sea escalable y fácil de utilizar. Para lograr dicho objetivo, se decidió crear una base de datos relacionales, que es aquella donde se representan los datos y las relaciones entre los datos mediante una colección de tablas, cada una con un nombre único y donde cada fila de la tabla representa una relación entre un conjunto de valores (15). Además, se caracteriza por proporcionar mecanismos para definir reglas de integridad, ya sean inherentes al modelo, como propias de los usuarios (16).

4.2.2 EXPORTACIÓN DE LAS BASES DE DATOS

A partir de la base de datos depurada correspondiente a las respuestas del cuestionario se seleccionaron las variables correspondientes a cada una de las tablas correspondientes con el diseño de la base relacional propuesto y se crearon 4 *data frame* independientes almacenados en una lista. Para la exportación de dichas bases de datos de R a Microsoft Excel se utilizó el paquete “XLConnect”.

4.2.3 DIGITALIZACIÓN DEL CUESTIONARIO

CALIDAD DE LOS DATOS RECOGIDOS

La recogida de los datos es un proceso iterativo que requiere recolectar, transformar, almacenar, depurar y analizar la información disponible. Generalmente, este procedimiento se lleva a cabo por una o varias personas de la misma o diferentes organizaciones, durante un intervalo amplio de espacio y tiempo (8).

Arts et al. muestran una lista de recomendaciones con el fin de garantizar la calidad de los datos recogidos (17). Está basada en previos casos de insuficiencia de datos y experiencias descritas en la literatura. Entre ellas, se encuentran las siguientes sugerencias: tener el menor conjunto de datos posibles a rellenar; definir los datos y sus características en un diccionario de datos; diseñar un protocolo o guía; y explicar las posibles dificultades que pueda encontrarse el usuario.

Por otro lado, se recomienda también crear un diseño digitalizado del cuestionario, ya que la introducción manual de los datos tiene un alto riesgo de cometer errores. Una técnica habitual para evitar errores de este tipo es especificar y mantener la base de datos restringida, incluyendo comprobaciones de datos, límites en los valores numéricos y detección de datos que no existen (8). Sin embargo, forzar a cumplir ciertas condiciones puede crear frustración en el usuario encargado de transcribir las respuestas del cuestionario y consecuentemente, los datos estén inventados o no estén introducidos de la manera correcta. Por ello, es una labor muy importante lograr el equilibrio en las restricciones implementadas (8).

CUESTIONARIO ELECTRÓNICO

Para la digitalización del cuestionario de Kilisun se utilizó Epi Info TM versión 7.2.2.2 (de aquí en adelante se denomina como “Epi Info”). Se trata de un conjunto de herramientas

de software gratuito, diseñadas para los investigadores dedicados a la salud pública (18). Incluye herramientas de gestión, análisis y presentación de los resultados, proporcionando al usuario un entorno simple y a su vez, poderoso para llevar a cabo procesos de análisis de datos (19). Además, tiene la capacidad de almacenamiento de los datos centralizado y el acceso simultáneo de múltiples usuarios (18).

Al crear un proyecto en Epi Info, se pueden establecer tantos cuestionarios en él como se desee. Además, en estos cuestionarios se introducen las páginas que sean necesarias para poder abarcar toda la información que sea imprescindible y posteriormente, crear automáticamente una tabla por cada página del cuestionario del proyecto.

Una vez se han registrado los datos, existen diferentes posibilidades de análisis, de las cuales cabe mencionar la creación de tablas de frecuencias, tablas de contingencia 2x2, medias, gráficos, etc. dentro de la sección de “estadísticas básicas”; así como regresión lineal, regresión logística o análisis de supervivencia de Kaplan-Meier, dentro de las “estadísticas avanzadas” (20). Aun así, el cuestionario electrónico permite exportar y analizar los datos desde una perspectiva estadística en otras plataformas o lenguajes para un posterior análisis y tratamiento estadístico de los datos

Otra posibilidad que ofrece Epi Info es completar los cuestionarios en una aplicación gratuita para dispositivos Android e iOS. Resulta de gran utilidad para la recogida de datos en lugares sin grandes infraestructuras tecnológicas ya que no es necesaria una red inalámbrica, si no que basta con conectar el dispositivo mediante un cable USB al ordenador donde se encuentra la base de datos general, y transferir los nuevos registros (21).

5. RESULTADOS

5.1 ANÁLISIS DE LAS VARIABLES DEMOGRÁFICAS

Los datos resultantes de la adaptación de la base de datos integrada al análisis de supervivencia se resumen en el diagrama de flujo de la Figura 10. El número de participantes se vio reducido a un total 2669, al introducir las restricciones comentadas en la metodología, tres de las regiones de Tanzania no tenían ningún paciente incluido en la base de datos integrada para el análisis de supervivencia.

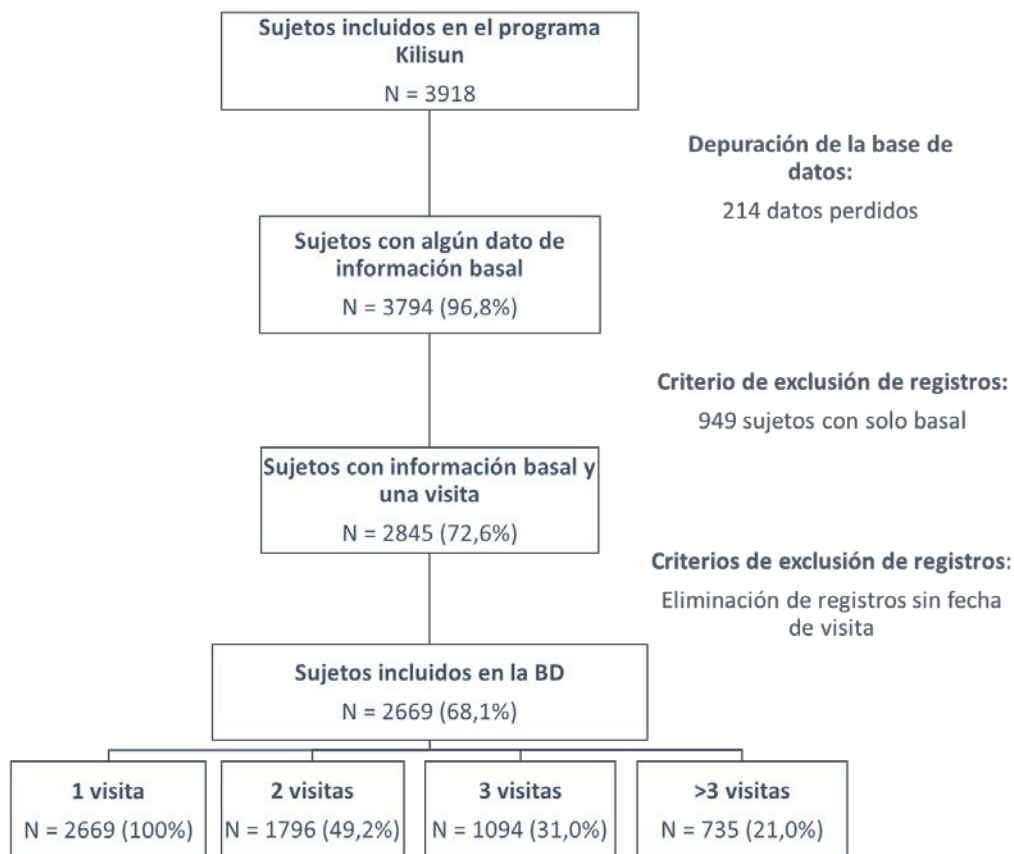


Figura 10. Diagrama de flujo de los pacientes registrados en la bases de datos integrada para el análisis de supervivencia.

Se realizó un estudio exploratorio previo de las variables que recogían la información de la edad y el sexo. Un análisis de frecuencias mostró que un 52 % de los sujetos eran mujeres, mientras que un 48% resultaron ser hombres, no observándose una descompensación entre sexos. La distribución de la edad (Figura 11) mostró no corresponderse con una distribución normal, el primer cuartil de los datos se situó en 10

años, la mediana en 16 años y el tercer cuartil en 25 años, siendo la media 19,8 años. Se observó en el histograma cómo la densidad de datos era muy superior para edades inferiores a 20 años respecto a los mayores de 20 años.

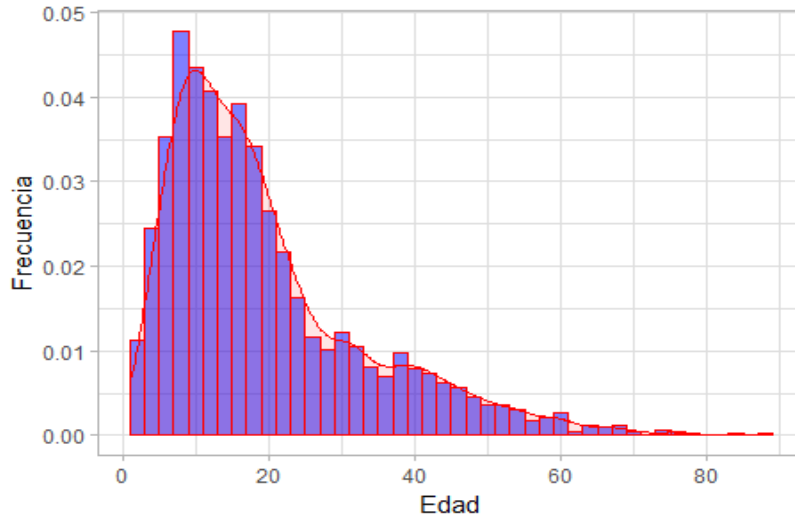


Figura 11. Histograma de la variable edad en los sujetos de la base de datos para el análisis de supervivencia. Se superpone la función de densidad de la variable edad.

A la vista del comportamiento de la distribución de la edad se decidió categorizar la variable edad en menores de 20 años (<20) y mayores o iguales a 20 años (≥ 20) ya que hasta 20 años se obtenía un 60% de la población.

5.1.1 CURVAS DE SUPERVIVENCIA DE KAPLAN-MEIER

La curva de supervivencia estimada por el método de Kaplan-Meier realizada en R (Figura 12) muestra que tiempo mediano de las visitas es de 635 días, es decir, el 50% de las visitas se han presentado a los 635 días.

De igual forma se analizaron las curvas de supervivencias para los distintos niveles de las variables sexo y edad categorizada por el Método de Kaplan Meier.

El gráfico de supervivencia en función del sexo (Figura 13) mostró que las curvas de supervivencia estaban superpuestas, siendo indistinguibles en todo momento. Esto se ve reflejado en el p-valor del test de Log-Rank (p-valor = 0.89) lo que indica que no existen diferencias entre hombres y mujeres. De igual forma el número de pacientes en riesgo se modifica a lo largo del tiempo de manera equivalente en ambos grupos tomando valores similares.

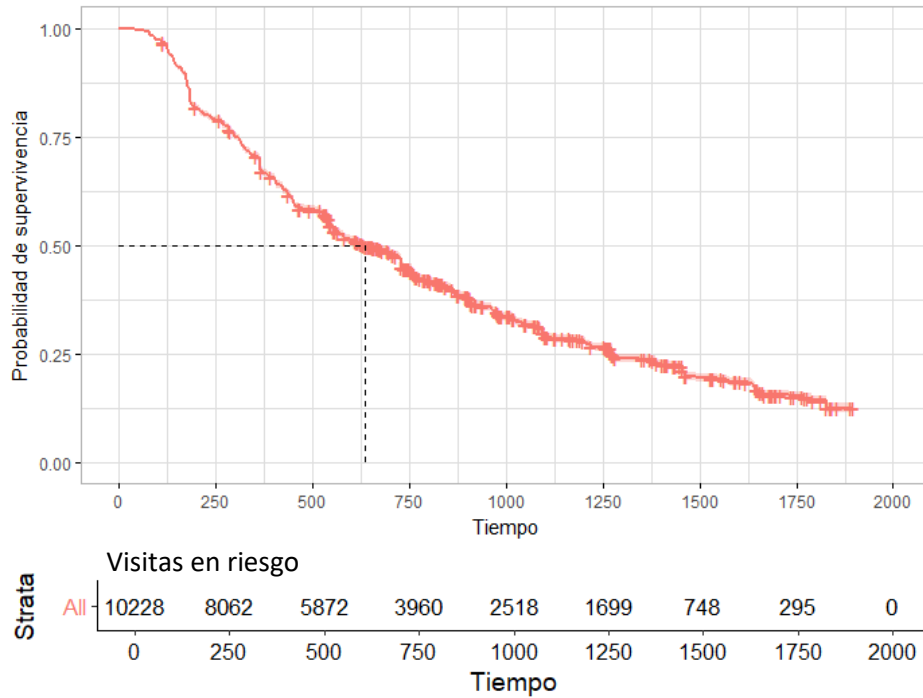


Figura 12. Curva de supervivencia para las visitas de los sujetos realizada en R. Se indica en la parte inferior el número de pacientes en riesgo.

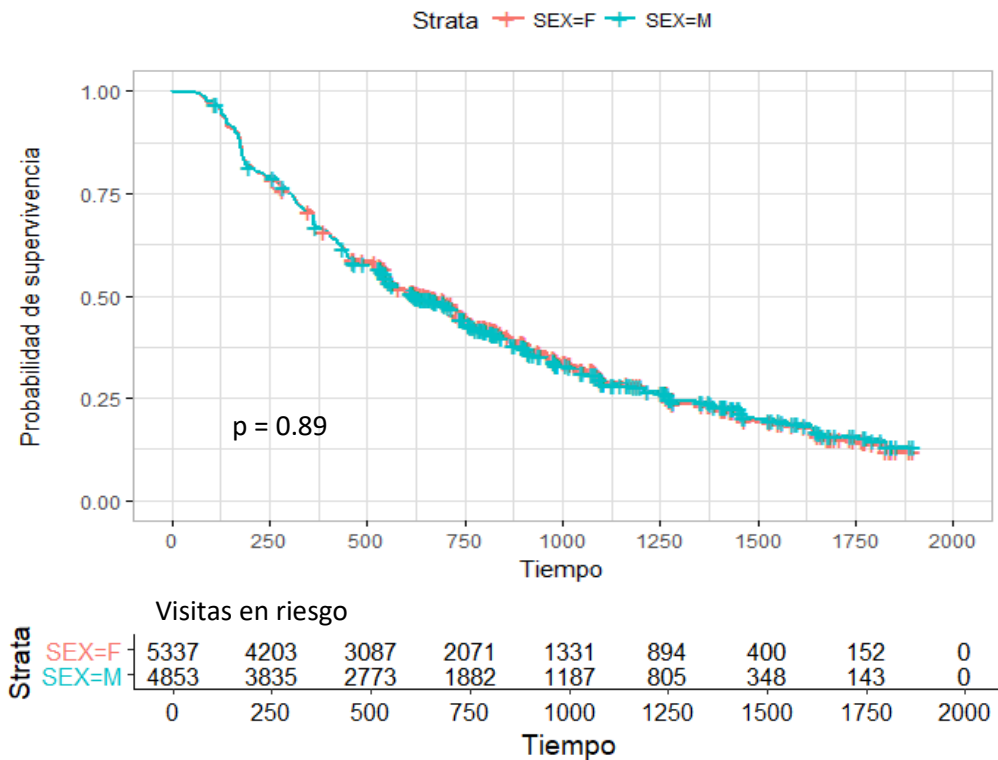


Figura 13. Curva de supervivencia para las visitas de los sujetos por sexo (M = hombre, F = mujer) realizada en R. Se indica en la parte inferior el número de pacientes en riesgo. Se indica el p-valor del test de Log-Rank

Al realizar el gráfico de supervivencia en función de la edad, categorizada de acuerdo con lo antes expuesto (Figura 14), se observó que las curvas de supervivencia eran distinguibles entre ellas, situándose la correspondiente a los menores de 20 años por debajo de la curva de los mayores de 20. El p-valor del test de Log-Rank era inferior a 0.05, indicando que existen diferencias entre los niveles de la variable.

De manera global estas curvas no se cruzan entre ellas a lo largo de la línea de tiempo, salvo en el período final de observación, período en el cual el número de eventos se reduce. Por lo tanto, la variable edad cumple la condición de riesgos proporcionales. Además, la tabla de individuos en riesgo indica cómo de partida hay 6115 sujetos con menos de 20 años y 4113 con 20 o más años.

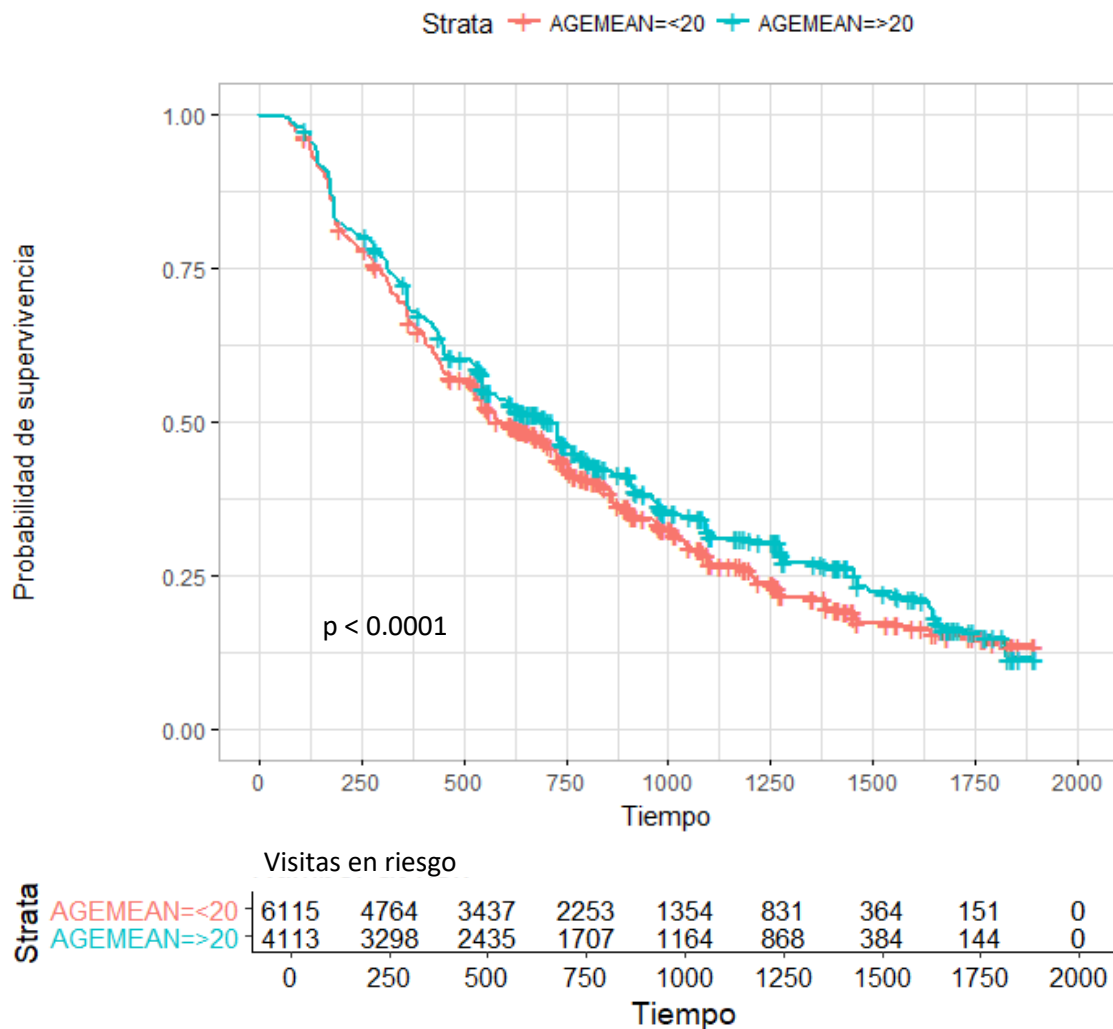


Figura 14. Curva de supervivencia para las visitas de los participantes por nivel de edad realizada en R. Se indica en la parte inferior el número de pacientes en riesgo. Se indica el p-valor del test de Log-Rank.

5.1.2 MODELO DE REGRESIÓN DE COX

Para la obtención del modelo de supervivencia de eventos recurrentes iguales con orden relevante se realizó el modelo univariante para la variable de edad categorizada al ser la única que mostró cumplir la hipótesis de riesgos proporcionales en los gráficos de las funciones de supervivencia (Tabla 3).

En el modelo la edad mostró una asociación significativa con la supervivencia, confirmando lo observado en las curvas de supervivencia, el intervalo de confianza del HR no incluyó a la unidad y el p-valor fue inferior a 0.05, lo que indica que tener 20 o más años reduce el riesgo de producirse un nuevo evento un 17% respecto a tener menos de 20 años entre sujetos en un mismo número de visitas (HR = 0.83). En otras palabras, los sujetos con menos de 20 años tienen una predisposición de realizar una nueva visita un 20% mayor que aquellos individuos con más de 20 años.

| Variable | n (visitas) | Estimación (SE robusto) | HR univariante (IC 95%) | p valor |
|-------------------|----------------|----------------------------|----------------------------|---------|
| Edad (20+ vs <20) | 10228 | -0.18 (0.02) | 0.83 (0.79,0.87) | <0.001 |

Tabla 3. Análisis de riesgos proporcionales de Cox univariante para eventos recurrentes estratificado por el número de visita.

Los resultados de los análisis univariantes están en concordancia con lo observado en las curvas de supervivencia. No se realizó un análisis multivariante al tener sólo una variable significativa.

De igual forma se estudió la posibilidad de que la variable sexo fuese una variable confusora incluyéndose en el modelo, el criterio utilizado para considerar la variable como confusora es la modificación de HR de la variable edad en más de un 10%. Al realizar el nuevo modelo no se observó una variación superior al 10% del HR por lo que se concluye que el sexo no es una variable confusora de la edad en este modelo.

Los modelos y salidas utilizadas en R, así como el diagnóstico del modelo se pueden consultar en Anexos (*D. Modelos de Cox y diagnóstico*).

5.2 EVALUACIÓN DEL TIEMPO MEDIO ENTRE VISITAS

Se evaluó la distribución del tiempo medio transcurrido entre dos visitas para los sujetos incluidos en el programa Kilisun (Figura 15). La media de este tiempo es de 230 días, lo

que sería algo más de 7 meses y medio, no obstante, existen valores de tiempo extremos que podrían ser *outliers* o datos erróneos por la falta de datos de visitas intermedias.

Para evitar el efecto que tienen estos datos sobre la medida se calculó la mediana del tiempo transcurrido entre visitas, siendo este valor de 189 días, aproximadamente medio año (183 días) entre dos visitas. Adicionalmente el primer y tercer cuartil recogen los datos entre 149 y 273 días. Si se comparan estos valores con el de la mediana se puede reconocer una asimetría hacia la derecha de los datos.

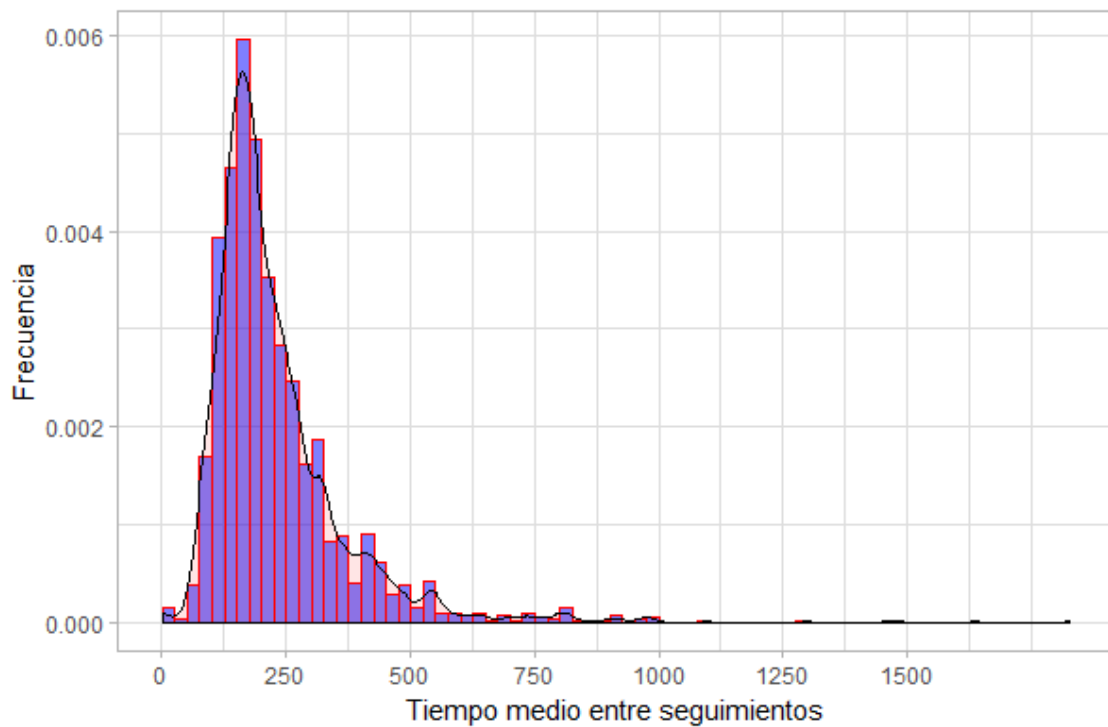


Figura 15. Histograma del tiempo medio transcurrido entre visitas en los sujetos del proyecto Kilisun. Se superpone la función de densidad de dicha variable.

Los datos del tiempo medio entre visitas por cada región de Tanzania (Figura 16) mostraron comportamientos heterogéneos. Algunas regiones como Manyara, Morogoro o Pwani tienen un tamaño muestral pequeño, además en estas regiones se reduce la dispersión de los datos, las razones de este comportamiento se comentan en el apartado de Discusión. También se observa como el número de outliers aumenta en las regiones que poseen un tamaño muestral elevado como Kilimanjaro o Mwanza.

Los datos de los estadísticos descriptivos utilizados (media, mediana, cuartiles, desviación estándar, error estándar, máximos y mínimos) para cada región se pueden consultar en Anexos (*E. Estadísticos descriptivos del tiempo entre visitas*).

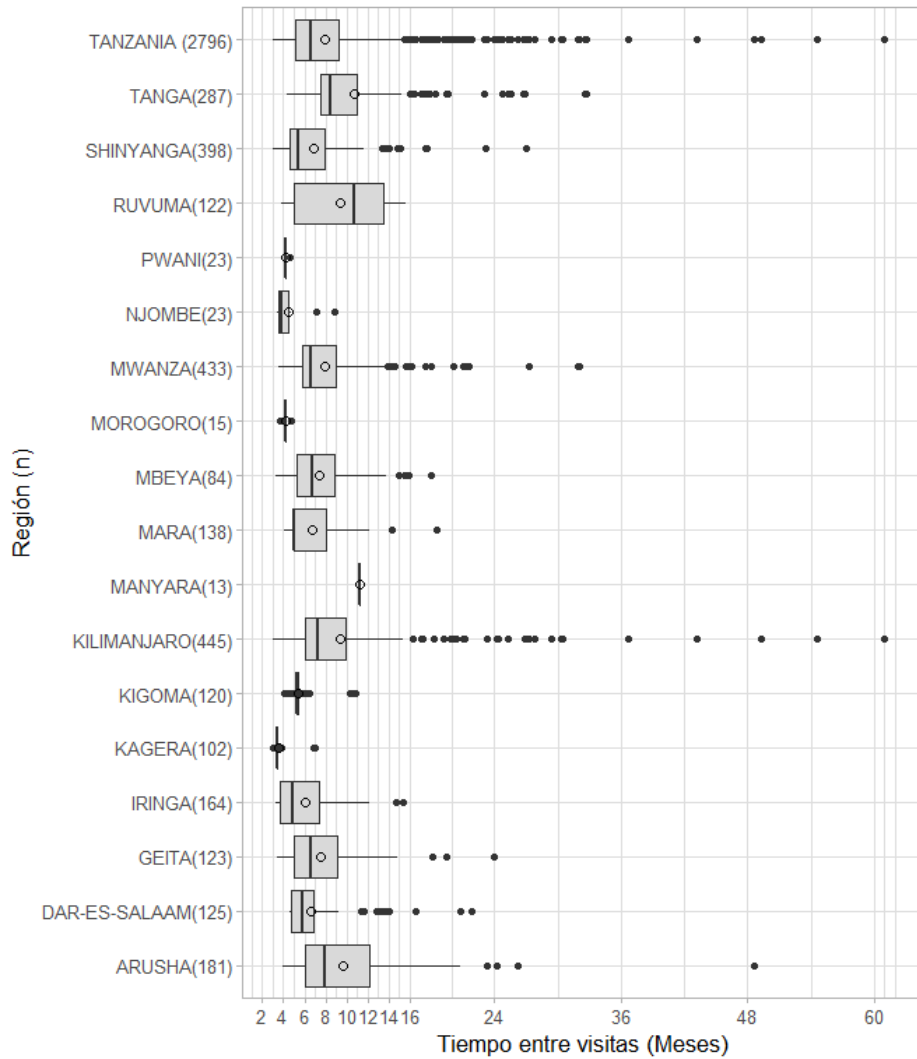


Figura 16. Diagrama de cajas para el tiempo medio transcurrido entre visitas de los participantes de las distintas regiones de Tanzania. Se indica con un punto sin rellenar el valor de la media en cada caso.

5.3 PLATAFORMA DE RECOGIDA DE DATOS

El cuestionario realizado en Epi Info es muy similar al realizado en papel para que la introducción de los datos sea sencilla y rápida para el usuario. Además, se ha hecho uso del lenguaje de programación de Epi Info con la finalidad de evitar errores de formato e incoherencias en la entrada de datos.

5.3.1 CODIFICACIÓN DEL CUESTIONARIO PILOTO

El primer paso fue realizar un cuaderno anotado, donde conjuntamente se unificó el nombre de las variables y los identificadores del cuestionario piloto de formato papel. A la hora de transcribirlo a Epi Info, se tuvieron en cuenta las 3 secciones descritas previamente en *Fuentes de datos* en el apartado de *Cuestionario de Kilisun Care*.

La primera de ellas fue la sección de seguimiento (*screening visit / enrollment visit y follow-up*), corresponde a la recogida de datos general de cada participante y como ya se ha dicho, se completa en cada visita que se realiza. Esta sección puede separarse en 4 módulos, en función del tema de interés:

- **DEMOG**: contiene los datos demográficos de cada paciente. Si observamos el cuestionario en formato papel, incluye el apartado de *PWA GENERAL INFORMATION* (información general de las personas con albinismo) y *MEDICAL HISTORY FOR A PWA/XP* (historial médico de personas con albinismo o XP).
- **SHABITS**: corresponde al hábito de protección solar. Es el apartado *SUN PROTECTION HABITS* (hábitos de protección solar).
- **DISTRIBUTION**: distribución de los botes Kilisun. En el cuestionario es el apartado *SUN PROTECTIVE MEASURE GIVEN* (medida de los protectores que se proporcionan). Además, cuando se trata de follow up, se incluye también *SUN PROTECTIVE MEASURE RETURNED* (medida de los protectores que se devuelven).
- **EDUCATION**: educación que se les proporciona a los pacientes. Está formado por la sección del cuestionario *EDUCATION PROVIDED* (educación proporcionada).

De esta manera, tanto el cuestionario de *screening visit* como el de *follow-up* pueden reconstruirse mediante estos 4 módulos, ya que la mayoría de las preguntas son comunes en ambos cuestionarios.

Ahora, teniendo en cuenta que *DEMOG* solamente ha de rellenarse cuando un nuevo paciente se inscribe en el programa (porque está constituido por los datos demográficos del participante), se decide distribuir estos 4 módulos en 2 cuestionarios de Epi Info; por un lado *DEMOG* (Figura 17), que se completa cuando el individuo no tiene ningún registro en la base de datos hasta el momento; por otro lado, se agrupan *SHABITS*, *DISTRIBUTION* y *EDUCATION* en otro nuevo cuestionario llamado *QUESTIONNAIRE* (Figura 18), y que se debe completar siempre, independientemente de la visita de la que se trate.

En resumen, el proyecto de Epi Info que se ha creado (*Cuestionario.prj*) consta de un total de 4 cuestionarios, incluyendo el cuestionario de satisfacción (Figura 19) y de discontinuación (Figura 20).

Figura 17. Cuestionario DEMOG del proyecto de Epi Info.

Figura 18. Cuestionario QUESTIONNAIRE del proyecto de Epi Info compuesto por: SHABITS (A), DISTRIBUTION (B) y EDUCATION (C).

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE
OPTIONAL VISIT SATISFACTION

Visit type: Visit number: PATIENT CODE:

DD MM YYYY
 DATE OF VISIT:

The time needed for the daily application is appropriated

The treatment does not limit my daily activities

The treatment does not limit my work/school commitments

My skin looks better after I started using Kilisun

I am more confident as a result of the treatment

As a result of the treatment, I am not worried that my skin condition will get worse

I consider the improvement in the condition of my skin to be good

I changed by sun protection behaviour due to the education I receive

I would recommend the product

Filled by:

Figura 19. Cuestionario *SATISFACTION* del proyecto de Epi Info.

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE
DISCONTINUATION VISIT

Visit type: Visit number: PATIENT CODE:

DD MM YYYY
 DATE OF VISIT:

If the patient has missed more than 2 visits in a row please fill out the information below.

Reason for program discontinuation:

- Too expensive to reach the place
- Too frequent clinics
- Lack of awareness
- Lack of time
- Lack of efficacy
- Migration to areas where Kilisun does not reach
- Death, indicate date: Died of:
- Other reason

Filled by:

Figura 20. Cuestionario *DISCONTINUATION* del proyecto de Epi Info.

5.3.2 MODIFICACIONES DEL CUESTIONARIO

Al tener un mismo cuestionario (*QUESTIONNAIRE*) para la visita de inscripción y para la de seguimiento, en cada nuevo cuestionario se han incluido dos nuevas variables para identificar tanto el tipo como el número de visita: *Visit type* y *Visit number*. En el caso del tipo de visita se selecciona entre *screening visit*, *follow-up*, *discontinuation* o *satisfaction*. En el número de visita, se asigna un número en función del tipo de visita que sea (Tabla 4).

| Cuestionario | <i>Visit type</i> (Tipo de visita) | <i>Visit number</i> (Número de visita) |
|-------------------------------|---|---|
| <i>DEMOG</i> | <i>SCREENING VISIT</i> | 0 |
| <i>QUESTIONNAIRE</i> | <i>ENROLLMENT VISIT</i> | 0 |
| | <i>FOLLOW UP</i> | [1,98] |
| <i>DISCONTINUATION</i> | <i>DISCONTINUATION</i> | 99 |
| <i>SATISFACTION</i> | <i>SATISFACTION</i> | [1,98] |

Tabla 4. Reglas de asignación del número de visita en función del cuestionario y tipo de la visita.

En el caso del cuestionario *SATISFACTION*, como no se completa cada vez que un paciente asiste a la clínica, se asignará el mismo número de visita que corresponda a *QUESTIONNAIRE* en el momento que se rellena.

Las variables *VISIT TYPE* y *VISIT NUMBER* que recogen el tipo de visita junto con el código del paciente (*PATIENT CODE*) son necesarias en todos los cuestionarios, de tal forma que cada registro quede perfectamente identificado y sea un registro consistente en la base de datos.

Por otro lado, se creó una nueva variable que recoge la edad de ingreso en el programa del participante. Esta casilla se rellena automáticamente sólo si previamente se ha introducido la fecha de la visita y fecha de nacimiento. Puede verse en la Figura 17. Finalmente, en la base de datos proporcionada por Kilisun se incluían otras dos variables que no pueden encontrarse en el cuestionario de formato papel: ocupación del paciente y su estado civil. Ambas fueron incluidas en el cuestionario *DEMOG* (Figura 17).

5.3.3 DESCRIPCIÓN DE LOS CUESTIONARIOS DE EPI INFO

Una de las características de Epi Info es la posibilidad de personalizar el cuestionario en función de las necesidades del usuario. Se ha mencionado que muchos errores en la base de datos provienen de la introducción manual de datos, entre ellas, el formato incorrecto de las variables. En este apartado se explicarán las normas y restricciones que se han implementado en el cuestionario para tratar de reducir al máximo la cantidad de errores de este tipo.

BÚSQUEDA AUTOMÁTICA

Como se ha mencionado antes, al crear el cuestionario en Epi Info se define la variable que recoge el número de visita del paciente. Esta información se desconoce a la hora de completar el cuestionario en papel o introducirlo en la base de datos. Para facilitar esta búsqueda al usuario que completa el cuestionario en Epi Info, tras indicar el tipo de visita del participante, salta en pantalla un cuadro de texto donde se muestran todos los registros del paciente y las visitas que ha realizado hasta el momento. Por lo tanto, bastaría con meter el número de la siguiente visita que le corresponda. La Figura 21 es un caso particular de búsqueda automática.

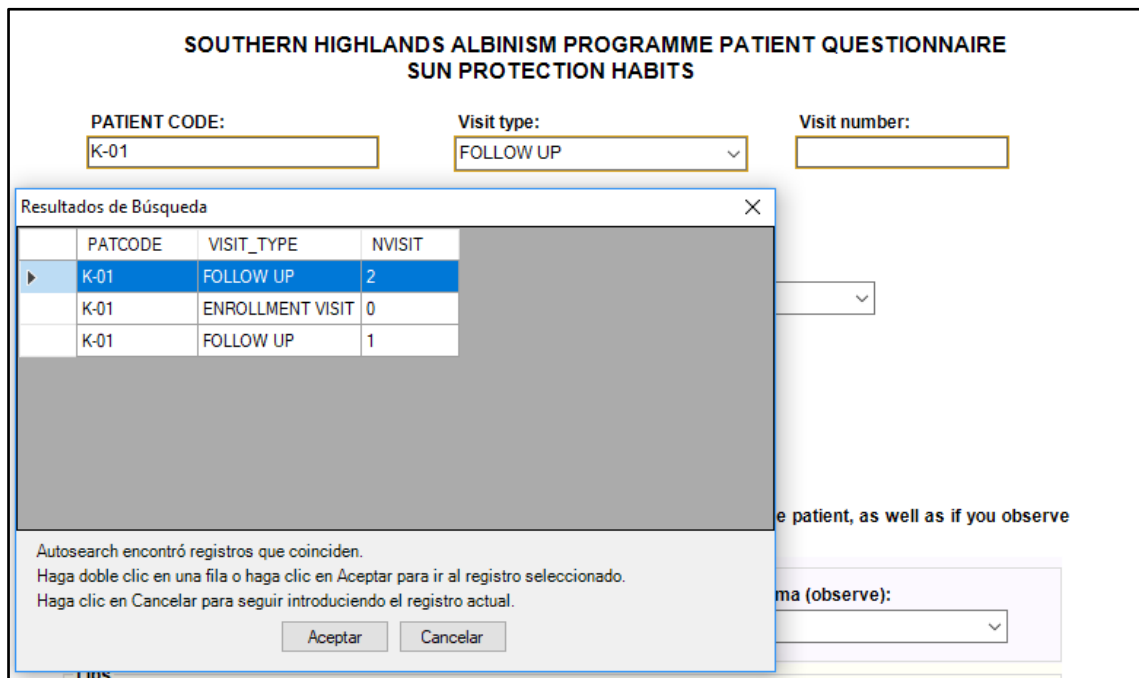


Figura 21. Búsqueda automática de registros del paciente K-01. El tipo de visita indicado es de seguimiento (*follow-up*), y el número correspondiente de la visita sería el 3, ya que el último registro fue la 2.

AUTORELLENO

Existen variables para las cuales se conoce la respuesta sin que el usuario tenga que introducirla. Por ejemplo, en el caso del cuestionario *DEMOG*, el tipo de visita es siempre de inscripción, es decir, *screening visit*, y el número de la visita, al tratarse de la basal, es 0. Lo mismo pasa en el cuestionario de discontinuación. La casilla está previamente programada para que se rellene automáticamente. De hecho, en la Figura 20 podemos ver cómo las casillas correspondientes al tipo de visita y al número de visitas están completadas al abrir el cuestionario. De esta manera, no sólo evitamos errores en la introducción de los datos si no que agilizamos el proceso de registro.

BLOQUEO DE CASILLAS

Otro recurso para evitar datos incoherentes es el bloqueo de casillas. Estas casillas sólo están habilitadas cuando se cumplen ciertas condiciones. Por ejemplo, si la pregunta es si ha tenido algún diagnóstico de cáncer de piel previamente y la respuesta es “no”, entonces la pregunta “¿Cuándo?” queda deshabilitada y no pueden introducirse datos en la casilla correspondiente.

Este método también es utilizado al realizar la selección del tipo de visita. Es cierto que aparentemente los cuestionarios iniciales de inscripción y seguimiento son similares, exceptuando la sección donde se recogen los datos demográficos. Pero al haber creado un cuestionario general (*QUESTIONNAIRE*, Figura 18), existen preguntas que no son comunes para la visita de inscripción y la de seguimiento y que, sin embargo, aparecen siempre en el cuestionario. Un ejemplo claro son las preguntas correspondientes a los botes que el paciente lleva de vuelta a la clínica. Esta pregunta no puede contestarse si es la primera vez que se participa en el proyecto Kilisun. Para solucionar este problema, se introduce una función en el lenguaje de programación de Epi Info para que cuando el usuario introduzca el tipo de visita, el formulario se adecúe automáticamente y se deshabiliten o desaparezcan de pantalla las preguntas que no deben contestarse.

RESPUESTAS LIMITADAS

Se ha visto en el apartado de metodología que una estrategia para reducir los errores es restringir las respuestas del cuestionario. Por esta razón, en muchas casillas hay que elegir entre una serie de opciones predeterminadas. Se utiliza en variables Sí/No, género, educación, tipo de visita y tipo de clínica, entre otras.

FECHAS

Uno de los problemas principales de la base de datos inicial son los distintos formatos de fecha. Para evitarlo, se muestra en el cuestionario de Epi Info la manera correcta de introducirlo: dd/mm/aaaa. Además, para cada dato se dispone de una casilla con número máximo de caracteres a introducir: en caso del día y mes es 2; en el caso del año 4.

5.3.4 PARTICIPANTES

Debido a la estructura de las dos bases de datos, resulta muy complicada su unión y de cara a incluirla en Epi Info, se tomó la decisión de considerar solamente la base de datos piloto.

Todos los pacientes de la base de datos básica vienen incluidos en la correspondiente al cuestionario piloto con la visita de inscripción ya completada, pero sin ningún registro de seguimiento. Bajo esta condición, la distribución del número de pacientes por visita viene representada en la Tabla 4.

| Cuestionario | Basal | Visita 1 | Visita 2 | Total |
|----------------------|--------------|-----------------|-----------------|--------------|
| <i>DEMOG</i> | 790 | - | - | 790 |
| <i>QUESTIONNAIRE</i> | 790 | 319 | 252 | 1361 |

Tabla 4. Distribución de los pacientes, en función de la visita y cuestionario incluidos en Epi Info. En la visita basal se toman en cuenta todos los participantes, y en las de seguimiento solamente aquellos que han realizado la visita.

5.3.5 ANÁLISIS DESCRIPTIVO

En Epi Info cabe la posibilidad de realizar un análisis estadístico básico que se actualiza cuando se ingresan nuevos registros. En este proyecto, se proporciona a los usuarios uno descriptivo donde poder visualizar la información recogida de ciertos aspectos como la distribución de los botes y la utilidad de la educación proporcionada. Existen 2 archivos con análisis estadístico: *DEMOGRAPHY*, que corresponde al cuestionario *DEMOG*; y *QUESTIONNAIRE*, que como su nombre dice, corresponde al cuestionario *QUESTIONNAIRE*. En ellos están expuestos las frecuencias y los histogramas de las siguientes variables:

- *DEMOGRAPHY*: Edad, sexo, diagnóstico previo de cáncer de piel, familiares con albinismo y nivel de educación

- QUESTIONNAIRE:
 - ¿La cantidad de botes era suficiente?
 - Si no se devuelve ningún bote, ¿cuál es la razón?
 - ¿Han recibido educación?
 - Utilidad de la información recibida sobre el albinismo, fotoprotección y uso de Kilisun.

Esta herramienta informática se concede a los encargados de los cuestionarios para tener bajo control ciertos aspectos como la insuficiencia de botes o la educación que adquieren los participantes, y poder hacer hincapié sobre ello si los resultados no son los esperados. Aun así, ha de tenerse en cuenta que los resultados son meramente descriptivos.

5.3.6 GUÍA PARA EL USUARIO

Con el fin de facilitar el uso del cuestionario y siguiendo las recomendaciones descritas en Metodología, se ha creado una guía de uso que puede consultarse en Anexos (*F. Manual de Usuario de Epi Info*). Está formada por los apartados que se muestran a continuación:

1. Instalación del software Epi Info.
2. Guía para la entrada de datos. Se explica cómo abrir el formulario, crear un nuevo registro y buscar o editar cuestionarios ya realizados con anterioridad.
3. Guía para el análisis descriptivo.
4. Diccionario de datos.

5.3.7 APLICACIÓN PARA DISPOSITIVOS MÓVILES

Se modificó el formato de algunas variables para exportar el cuestionario a los dispositivos, ya que la aplicación móvil no dispone todavía de todas las opciones que la aplicación de escritorio posee. Un ejemplo es la búsqueda automática del número de la visita de los participantes, que no es posible en la aplicación móvil. Aun así, no supone un problema grave porque también puede hacerse manualmente. En la Figura 22 se muestra la interfaz del cuestionario en la aplicación móvil.

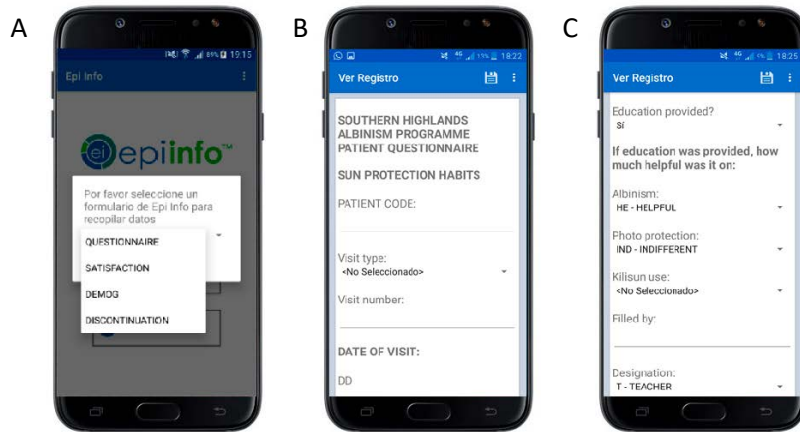


Figura 22. Cuestionario de Epi Info en un dispositivo móvil. En (A) se muestra la pantalla de elección de cuestionario. Las imágenes (B) y (C) corresponden a *QUESTIONNAIRE*: la de la izquierda es el inicio de *SHABITS*; la de la derecha forma parte del apartado de *EDUCATION*.

La razón de la elección del software Epi Info y sus ventajas se tratan en la siguiente sección de *Discusión*.

6. DISCUSIÓN

6.1 EVALUACIÓN DE LAS BASES DE DATOS

Como se ha comentado en el apartado “Evaluación de la calidad de la base de datos” en las bases de datos existían numerosos errores e inconsistencias difíciles de solucionar para realizar los objetivos planteados inicialmente. Un aspecto importante en este punto es analizar las posibles causas de este problema con el objetivo de minimizar las inconsistencias de las futuras bases de datos que lleve a cabo Kilisun.

Una de las principales causas de inconsistencias es el método de recogida que fue llevado a cabo, la realización de encuestas en papel conlleva transcribir los datos a mano en la base de datos de Microsoft Excel, este tipo de errores por acción humana son difíciles de solucionar, no obstante eliminar un paso de transcripción de datos reduciría los errores mecanográficos de las bases de datos. Por tanto, la solución más inmediata es eliminar la recogida de datos en papel y realizarla directamente desde un dispositivo electrónico.

No todas las inconsistencias encontradas fueron errores mecanográficos, en el análisis de los datos destacó el elevado número de datos faltantes y no disponibles dentro de las tres bases de datos, estos errores son debidos por un lado a la falta de previsión a la hora de crear la base de datos en Excel y por otro lado a la falta de supervisión en la introducción de datos. En la Figura 23 se muestra el tipo de inconsistencias observadas en la base de datos básica. La base de datos está en formato ancho, el principal problema observado fue que los datos recogidos no tienen por qué coincidir con el valor indicado en el encabezado de la columna, por ejemplo, los pacientes 4,5 y 6 de la Figura 23 no tienen los datos basales en la primera visita. Por tanto, no existe certeza de que los datos faltantes sean realmente *missing* o un error a la hora de colocar los datos en una columna incorrecta. De igual forma, los datos perdidos correspondientes a las últimas visitas, como en el caso del Paciente 3, no pueden ser distinguidos de una discontinuación del paciente con el proyecto (ausencia de datos de discontinuación). Distintas combinaciones de estos errores se encontraron en las bases de datos.

| Paciente | Basal | Visita 1 | Visita 2 | Visita 3 | Visita 4 | Visita 5 | Visita 6 | Visita 7 |
|----------|-------|----------|----------|----------|----------|----------|----------|----------|
| 1 | B | X | X | X | X | X | X | X |
| 2 | B | X | X | X | X | | X | X |
| 3 | B | X | X | | | | | |
| 4 | | B | X | X | X | X | X | X |
| 5 | | B | X | X | | | X | X |
| 6 | | | | B | X | X | X | |

Figura 23. Ejemplo de distribución de los datos a lo largo de las visitas. El valor “B” indica que es la visita basal, el valor “X” indica cualquier otra visita.

Es posible que estos errores sean producidos por la falta de conocimiento del personal encargado de la recogida de datos, siendo necesaria una formación previa de cómo han de introducirse los datos en una base de datos de formato ancho. Otra opción que podría haber reducido estos errores sería crear la base de datos en formato longitudinal en Excel (ordenando por la columna identificadora de paciente), de esta forma se eliminarían los errores producidos por incluir los datos en columnas equivocadas, únicamente quedando los verdaderos datos *missing*. De igual forma la anotación de la fecha e identificador de los pacientes cuando dejan de realizar visitas debieron ser recogidos para poder cerrar los registros de aquellos pacientes que no van a tener más visitas.

Ante la falta de datos consistentes, una prioridad del presente trabajo fue realizar un plan de recogida de datos adaptado a las necesidades y limitaciones del proyecto de Kilisun que minimizará los errores de la base de datos. Para la realización de dicho plan se decidió utilizar Epi Info como plataforma para la gestión de bases de datos de cuestionarios.

6.2 CUESTIONARIO EN EPI INFO

6.2.1 SOFTWARE PARA EL CUESTIONARIO

Los recursos disponibles de forma gratuita han evolucionado notablemente hasta día de hoy, existiendo todo un arsenal de herramientas que incluso pueden llegar a superar, en ciertos aspectos, a los productos comerciales más conocidos (19). Por esta razón, el software estadístico gratuito se presenta como opción principal para este proyecto debido a las limitaciones económicas de las que se disponen.

Los programas estadísticos gratuitos y más atractivos para la investigación mediante encuestas son Epi Info, WinIDAMS, ViSta y OS3 (19). Pero ¿por qué elegir Epi Info para llevar a cabo el proyecto?

En primer lugar, Epi Info se comunica fácilmente con bases de datos de diferentes formatos: Excel, Parados, dBase e incluso tablas HTML y procedimientos SQL. Para el interés de este proyecto, trabaja preferentemente con bases de datos relacionales, lo que es difícil de encontrar en otros programas de análisis de datos (20).

En segundo lugar, está disponible de manera gratuita y completa en Internet, sin restricciones del programa ni costos. Además, se puede acceder gratuitamente a las guías, manuales de referencia y tutoriales del software desde su página web y en diferentes idiomas (20). Este factor resultó ser determinante, ya que Kilisun desconocía el uso y existencia del programa Epi Info y era necesario un aprendizaje rápido del mismo. Se encuentra en una situación opuesta el programa ViSta, ya que la documentación es incompleta y desactualizada, lo que, sin lugar a duda, retrasaría el proceso de creación del formulario. Se debe mencionar además que sus versiones más recientes sólo están disponibles para el sistema operativo Windows y en inglés (19).

En cuanto a su interfaz, ofrece un entorno estructurado gráficamente, resultando muy sencillo de utilizar y entender incluso para usuarios no familiarizados con él. Si lo comparamos con WinIDAMS, este carece de una interfaz interactiva (19), por lo que automáticamente se eliminó de la lista de candidatos.

Finalmente, otra razón de peso para la elección de Epi Info fue que está muy extendido y utilizado en instituciones relacionadas con la salud. Su función principal es la gestión y análisis de datos provenientes de cuestionarios, pero, además, tiene la posibilidad de almacenar todos los datos en única base datos, dando acceso simultáneo a múltiples usuarios (18).

No obstante, su desventaja es la limitación en técnicas de visualización de datos. De hecho, al realizar el análisis descriptivo para el usuario, no permite representar los histogramas en función de porcentajes, por lo que es necesario acompañarlo siempre de una tabla de frecuencias para no dar lugar a errores. Tampoco da opción a crear diagramas de cajas, que podrían resultar interesantes para observar la distribución de ciertas variables a lo largo de las visitas, por ejemplo, el número medio de horas diarias al sol.

En consecuencia, para un análisis estadístico más completo, se recomienda sin duda el uso de otro software que permita mayor versatilidad, por ejemplo, R.

6.2.2 BASE DE DATOS

La base de datos que se importó a Epi Info fue solamente la correspondiente al cuestionario piloto. Realmente, la base de datos básica no aporta datos más allá de los demográficos (que se mantienen también en la del cuestionario piloto) y datos relacionados con la cantidad de botes de crema solar que reciben y devuelven en cada visita (diferencias entre bases de datos en Figura 3). Sin embargo, al prescindir de la base de datos básica se pierde el número de visitas que se han realizado hasta el momento en el que se implanta el nuevo cuestionario piloto.

Aunque de esta manera se pierda información acerca del número de visitas de los pacientes, el cuestionario electrónico se crea con intención de ponerlo en marcha en futuros proyectos, por lo que este inconveniente no surgirá cuando se implante esta nueva metodología desde el inicio de un estudio.

6.3 ANÁLISIS DE SUPERVIVENCIA

Se escogió un análisis de supervivencia de eventos recurrentes iguales debido al carácter recurrente del seguimiento de los participantes. En este punto se valoró si tomar como relevante el orden de las medidas repetidas (estratificación por el número de seguimiento en el modelo de Cox), se concluyó que no era comparable un paciente con una única visita con otro sujeto al que se le han realizado numerosas visitas ya que a medida que aumentan los reconocimientos médicos aparece un comportamiento de hábito en el participante, inexistente en aquellos que acaban de comenzar en el programa.

En este contexto el enfoque elegido para tratar el problema fue el método condicional 1 al ser el mejor modelo para contestar cómo varía el riesgo de evento desde que el individuo entra en el estudio hasta que sale de él ya que las observaciones temporales están referidas al momento de inicio. Otros modelos como el modelo condicional 2, que analiza el tiempo de cada evento desde el tiempo del último evento no respondían adecuadamente a los objetivos planteados.

Si los datos iniciales de Kilisun hubiesen proporcionado datos de discontinuación se podría haber realizado un análisis de supervivencia dentro del proyecto Kilisun, tomando

como evento la discontinuación y como censura permanecer en el programa. Este análisis podría aportar respuestas clave para comprender qué participantes son más propensos a abandonar el programa.

Únicamente se tuvieron en cuenta las variables Edad y Sexo, los resultados obtenidos en el modelo final concuerdan con lo esperado al observar las funciones de supervivencia. La edad no seguía una distribución normal, aunque esto no influya en el análisis realizado sí que se puede relacionar con el tipo de pacientes que recoge este estudio, las personas con albinismo tienen una esperanza de vida reducida lo que aumenta la densidad de pacientes jóvenes, no obstante, alguno albinos pueden alcanzar edades avanzadas como se ve reflejado en la Figura 11. En este análisis no se incluyó la variable categórica región ya que, al tenerla en cuenta, el número de eventos por cada nivel era muy reducido en las regiones con menos pacientes y por tanto la fiabilidad de los resultados sería muy baja.

El modelo final concluye que la edad, de acuerdo con la categorización realizada, es una variable predictora de la predisposición de realizar una nueva visita, esto podría ser tenido en cuenta por parte de Kilisun para elaborar un protocolo sobre este tipo de participantes para mejorar su seguimiento y poder así ampliar los beneficios potenciales en la salud del paciente. Por otro lado, el hecho de que los participantes más jóvenes tienen una mayor predisposición de realizar una nueva visita podría deberse a que los registros se realizan en las propias escuelas y centros de educación, facilitando la realización de las visitas.

6.4 VISITAS

Los valores medianos entre visitas indican que se realizaron en torno a cada seis meses, de acuerdo con lo que tenía planeado Kilisun, no obstante, se observó una gran dispersión de los datos y heterogeneidad entre las distintas regiones. Parte de esta variabilidad puede ser explicada por la presencia de datos erróneos que no pudieron ser corregidos en la depuración de la base de datos y *outliers*. La presencia de asimetría, además de los valores extremos, explica que el valor de la media (230 días) sea superior en 41 días a la mediana (189 días).

Por otro lado, al determinar los tiempos medianos entre visitas se observa como ya se comentó en el apartado de Resultados que aquellas regiones con menor número de participantes tuvieron valores inferiores de dispersión. Esto en parte es un artefacto ya que es posible que algunos centros médicos o sitios en los que se realizaban los

cuestionarios planificaran días concretos para realizar la visita, un ejemplo claro de esto sería la región de Manyara en la que los 13 pacientes tuvieron el mismo tiempo de seguimiento.

Los datos por regiones pueden ser utilizados por Kilisun para focalizar los esfuerzos en aquellas regiones con patrones de visitas que difieren más con los tiempos establecidos y así optimizar el uso de los recursos que poseen.

6.5 EL FUTURO DEL PROYECTO KILISUN:

6.5.1 NECESIDAD DE APOYO ESTADÍSTICO

Los especialistas que forman la organización *Kilisun Care* son principalmente farmacéuticos e ingenieros bioquímicos, pero hasta el plan de elaboración de un nuevo cuestionario no se recurrió a un estadístico. Desde el inicio se recogieron los datos en Microsoft Excel, que más que un software para la gestión de datos se trata de una hoja de cálculo, lo que desencadenó en una gran cantidad de datos perdidos e incoherentes. Todo ello lleva a pensar que un buen diseño de estudio previo hubiera reducido considerablemente los errores.

Después de la gran cantidad de dificultades que se han encontrado en el estudio a la hora de manejar las bases de datos, el soporte estadístico de cara al futuro es imprescindible. La gestión de los datos no debe suponer un problema, ya que, en este caso, no ha favorecido la realización de análisis estadísticos más exhaustivos y lograr así cumplir con los objetivos iniciales.

6.5.2 PERSPECTIVAS DE FUTURO

IMPLANTACIÓN DE EPI INFO

Hoy en día y en un futuro próximo, la metodología de recogida de datos seguirá siendo primero en formato papel y después su introducción manual en la base de datos electrónica. Aun así, se propone un nuevo plan de recogida de datos, donde el uso de nuevas tecnologías cobra un papel fundamental.

La implantación del cuestionario de Epi Info no supone un coste adicional y, además, se ha proporcionado una guía para su instalación y uso del mismo. Con esta metodología, se espera reducir notablemente la cantidad de errores de la base de datos, así como unificar

las variables de tipo fecha y texto. Asimismo, al almacenarse los datos automáticamente en una tabla, se mantendrán ordenados y sin visitas faltantes, facilitando al investigador la depuración de la base de datos previa al análisis estadístico. Este inconveniente ha ralentizado considerablemente el avance en el proyecto.

En cuanto a la implantación de una aplicación móvil en el programa de recogida de datos, hay que tener en mente las dificultades del uso de tecnologías avanzadas en África. Uno de los problemas principales es la inaccesibilidad actual a Internet en cualquier lugar y momento, de hecho, sólo el 21.8% de la población de africana utiliza Internet (22). Aun así, ya existe una iniciativa similar de una aplicación móvil para la recogida de datos de pacientes con cáncer de esófago que se está llevando a cabo en Moshi, Tanzania y Malawi (23). Por eso, aunque aparentemente sea un proyecto complejo y a largo plazo, se plantea un sistema donde se pueda en un futuro recoger datos tanto en un ordenador como en un dispositivo móvil (teléfono, tablet, etc.) y tener la base de datos sincronizada en tiempo real.

PROPÓSITOS DE CARA AL FUTURO

Este año 2018, *Kilisun Care* planea pilotar su programa en Malawi y Ruanda, lanzar una fórmula mejorada de protección solar, realizar un estudio de viabilidad en Fiji y llevar a cabo un estudio de impacto de 4 años. En 2019, su principal objetivo es presentar estudios piloto a las agencias de la ONU y otras organizaciones estratégicas, evaluar la posibilidad de implantar el programa en otros dos países y construir una unidad de producción en Malawi. Para 2020, el propósito es inaugurarla y lanzar el programa a nivel nacional en Malawi. Para entonces, la organización espera que 10.000 personas con albinismo reciban servicios de Kilisun en 3 países de África subsahariana.

En cuanto al análisis estadístico aún queda mucho camino por recorrer. Por un lado, se propone elaborar un plan de análisis en el que se defina la mejor estrategia para tratar los datos perdidos ya que con la información recogida hasta ahora no ha sido posible (ver Anexo C. *Datos no disponibles (Cuestionario piloto)*). Por otro lado, el objetivo se enfoca en definir los indicadores claves para obtener respuestas sobre los diferentes asuntos de interés que se presentaron al inicio del proyecto: la capacidad de los participantes para tomar medidas de prevención solar en sus cuidados; la evolución en la prevalencia de condiciones médicas asociadas con el desarrollo de cáncer de piel en las personas con albinismo; y el nivel de satisfacción de los pacientes con el programa.

7. CONCLUSIONES

La elaboración del presente Trabajo Fin de Máster ha permitido alcanzar una serie de conclusiones que se detallan a continuación:

- Los plataforma de recogida de datos utilizada hasta día de hoy no era la adecuada. Esto supuso una depuración de la base de datos muy laboriosa, que redujo el número de sujetos de 3919 a un total de 2669, reduciendo un 32% el volumen de los registros.
- La base de datos básica llevaba 4 años en curso, pero estaba compuesta por datos muy generales. Sin embargo, la base de datos piloto tenía información más específica, pero llevaba muy poco tiempo puesta en marcha. Dada la situación, no fue posible responder a los objetivos iniciales sobre la capacidad de los pacientes de tomar medidas de protección solar, la prevalencia del cáncer de piel, el patrón de datos perdidos y el nivel de satisfacción con el proyecto. En consecuencia, se decidió replantear el rumbo del proyecto estableciendo nuevos objetivos.
- Se propuso analizar el efecto de las variables demográficas sobre el riesgo de realizar una nueva visita. La variable edad fue significativa en el modelo de supervivencia propuesto. La predisposición de realizar una nueva visita se veía incrementada un 20% en aquellos participantes con menos de 20 años respecto al resto de la población. Esto podría darse por la gran involucración de Kilisun con colegios y centros de menores. En cuanto al sexo de los participantes, no influyó en el riesgo de realizar una nueva visita y además, no resultó ser una variable confusora de la edad.
- Para evaluar el seguimiento de los pacientes en el programa, se realizó un análisis descriptivo del tiempo entre visitas para el conjunto de todos los pacientes registrados y por región. El tiempo mediano transcurrido entre dos visitas fue de 189 días, aproximadamente medio año. Este valor corresponde con las previsiones de Kilisun, no obstante, la asimetría observada en los datos muestra una tendencia a realizar visitas más distanciadas de lo esperado.

- Si Kilisun ya era una organización referente por su empuje y sacrificio en la mejora de la calidad de vida de sus pacientes, con la nueva plataforma informática creada en Epi Info se convertirá en un paquete completo. De esta manera, se llenará el vacío estadístico que se disponía hasta día de hoy, y progresará como organización cuando se expandan a otros países y puedan aprovecharse de esta herramienta.

Con esta metodología se establece un nuevo protocolo de recogida y gestión de los datos con múltiples ventajas, tanto para la persona encargada de transcribir los datos como para el estadístico que realiza el análisis estadístico posterior.

Entre otras cosas, se proporciona al usuario la capacidad de introducir y almacenar los datos en un ordenador de una manera más dinámica, de forma que no interactúe directamente con la base de datos, sino que su labor tan solo sea transcribir los datos del cuestionario en formato papel al cuestionario digital, muy similares entre sí. A su vez también se evitan los habituales errores humanos ya que, en la medida de lo posible, el cuestionario digital restringe las respuestas permitiendo sólo ciertas opciones. Este tipo de errores ha requerido mucho tiempo al depurar las bases de datos de este proyecto por lo que se agilizará el proceso de limpieza.

Por otro lado, permite obtener estadísticas en tiempo real a los empleados de Kilisun, pudiendo detectar de manera inmediata cambios relevantes en variables indicadoras de la calidad del programa de Kilisun. Si el análisis estadístico proporcionado por Epi Info no fuera suficiente, la exportación de los datos a otra plataforma estadística más versátil resulta muy sencilla.

Finalmente, la adaptación del cuestionario electrónico para dispositivos móviles será un avance espectacular a medida que Internet sea un recurso accesible para la mayoría de la población africana.

8. BIBLIOGRAFÍA

1. Meredith P, Riesz J. Radiative relaxation quantum yields for synthetic eumelanin. *Photochem Photobiol.* 2004;79(2):211-6.
2. Brenner M, Hearing VJ. The protective role of melanin against UV damage in human skin. *Photochem Photobiol.* 2008;84(3):539-49.
3. Grønskov K, Ek J, Brøndum-Nielsen K. Oculocutaneous albinism. *Orphanet J Rare Dis.* 2007;2(1):1-8.
4. Standing Voice [Internet]. [citado 11 de junio de 2018]. Disponible en: <http://www.standingvoice.org/albinism>
5. Kilisun | Apoyo albinismo en África [Internet]. [citado 11 de junio de 2018]. Disponible en: <http://ponteensupiel.com/>
6. Velton R. The «silent killer» of Africa's albinos. *BBC* [Internet]. 2017 [citado 11 de junio de 2018]; Disponible en: <http://www.bbc.com/future/story/20170425-the-silent-killer-of-africas-albinos>
7. KiliSun Care [Internet]. [citado 11 de junio de 2018]. Disponible en: <https://www.onlinevolunteering.org/es/kilisun-care>
8. Hellerstein JM. Quantitative data cleaning for large databases. *United Nations Econ Comm Eur* [Internet]. 2008;42. Disponible en: <http://db.cs.berkeley.edu/jmh/cleaning-unece.pdf%5Cnpapers2://publication/uuid/DC7173AB-6B26-4B8B-AEC3-4C7E65CEEFED>
9. de Jonge E, van der Loo M. An introduction to data cleaning with R [Internet]. *Statistics Netherlands*; 2013. 53 p. Disponible en: http://cran.vinastat.com/doc/contrib/de_Jonge+van_der_Loo-Introduction_to_data_cleaning_with_R.pdf
10. Kleinbaum DG, Klein M. *Survival Analysis. A Self-Learning Text.* 2.^a ed. Gail M, Krickeberg K, Samet J, Tsiatis A, Wong W, editores. Vol. 53. Nueva York:

- Springer International Publishing; 2010. 590 p.
11. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. Source J Am Stat Assoc. 1958;53(282):457-81.
 12. Hosmer DW, Lemeshow S, May S. Applied Survival Analysis: Regression Modeling of Time to Event Data [Internet]. 2.^a ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2008 [citado 11 de junio de 2018]. (Wiley Series in Probability and Statistics). Disponible en: <http://doi.wiley.com/10.1002/9780470258019>
 13. Cox DR. Regression Models and Life-Tables. J R Stat Soc Ser B. 1972;34(2):187-220.
 14. Andersen PK. Statistical models based on counting processes. 1.^a ed. Nueva York: Springer-Verlag; 1993. 784 p.
 15. Hueso Ibañez L. Gestión de bases de datos [Internet]. 2.^a ed. Madrid: Ra-Ma Editorial; 2014 [citado 11 de junio de 2018]. 314 p. Disponible en: <http://ebookcentral.proquest.com/lib/univcomplutensesp/detail.action?docID=3228876>
 16. Rodríguez González ME. Gestión de datos: bases de datos y sistemas gestores de bases de datos [Internet]. Barcelona: Editorial UOC; 2013. Disponible en: <http://ebookcentral.proquest.com/lib/univcomplutensesp/detail.action?docID=3219201>
 17. Arts DGT, De Keizer NF, Scheffer G-J. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc [Internet]. [citado 11 de junio de 2018];9(6):600-11. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/12386111>
 18. Epi InfoTM | CDC [Internet]. [citado 11 de junio de 2018]. Disponible en: <https://www.cdc.gov/epiinfo/index.html>
 19. Ledesma R. Sistemas estadísticos de propósitos múltiples: una revisión de programas gratuitos. 2004 [citado 11 de junio de 2018];6(2):105-17. Disponible en: <http://casus.usal.es/pkp/index.php/MdE/article/viewFile/956/897>
 20. Paloma V, Manzano V. Epi Info: del mundo de la salud a la metodología de

- encuestas. 2004 [citado 11 de junio de 2018];6(2):177-82. Disponible en:
<http://casus.usal.es/pkp/index.php/MdE/article/viewFile/962/903>
21. Companion for Android: Collecting and Analyzing Data on an Android Device. En: Epi Info™ 7 User Guide [Internet]. 2016 [citado 11 de junio de 2018]. Disponible en: <https://www.cdc.gov/epiinfo/pdfs/UserGuide/EI7Full.pdf>
 22. ITU. ICT Facts and Figures 2017 [Internet]. [citado 12 de junio de 2018]. Disponible en: <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf>
 23. Pritchett N, Murphy G, Middleton D, McCormack V. Using a Mobile Health Application for Data Collection in Esophageal Cancer Case-Control Studies in Africa. *J Glob Oncol* [Internet]. 15 de marzo de 2018 [citado 13 de junio de 2018];(4_suppl):17s-17s. Disponible en: <http://ascopubs.org/doi/10.1200/jgo.18.90000>

ANEXOS

A. CUESTIONARIO PILOTO DE KILISUN

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE

SCREENING VISIT (1/1)

PWA GENERAL INFORMATION:

DATE OF VISIT: __/__/____ Condition: Albinism XP PATIENT CODE:

Name of the client (Last name/middle name/first name): _____

Clinic Registered: Hospital Health Centre/Dispensary School PWA Care Centre Others

Name of the Centre/clinic: _____ Patient telephone number : _____

Address: Region _____ District: _____ Ward/Village: _____

Sex: Male Female Year of Birth: _____

Educational Level: Primary Secondary College Never attended school Others

MEDICAL HISTORY FOR A PWA/XP:

Any direct relative with albinism: Yes No

Previous diagnoses of Skin Cancer? Yes No When (year): _____

SUN PROTECTION HABITS

Average n° daily sun hours: ____ Sunglasses use? Yes No

Has the patient ever used sunscreen in his/her life? Yes No Is he/she using sunscreen now? Yes No

If you use sunscreen (you responded yes in last question), how often do you apply it? Check all that apply:

Morning Afternoon Evening Additional Times

Please report below your observations on Sunscreen and clothing protection by the patient, as well as if you observe erythema

| | | |
|-------------------------------|---|---|
| | Face | Lips |
| Sunscreen use (ask) | <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection (observe) | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema (observe) | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |
| | Ears | Neck |
| Sunscreen use (ask) | <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection (observe) | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema (observe) | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |
| | Upper Limbs | Lower limbs |
| Sunscreen use (ask) | <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection (observe) | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema (observe) | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |

Filled by: _____ Signature: _____

Designation: KSPU staff Consultant Dermatologist DVO Nurse Teacher Others

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE

ENROLLMENT VISIT (1/1)

DATE OF VISIT: __/__/__

PATIENT CODE:

Name of the client (Last name/middle name/first name): _____

SUNPROTECTIVE MEASURE GIVEN

Name of sunscreen: Kilisun Other: _____

Total amount dispensed: 0 1B 2B 3B 4B 5B 1S 2S 3S 4S 5S

Batch No. of Containers: 1) _____ 2) _____ 3) _____ 4) _____ 5) _____

Protection elements: Hat Sunglasses Other _____

EDUCATION PROVIDED? Yes No

If education was provided, How much helpful was it on:

| | | | |
|------------------|-------------------------------|-----------------------------------|-----------------------------------|
| Albinism | <input type="radio"/> Helpful | <input type="radio"/> Indifferent | <input type="radio"/> Not helpful |
| Photo protection | <input type="radio"/> Helpful | <input type="radio"/> Indifferent | <input type="radio"/> Not helpful |
| Kilisun Use | <input type="radio"/> Helpful | <input type="radio"/> Indifferent | <input type="radio"/> Not helpful |

Filled by: _____

Signature: _____

Designation: KSPU staff Consultant Dermatologist ADDV Nurse TAS leader Teacher Others

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE

FOLLOW UP VISIT (1/1)

DATE OF VISIT: __/__/____

PATIENT CODE:

Name of the client (Last name/middle name/first name): _____

SUN PROTECTION HABITS

Average nº daily sun hours: ____ Sunglasses? Yes No

How often did you apply the sunscreen? Check all that apply:

Morning Afternoon Evening Additional Times

Please report below your observations on Sunscreen and clothing protection as well as if you observe erythema:

| | Face | Lips |
|---------------------|---|---|
| Sunscreen use | <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |
| | Ears | Neck |
| Sunscreen use | <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |
| | Upper Limbs | Lower limbs |
| Sunscreen Use | <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |

Actinic Keratosis (evaluated only in clinic distributions) Present Absent **Cryotherapy done:** Yes No

Suspicious malignancies: Yes No

SUNPROTECTIVE MEASURE RETURNED

No of Containers Returned: 0 1 2 3 4 5

If no containers return mark why: Forgot Lost not finished not aware others

Was the sunscreen provided sufficient? Yes No **If not, how long where you without sunscreen (weeks)?**

SUNPROTECTIVE MEASURE GIVEN

Name of sunscreen dispensed: Kilisun Other: _____

Sunscreen dispensed: 0 1B 2B 3B 4B 5B 1S 2S 3S 4S 5S

Batch No. of Containers dispensed: 1) _____ 2) _____ 3) _____ 4) _____

Protection elements: Hat Sunglasses Other _____

EDUCATION PROVIDED? Yes No

If education was provided, How much **helpful** was it on:

| | | | |
|------------------|-------------------------------|-----------------------------------|-----------------------------------|
| Albinism | <input type="radio"/> Helpful | <input type="radio"/> Indifferent | <input type="radio"/> Not helpful |
| Photo protection | <input type="radio"/> Helpful | <input type="radio"/> Indifferent | <input type="radio"/> Not helpful |
| Kilisun Use | <input type="radio"/> Helpful | <input type="radio"/> Indifferent | <input type="radio"/> Not helpful |

Filled by: _____

Signature: _____

Designation: KSPU staff Consultant Dermatologist ADDV Nurse TAS leader Teacher Others

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE

DISCONTINUATION VISIT (1/1)

DATE OF VISIT: __/__/____

PATIENT CODE:

If the patient has missed more than 2 visits in a row please fill out the information below:

Reason for program discontinuation:

- Too expensive to reach the place
- Too frequent clinics
- Lack of awareness
- Lack of time
- Lack of efficacy
- Migration to areas where Kilisun does not reach
- Death, Indicate date __/__/____ Died of: Skin Cancer Others
- Other reason

Filled by: _____

Signature: _____

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE

OPTIONAL VISIT SATISFACTION (1/1)

DATE OF VISIT: __/__/____

PATIENT CODE:

| | | Agree | Undecided | Disagree |
|---|--|-------|-----------|----------|
| | | 0 | 1 | 2 |
| 1 | The time needed for the daily application is appropriated? | | | |
| 2 | The treatment does not limit my daily activities | | | |
| 3 | The treatment does not limit my work /school commintments | | | |
| 4 | My skin looks better after I started using Kilisun | | | |
| 5 | I am more confident as a result of the treatment | | | |
| 6 | As a result of the treatment, I am not worried that my skin condition will get worse | | | |
| 7 | I consider the improvement in the condition of my skin to be good | | | |
| 8 | I cahnged by sun protection behaviour due to the education I receive | | | |
| 9 | I would recommend the product | | | |

Filled by: _____

Signature: _____

B. PAQUETES Y FUNCIONES DE R

A continuación, se muestran los paquetes utilizados que no se encuentran en el paquete base de R, así como la utilidad que se le dio:

- Paquete: *ggplot2*. Utilización de la función `ggplot()` para la creación de histogramas, gráficos de frecuencias y otros gráficos.
- Paquete: *ggmap*. Paquete complementario a *ggplot2* para la creación de mapas.
- Paquete: *pgdal*. Utilización de la función `readOGR()` para la lectura de datos que recogen información espacial en vectores. Lectura de mapas.
- Paquete: *survival*. Utilización de la función `survfit()` para la creación de curvas de supervivencia a partir de un modelo Cox previamente ajustado. Utilización de la función `coxph()` para el ajuste de modelos de regresión de Cox.
- Paquete: *survminer*. Utilización de la función `ggsurvplot()` para la creación de gráficos de funciones de supervivencia..
- Paquete: *readxl*. Utilización de la función `read_excel()` para la lectura de bases de datos en Excel.
- Paquete: *XLConnect*. Utilización de las funciones `loadWorkbook()`, `createSheet()`, `writeWorksheet()` y `saveWorkbook()` para la exportación de datos desde R a Microsoft Excel.

C. DATOS NO DISPONIBLES (CUESTIONARIO PILOTO)

Datos perdidos en las variables correspondientes a los datos demográficos (N=790):

| VARIABLE | GLOBAL | |
|-------------|-------------|-------|
| | N (missing) | % |
| PATNAME | 4 | 0,5% |
| SEX | 7 | 0,9% |
| DOB | 60 | 7,6% |
| CLINIC_NAME | 104 | 13,2% |
| REG | 243 | 30,8% |
| DIST | 359 | 45,4% |
| WARD | 62 | 7,8% |
| TLF. | 397 | 50,3% |
| STAT | 522 | 66,1% |

| VARIABLE | GLOBAL | |
|-------------|-------------|--------|
| | N (missing) | % |
| OCCUP | 525 | 66,5% |
| EDU | 523 | 66,2% |
| COND | 512 | 64,8% |
| REL_HIST | 572 | 72,4% |
| CANCER | 566 | 71,6% |
| YOC | 783 | 99,1% |
| DOV | 756 | 95,7% |
| CLINIC_TYPE | 790 | 100,0% |

Datos perdidos en las variables correspondientes a las visitas de los participantes (N=2370, 790 por visita):

| VARIABLE | GLOBAL | | VISITA BASAL | | VISITA 1 | | VISITA 2 | |
|------------|-------------|--------|--------------|--------|-------------|--------|-------------|--------|
| | N (missing) | % | N (missing) | % | N (missing) | % | N (missing) | % |
| SUNSCREEN | 1772 | 74,8% | 756 | 95,7% | 474 | 60,0% | 542 | 68,6% |
| DISPENSED | 1767 | 74,6% | 756 | 95,7% | 472 | 59,7% | 539 | 68,2% |
| JAR1 | 1785 | 75,3% | 757 | 95,8% | 479 | 60,6% | 549 | 69,5% |
| JAR2 | 1870 | 78,9% | 758 | 95,9% | 483 | 61,1% | 629 | 79,6% |
| JAR3 | 1923 | 81,1% | 768 | 97,2% | 512 | 64,8% | 643 | 81,4% |
| JAR4 | 2365 | 99,8% | 790 | 100,0% | 787 | 99,6% | 788 | 99,7% |
| JAR5 | 2370 | 100,0% | 790 | 100,0% | 790 | 100,0% | 790 | 100,0% |
| HAT | 1769 | 74,6% | 756 | 95,7% | 474 | 60,0% | 539 | 68,2% |
| SUNGLASSES | 1769 | 74,6% | 756 | 95,7% | 474 | 60,0% | 539 | 68,2% |
| OTHER | 1778 | 75,0% | 756 | 95,7% | 474 | 60,0% | 548 | 69,4% |
| RETURNED | 1800 | 75,9% | 790 | 100,0% | 471 | 59,6% | 539 | 68,2% |

| VARIABLE | GLOBAL | | VISITA BASAL | | VISITA 1 | | VISITA 2 | |
|----------------|-------------|--------|--------------|--------|-------------|--------|-------------|--------|
| | N (missing) | % | N (missing) | % | N (missing) | % | N (missing) | % |
| REASON | 1973 | 83,2% | 790 | 100,0% | 563 | 71,3% | 620 | 78,5% |
| SUFFICIENT | 1802 | 76,0% | 790 | 100,0% | 473 | 59,9% | 539 | 68,2% |
| WEEKS | 2329 | 98,3% | 790 | 100,0% | 765 | 96,8% | 774 | 98,0% |
| SUNSCREEN_OTH | 2370 | 100,0% | 790 | 100,0% | 790 | 100,0% | 790 | 100,0% |
| PROTECTION_OTH | 2370 | 100,0% | 790 | 100,0% | 790 | 100,0% | 790 | 100,0% |

Datos perdidos en las variables correspondientes a los datos de referentes a la educación de los participantes (N=2370, 790 por visita):

| VARIABLE | GLOBAL | | VISITA BASAL | | VISITA 1 | | VISITA 2 | |
|------------|-------------|-------|--------------|-------|-------------|-------|-------------|-------|
| | N (missing) | % | N (missing) | % | N (missing) | % | N (missing) | % |
| PROVIDED | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| ALBINISM | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| PROTECTION | 1772 | 74,8% | 756 | 95,7% | 476 | 60,3% | 540 | 68,4% |
| USE | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |

Datos perdidos en las variables correspondientes a los hábitos de salud (N=2370, 790 por visita):

| VARIABLE | GLOBAL | | VISITA BASAL | | VISITA 1 | | VISITA 2 | |
|----------------|-------------|-------|--------------|-------|-------------|--------|-------------|--------|
| | N (missing) | % | N (missing) | % | N (missing) | % | N (missing) | % |
| DOV | 1765 | 74,5% | 756 | 95,7% | 471 | 59,6% | 538 | 68,1% |
| HOURS | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| SUNGLASSES | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| SUNSCREEN_LIFE | 2336 | 98,6% | 756 | 95,7% | 790 | 100,0% | 790 | 100,0% |
| SUNSCREEN_NOW | 2336 | 98,6% | 756 | 95,7% | 790 | 100,0% | 790 | 100,0% |
| MORNING | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |

| VARIABLE | GLOBAL | | VISITA BASAL | | VISITA 1 | | VISITA 2 | |
|-----------|----------------|--------|----------------|--------|----------------|--------|----------------|--------|
| | N (missing) | % | N (missing) | % | N (missing) | % | N (missing) | % |
| AFTERNOON | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| EVENING | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| ADDIT | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| C_FACE | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| C_LIPS | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| C_EARS | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| C_NECK | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| C_UL | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| C_LL | 1770 | 74,7% | 756 | 95,7% | 476 | 60,3% | 538 | 68,1% |
| S_FACE | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| S_LIPS | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| S_EARS | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| S_NECK | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| S_UL | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| S_LL | 1772 | 74,8% | 756 | 95,7% | 477 | 60,4% | 539 | 68,2% |
| E_FACE | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| E_LIPS | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| E_EARS | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| E_NECK | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| E_UL | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| E_LL | 1771 | 74,7% | 756 | 95,7% | 476 | 60,3% | 539 | 68,2% |
| AUTHOR | 1767 | 74,6% | 756 | 95,7% | 472 | 59,7% | 539 | 68,2% |
| KERATOSIS | 1806 | 76,2% | 790 | 100,0% | 476 | 60,3% | 540 | 68,4% |
| CRYO | 1806 | 76,2% | 790 | 100,0% | 476 | 60,3% | 540 | 68,4% |
| MALIG | 1805 | 76,2% | 790 | 100,0% | 476 | 60,3% | 539 | 68,2% |
| DESIG | 2370 | 100,0% | 790 | 100,0% | 790 | 100,0% | 790 | 100,0% |

D. MODELOS DE COX Y DIAGNÓSTICO

Modelos de Cox realizados en el análisis de supervivencia:

- Modelo de regresión de Cox univariante para el sexo (no cumple PH):

```
> modelo <- coxph(Surv(START,STOP,EVENT)~ SEX + strata(INTERVAL),data = DATOS,
  robust = TRUE)

Call:
coxph(formula = Surv(START,STOP,EVENT) ~ SEX + strata(INTERVAL),data = DATOS,
  robust = TRUE)

n= 10190, number of events= 7536
(38 observations deleted due to missingness)

              coef exp(coef) se(coef) robust se      z Pr(>|z|)
SEXM -0.03372    0.96684  0.02310   0.02304 -1.463   0.143

      exp(coef) exp(-coef) lower .95 upper .95
SEXM    0.9668      1.034    0.9241    1.012

Concordance= 0.503 (se = 0.007 )
Rsquare= 0 (max possible= 1 )
Likelihood ratio test= 2.13 on 1 df,  p=0.1
Wald test               = 2.14 on 1 df,  p=0.1
Score (logrank) test = 2.13 on 1 df,  p=0.1,  Robust = 2.14 p=0.1

(Note: the likelihood ratio and score tests assume independence of
  observations within a cluster, the Wald and robust score tests do not)
```

- Modelo de regresión de Cox univariante para la edad (*modelo final*):

```
> modelo <- coxph(Surv(START,STOP,EVENT)~ AGE + strata(INTERVAL),data = DATOS,
  robust = TRUE)

Call:
coxph(formula = Surv(START,STOP,EVENT) ~ AGE + strata(INTERVAL),data = DATOS,
  robust = TRUE)

n= 10228, number of events= 7559

              coef exp(coef) se(coef) robust se      z Pr(>|z|)
AGE>20 -0.18723    0.82926  0.02362   0.02335 -8.017 1.08e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
AGEMEAN>20    0.8293      1.206    0.7922    0.8681

Concordance= 0.524 (se = 0.007 )
Rsquare= 0.006 (max possible= 1 )
Likelihood ratio test= 63.55 on 1 df,  p=2e-15
Wald test               = 64.28 on 1 df,  p=1e-15
Score (logrank) test = 63.03 on 1 df,  p=2e-15,  Robust = 64.36 p=1e-15

(Note: the likelihood ratio and score tests assume independence of
  observations within a cluster, the Wald and robust score tests do not).
```

- Modelo de regresión de Cox multivariante para edad y sexo (comprobación de confusión)

```

> modelo <- coxph(Surv(START,STOP,EVENT)~ SEX + AGE + strata(INTERVAL),data =
  DATOS,robust = TRUE)

Call:
coxph(formula = Surv(START, STOP, EVENT) ~ SEX + AGEM + strata(INTERVAL),
  data = DATOS, robust = TRUE)

n= 10190, number of events= 7536
(38 observations deleted due to missingness)

              coef exp(coef) se(coef) robust se      z Pr(>|z|)
SEXM          -0.02839  0.97201  0.02311  0.02306 -1.231  0.218
AGEMEAN>20    -0.18773  0.82884  0.02367  0.02338 -8.031 9.68e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

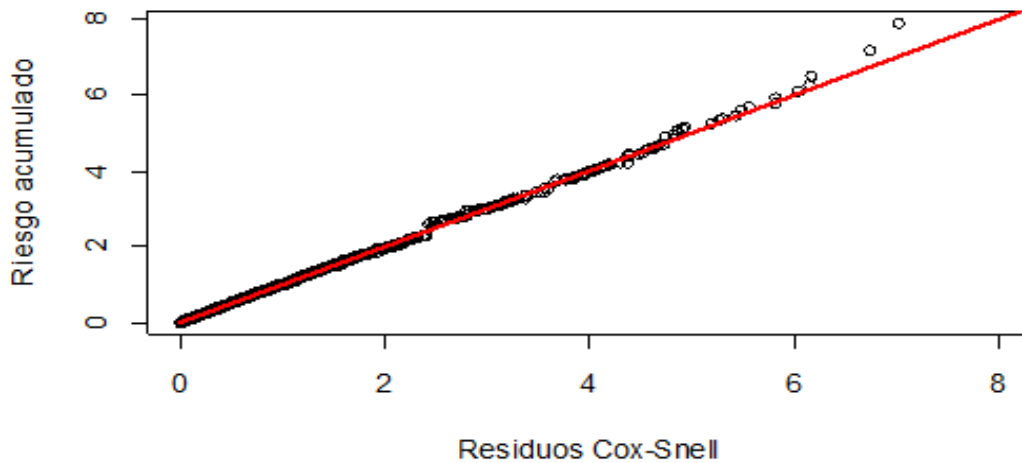
              exp(coef) exp(-coef) lower .95 upper .95
SEXM           0.9720    1.029    0.9291    1.0169
AGEMEAN>20     0.8288    1.207    0.7917    0.8677

Concordance= 0.526 (se = 0.008 )
Rsquare= 0.006 (max possible= 1 )
Likelihood ratio test= 65.73 on 2 df,  p=5e-15
Wald test               = 65.83 on 2 df,  p=5e-15
Score (logrank) test = 65.2 on 2 df,  p=7e-15,  Robust = 65.91 p=5e-15

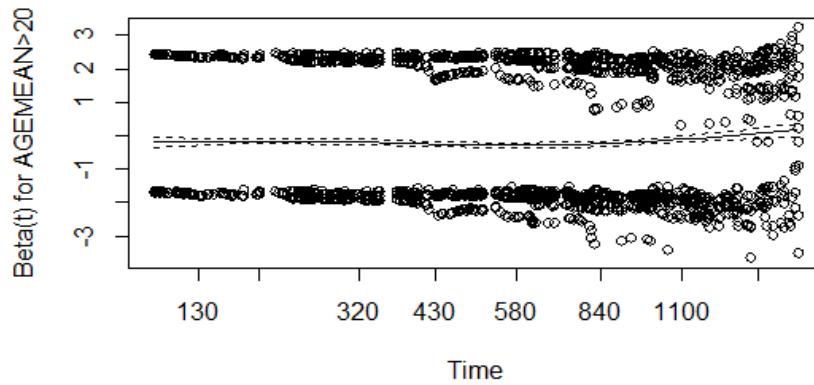
(Note: the likelihood ratio and score tests assume independence of
  observations within a cluster, the Wald and robust score tests do not).

```

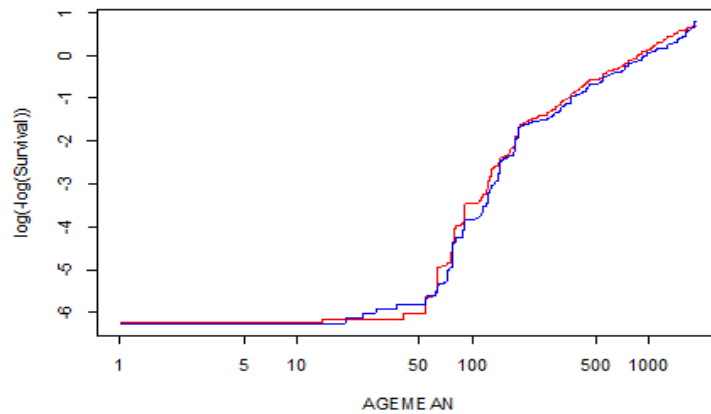
Se realizó el diagnóstico para el modelo final. El diagnóstico global del modelo se realizó a partir del gráfico de riesgo acumulado frente a los residuos de Cox-Snell. Un diagnóstico correcto presentaría los valores en torno a una pendiente de valor 1 que corta el origen de coordenadas. El gráfico resultante fue correcto.



La condición de riesgos proporcionales impuesta por el modelo de Cox se evalúa mediante el gráfico de residuos de Schoenfeld, la variable cumplirá PH si la pendiente observada es nula, en el caso de la variable edad lo es, por lo que en principio cumpliría PH.



Para no dejar opción a la duda, la condición de riesgos proporcionales también se puede evaluar mediante el gráfico $\log(-\log(S(t)))$ vs t al ser la edad una variable categórica y mediante un test de *Proportional Hazards*, que proporciona un valor p global del modelo. Las líneas del gráfico parecen cruzarse, aunque están muy próximas, en cualquier caso el valor p del test es superior a 0.05 por lo que el modelo cumple la condición de riesgos proporcionales.



```
> sch_rest <- cox.zph(modelo); sch_rest
      rho chisq      p
AGE>20 0.0102 0.766 0.381
```

E. ESTADÍSTICOS DESCRIPTIVOS DEL TIEMPO ENTRE VISITAS

Se muestran los estadísticos descriptivos del tiempo entre visitas para las regiones de Tanzania que disponían al menos dos visitas (incluyendo la visita basal):

| Región | N | Media | SD | Mediana | Min- Máx | SE | Q25 | Q75 |
|--------------------|-----|--------|--------|---------|----------|-------|-------|-------|
| Arusha | 181 | 289,01 | 158,38 | 236,3 | 72-1458 | 11,77 | 182,3 | 364,5 |
| DarEsSalaam | 125 | 197,35 | 90,26 | 173,7 | 1-656 | 8,07 | 143,0 | 203,0 |
| Geita | 123 | 220,34 | 108,77 | 198,5 | 63-717 | 9,81 | 145,8 | 271,3 |
| Iringa | 164 | 165,11 | 89,75 | 139,0 | 1-460 | 7,01 | 105,0 | 220,5 |
| Kagera | 102 | 105,19 | 23,94 | 103,5 | 88-211 | 2,37 | 94,6 | 103,5 |
| Kigoma | 120 | 163,41 | 40,03 | 159,5 | 122-325 | 3,65 | 159,4 | 160,0 |
| Kilimanjaro | 445 | 274,99 | 196,78 | 212,3 | 1-1827 | 9,33 | 182,0 | 293,7 |
| Manyara | 13 | 335,00 | 0,00 | 335,0 | 335-335 | 0,00 | 335,0 | 335,0 |
| Mara | 138 | 202,60 | 83,79 | 149,3 | 124-554 | 7,13 | 149,0 | 242,3 |
| Mbeya | 84 | 200,73 | 101,60 | 180,3 | 1-539 | 11,09 | 134,6 | 242,0 |
| Morogoro | 15 | 127,27 | 7,96 | 125,0 | 111-141 | 2,06 | 125,0 | 125,0 |
| Mwanza | 433 | 237,05 | 107,68 | 199,0 | 108-961 | 5,17 | 173,6 | 270,0 |
| Njombe | 23 | 131,48 | 43,64 | 115,0 | 77-267 | 9,10 | 107,6 | 134,5 |
| Pwani | 23 | 127,91 | 4,26 | 126,0 | 126-137 | 0,89 | 126,0 | 126,0 |
| Ruvuma | 122 | 281,66 | 122,33 | 320,7 | 115-467 | 11,08 | 149,0 | 406,0 |
| Shinyanga | 398 | 184,39 | 106,41 | 151,3 | 75-811 | 5,33 | 131,5 | 206,3 |
| Tanga | 287 | 320,90 | 167,41 | 254,7 | 75-982 | 9,88 | 226,0 | 327,0 |

F. MANUAL DE USUARIO DE EPI INFO

(Pag. Siguiete)

Epi Info™ Manual de usuario
Proyecto Kilisun Care

Miguel Ángel Calvo y Onintze Zaballa

Epi Info™ Manual de usuario

Contenido

| | |
|---|----|
| Instalación del software Epi Info™ | 3 |
| Guía para la entrada de datos..... | 4 |
| Abriendo el formulario de entrada de datos | 4 |
| Entrada de datos | 6 |
| Guardar datos y entrada de un nuevo registro..... | 8 |
| Búsqueda o edición de registros | 9 |
| Guía para el análisis descriptivo..... | 11 |
| Diccionario de datos..... | 13 |

Instalación del software Epi Info™

Epi info™ es un paquete de software libre desarrollado por el CDC (Centro para el Control y la Prevención de Enfermedades de los Estados Unidos) que permite la creación de cuestionarios, así como una recogida de datos personalizada y la posibilidad de hacer un análisis estadístico sencillo. Este software puede descargarse en la siguiente dirección: <https://www.cdc.gov/epiinfo/support/downloads.html>.

Una vez descargada la última versión de Epi Info™ 7, abrir el archivo y seguir las instrucciones hasta completar la instalación. Podemos encontrar las instrucciones detalladas de instalación y recursos adicionales en la web de CDC: <https://www.cdc.gov/epiinfo/index.html>.

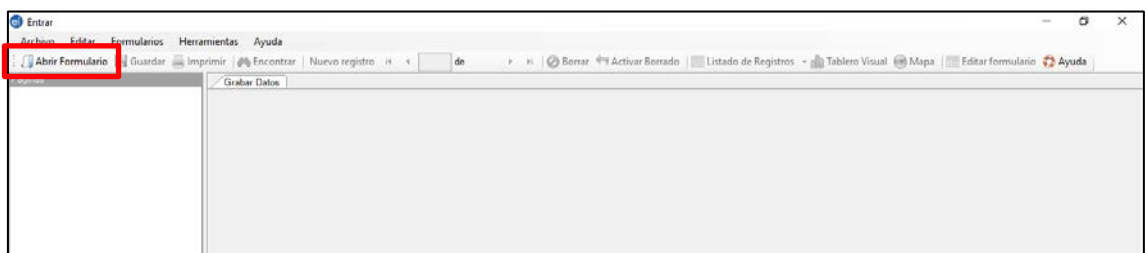
Guía para la entrada de datos

Abriendo el formulario de entrada de datos

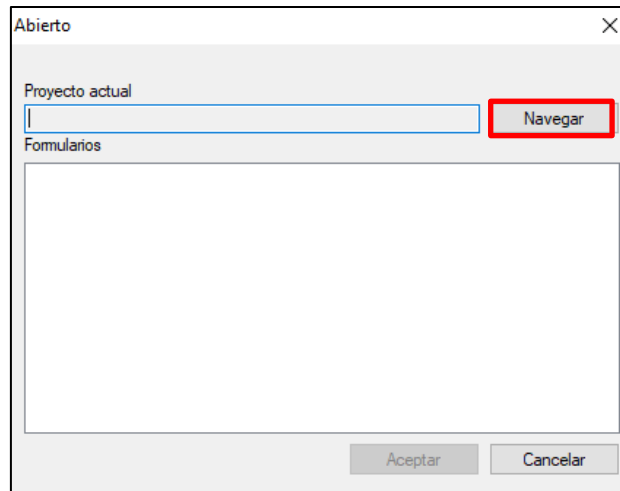
1. La primera ventana que nos encontramos al abrir el programa Epi Info™ es la que se muestra a continuación. Hacer clic sobre **ENTRAR DATOS**.



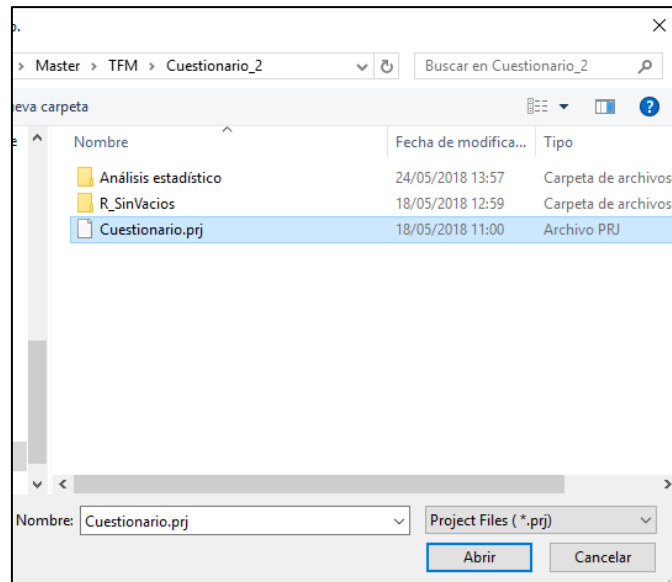
2. Abrir el cuestionario de entrada de datos haciendo clic en el botón **Abrir formulario** en la esquina superior izquierda.



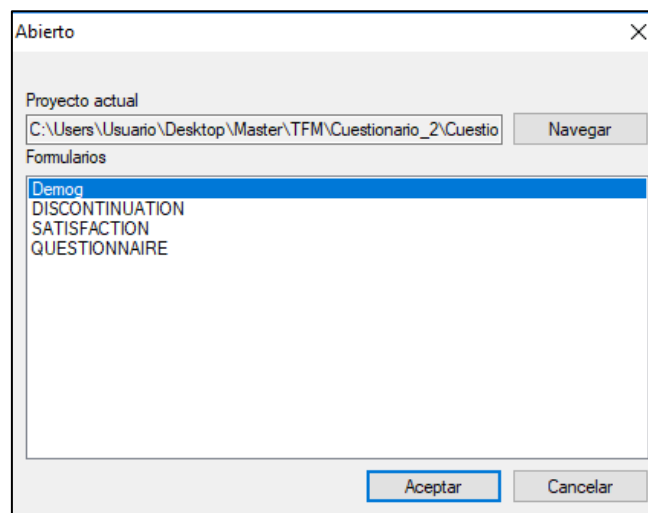
3. Hacer clic sobre el botón **Navegar** para buscar el proyecto donde se encuentra el cuestionario.



4. Seleccionar **Cuestionario.prj**.



5. Dentro del proyecto, escoger el cuestionario que se quiere rellenar.



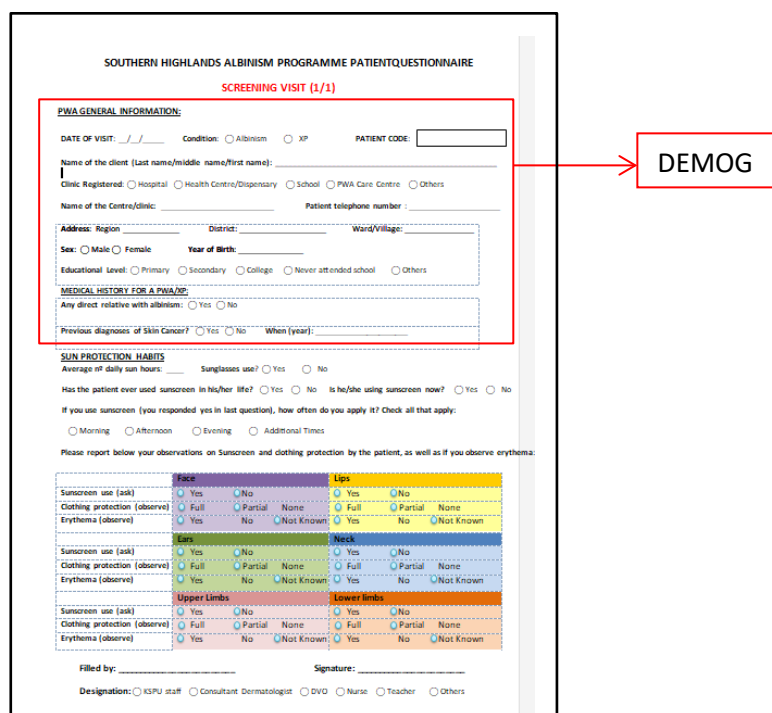
Entrada de datos

El cuestionario completo de formato papel se ha redistribuido en 4 cuestionarios dentro de Epi Info™, pero esta modificación no afecta al orden de la entrada de los datos:

1. DEMOG

Se rellena cada vez que un nuevo participante entra en el programa y es aquí donde se les asigna un código de identificación. En este módulo se recogen los datos generales de los individuos, por ejemplo, el nombre, sexo, número de teléfono, etc.

Es obligatoria la introducción del código del paciente (PATIENT CODE) para poder guardar el registro.



SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE

SCREENING VISIT (1/1)

PWA GENERAL INFORMATION:

DATE OF VISIT: ___/___/___ Condition: Albinism XP PATIENT CODE:

Name of the client (Last name/middle name/first name): _____

Clinic Registered: Hospital Health Centre/Dispensary School PWA Care Centre Others

Name of the Centre/Clinic: _____ Patient telephone number: _____

Address: Region _____ District: _____ Ward/Village: _____

Sex: Male Female Year of birth: _____

Educational Level: Primary Secondary College Never attended school Others

MEDICAL HISTORY FOR A PWA/XP:

Any direct relative with albinism: Yes No

Previous diagnoses of skin cancer? Yes No When (year): _____

SUN PROTECTION HABITS

Average # of daily sun hours: _____ Sunglasses use? Yes No

Has the patient ever used sunscreen in his/her life? Yes No Is he/she using sunscreen now? Yes No

If you use sunscreen (you responded yes in last question), how often do you apply it? Check all that apply:

Morning Afternoon Evening Additional Times

Please report below your observations on Sunscreen and clothing protection by the patient, as well as if you observe erythema:

| Face | Lips |
|---|---|
| Sunscreen use (ask) <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection (observe) <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema (observe) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |
| Ears | Neck |
| Sunscreen use (ask) <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection (observe) <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema (observe) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |
| Upper Limbs | Lower Limbs |
| Sunscreen use (ask) <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection (observe) <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema (observe) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |

Filled by: _____ Signature: _____

Designation: KSPU staff Consultant Dermatologist DVO Nurse Teacher Others

2. QUESTIONNAIRE

Se rellena en cada visita que se realiza (incluida la primera visita) y está formado por 3 páginas:

- SHABITS: datos relacionados con los hábitos de protección solar de los participantes.
- DISTRUBUTION: datos de la recogida y distribución de los botes que se les proporciona.
- EDUCATION: educación que reciben por parte del proyecto Kilisun.

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE
FOLLOW UP VISIT (1/1)

DATE OF VISIT: / / PATIENT CODE: _____

Name of the client (Last name/middle name/first name): _____

SUN PROTECTION HABITS

Average # of daily sun hours: _____ Sunglasses? Yes No

How often did you apply the sunscreen? Check all that apply:

Morning Afternoon Evening Additional Times

Please report below your observations on Sunscreen and clothing protection as well as if you observe erythema:

| | Face | Lips |
|---------------------|---|---|
| Sunscreen use | <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |
| | Ears | Neck |
| Sunscreen use | <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |
| | Upper Limbs | Lower limbs |
| Sunscreen Use | <input type="radio"/> Yes <input type="radio"/> No | <input type="radio"/> Yes <input type="radio"/> No |
| Clothing protection | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None | <input type="radio"/> Full <input type="radio"/> Partial <input type="radio"/> None |
| Erythema | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not Known |

Actinic Keratosis (evaluated only in clinic distributions) Present Absent Cryotherapy done: Yes No

Suspicious melanocytic: Yes No

SUNPROTECTIVE MEASURE RETURNED

No of Containers Returned: 0 1 2 3 4 5

If no containers return mark why: Forgot Lost not finished not aware others

Was the sunscreen provided sufficient? Yes No If not, how long where you without sunscreen (weeks)? _____

SUNPROTECTIVE MEASURE GIVEN

Name of sunscreen dispensed: Kilisun Other: _____

Sunscreen dispensed: 0 1B 2B 3B 4B 5B 15 25 35 45 55

Batch No. of Containers dispensed: 1) _____ 2) _____ 3) _____ 4) _____

Protection elements: Soap Sunscreens Other: _____

EDUCATION PROVIDED? Yes No

If education was provided, How much **helpful** was it on:

| | |
|------------------|---|
| Albinism | <input type="radio"/> helpful <input type="radio"/> indifferent <input type="radio"/> Not helpful |
| Photo protection | <input type="radio"/> helpful <input type="radio"/> indifferent <input type="radio"/> Not helpful |
| Widow Use | <input type="radio"/> helpful <input type="radio"/> indifferent <input type="radio"/> Not helpful |

Filled by: _____ Signature: _____

SHABITS

DISTRIBUTION

EDUCATION

Existen datos en QUESITONNAIRE que son obligatorios de introducir para poder completar y guardar un registro:

- **PATIENT CODE.** Código del paciente.
- **VISIT TYPE.** Tiene dos opciones: *ENROLLMENT VISIT* y *FOLLOW UP*. Cuando el participante sea recién ingresado se introducirá *ENROLLMENT VISIT*. En otro caso, *FOLLOW UP*.
- **VISIT NUMBER.** Se ha creado esta variable para guardar un orden de visitas. En el caso de ser *VISIT TYPE=ENROLLMENT VISIT*, esta casilla se rellenará con un 0 automáticamente, ya que se trata de datos basales. Sin embargo, si en *VISIT TYPE* se selecciona la opción *FOLLOW UP*, se mostrará en pantalla el siguiente recuadro donde se encuentran los registros que se tienen guardados de un paciente en particular. De esta manera, sabremos cuál es el número de visita que hemos de introducir en la casilla de *VISIT NUMBER*. En el caso que se muestra a continuación introduciríamos el 3. Hacemos clic sobre cancelar e introducimos en la casilla de *VISIT NUMBER* el correspondiente número de visita.

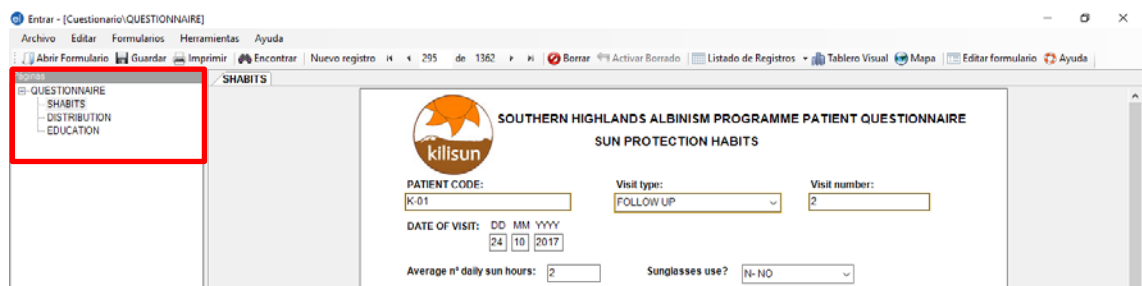
Resultados de Búsqueda

| | PATCODE | VISIT_TYPE | NVISIT |
|---|---------|------------|--------|
| ▶ | K-01 | FOLLOW UP | 2 |
| | K-01 | FOLLOW UP | 1 |

Autosearch encontró registros que coinciden.
Haga doble clic en una fila o haga clic en Aceptar para ir al registro seleccionado.
Haga clic en Cancelar para seguir introduciendo el registro actual.

Aceptar Cancelar

Recordamos que QUESTIONNAIRE está formado por tres páginas. Basta con hacer clic en la parte izquierda de la pantalla para movernos de una a otra.



Todos aquellos datos que no se rellenen serán considerados valores “missing”.

3. SATISFACTION

Cuestionario de satisfacción de los participantes. Se muestra tal y como es en el formato papel.

4. DISCONTINUATION

En este cuestionario el paciente muestra su razón de abandono dentro de una serie de opciones. Puede escoger tantas como considere.

Guardar datos y entrada de un nuevo registro

1. A medida que el usuario pasa de una página a otra, Epi info™ guarda de forma automática los datos. Además, también se puede guardar el cuestionario manualmente haciendo clic en **Guardar**.

Entrar - [Cuestionario/QUESTIONNAIRE]

Archivo Editar Formularios Herramientas Ayuda

Abrir Formulario Guardar Imprimir Encontrar Nuevo registro 295 de 1362 Borrar Activar Borrado Listado de Registros Tablero Visual Mapa Editar formulario Ayuda

QUESTIONNAIRE
SHABITS
DISTRIBUTION
EDUCATION

SHABITS

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE
SUN PROTECTION HABITS

PATIENT CODE: K-01 Visit type: FOLLOW UP Visit number: 2

DATE OF VISIT: DD MM YYYY
24 10 2017

Average n° daily sun hours: 2 Sunglasses use? N- NO

2. Para crear un nuevo registro hacer clic en **Nuevo registro** en la barra de herramientas.

Entrar - [Cuestionario/QUESTIONNAIRE]

Archivo Editar Formularios Herramientas Ayuda

Abrir Formulario Guardar Imprimir Encontrar Nuevo registro 295 de 1362 Borrar Activar Borrado Listado de Registros Tablero Visual Mapa Editar formulario Ayuda

QUESTIONNAIRE
SHABITS
DISTRIBUTION
EDUCATION

SHABITS

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE
SUN PROTECTION HABITS

PATIENT CODE: K-01 Visit type: FOLLOW UP Visit number: 2

DATE OF VISIT: DD MM YYYY
24 10 2017

Average n° daily sun hours: 2 Sunglasses use? N- NO

Búsqueda o edición de registros

1. Para eliminar un registro hacer clic sobre el botón **Borrar**.

Entrar - [Cuestionario/QUESTIONNAIRE]

Archivo Editar Formularios Herramientas Ayuda

Abrir Formulario Guardar Imprimir Encontrar Nuevo registro 295 de 1362 Borrar Activar Borrado Listado de Registros Tablero Visual Mapa Editar formulario Ayuda

QUESTIONNAIRE
SHABITS
DISTRIBUTION
EDUCATION

SHABITS

SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE
SUN PROTECTION HABITS

PATIENT CODE: K-01 Visit type: FOLLOW UP Visit number: 2

DATE OF VISIT: DD MM YYYY
24 10 2017

Average n° daily sun hours: 2 Sunglasses use? N- NO

2. Para buscar un registro, la manera más fácil es hacerlo mediante el código del paciente (*PATIENT CODE*). Para ello, comenzamos haciendo clic sobre **Encontrar**.

Entrar - [Cuestionario/QUESTIONNAIRE]

Archivo Editar Formularios Herramientas Ayuda

Abrir Formulario Guardar Imprimir Encontrar Nuevo registro 295 de 1362 Borrar Activar Borrado Listado de Registros Tablero Visual Mapa Editar formulario Ayuda

QUESTIONNAIRE
SHABITS
DISTRIBUTION
EDUCATION

SHABITS

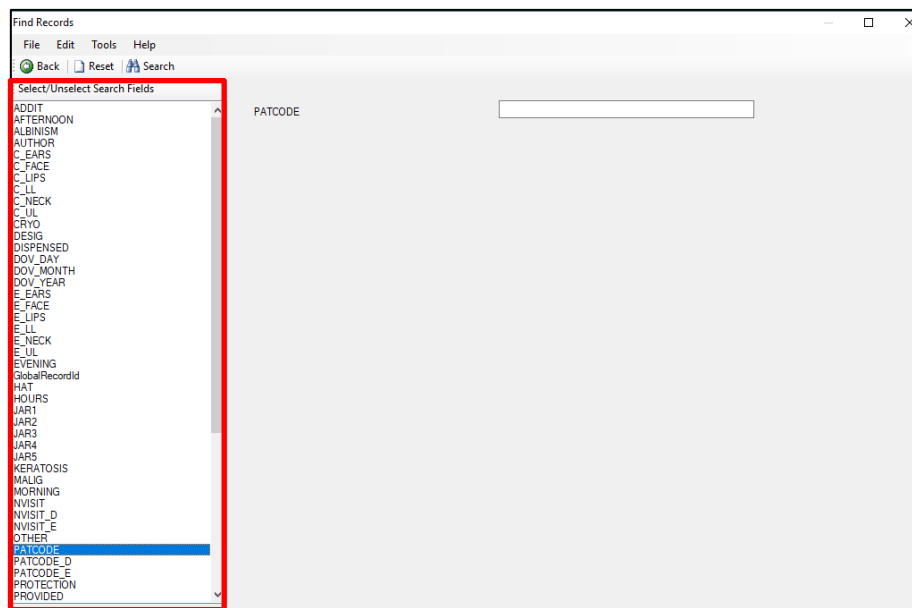
SOUTHERN HIGHLANDS ALBINISM PROGRAMME PATIENT QUESTIONNAIRE
SUN PROTECTION HABITS

PATIENT CODE: K-01 Visit type: FOLLOW UP Visit number: 2

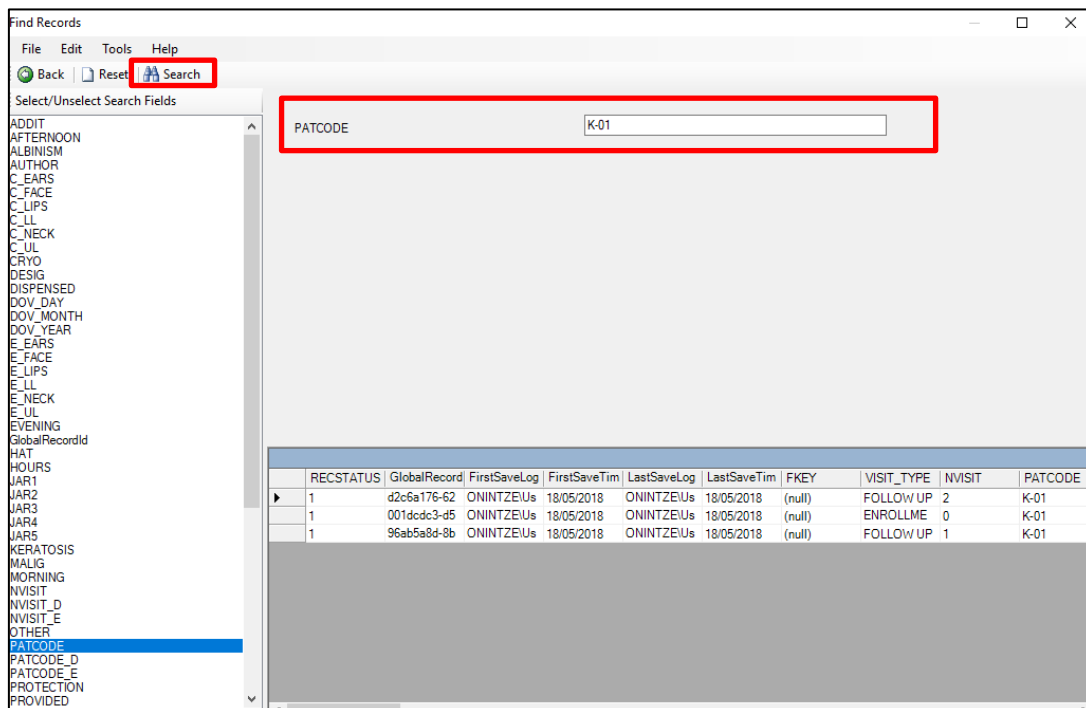
DATE OF VISIT: DD MM YYYY
24 10 2017

Average n° daily sun hours: 2 Sunglasses use? N- NO

Buscamos y hacemos clic sobre “PATCODE” en la lista de la izquierda.



Introducimos el código del participante en la casilla y hacemos clic sobre **Search**. Esta opción nos da acceso a todos los registros del participante que deseemos.



En caso de querer hacer una búsqueda más exhaustiva, se permite introducir tantos campos como se quiera. Finalmente, se hace doble clic sobre el registro que se quiere abrir, tanto para editar, consultar o borrar.

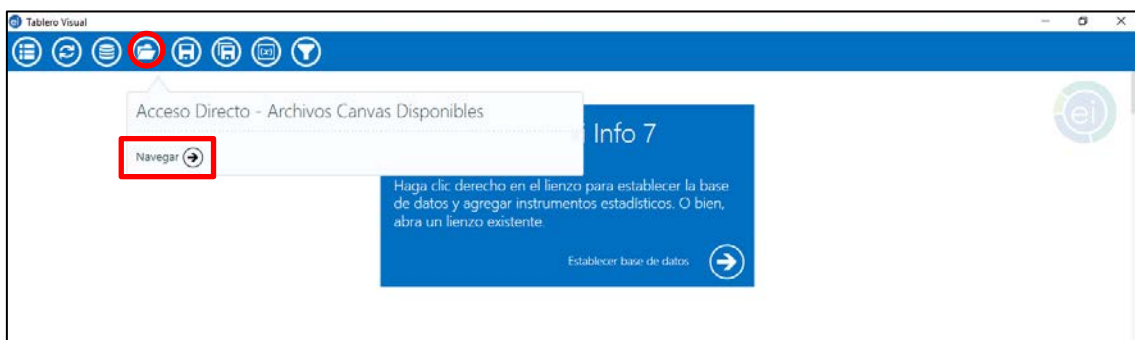
Guía para el análisis descriptivo

Epi Info da la opción de realizar un análisis estadístico básico, que se actualiza cuando se almacenan registros nuevos.

1. En la ventana principal de Epi info™ hacer clic sobre el botón **TABLERO VISUAL**.



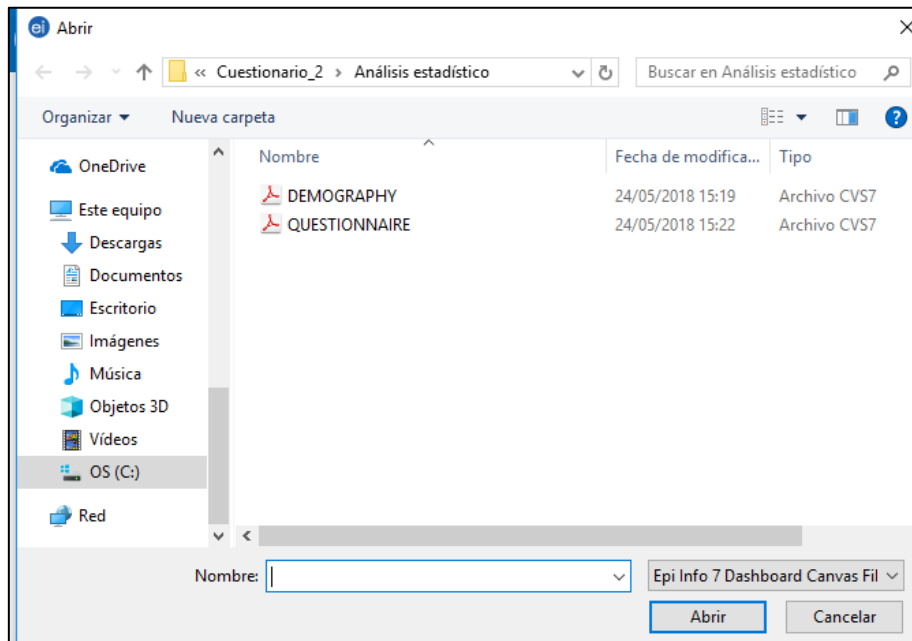
2. Seleccionar la opción de **Abrir** y después hacemos clic sobre **Navegar**



3. Escogemos el archivo que queremos abrir.

- **DEMOGRAPHY**: corresponde al cuestionario de demog, donde se encuentran los datos generales de los pacientes
- **QUESTIONNAIRE**: se muestra un análisis descriptivo básico de algunas variables de

DISTRIBUTION y EDUCATION.



Diccionario de datos

| Cuestionario | Página | Nombre de la variable | Descripción de variable | Tipo de variable |
|--------------|--------|-----------------------|------------------------------|---------------------------|
| DEMOG | — | PATCODE | ID del paciente | Texto |
| DEMOG | — | VISIT_TYPE | Tipo de visita | Elenco desplegable: texto |
| DEMOG | — | NVISIT | Número de visita | Numérico |
| DEMOG | — | DOV_DAY | Día de visita | Numérico |
| DEMOG | — | DOV_MONTH | Mes de visita | Numérico |
| DEMOG | — | DOV_YEAR | Año de visita | Numérico |
| DEMOG | — | COND | Condición | Elenco desplegable: texto |
| DEMOG | — | PATNAME | Nombre | Línea de texto |
| DEMOG | — | CLINIC_TYPE | Tipo de clínica | Elenco desplegable: texto |
| DEMOG | — | CLINIC_NAME | Nombre de la clínica | Texto |
| DEMOG | — | TLF | Teléfono del paciente | Texto |
| DEMOG | — | REG | Región | Texto |
| DEMOG | — | DIST | Distrito | Texto |
| DEMOG | — | WARD | Pueblo | Texto |
| DEMOG | — | SEX | Sexo | Elenco desplegable: texto |
| DEMOG | — | DOB | Año de nacimiento | Numérico |
| DEMOG | — | AGE | Edad al inscribirse | Numérico |
| DEMOG | — | EDU | Nivel de educación | Elenco desplegable: texto |
| DEMOG | — | OCCUP | Ocupación | Texto |
| DEMOG | — | STATUS | Estado civil | Elenco desplegable: texto |
| DEMOG | — | REL_HIST | Familiar con albinismo | Booleano (Sí/No) |
| DEMOG | — | CANCER | Diagnóstico de cáncer previo | Booleano (Sí/No) |

| DEMOG | — | YOC | Año del diagnóstico | Número |
|---------------|---------|----------------|--|---------------------------|
| QUESTIONNAIRE | SHABITS | PATCODE | ID del paciente | Texto |
| QUESTIONNAIRE | SHABITS | VISIT_TYPE | Tipo de visita | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | NVISIT | Número de visita | Numérico |
| QUESTIONNAIRE | SHABITS | DOV_DAY | Día de visita | Numérico |
| QUESTIONNAIRE | SHABITS | DOV_MONTH | Mes de visita | Numérico |
| QUESTIONNAIRE | SHABITS | DOV_YEAR | Año de visita | Numérico |
| QUESTIONNAIRE | SHABITS | HOURS | Nº medio de horas al sol diarias | Numérico |
| QUESTIONNAIRE | SHABITS | SUNGLASSES | Uso de gafas | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | SUNSCREEN_LIFE | Alguna vez ha usado crema solar en su vida | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | SUNSCREEN_NOW | Uso de crema solar en la visita | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | MORNING | Crema solar por la mañana | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | AFTERNOON | Crema solar al mediodía | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | EVENING | Crema solar por la tarde | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | ADDIT | Crema solar en otro momento | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | S_FACE | Crema solar en la cara | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | C_FACE | Protección textil en la cara | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | E_FACE | Eritema en la cara | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | S_LIPS | Crema solar en los labios | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | C_LIPS | Protección textil en los labios | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | E_LIPS | Eritema en los labios | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | S_EARS | Crema solar en las orejas | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | C_EARS | Protección textil en las orejas | Elenco desplegable: texto |

| | | | | |
|---------------|--------------|--------------|-------------------------------------|---------------------------|
| QUESTIONNAIRE | SHABITS | E_EARS | Eritema en las orejas | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | S_NECK | Crema solar en el cuello | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | C_NECK | Protección textil en el cuello | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | E_NECK | Eritema en el cuello | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | S_UL | Crema solar en los brazos | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | C_UL | Protección textil en los brazos | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | E_UL | Eritema en los brazos | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | S_LL | Crema solar en las piernas | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | C_LL | Protección textil en las piernas | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | E_LL | Eritema en las piernas | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | KERATOSIS | Queratosis actínica | Elenco desplegable: texto |
| QUESTIONNAIRE | SHABITS | CRYO | Crioterapia | Booleano (Sí/No) |
| QUESTIONNAIRE | SHABITS | MALIG | Malignidades sospechosas | Booleano (Sí/No) |
| QUESTIONNAIRE | DISTRIBUTION | PATCODE_D | ID del paciente | Texto |
| QUESTIONNAIRE | DISTRIBUTION | VISIT_TYPE_D | Tipo de visita | Elenco desplegable: texto |
| QUESTIONNAIRE | DISTRIBUTION | NVISIT_D | Número de vista | Numérico |
| QUESTIONNAIRE | DISTRIBUTION | RETURNED | Nº de botes devueltos | Numérico |
| QUESTIONNAIRE | DISTRIBUTION | REASON | Razón de los que no se han devuelto | Elenco desplegable: texto |
| QUESTIONNAIRE | DISTRIBUTION | SUFFICIENT | Suficiente crema solar | Booleano (Sí/No) |
| QUESTIONNAIRE | DISTRIBUTION | WEEKS | Nº de semanas sin crema solar | Numérico |
| QUESTIONNAIRE | DISTRIBUTION | SUNSCREEN | Nombre de la crema solar | Elenco desplegable: texto |

| | | | | |
|---------------|--------------|---------------|--------------------------------------|---------------------------|
| QUESTIONNAIRE | DISTRIBUTION | SUNSCREEN_OTH | Otra crema solar | Texto |
| QUESTIONNAIRE | DISTRIBUTION | DISPENSED | Nº de botes proporcionados | Elenco desplegable: texto |
| QUESTIONNAIRE | DISTRIBUTION | JAR1 | ID bote 1 | Texto |
| QUESTIONNAIRE | DISTRIBUTION | JAR2 | ID bote 2 | Texto |
| QUESTIONNAIRE | DISTRIBUTION | JAR3 | ID bote 3 | Texto |
| QUESTIONNAIRE | DISTRIBUTION | JAR4 | ID bote 4 | Texto |
| QUESTIONNAIRE | DISTRIBUTION | JAR5 | ID bote 5 | Texto |
| QUESTIONNAIRE | DISTRIBUTION | HAT | Protección con sombrero | Booleano (Sí/No) |
| QUESTIONNAIRE | DISTRIBUTION | SUNGLASSES1 | Protección con gafas de sol | Booleano (Sí/No) |
| QUESTIONNAIRE | DISTRIBUTION | OTH | Otro tipo de protección | Texto |
| QUESTIONNAIRE | EDUCATION | PATCODE_E | ID del paciente | Texto |
| QUESTIONNAIRE | EDUCATION | VISIT_TYPE_E | Tipo de visita | Elenco desplegable: texto |
| QUESTIONNAIRE | EDUCATION | NVISIT_E | Nº de visita | Numérico |
| QUESTIONNAIRE | EDUCATION | PROVIDED | Educación proporcionada | Booleano (Sí/No) |
| QUESTIONNAIRE | EDUCATION | ALBINISM | Educación en albinismo | Elenco desplegable: texto |
| QUESTIONNAIRE | EDUCATION | PROTECTION | Educación en fotoprotección | Elenco desplegable: texto |
| QUESTIONNAIRE | EDUCATION | USE | Educación en uso de Kilisun | Elenco desplegable: texto |
| QUESTIONNAIRE | EDUCATION | AUTHOR | Autor del cuestionario | Texto |
| QUESTIONNAIRE | EDUCATION | DESIG | Designación | Elenco desplegable: texto |
| SATISFACTION | — | VISIT_TYPE | Tipo de visita | Elenco desplegable: texto |
| SATISFACTION | — | NVISIT | Nº de visita | Numérico |
| SATISFACTION | — | PATCODE | ID del paciente | Texto |
| SATISFACTION | — | DOV_DAY | Día de visita | Numérico |
| SATISFACTION | — | DOV_MONTH | Mes de visita | Numérico |
| SATISFACTION | — | DOV_YEAR | Año de visita | Numérico |
| SATISFACTION | — | SATIS1 | El tiempo de aplicación es apropiado | Elenco desplegable: texto |

| | | | | |
|-----------------|---|------------|---|---------------------------|
| SATISFACTION | — | SATIS2 | El tratamiento no condiciona sus actividades | Elenco desplegable: texto |
| SATISFACTION | — | SATIS3 | El tratamiento no limita su trabajo/estudios | Elenco desplegable: texto |
| SATISFACTION | — | SATIS4 | Mejora de la piel con el uso de Kilisun | Elenco desplegable: texto |
| SATISFACTION | — | SATIS5 | Más seguro de si mismo | Elenco desplegable: texto |
| SATISFACTION | — | SATIS6 | Menos preocupación de que la piel empeore | Elenco desplegable: texto |
| SATISFACTION | — | SATIS7 | Considera que la mejora de su piel es buena | Elenco desplegable: texto |
| SATISFACTION | — | SATIS8 | Cambio en los hábitos de protección gracias a la educación recibida | Elenco desplegable: texto |
| SATISFACTION | — | SATIS9 | Recomienda el producto | Elenco desplegable: texto |
| SATISFACTION | — | AUTHOR | Autor del cuestionario | Texto |
| DISCONTINUATION | — | VISIT_TYPE | Tipo de visita | Elenco desplegable: texto |
| DISCONTINUATION | — | NVISIT | Nº de visita | Numérico |
| DISCONTINUATION | — | PATCODE | ID del paciente | Texto |
| DISCONTINUATION | — | DOV_DAY | Día de visita | Numérico |
| DISCONTINUATION | — | DOV_MONTH | Mes de visita | Numérico |
| DISCONTINUATION | — | DOV_YEAR | Año de visita | Numérico |
| DISCONTINUATION | — | DIS1 | Dificultad económica para llegar al lugar | Booleano (Sí/No) |
| DISCONTINUATION | — | DIS2 | Visitas con mucha frecuencia | Booleano (Sí/No) |
| DISCONTINUATION | — | DIS3 | Falta de preocupación | Booleano (Sí/No) |
| DISCONTINUATION | — | DIS4 | Falta de tiempo | Booleano (Sí/No) |
| DISCONTINUATION | — | DIS5 | No tiene eficacia | Booleano (Sí/No) |

| | | | | |
|-----------------|---|----------|---|---------------------------|
| DISCONTINUATION | — | DIS6 | Cambio de residencia donde no llega Kilisun | Booleano (Sí/No) |
| DISCONTINUATION | — | DIS7 | Fallecimiento | Booleano (Sí/No) |
| DISCONTINUATION | — | DIS8 | Otra razón | Booleano (Sí/No) |
| DISCONTINUATION | — | OTH | Razón | Texto |
| DISCONTINUATION | — | DOD | Fecha de fallecimiento | Fecha DD/MM/AAAA |
| DISCONTINUATION | — | D_REASON | Razón de fallecimiento | Elenco desplegable: texto |
| DISCONTINUATION | — | AUTHOR | Autor del cuestionario | Texto |