

UNIVERSIDAD NACIONAL DE CÓRDOBA
FACULTAD DE CIENCIAS MÉDICAS
SECRETARÍA DE GRADUADOS EN CIENCIAS DE LA SALUD
DOCTORADO EN CIENCIAS DE LA SALUD

**Sesgos en el contexto de la Epidemiología Nutricional del Cáncer: Enfoque Integral de su
cuantificación en el marco de los estudios caso-control en Córdoba**

Trabajo de Tesis para optar al
Título de Doctor en Ciencias de la Salud

Julia Becaria Coquet

Córdoba, Argentina

Año 2017

Comisión de Seguimiento de Tesis

Directora

Dra. María del Pilar Díaz

Co-Directora

Dra. Sonia Edith Muñoz

Integrantes

Dra. Silvina Berra

Dra. Sonia Edith Muñoz

Dra. María del Pilar Díaz

“LA FACULTAD DE CIENCIAS MÉDICAS NO SE HACE SOLIDARIA CON LAS OPINIONES DE LA
TESIS” (Art. 43°. Reglamento de la Carrera de Doctorado en Ciencias de la Salud)

Agradecimientos

A Pilar por acompañarme y apoyarme en estos años de formación. Por todo lo que aprendí y el tiempo compartido.

A Sonia, por la energía y la pasión que transmitís por el trabajo.

A la Dra. Silvina Berra, al Dr. Oscar Bottasso y al Dr. Arnaldo Mangeaud por los valiosos aportes que realizaron a este trabajo.

A CONICET, FONCYT y SECYT UNC por el apoyo económico a esta investigación.

A Beto y toda su familia por el cariño y la hospitalidad, los recuerdo siempre (incluyo a Tutin). A Beto, en especial, por los aportes a este trabajo, sin ellos este recorrido hubiese sido más complejo y me hubiese costado más llegar a festejar con un ¡viva el p...!

A Lau, por haberme invitado al trabajar en el equipo y a Info. Por la buena onda, predisposición y amistad.

A Soni, Dolo, Cami, Nati, Gaby, Leo por la compañía, alegría, amistad, por dejarme aprender de ustedes todos los días y por incluirme en todo aún en las rondas de mate jajaja (Cami sobretodo)...

A Anita y Ger por acompañarme en estos años de formación en la universidad y fuera de ella mucho más! Ñoñis forever jajaja los quiero!

A Ceci, Valen, Pola, Agus, Elio por la compañía, por el humor y las charlas. A Ceci especialmente por haberme aceptado como roommate por un corto período jajaja

A Marce y Yanet por estar, por la predisposición y las charlas compartidas.

Al Instituto de investigaciones en Ciencias de la Salud (INICSA) por brindarme un espacio físico para desarrollar esta tesis y a todos sus integrantes.

A mi familia, a mi mamasita, mi papa, mis hermanos, J, Paula y mi sobrino Tomás, gracias a todos por apoyarme, acompañarme y bancarme siempre, los amo.

A mis prim@s, tí@s y abuel@s..

A mis amigas de siempre Coco, Flor, Juliana, Desi, Cuchi, Meli, e incluyo acá a las primiamigas Saru, Can, Euge y Lucía, gracias por estar siempre conmigo.

A Cruz =)

A mis amigas colegas Mari, Romi Mu, Romi Ro, Andre, Nady, Made, Ana por estar siempre y acompañarme durante esta última década :O ...las quiero.

A mis colegas María Cecilia Camisasso, María Eugenia Bovero, María Florencia Gonzalez, Victor Caballero, Florencia Demarchi, Maite Figueroa, Yamila Lescano, Natalia Amse, Athina Nieto por colaborar con el trabajo de campo y permitirme aprender junto a ustedes en la última instancia de la carrera.

A todas las personas que aceptaron participar en esta investigación y brindaron su valioso tiempo.
A los profesionales que facilitaron el trabajo de campo.

Índice

Resumen.....	9
Summary.....	12
CAPÍTULO 1. INTRODUCCIÓN.....	14
1.1. Epidemiología del Cáncer.....	16
1.2. Alimentación y Cáncer.....	18
1.3. Estudios caso-control y errores sistemáticos.....	21
1.4. Hipótesis.....	29
1.5. Objetivos.....	29
CAPÍTULO 2. MÉTODOS.....	30
2.1. Metodología para la valoración de errores sistemáticos.....	30
2.1.1. Introducción.....	30
2.1.2. Confusión.....	31
2.1.3. Sesgo de Selección.....	34
2.1.4. Sesgo de Información.....	35
2.1.4.1. Sesgo de información derivado de datos faltantes.....	35
2.1.4.2. Sesgo de información por error de clasificación.....	36
2.1.4.2.1. Una propuesta para analizar el error de clasificación de las frecuencias de consumo alimentario.....	37
2.2. Materiales.....	39
2.2.1. Estudios caso control para cáncer en Córdoba.....	39
2.3. Análisis estadísticos.....	47
2.3.1. Confusión.....	47
2.3.2. Sesgo de Selección.....	48
2.3.3. Sesgo de Información.....	49
CAPÍTULO 3. RESULTADOS.....	58

3.1. Análisis de sesgos por fenómeno de confusión	58
3.2. Análisis de sesgo de selección	69
3.3. Análisis de sesgo de información	71
3.3.1. Análisis de sesgo de información por datos faltantes	71
3.3.2. Análisis de sesgo de información por error de clasificación en la exposición	74
CAPÍTULO 4. DISCUSIÓN	83
Conclusiones.....	99
CAPÍTULO 5. BIBLIOGRAFIA	102
CAPÍTULO 6. APÉNDICES Y ANEXOS	126
APÉNDICE 1. Grupos de alimentos utilizados para la construcción de los patrones alimentarios de la población de Córdoba.	127
APÉNDICE 2. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas del principal sostén del hogar, estudio caso-control de cáncer de mama, Córdoba, Argentina 2008-2015.....	128
APÉNDICE 3. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas del principal sostén del hogar, estudio caso-control de cáncer colorrectal, Córdoba, Argentina 2010-2016.....	130
APÉNDICE 4. Parámetros para el análisis del error de clasificación	132
ANEXO 1. Consentimiento Informado estudios caso-control de cáncer en Córdoba.....	134
ANEXO 2. Registro Provincial de Investigaciones en Salud (RePIS N°044/10).....	136
ANEXO 3. Instrumento Encuesta Alimentaria	140
ANEXO 4. Categorización de la variable Nivel Socioeconómico	153
ANEXO 5. Categorización de la variable Actividad Física	155
ANEXO 6. Registro Provincial de Investigaciones en Salud (RePIS N°159/13)	157
ANEXO 7. Publicaciones científicas	161

Resumen

Existe evidencia acerca del rol que tiene la alimentación, como también otros hábitos modificables, en la salud de la población. Para la identificación de factores de riesgo asociados a patologías con largos períodos de inducción, como el cáncer, los estudios tipo caso-control constituyen uno de los más utilizados a nivel mundial y en Argentina. Éstos requieren una adecuada planificación para evitar sesgos y obtener estimaciones válidas de riesgo, generalizables, sobre el efecto de la alimentación de la población en la ocurrencia de enfermedades. El objetivo de la presente tesis fue construir un marco metodológico para la valoración cuantitativa de los errores sistemáticos presentes en los estudios epidemiológicos de tipo caso-control en cáncer de mama y colorrectal, llevados a cabo en la población de adultos de la provincia de Córdoba en el período 2008-2016.

Fueron utilizados dos estudios caso-control y uno de cohorte conducidos por el Grupo de Epidemiología de Cáncer y otras enfermedades en Córdoba (GEACC, Facultad de Ciencias Médicas, Universidad Nacional de Córdoba) en la provincia de Córdoba, Argentina. Los primeros correspondieron a cáncer de mama (CM, n=844, 318/526 casos/controles, 2008-2015) y colorrectal (CCR, n=490, 161/329 casos/controles, 2010-2016). El fenómeno de confusión fue analizado aplicando el enfoque de modelos de regresión para variables observadas y no observadas, es decir, no registradas y reconocidas previamente asociadas a la exposición y *outcome* estudiados. El sesgo de selección fue indagado usando escenarios determinísticos y también probabilísticos, asignando *a priori* probabilidades de participación a los grupos en estudio. El sesgo de información, derivado del problema de datos faltantes en las covariables, se afrontó aplicando imputación múltiple de datos por ecuaciones encadenadas, considerando como mecanismo de datos faltantes la modalidad al azar. Se estimaron modelos de regresión logística múltiple para las estimaciones de riesgo usando datos completos e imputación múltiple. Por último, a través del diseño del estudio de validación, se evaluó la posible presencia del error de clasificación de la exposición mediante un estudio de cohorte anidado en CM (n=57, 20/37 casos/controles, 2015-2016). Se utilizó como estándar de referencia el recordatorio de 24 horas, y se realizaron análisis comparativos respecto a la calidad de los sistemas de clasificación.

Para CM, no existió efecto confundente de las variables edad, lactancia materna, hijos, nivel socioeconómico, edad de la menarca, IMC, actividad física y estado ginecológico, en el estudio de cáncer de mama. En cambio, en el estudio de CCR, se observaron efectos

confundentes del sexo y nivel socioeconómico (mostrando cambios en las estimaciones en 34 y 10%, respectivamente). Las modificaciones en los riesgos estimados, según el efecto de confusión de variables no medidas, no fueron importantes, existiendo apenas un posible efecto de glucemia y aumento de peso durante la vida adulta (no registradas y propuestas en esta tesis) para CM y CCR respectivamente. Los cambios en las estimaciones, entre un 40 y 70%, ocurrieron al imponer gran variabilidad y correlación con la variable medida asociada a la variable no medida. En cuanto al sesgo de selección las diferencias respecto a las estimaciones convencionales fueron pequeñas, afectando, en parte, la magnitud de las asociaciones y la precisión de la estimación por intervalo, no así la dirección de las estimaciones. Situación similar ocurrió en ambos estudios. En CM, un 32% de mujeres presentó información completa; un 83% de la base fue considerada en la aplicación de Imputación Múltiple. Ambos enfoques mostraron efecto promotor del patrón Tradicional (datos completos OR: 1,33; IC 95%:1,015-1,755; con imputación OR: 1,4; IC 95%: 1,184-1,657), no obstante los efectos de otras variables, como Índice de Masa Corporal (OR: 1,03; IC 95%: 1,004-1,067) y la práctica de lactancia materna (OR: 0,54; IC 95%: 0,365-0,813) solo se evidenció bajo esquema de Imputación Múltiple. En CCR, un 51% de los sujetos fueron incluidos en análisis de datos completos y un 71% al realizar la imputación. Un efecto promotor significativo del patrón dietario Cono Sur se observó en ambos enfoques (datos completos OR: 1,56; IC 95%:1,164-2,116; con imputación OR: 1,44; IC 95%: 1,115-1,863), además de asociaciones significativas con sexo, IMC, antecedentes primarios de CCR y nivel socioeconómico. El análisis de sensibilidad de los modelos de imputación propuestos mostró un desempeño satisfactorio en ambos estudios. Se obtuvo una clasificación de la exposición aceptable para los grupos de consumo de carnes procesadas y productos de pastelería analizados, con valores de sensibilidad, de especificidad y valores predictivos positivos y negativos entre 63-86%, 42-84%, y 63-77%, respectivamente.

El análisis de sensibilidad desarrollado para los estudios caso-control que se conducen en Córdoba del grupo GEACC, mostró la necesidad de evaluar errores sistemáticos para mejorar las estimaciones por confusión, sesgo de selección y de información. Entonces, a los fines de reducir el riesgo de ocurrencia cáncer en la región y en el mundo, la presente investigación corrobora la necesidad de dirigir esfuerzos hacia la obtención de estimaciones confiables derivadas de los estudios caso-control y observacionales en general, con el fin de obtener recomendaciones alimentarias válidas y específicas, en especial para diferentes

grupos poblacionales en riesgo. Para ello, se incita a investigadores a llevar a cabo análisis de sensibilidad cuantitativos en sus estudios.

Palabras clave: alimentación; Argentina; errores sistemáticos; imputación múltiple por ecuaciones encadenadas; neoplasia; regresión logística múltiple; validación.

Summary

Diet plays a recognized role in population's health, like other modifiable habits. To identify risk factors associated with long induction period diseases like cancer, case-control studies are one of the most used worldwide and in Argentina. These require an adequate planning to avoid bias and to obtain valid and generalizable risk estimates of the effect of the dietary consumption of the population on the occurrence of diseases. The objective of this thesis was to construct a methodological framework to quantitatively assess systematic errors in case-control epidemiologic studies of breast and colorectal cancer, carried out in adult population of Córdoba province in 2008-2016 period.

Two case-control studies and a cohort study were conducted by the Group of Environmental cancer and other chronic diseases in Córdoba (GEACC, Faculty of Medical Sciences, National University of Córdoba). The former corresponded to breast cancer (BC; n= 844, 318/526 cases/controls, 2008-2015) and colorectal (CRC, n=490, 161/329 cases/controls, 2010-2016). Confounding was analyzed applying the regression models approach for observed and unobserved variables, meaning there were not registered and previously recognized to be associated with the exposure and outcome in study. Selection bias was investigated utilizing deterministic and probabilistic scenarios, assigning *a priori* participation probabilities to study groups. Information bias, derived from missing data in covariates, was handled applying multiple imputation by chained equations, considering the data was missing at random. Multiple logistic regression models were estimated using complete case analysis and multiple imputation. A specific sensibility analysis was carried out to evaluate the quality of the final imputation models. Finally, through a validation study, misclassification of the exposure was assessed using the cohort study of BC (n=57, 20/37 cases/controls, 2015-2016). The *Gold Standard* was the 24 hour-dietary recall, and comparative analyses were performed regarding the quality of the classification systems.

For BC, no confounding effect of age, breastfeeding, having children, socioeconomic level, age of menarche, BMI, physical activity and gynecologic status was found, in the BC case-control study. On the CRC study, confounding effects of sex and socioeconomic status were found (showing estimate changes of 34 and 10% respectively). Regarding the effect of not measured confounding variables, the modifications in the risk estimates were not relevant. A possible effect of glycaemia and weight gain during adult life (not registered and proposed in this thesis) for BC and CRC respectively, of about a 40 and 70% when a great

variability and correlation with the observed variable associated with the not measure one were imposed. In respect of selection bias, differences regarding conventional estimates were small, affecting, in part, the magnitude of the associations and the precision of the interval estimation not the direction. Similar situations happened in both studies. In BC, 32% had complete information and 83% was considered when Multiple Imputation (MI) was applied. Both approaches showed a promoting effect of the Traditional pattern (complete case OR: 1.33; CI 95%:1.015-1.755; MI OR: 1.4; CI 95%: 1.184-1.657). Nevertheless, the effect of other variables was observed only with MI, like BMI (OR: 1.03; CI 95%: 1.004-1.067) and breastfeeding practice (OR: 0.54; CI 95%: 0.365-0.813). In CRC, a 51% of the subjects were included in the complete case analysis and 71% with MI. A promoting effect of the South Cone dietary pattern was observed with both approaches (complete case OR: 1.56; CI 95%:1.164-2.116; MI OR: 1.44; CI 95%: 1.115-1.863). In addition, significant associations with sex, BMI, family history of CRC and socioeconomic status were observed. An acceptable classification of the exposure was obtained for processed meat consumption and bakery products groups analyzed, with sensibility, specificity, and positive and negative predictive values between 63-86%, 42-84% and 63-77%, respectively.

Sensitivity analysis developed for case-control studies conducted in Córdoba by GEACC, showed the necessity to evaluate systematic errors associated with selection, information bias and confounding to improve risk estimates. Hence, to reduce the risk of cancer occurrence in the region and the world, the present research reaffirms the need to direct efforts towards obtaining reliable estimates derived from case-control studies, and observational studies in general. In order to obtain valid and specific dietary recommendations, especially for different group populations at risk. Researchers are encouraged to undertake quantitative sensitivity analyses in their studies.

Key words: dietary; Argentina; systematic errors; Multiple Imputation by Chained Equations; neoplasm; multiple logistic regression; validation.

CAPÍTULO 1. INTRODUCCIÓN

La salud, tanto en su concepción individual como colectiva, es producto de complejas interacciones entre procesos biológicos, ecológicos, culturales y socioeconómicos. Por lo tanto, las causas de enfermedad, dependen de factores que influyen en la salud individual, que interactuando en diferentes niveles de organización, determinan el estado de salud de la población (Organización Panamericana de la Salud [OPS], 2005). Se entiende, entonces, que existen múltiples determinantes de los procesos de salud-enfermedad de los sujetos, y que su estudio necesita considerar esta multiplicidad de factores.

En los últimos años, las enfermedades no transmisibles se han convertido en una epidemia en aumento, siendo las principales representantes de este grupo de patologías el cáncer, las enfermedades cardiovasculares, la diabetes y las enfermedades respiratorias crónicas (Organización Mundial de la Salud [OMS] 2017 c). En la Argentina, como en el mundo, el cáncer se encuentra entre las principales causas de mortalidad junto a las enfermedades cardiovasculares (ECV) (Dirección de Estadísticas e Información de Salud, 2013; Ferrante *et al*, 2011). Estas patologías y otras comorbilidades asociadas (como diabetes y obesidad), comparten una forma de reacción inmune que implica una inflamación sistémica, de bajo grado (Hotamisligil, 2006), que subyace en la mayoría de las enfermedades crónicas (Zheng *et al*, 2017; Cooke *et al*, 2016; Aballay *et al*, 2013; Libby, 2007). Este estado inflamatorio se reconoce asociado a factores comunes relacionados con estilos de vida, como la alimentación, entre otros (Nasef *et al*, 2017; Bosma-den Boer *et al*, 2012). Aunque los antecedentes familiares y la malnutrición son factores de riesgo importantes para las enfermedades crónicas, entre ellas el cáncer, los efectos adversos en la salud responden a complejas interacciones genéticas y epigenéticas. Por lo tanto, una prioridad principal para la investigación en salud es identificar y esclarecer, de manera confiable, las vías y mecanismos biológicos de acción de factores culturales y ambientales que puedan contribuir a la prevención de esta enfermedad.

Para tratar de comprender cómo ciertos aspectos de la alimentación de las poblaciones intervienen como determinantes en estos procesos, se ha ido desarrollando a lo largo del tiempo, un campo de estudio específico, denominado Epidemiología Nutricional, derivado de la clásica disciplina Epidemiología, cuyos primeros antecedentes datan de Hipócrates (siglo IV a.C.) (Royo Bordonada y Moreno, 2009). La Epidemiología Nutricional es

considerada una disciplina más contemporánea, desarrollada en los últimos dos o tres siglos aproximadamente, con un importante crecimiento en las décadas más recientes. El estudio de Lind realizado en 1747, sobre escorbuto en tripulantes de un barco inglés, puede considerarse como el primer estudio de epidemiología nutricional (Kac *et al*, 2007). Lind, junto con Takaki y Goldberger (estudiosos del beriberi y la pelagra, respectivamente) son considerados los pioneros en el estudio de la relación entre dieta y enfermedades. Si bien, los orígenes de esta disciplina estuvieron orientados a las enfermedades relacionadas con las deficiencias nutricionales (Kac *et al*, 2007), desde hace más de 40 años el interés ha girado en torno a las enfermedades crónicas más prevalentes en las civilizaciones occidentales, particularmente enfermedades cardiovasculares y cáncer, proceso similar al que atravesó la epidemiología clásica (Willet, 1998).

La evolución de la metodología epidemiológica, junto con las transiciones epidemiológicas y nutricionales experimentadas por las poblaciones de diferentes regiones del mundo, han contribuido a que se amplíe el interés de la epidemiología, incorporando el estudio del efecto de la dieta sobre la ocurrencia de enfermedades no transmisibles. Así, se reconoce esta nueva disciplina llamada Epidemiología Nutricional, como la ciencia que investiga el efecto de la dieta sobre la ocurrencia de enfermedades específicas (Kac *et al*, 2007).

En los últimos años, la Epidemiología Nutricional ha tenido un importante crecimiento. Sin embargo, existe aún bastante escepticismo, que deriva en cuestionamientos sobre las prácticas de investigación en el área, generando muchos debates y poniendo en tela de juicio la credibilidad de los resultados derivados de estos estudios (Schoenfeld y Ioannidis, 2013; Archer *et al*, 2013; Potter, 2015; Satija *et al*, 2015). Esto se debe, principalmente, a que la evaluación de la exposición, en este caso la alimentación, es compleja de llevar a cabo y puede existir error de medición (Lachat *et al*, 2016). A pesar del desafío de enfrentar dicha tarea, los conocimientos derivados de estos estudios epidemiológicos no deben desestimarse. En los últimos tiempos, muchos esfuerzos han generado importantes iniciativas para reportar adecuadamente información derivada de estudios observacionales (von Elm *et al*, 2007) y estudios epidemiológicos nutricionales específicamente (Lachat *et al*, 2016), derivando en mejoras en la calidad de las publicaciones (Fung *et al*, 2009). En la actualidad, muchos de los conocimientos sobre las relaciones entre alimentación y salud se obtienen a través de estudios observacionales, por lo que los esfuerzos deben dirigirse a perfeccionar los métodos

de recolección de datos, los procedimientos analíticos y las interpretaciones de los resultados que deriven de los mismos (Willet, 1998; Lachat *et al*, 2016).

1.1. Epidemiología del Cáncer

El cáncer abarca un grupo de más de cien enfermedades caracterizadas por un crecimiento celular descontrolado como resultado de cambios en la información genética de las células. Es una enfermedad de origen monoclonal, es decir, que todos los cánceres comienzan en una sola célula que pierde el control de su crecimiento normal y de los procesos de replicación (World Cancer Research Foundation/American Institute for Cancer Research [WCRF/AICR], 2007). La mayoría de los cánceres se vuelven clínicamente detectables luego de años o décadas de iniciado el daño genético del ADN. El desarrollo tumoral requiere de un conjunto de modificaciones celulares; es un proceso multietápico causado por acumulación de errores en los genes que controlan los procesos celulares. La carcinogénesis consta de tres etapas: iniciación, promoción y progresión. La última de estas etapas, es exclusiva de la transformación maligna e implica la capacidad de invadir tejidos vecinos o a distancia (Civetta 2011). La capacidad de una célula de lograr una prevención efectiva frente al cáncer o reparación depende del microambiente extracelular, que incluye la disponibilidad energética y la presencia adecuada de macro y micronutrientes (WCRF/AICR 2007).

El cáncer de mama ocurre principalmente en las mujeres y se origina en las células mamarias. La mayoría de los cánceres de seno son adenocarcinomas, originándose en células epiteliales glandulares y suelen ocurrir en los conductos galactóforos (carcinomas ductales) o en los lobulillos (carcinomas lobulillares) (American Cancer Society, 2016 *b*).

El cáncer colorrectal se origina en la última porción del tracto digestivo, en colon o recto. La mayoría de los cánceres colorrectales comienza como un crecimiento en el revestimiento interno de esta zona denominado pólipo. Existen dos tipos principales de pólipos, los adenomatosos, que pueden transformarse en tumores malignos o los pólipos inflamatorios o hiperplásicos, que son más frecuentes y en general no suelen ser precancerosos (American Cancer Society, 2016 *a*).

El cáncer es la segunda causa de mortalidad en el mundo, ocasionando 8,8 millones de defunciones en el año 2015, según la Organización Mundial de la Salud (OMS, 2017a). De acuerdo al último reporte de la International Agency for Research on Cancer (IARC), en el año 2012, se registraron alrededor de 14 millones de nuevos casos (Ferlay *et al*, 2012) y se prevé

que este número aumente en un 70% en los próximos 20 años. Además, se ha observado que el número de casos incidentes y de muertes por cáncer varían según sexo (siendo ligeramente superior en hombres que en mujeres) (Ferlay *et al*, 2012; Stewart y Wild, 2014).

En la Argentina, el cáncer es la segunda causa de muerte por enfermedad, sólo precedida por las enfermedades cardiovasculares (Dirección de Estadísticas e Información de Salud, 2013). En el año 2012, se registraron 115162 casos nuevos y 66433 muertes por la patología (Ferlay *et al*, 2012). En el país los cinco sitios tumorales más frecuentes fueron mama, colorrectal, pulmón, próstata y cuello de útero. En la provincia de Córdoba, aproximadamente 9000 personas por año desarrollan cáncer y cerca de 4500 mueren por esta causa. El Registro Provincial de Tumores reportó que se registraron 54296 nuevos casos y 29513 muertes por la patología en el período 2004-2009, de las cuales 61,4% correspondieron a hombres (Registro Provincial de Tumores 2013). Díaz y cols. (2009, 2010), describieron la distribución espacial de la incidencia y estudiaron la incidencia de cáncer en la provincia, reportando tasas de incidencia totales estandarizadas y ajustadas por edad, iguales a 263,53 DE=138,34 y 200,45 DE=98,30 (x100000 hab/año).

En la actualidad, el cáncer de mama en la población femenina a nivel mundial es el más incidente (25,2% del total), seguido por el colorrectal (9,2%), pulmón (8,7%), cuello de útero (7,9%) y estómago (4,8%). Estos sitios tumorales también representan las causas de muerte por cáncer más frecuentes, entre las cuales, la importancia relativa del cáncer de mama desciende a un 14,7%, seguido por el cáncer de pulmón con 13,8%. El cáncer de mama es la quinta causa de muerte entre todos los cánceres y, aunque es la causa de muerte de cáncer más frecuente en mujeres en regiones menos desarrolladas, pasó a ser la segunda causa de muerte en regiones más desarrolladas luego del cáncer de pulmón (Ferlay *et al*, 2012). Argentina, junto a Brasil y Uruguay, fueron los países de América Central y del Sur con mayores tasas de incidencia de cáncer de mama en la población femenina en los últimos 5 años (67,7-71,9). También presentó altas tasas de mortalidad por la enfermedad (14,9-20,5) junto con Uruguay y Cuba (x100000 hab/año) (Di Sibio *et al*, 2016). Niclis *et al* (2010) estudiaron las tendencias de mortalidad por cáncer de mama en Córdoba y sus patrones espaciales, reportando un aumento en las tasas de mortalidad estandarizadas por edad desde 1986 a 1997, que luego declinaron, principalmente en los departamentos más urbanizados.

En Argentina, el cáncer de mama es el más incidente representando un 16,8% del total, y a nivel provincial, un estudio realizado en el año 2010 revela los cocientes sitio-

específico (*specific-site ratio*), respecto de la incidencia total por sexo, para los cánceres más incidentes reportando un 0.25 ± 0.19 para cáncer de mama en mujeres (Diaz *et al*, 2010).

A nivel mundial, los cánceres más incidentes en la población masculina son pulmón (16,7% del total), próstata (15%), colorrectal (10%), estómago (8,5%) e hígado (7,5%). Estos también representan las causas más frecuentes de muerte por la patología, siendo el cáncer de pulmón la primera con un 23,6% del total de muertes, seguido por el cáncer de hígado (11,2%) y de estómago (10,1%). En la población total, los cinco cánceres más frecuentes son pulmón (13%), mama (11,9%), colorrectal (9,7%), próstata (7,9%) y estómago (6,8%), constituyendo éstos la mitad de la carga total mundial de muerte por cáncer (Ferlay *et al*, 2012, Stewart y Wild, 2014).

Por su parte, el cáncer colorectal, a nivel mundial, se reporta como el tercer cáncer más frecuente en hombres (representando un 10% del total) y el segundo en mujeres (9,2% del total). La mortalidad respecto a este cáncer es baja (un 8,5% del total) y la mayoría de las muertes (52%) ocurren en países menos desarrollados, lo que refleja una peor supervivencia en estas regiones (Ferlay *et al*, 2012). En Argentina, es el tercer cáncer de mayor incidencia con un 13,2% (Sierra y Forman, 2016; Ferlay *et al*, 2012). Respecto a las tasas de incidencias totales estandarizadas por edad según sexo en la provincia de Córdoba, se reportaron tasas para cáncer colorrectal en la población femenina de 12,45 (x100.000 hab/año) y de 19,43 (x100.000 hab/año) para la población masculina. Ambos reportados en el tercer lugar por tasas de incidencia según sexo (Diaz *et al*, 2010).

1.2. Alimentación y Cáncer

Las primeras publicaciones sobre los aportes ambientales al desarrollo del cáncer datan de los '60 (Doll, 1967), siendo ampliamente reconocidos e involucrando factores tan diversos como la dieta, el consumo de tabaco y alcohol, comportamientos sexuales y reproductivos, exposiciones ocupacionales, contaminantes ambientales, terapias médicas, factores geofísicos y agentes infecciosos (Colditz *et al*, 2006; Boffeta, 2004; Doll y Peto, 1981).

Respecto a la alimentación, muchos estudios observacionales y experimentales se han realizado sobre las asociaciones entre este hábito modificable y el proceso de carcinogénesis, y se estima que la alimentación estaría asociada a aproximadamente más de un tercio de los cánceres en adultos (WCRF/IACR, 2007). El papel patogénico de la dieta tendría especial importancia en los tumores del tracto gastrointestinal, mama, próstata y endometrio. Algunos componentes dietarios han sido asociados de manera convincente a algunos

cánceres (WCRF/IARC, 2007). El alcohol es uno de ellos, encontrándose un aumento del 10% del riesgo de ocurrencia de cáncer de mama por cada incremento en el consumo de 10 gramos de etanol al día (Hammerling *et al*, 2016; Liu *et al* 2015; Scocciati *et al*, 2014). A pesar de la plausibilidad biológica entre el consumo de grasas y el riesgo de cáncer de mama, todavía existe controversia sobre esta relación. Sin embargo, la evidencia científica sugiere una modesta aunque directa asociación entre el consumo total de lípidos y el riesgo de presentar la patología (Mourouti *et al*, 2015; Chajès y Romieu, 2014). En cuanto al consumo de fibra dietética, algunas revisiones sistemáticas han asociado un mayor consumo de esta con un menor riesgo de presentar la patología, un 5% menos de riesgo por cada incremento adicional de 10 gramos de fibra (Harvie *et al*, 2015; Aune *et al*, 2012 *a*). Otros nutrientes, entre ellos algunas vitaminas (ácido fólico, carotenoides), minerales (selenio), hidratos de carbono y lípidos (ácidos grasos omega 3, EPA y DHA) han demostrado un efecto en el riesgo de esta patología, pero la evidencia es aún controversial (Harvie *et al*, 2015; Mourouti *et al*, 2015; Baena Ruiz y Salinas, 2015; Chajès y Romieu, 2014; Babaknejad *et al*, 2014). Además, estudios sobre el efecto de alimentos y grupos de alimentos ponen en evidencia similares resultados a lo observado derivado de los estudios sobre nutrientes, sugiriendo un efecto protector de las frutas, los vegetales (fuente de algunos antioxidantes, como minerales y vitaminas), legumbres (especialmente la soja) y pescados (fuente de ácidos grasos omega 3 y ciertas vitaminas). Por otro lado, un mayor consumo de carnes rojas y procesadas (alimentos fuente de grasas saturadas, hierro, aminos heterocíclicos, hidrocarburos aromáticos o compuestos N-nitrosos) se asocia a un mayor riesgo de ocurrencia de la enfermedad. Otros grupos de alimentos con alto contenido de grasas animales y azúcares también se encontraron asociados positivamente al riesgo de este cáncer (Harvie *et al*, 2015; Baena Ruiz y Salinas, 2015; Chajès y Romieu, 2014; Rossi *et al*, 2014; Vera-Ramirez *et al*, 2013).

En cuanto al cáncer colorrectal, la evidencia publicada en relación a factores alimentarios es similar a lo mencionado en párrafos anteriores. El consumo moderado de etanol (25-30 gramos/día) aumenta el riesgo en un 16%, siendo la evidencia en hombres más fuerte que en mujeres (Baena Ruiz y Salinas, 2015). El consumo habitual de fibra disminuye el riesgo en un 10% por cada incremento adicional de 10 gramos de consumo (WCRF/AICR, 2011). Otros nutrientes han sido relacionados con esta patología, entre ellos el calcio, probablemente asociado de manera inversa al riesgo de presentar cáncer colorrectal (Song *et al*, 2015 *a*). Respecto al consumo de lácteos, (fuente de calcio, otros nutrientes y elementos bioactivos), la evidencia derivada de revisiones y meta-análisis sugiere una disminución del

riesgo de ocurrencia de este cáncer (Aune *et al*, 2012 *b*). El consumo de carnes rojas y procesadas ha sido asociado de manera convincente a un incremento en el riesgo de la patología, especialmente las procesadas (Aykan, 2015; Abid *et al*, 2014; Navarro *et al*, 2003). Otros nutrientes y alimentos (selenio, vitaminas B, ajo) también se han asociado al cáncer colorrectal, pero los resultados siguen siendo controvertidos (Chiavarini *et al*, 2015; Crosara Texeira *et al*, 2014; Aune *et al*, 2012 *b*).

Desde un punto de vista biológico, los nutrientes contribuyen a la modulación epigenética y podrían tener efectos anticancerígenos y cancerígenos basados en la posible influencia sobre la oxidación, la metilación, la reparación de ADN y otras capacidades, posiblemente afectando la iniciación, la promoción o la progresión del cáncer (WCRF/IARC, 2007; Michels, 2003). La mayoría de los estudios epidemiológicos nutricionales que estudian el riesgo de presentar cáncer han aplicado metodologías con enfoque en un nutriente o alimento, aunque en los últimos años, el enfoque fue girando en torno a la utilización de nuevas metodologías, como es el caso del Análisis Factorial de Componentes Principales (Kac *et al*, 2007), se trata de identificar patrones de consumo alimentarios y su asociación con patologías como el cáncer. Este análisis de patrones alimentarios es una alternativa muy usada, la cual permite la integración de varias exposiciones alimentarias al mismo tiempo y superar algunas limitaciones de los enfoques anteriores (Pou *et al*, 2014, Baglietto *et al*, 2011, Edefonti *et al*, 2009).

En Argentina, existe extensa bibliografía que ha identificado varios factores de riesgo socio-culturales y biológicos asociados con los cánceres más incidentes del país (Nclis *et al*, 2015; Pou *et al*, 2014; Román *et al*, 2014; Tumas *et al*, 2014). La mayoría de estos trabajos han sido llevados a cabo por el Grupo de Epidemiología Ambiental del cáncer y enfermedades crónicas (GEACC). Entre estos factores de riesgo estudiados, la alimentación es un reconocido factor modificable asociado con los cánceres más incidentes en el país. La dieta tradicional argentina está caracterizada por un elevado consumo de proteína y lípidos animales (obtenidos principalmente de carne de vaca), y bajos consumos de pescado, frutas y vegetales (Pou *et al*, 2014; Navarro *et al*, 2003). Además, en el país ya se han publicado estudios sobre las relaciones entre la dieta, el cáncer de mama y cáncer colorrectal (Becaria Coquet *et al*, 2014; De La Quintana *et al*, 2014; Tumas *et al*, 2014; Pou *et al*, 2014, 2012).

1.3. Estudios caso-control y errores sistemáticos

Para abordar la temática sobre alimentación y cáncer varios diseños de estudios epidemiológicos pueden adoptarse para la recolección de la información, siendo el tipo caso-control uno de los más utilizados. Éstos son estudios epidemiológicos analíticos observacionales, requieren bastante planificación para evitar sesgos y la información derivada de estos estudios es importante para identificar factores de riesgo asociados a patologías raras con largos períodos de inducción como el cáncer (Breslow y Day, 1980). Además, estos estudios son más económicos que los estudios de cohorte, por lo que son frecuentemente utilizados en Argentina y la región de América Latina en general. Una de las ventajas de los estudios caso-control es que la medida de asociación derivada de estos estudios, el Odds Ratio (OR), se aproxima al riesgo relativo (RR) cuando se estudia un evento de salud muy raro o poco frecuente a nivel poblacional (como lo es el cáncer) y cuando los valores de las estimaciones son bajos (Cerdeira *et al*, 2013), y por ende, pueden ser intercambiables. En la presente tesis, y a partir de ahora, esta medida de asociación estimada (OR) se la mencionará directamente como riesgo.

Globalmente, el diseño de los estudios caso-control implica detectar un número de personas con la enfermedad bajo estudio, los casos, y seleccionar un grupo de personas “comparables” pero que no presenten la enfermedad, los controles (Breslow y Day, 1980). Sujetos denominados casos y controles son indagados retrospectivamente para observar si han estado expuestos o no a potenciales factores de riesgo de la enfermedad en el pasado, y determinar por ende las diferencias o similitudes (Rothman y Greenland, 1998; Breslow & Day, 1980). Para que las comparaciones entre los grupos sean válidas, los sujetos no expuestos debieran presentar la proporción de incidencia del evento de salud semejante a aquella de los individuos expuestos si éstos no hubiesen presentado la enfermedad. Entonces, la validez de las mediciones dependerá principalmente, de la calidad del grupo de comparación (controles) (Greenland y Robins, 1986).

Distintos estudios comenzaron a indagar principalmente el enfoque de la multicausalidad en enfermedades crónicas y, el campo de la epidemiología fue transformándose en una disciplina más compleja. Por lo general, las estimaciones de efecto obtenidas de estudios epidemiológicos, pueden ser susceptibles a dos tipos de error, el aleatorio y el sistemático. El primero puede afectar la precisión de las estimaciones, en cambio, el error sistemático influye sobre la validez de los mismos. Cuando se realiza una evaluación cuantitativa sobre el error sistemático de una estimación de efecto, se lleva a cabo

un análisis de sensibilidad (Lash *et al*, 2009). Este análisis se puede definir como el método utilizado para determinar la robustez de una evaluación examinando la medida en que los resultados se ven afectados por cambios en los métodos, modelos, valores de variables no medidas o algunos supuestos (Schneeweiss, 2006). El primer análisis de sensibilidad en el área de la salud se realizó en un estudio observacional, que indagaba el hábito de fumar como posible causa de cáncer de pulmón en el año 1959 (Cornfield *et al*, 1959). En términos generales, un análisis de sensibilidad cuantitativo puede proveer información muy valiosa sobre la importancia de algunas fuentes de error y puede ayudar a una mejor evaluación de la incertidumbre de los resultados del estudio (Rothman y Greenland, 1998).

A pesar de la importancia de abordar los errores sistemáticos o sesgos, los métodos cuantitativos que consideran estos sesgos han experimentado menos desarrollo que los métodos que afrontan el error aleatorio. Los países desarrollados principalmente, han logrado metodologías que evalúan la validez de las estimaciones obtenidas (Orsini *et al*, 2008; Fox *et al*, 2005; Lash y Fink, 2003; Little y Rubin 2002; Lash y Silliman, 2000; Greenland, 1996) y su aplicación en estudios epidemiológicos sobre nutrición y cáncer (Agogo *et al*, 2016; Boyle *et al*, 2014; Sachin *et al*, 2008; Mezei y Kheifets, 2006; Vrijheid *et al*, 2006; Rosenbaum, 2005; Maclure y Hankinson, 1990). Sin embargo, en América del Sur los análisis de sensibilidad en estudios epidemiológicos nutricionales son aún limitados y poco frecuentes (Niclís *et al*, 2015; Pou *et al*, 2014).

Dependiendo del momento o la etapa del estudio en el que se originan, los errores sistemáticos que pueden interferir con la validez interna de un estudio pueden clasificarse en tres grandes grupos (Rothman y Greenland, 1998). Los sesgos en las estimaciones derivados del fenómeno de confusión, originados por la imposibilidad de asignar la exposición de manera aleatoria en los estudios observacionales, son uno de éstos. Otro grupo abarca a los sesgos de selección, que son los errores que se introducen durante la selección de los sujetos para participar en los grupos casos y controles en el caso de los estudios caso-control o en el seguimiento de la población en estudio de cohorte. El tercer grupo se refiere a los sesgos de información, que surgen al momento de medir a la población en estudio. Todos los estudios epidemiológicos, en mayor o menor medida, son susceptibles de presentar estos errores sistemáticos. Por esto, los investigadores deben esforzarse para evitar o disminuir al máximo la presencia de estos sesgos. A continuación se describe cada uno de ellos.

El fenómeno de confusión se puede definir como una asociación no causal entre una exposición determinada y un *outcome*, como resultado de la influencia de una tercer variable

(o conjunto de variables) (Hernán y Robins, 2018; Szklo y Nieto, 2007). Esta variable, denominada confundente, debe estar asociada a la exposición, asociada de manera “causal” al evento de salud bajo estudio y no debe ser una variable intermedia entre la exposición y el resultado (Szklo y Nieto, 2007; Weisberg, 2010). Este fenómeno puede generar estimaciones sesgadas, presentándose con más frecuencia en estudios observacionales que en experimentales, debido a que la asignación aleatoria disminuye la probabilidad de que los grupos a comparar difieran en relación a las variables confundentes (Szklo y Nieto, 2007). Este fenómeno es considerado una posible fuente de sesgos en la planificación y diseño de un estudio; ya que, sin su control o ajuste analítico, la interpretación de los resultados obtenidos es errónea (Lash *et al.*, 2009, Rothman y Greenland, 1998).

Existen dos grandes situaciones en las que el investigador debe enfrentar el problema de confusión. Una acontece cuando la variable que “confunde” la asociación entre la exposición y el resultado se conoce y ha sido registrada (es decir, fue observada). La otra, ocurre cuando la variable confundente no ha sido medida (Lash *et al.*, 2009).

El sesgo de selección sucede cuando dos variables entre las cuales su asociación se encuentra bajo estudio, generalmente una exposición y un resultado, afectan la participación de los sujetos en el estudio (Lash *et al.*, 2009). Según Last (2001), el sesgo de selección se define como el error debido a diferencias sistemáticas en las características de los sujetos que participan en el estudio y los que no lo hacen. En estos casos, la asociación entre la exposición y la respuesta (*outcome*) entre los sujetos seleccionados para el análisis difiere de la “real” asociación entre los elegibles. En otras palabras, el sesgo de selección ocurre cuando se comete un error sistemático al seleccionar uno o varios de los grupos que se van a comparar en el estudio. Esto provoca estimaciones de riesgo sesgadas y conduce a inferencias inválidas entre la exposición y la enfermedad, por ejemplo (Last, 2001). Los estudios de tipo caso-control son particularmente vulnerables a este sesgo, ya que los casos y los controles se seleccionan sobre la base de la ocurrencia del evento. Por esto, se recomienda trabajar con casos incidentes y evitar que la exposición o variables asociadas a ésta determinen o condicionen la participación en el estudio, como caso o control (Lazcano-Ponce *et al.*, 2001).

En todos los estudios comparativos existen dos variables críticas que deben ser recolectadas, como se ha mencionado anteriormente, éstas son la exposición y el *outcome* o resultado. Cualquier error que surja de la medición de éstas puede resultar en estimaciones de riesgo sesgadas (Weisberg, 2010).

El término sesgo de información también suele utilizarse para describir errores de medición en las covariables. Este sesgo se define como la diferencia que existe entre el efecto observado (distorsionado) y el verdadero efecto medido sin ninguna distorsión. El “efecto verdadero” es la asociación estadística que se hubiese medido sin error (Weisberg, 2010).

En los estudios epidemiológicos existe la posibilidad de que aparezcan errores de medición. Cuando las variables son categóricas, estos errores se denominan errores de clasificación. Existen dos formas en que estos errores pueden surgir: cuando la información no se registra correctamente en la base de datos (por instrumentos defectuosos, cuando los participantes no responden con la verdad a temas sensibles o cuando hay errores en las fuentes de información como las historias clínicas), o cuando el origen del error de clasificación es conceptual. Este último ocurre cuando hay una discordancia entre la definición de una variable en el estudio y la real definición de la variable (por ejemplo cuando una variable de exposición se define incorrectamente en la operacionalización por su duración, dosis, periodo de inducción, etc.) (Lash *et al*, 2009).

Los estudios caso-control pueden presentar este sesgo que se introduce en el proceso de medición de la exposición, ya que esta se mide de manera retrospectiva, luego del inicio de la enfermedad o del evento en estudio. En estos estudios, la información se obtiene frecuentemente por medio de cuestionarios aplicados por entrevistadores y sucede que las personas puedan presentar problemas para recordar la información exacta sobre alguna exposición pasada. Lo importante en estas situaciones es tratar de minimizar estas diferencias entre casos y controles, ya que los primeros pueden llegar a tener más motivación para recordar la exposición pasada comparada con los controles, en los cuales el recuerdo puede ser más inexacto. Este sesgo se reconoce como sesgo de memoria o *recall bias* (Lazcano-Ponce *et al*, 2001; Gordis, 2009).

El sesgo de información puede surgir en ciertas situaciones donde la información no se consigue recolectar de manera adecuada en todos los sujetos participantes del estudio, por lo que se debe trabajar con bases de datos incompletas. Es decir, la información en una o varias variables de interés está ausente. Este problema de datos faltantes puede afectar a los estudios caso-control, especialmente la falta de datos en covariables. Este fenómeno en las variables predictoras puede conducir a estimaciones sesgadas de los parámetros de interés, o poco eficientes (errores estándar de gran magnitud) provocando intervalos de confianza poco precisos y prueba de hipótesis incorrectas (Ibrahim *et al*, 2012). En consecuencia, para obtener estimaciones de riesgo válidas y precisas, y con esto, evidencia para la realización de

recomendaciones a la población, mucho esfuerzo debe considerarse a la hora de analizar al conjunto de observaciones (Lash *et al*, 2009). Una manera de mejorar la validez es abordar el problema de datos faltantes, aspecto éste frecuente en las investigaciones biomédicas.

Para llevar a cabo inferencias estadísticas en presencia de datos faltantes se plantean algunos supuestos sobre el mecanismo por el cual los datos están faltando y, cada análisis estadístico se basa en alguno de estos supuestos. La validez de los resultados obtenidos, luego de aplicar algún método de imputación, depende del cumplimiento de los supuestos sobre el mecanismo de datos faltantes que se adoptó (White *et al*, 2011). La clasificación de Little y Rubin se utiliza frecuentemente, la misma propone que los datos faltantes son completamente al azar (*Missing Completely at Random*, MCAR), al azar (*Missing at Random*, MAR), o que faltan de manera no aleatoria (*Missing Not at Random*, MNAR) (Acock 2005; White *et al*, 2011).

El supuesto más fuerte es el MCAR; éste se puede explicar fácilmente si hay solo una variable con datos faltantes, denotada con Z. Supongamos que hay otro grupo de variables que siempre es observado (es decir, que no tiene faltantes), representado por el vector X. R_z es un indicador *dummy* de la variable Z que es igual 1 si es *missing* (es decir, tiene valor perdido) y 0 si se observa Z. El supuesto MCAR puede expresarse como:

$$Pr = (P_z = 1|X, Z) = Pr (R_z = 1),$$

Esto significa que la probabilidad de que Z no se observe no depende ni de las variables X ni de los valores de la misma Z. El supuesto MCAR se satisface si en el conjunto de variables X sólo se incluyen las variables que formarán parte del modelo. Esta idea es la misma para el caso que más de una variable del modelo tenga *missing*, aunque analíticamente es más complicado. En otras palabras este mecanismo establece que la probabilidad de que una variable tenga *missing* no depende de ninguna otra variable del modelo ni de los potenciales datos faltantes de la misma variable. Esta situación es poco probable que ocurra y sucede, por lo general, si por diseño se decide recolectar cierta información en un subconjunto de sujetos y no en todos los sujetos estudiados (Allison, 2009).

El segundo supuesto es el denotado por MAR. La idea básica de éste recae en que la probabilidad de que una variable tenga datos faltantes podría depender de cualquiera de las

observadas, aunque no puede depender de ninguno de los valores no observados de las variables con datos faltantes (luego de ajustar por los valores observados). Esto es, la

$$\Pr(R_Z = 1|X, Z) = \Pr(R_Z = 1 | X).$$

Los datos faltantes en Z podrían depender de X, pero no de la misma variable Z (luego de ajustar por X). Lamentablemente, el supuesto MAR no puede ser evaluado. Por ejemplo, se podría tener razones para asumir que la probabilidad de ausencia de datos dependa de los valores que están faltando, por ejemplo, en el caso que las personas con ingresos altos pueden llegar a no desear reportar sus ingresos, aunque ningún aspecto de los datos informará si este es el caso o no. No obstante, existe una forma para que este supuesto sea más plausible. El supuesto MAR refiere que los patrones de datos faltantes en Z no dependen de Z, una vez que se ajusta por las variables en X. Si se colocan varias variables en X, especialmente las que correlacionan altamente con Z, se podría reducir o eliminar la dependencia residual de los datos faltante con la variable Z (Allison, 2009). En el caso del ingreso, colocando en el vector X variables como edad, sexo, ocupación, educación se puede hacer que el supuesto MAR sea más razonable.

Por ejemplo, los datos faltantes sobre presión sanguínea cumplen con el supuesto MAR si individuos de mayor edad tienen más probabilidad de tener su presión registrada (y la edad está incluida en el análisis), aunque podrían llegar a presentar un mecanismo MNAR si los individuos con elevada presión tienen más probabilidades de que su presión sea registrada, respecto de que otros individuos de la misma edad (Sterne *et al*, 2009).

Si el supuesto MAR no se cumple, entonces los datos están ausentes de manera no aleatoria. Este tercer supuesto (MNAR) ocurre cuando la probabilidad de que una observación contenga un dato perdido depende de los datos faltantes (también puede depender de los datos observados). La mayoría de los software estadísticos utilizan por defecto el análisis de datos completos para resolver el problema de datos faltantes, excluyendo del análisis sujetos con información incompleta en por lo menos una variable. En las últimas décadas, varios métodos simples fueron desarrollados, comenzando con la sustitución por la media o mediana, y la regresión lineal. En estos casos, los datos faltantes se imputan con un solo valor y la variable completa se utiliza para realizar los análisis siguientes (Allison, 2009). La incertidumbre asociada con el valor imputado debe considerarse para obtener estimaciones válidas, ya que el valor imputado no es el verdadero valor que se hubiese observado. Para resolver esta cuestión, el método más exitoso fue la imputación múltiple de datos,

desarrollado por Rubin en 1976 (Rubin, 1976). Esta metodología, a través de un proceso estocástico, selecciona posibles valores para la información faltante y la utilización de dichos valores recoge el componente aleatorio del dato imputado (Alfaro y Fuenzalida, 2009). Considerando la incertidumbre en la estimación de los parámetros y siguiendo los lineamientos de Rubin, la imputación multivariada mediante ecuaciones encadenadas es uno de los métodos más usados, y se basa en crear múltiples bases de datos completas, asignando valores a los datos faltantes (Acock, 2005; White *et al*, 2011). Luego, cada base de datos completada se analiza como si fuese una base completa sin faltantes. Las estimaciones obtenidas de cada base de datos se combinan en un único resultado (promedio de las bases completadas). Es importante mencionar que los valores imputados son de naturaleza aleatoria, aspecto éste a considerar en los errores estándar de los estimadores de los parámetros (Ibrahim *et al*, 2012). Este método, conocido como MICE por sus siglas en inglés (*Multiple Imputation by Chained Equations*), propone que la imputación se realice de forma secuencial encadenada, esto es, imputando primero una variable específica y luego utilizando dichos valores como verdaderos para imputar el resto de las variables. Realizando este algoritmo reiteradas veces se obtienen las distintas bases imputadas (Azur *et al*, 2011; Alfaro y Fuenzalida, 2009). Una característica importante de la imputación múltiple de ecuaciones en cadena es la habilidad de manejar diferentes tipos de variables (continuas, binarias, categóricas nominales, ordinales) desde que cada variable es imputada utilizando su modelo de imputación específico. La idea básica detrás de la imputación múltiple se describe en la próxima sección de esta tesis.

Como se mencionó anteriormente, el sesgo de memoria o *recall bias* puede presentarse en los estudios caso-control e influir en las estimaciones derivadas de los mismos. Sin embargo, pocos ejemplos han demostrado el sesgo de memoria haya sido un problema y derivado en conclusiones erróneas en estos estudios. Esta escasez de ejemplos disponibles puede significar una baja ocurrencia de este sesgo, o que la información necesaria para demostrar la existencia del mismo es difícil de obtener o generalmente no está disponible (Gordis, 2009). No obstante, este potencial problema no debe desestimarse. Existen análisis de sensibilidad bastante sencillos que pueden utilizarse para evaluar la calidad de los esquemas de clasificación de la variable de interés, por ejemplo, la exposición, donde el sistema de clasificación utilizado en el estudio se compara con un sistema de clasificación considerado más preciso para clasificar la variable de interés, el cual suele llamarse el *Gold*

Standard. Este análisis intenta conocer cómo sería la distribución de la exposición de los sujetos estudiados si el esquema de clasificación utilizado no hubiese presentado errores (Lash *et al*, 2009).

Cuando es posible llevar a cabo un estudio de validación interno, es decir, comparar el sistema de clasificación utilizado en el estudio con un *Gold Standard*, y la selección de los sujetos se realiza de manera aleatoria (no considera la clasificación de la exposición ni el verdadero valor de exposición para la selección) (Lash *et al*, 2009), los parámetros que se utilizan frecuentemente en los análisis de sensibilidad son las mediciones de sensibilidad (S), especificidad (E), el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN). La S está asociada a la probabilidad de que un sujeto expuesto sea clasificado correctamente como expuesto. Asimismo, la probabilidad de que un sujeto no expuesto sea clasificado correctamente como no expuesto corresponde a la E del sistema de clasificación. Interesa también conocer, dado que un sujeto fue clasificado como expuesto (o no expuesto), la probabilidad de que sea correcto. Estos valores son el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN). El VPP calcula la probabilidad de ser verdaderamente expuesto habiendo sido clasificado como tal, mientras que el VPN calcula la probabilidad ser clasificado como no expuesto siendo realmente no expuesto (Lash *et al*, 2009).

El sesgo de información derivado del error de clasificación, es muy común en estudios epidemiológicos y recibió muy poca atención porque existe una importante generalización de que el error de clasificación no diferencial entre los grupos a comparar (es decir, entre casos y controles) sesga los resultados hacia el valor nulo (Lash *et al*, 2009). Esto significa que si se observara alguna asociación entre la exposición y el resultado y se sospechara de la existencia del sesgo de información, la estimación obtenida sería “conservadora”. Probablemente la magnitud de la asociación fuese mayor pero no se estaría observando ya que la presencia del sesgo estaría “corriendo” las estimaciones hacia el valor nulo.

En los siguientes apartados se describen detalladamente el diseño, metodologías y desarrollo del presente estudio llevado a cabo con el objetivo de valorar cuantitativamente los errores sistemáticos presentes en los estudios epidemiológicos de tipo caso-control en cáncer conducidos por el Grupo de Epidemiología Ambiental del Cáncer y otras enfermedades en Córdoba (GEACC).

De acuerdo a la evidencia existente acerca del rol que tiene la alimentación, como también otros hábitos modificables, en la salud de la población, y teniendo en cuenta las

dificultades con las que se enfrentan los investigadores para llevar a cabo estudios epidemiológicos nutricionales a nivel local, esta tesis pretende aportar a la calidad de los estudios caso-control del GEACC para obtener estimaciones de riesgo válidas, precisas y generalizables sobre el efecto de la alimentación característica de la población de la provincia de Córdoba en la ocurrencia del cáncer de mama y colorrectal.

1.4. Hipótesis

En los estudios caso-control para la identificación de factores de riesgo asociados al cáncer, el diseño del análisis de sesgo sistemático y su cuantificación permite obtener estimaciones de riesgo que difieren de las estimaciones clásicas.

1.5. Objetivos

Objetivo General

Construir un marco metodológico para la valoración cuantitativa de los errores sistemáticos presentes en los estudios epidemiológicos de tipo caso-control en cáncer, llevados a cabo en la población de adultos de la provincia de Córdoba.

Objetivos Específicos

- Construir un proceso de análisis que permita medir presencia de *confusión* respecto a los factores involucrados en la red causal del cáncer a partir de estudios casos y controles de cáncer colorrectal y de mama, y la consecuente corrección de las estimaciones de riesgo.

- Establecer un protocolo para la detección del *sesgo de selección* de los sujetos participantes en los estudios de casos y controles de cáncer colorrectal y de mama.

- Valorar la posible presencia de *sesgos de información* y diseñar estrategias metodológicas a fin de minimizarlos durante el proceso de medición de la exposición para cada tipo de cáncer.

CAPÍTULO 2. MÉTODOS

El presente apartado está dividido en tres secciones principales. En el inicio se presentan las metodologías para valorar los errores sistemáticos. Luego, se profundiza sobre los estudios en los cuales se aplican las metodologías. Y finalmente, se describe específicamente cómo se llevaron a cabo los análisis y esquemas de modelación en cada uno de los estudios epidemiológicos mencionados anteriormente.

2.1. Metodología para la valoración de errores sistemáticos

2.1.1. Introducción

El objetivo de llevar a cabo un análisis de sensibilidad es estimar cuantitativamente el error sistemático que permanece luego de implementar el diseño del estudio y los análisis correspondientes (Lash *et al*, 2009). El marco metodológico para valorar sesgos depende principalmente del estudio epidemiológico que se aborda. Las estimaciones de medidas de riesgo, derivadas de estudios epidemiológicos observacionales, como los estudios caso control, pueden ser susceptibles al error sistemático, además del error aleatorio. Los epidemiólogos deben intentar obtener estimaciones precisas y válidas del efecto de una exposición sobre la ocurrencia de un evento de salud. Así, debieran diseñar y llevar a cabo análisis con este objetivo, reconociendo que nunca se reduce el error por completo (Lash *et al*, 2009).

La presente Tesis Doctoral se enmarca en el proyecto sobre “Epidemiología Ambiental del Cáncer en Córdoba”, llevado a cabo, desde el año 2004, por el equipo de investigadores pertenecientes a varias unidades académicas de la Universidad Nacional de Córdoba. Una de las líneas que aborda este equipo es el estudio de factores de exposición de origen alimentario y su relación con los cánceres más prevalentes de la provincia de Córdoba, entre ellos cáncer de mama, próstata, colorrectal y uroteliales. Esto conlleva a la conducción de diferentes estudios de tipo caso-control en el territorio provincial. Los resultados derivados de este tipo de estudios son susceptibles a una cierta cantidad de errores sistemáticos, lo que los convierte en buenos candidatos para aplicar metodología para valorar cuantitativamente sesgos.

Esta investigación diseñó un esquema de análisis cuantitativos de sesgos en los estudios caso-control sobre cáncer colorrectal y cáncer de mama de la provincia de Córdoba,

abordando el fenómeno de confusión y la posible presencia de los sesgos de selección y de información. Para estudiar este último sesgo, además de llevar a cabo análisis sobre sesgo de información derivado de datos faltantes en los estudios caso-control, se condujo un estudio de validación para evaluar presencia de error de clasificación en la exposición, específicamente en cáncer de mama.

2.1.2. Confusión

El fenómeno de confusión debe considerarse en todo estudio observacional ya que sus resultados pueden estar potencialmente influenciados por este problema. La presencia del mismo puede resultar en una sobre o subestimación de la asociación real (Hernandez-Ávila *et al*, 2000). Como se mencionó anteriormente, pueden ocurrir dos situaciones respecto a este fenómeno, una sucede cuando la variable que “confunde” la asociación entre la exposición y el resultado se conoce y ha sido registrada (es decir, fue observada). Y la otra ocurre cuando la variable confundente no ha sido medida (Lash *et al*, 2009).

En los estudios caso-control, varias son las metodologías que se implementan para evitar el fenómeno de confusión. Una de éstas, ampliamente utilizada en epidemiología, es el apareamiento o *matching*. Por diseño, se condiciona a los sujetos controles a cumplir ciertas características similares a los casos para participar. Por ejemplo, en los estudios caso-control del grupo de investigación del cual deriva esta tesis, los controles deben tener la misma edad (± 5 años) y ser de la misma ciudad o zona de residencia que los casos. Si el apareamiento se realiza con variables confusoras entre la asociación de la exposición y el resultado, luego la variable de apareamiento o de *matching* deberá incluirse en los análisis para evitar posible sesgo (Rothman y Greenland, 1998). Además de esta forma de estratificación, para probar el efecto de una posible confundente, se puede aplicar análisis de estratificación (por método de Mantel-Haenszel), o utilizar métodos de regresión cuando se desea analizar, de manera simultánea, el efecto de más de una variable con posible efecto confundente.

A continuación se describen, brevemente, tres metodologías aplicables en situaciones donde la posible confundente ha sido registrada y cómo probar su efecto.

Una de las técnicas más simples para detectar si una variable ejerce un efecto confundente en la asociación entre una exposición determinada y un resultado o evento en salud, es la comparación de las estimaciones de riesgo cruda y ajustada por la posible confundente mediante la estratificación, usando el test de Mantel-Haenszel (Mantel y Haenszel, 1959). La asociación cruda se obtiene calculando la medida de asociación (en este

caso los *odds ratio*, OR) entre la exposición y el outcome. El análisis estratificado consiste en calcular los ORs entre las dos variables principales (exposición y outcome) para cada uno de los estratos de la variable confundente, y luego, la estimación de riesgo ponderada (de cada estrato de la variable confundente), es decir la estimación ajustada o global. Este OR ajustado es simplemente el promedio ponderado de los OR de cada estrato de la variable confundente. Partiendo de la disposición de información a través de una tabla de 2x2 (Tabla 2.1), la fórmula general del procedimiento es la siguiente:

$$OR_{MH} = \frac{\sum_{i=1}^K a_i \times d_i / n_i}{\sum_{i=1}^K b_i \times c_i / n_i}$$

Cada estrato está representado por la letra subíndice “i” en esta fórmula y K es la cantidad de estratos.

Tabla 2.1. Tabla de contingencia (2x2) de disposición de grupos expuestos y no expuestos en estudio caso-control.

<i>i</i>	Casos	Controles
Expuestos	A	b
No Expuestos	C	d

En el caso de que exista efecto de confusión, las asociaciones en los diferentes estratos de la variable confusora deberían ser de similar magnitud entre éstos, aunque diferentes a la estimación cruda. Esta situación no siempre ocurre, por lo que además, debe probarse el potencial efecto confundente comparando las estimaciones de riesgo globales o ajustadas con las crudas. Al demostrar que estas estimaciones difieren, se está probando el efecto confundente de la variable en cuestión (Szklo y Nieto, 2007).

Otro método que permite analizar el fenómeno de confusión, sin dejar de tener presente esquemas de causalidad como el caso anterior, es utilizando modelos de regresión. Los modelos de regresión con múltiples covariables o predictores ajustan por las variables confundentes propuestas, modelando la exposición y las potenciales confundentes en relación al resultado. Estos modelos de regresión estiman el efecto de la exposición mientras mantienen los niveles de las confundentes constantes (Kamangar, 2012). Además, esta metodología permite verificar el posible efecto confundente de varias variables simultáneamente. Este método presenta las características del algoritmo *stepwise*, en el que

se incluyen todas las variables observadas con posible efecto confundente. Se basa en el supuesto de que aunque no todas las variables que se seleccionen en el modelo sean confusoras, las confundentes más importantes estarán incluidas y consideradas en el análisis. Este modelo ordena a las posibles confusoras según el porcentaje de modificación de efecto de la estimación de riesgo de la exposición sobre el resultado. La información obtenida se presenta de una manera gráfica para demostrar las variaciones en los OR calculados a medida que se va incluyendo cada variable al modelo (Wang, 2007).

Un tercer método, similar al anterior, que puede aplicarse cuando la variable de exposición es continua, consiste en plantear diferentes modelos de regresión con todas las variables que se deseen evaluar como posibles confundentes y observar gráficamente cómo se modifican las magnitudes de las estimaciones de riesgo al combinarse todos los sets de variables confundentes incorporadas en el modelo. Esto permite observar la dirección y magnitud de la estimación de riesgo que se estudia y analizar la significación estadística del modelo propuesto (Draper y Smith, 1998; Wang, 2007).

Los métodos presentados hasta el momento se utilizan cuando las variables confundentes son registradas. Sin embargo, existen situaciones en las cuales el investigador no ha recabado determinada información sobre los sujetos estudiados. Esto puede deberse a varios motivos, por cuestiones de diseño, por falta de recursos, etc. En algunos casos, las variables que no se registraron pueden ser identificadas por el investigador como posibles confundentes, es decir, el investigador reconoce que esa variable no recolectada puede ejercer un efecto confusor en la estimación de riesgo observada, aunque no se ha podido registrar. A esta variable se la denomina confundente no observada o no registrada. En otros casos, el investigador no conoce a la posible confusora. En estos contextos es más difícil proponer supuestos razonables sobre las relaciones entre la confundente y la enfermedad, o su distribución en la población en estudio. En el caso de que la confundente sea reconocida por el investigador y no se haya registrado, existe metodología que propone el supuesto de posibles escenarios en los cuales se explora cómo hubiese sido la asociación entre la exposición y el *outcome* en el caso de que la variable hubiese podido registrarse y ajustarse, por ésta, en el modelo (Lash *et al*, 2009).

Los resultados derivados de modelos para variables dependiente binarias, como el caso de las regresiones logísticas, pueden ser influenciados por variables no observadas. La idea básica de este método es proponer *a priori* diferente variabilidad a la confundente no

observada, o más aún, diferentes distribuciones de la variable y observar el impacto que tendría esta confusora en las estimaciones de riesgo obtenidas (Buis, 2010). Además, se deben plantear *a priori* coeficientes de asociación entre la variable no observada y alguna variable registrada, asumiendo que pueden presentar distribuciones similares. Esta metodología permite explorar el impacto que hubiese tenido ajustar por la variable confusora no registrada, mediante la simulación de diferentes escenarios, que imponen distribuciones específicas de la variable y determina *a priori* la variabilidad de la confundente no observada.

2.1.3. Sesgo de Selección

Los estudios epidemiológicos caso-control retrospectivos pueden ser los más propensos a presentar este sesgo (Geneletti et al, 2009). Por esto, con el fin de evaluar la posible presencia de sesgo de selección y estudiar el impacto en las estimaciones de riesgo, en los estudios caso-control del GEACC, se llevó a cabo un análisis de sensibilidad, implementando un método determinístico y otro probabilístico. Este análisis consiste en especificar *a priori* distribuciones de probabilidad para los parámetros de sesgo. Al utilizar distribuciones de probabilidad plausibles, se pueden reportar resultados que incorporan la incertidumbre derivada de este error sistemático y evitar sobrevalorar la seguridad con la que se declara el efecto estudiado (Orsini et al, 2008). Este análisis de sensibilidad basado en simulaciones de Monte Carlo incorpora dos pasos iterativos: 1) a partir de funciones de densidad de probabilidad especificadas se extrae una muestra aleatoria de parámetros de sesgo, y 2) se calculan, de manera inversa, las estimaciones de riesgo ajustadas (“corregidas”) derivada de la muestra aleatoria de los parámetros de sesgo. Estos pasos se repiten varias veces para obtener la distribución de estimaciones de riesgo ajustadas por el sesgo.

Dependiendo del software utilizado para analizar este sesgo de selección, se pueden postular diferentes funciones de densidad probabilística, entre ellas, una constante, uniforme, triangular, trapezoidal, logística o normal (Orsini et al, 2009). En los estudios caso-control se conforman cuatro grupos de sujetos en base a la enfermedad y la exposición: casos expuestos, casos no expuestos, controles expuestos, controles no expuestos. En el método determinístico, a cada uno de estos grupos se le asigna una distribución de probabilidad específica de tipo trapezoidal para evaluar el impacto en las estimaciones de riesgo obtenidas. Es decir, se plantean diferentes escenarios imponiendo distribuciones de probabilidad distintas, donde algunos grupos presenten mayores probabilidades de participación en el estudio que otros, para luego evaluar el efecto que tendrían estas

situaciones en las estimaciones de OR. En el caso del método probabilístico, particularmente, se asigna una distribución log normal *a priori* al sesgo de selección, con media 0 y desviación estándar 0,21, que genera un 95% de probabilidad *a priori* de que el factor de sesgo se encuentre entre $\exp(0-1,96*0,21)=0,7$ y $\exp(0+1,96*0,21)=1,5$, como otros autores han propuesto (Orsini *et al*, 2009).

2.1.4. Sesgo de Información

A continuación se desarrollan metodologías para enfrentar el sesgo de información por falta de datos en covariables y un análisis de sensibilidad para abordar el posible error de clasificación de la exposición.

2.1.4.1. Sesgo de información derivado de datos faltantes

El problema de los datos faltantes en los estudios epidemiológicos ocurre muy frecuentemente y el método de imputación múltiple por ecuaciones en cadena (MICE) es uno que se utiliza ampliamente para resolver estas situaciones, como se mencionó anteriormente. A continuación se explica de forma sencilla, cómo se realiza la imputación mediante ecuaciones encadenadas. Este método imputa, es decir, completa valores faltantes en varios pasos. Inicialmente, todos los valores faltantes se completan de manera aleatoria, reemplazándolos por valores observados. La primera variable con datos faltantes, X_1 , se le aplica un método de regresión en base a todas las otras variables del modelo, condicionando solo entre las que tienen los datos observados de X_1 . Los datos faltantes de la variable X_1 son reemplazados mediante extracciones simuladas correspondientes a la distribución predictiva posterior de X_1 creada recientemente. Luego, la siguiente variable con datos faltantes, X_2 , se le aplica la regresión incluyendo el resto de las variables, restringiendo solo a los sujetos que tienen observaciones en la X_2 , utilizando los valores imputados para la X_1 . Otra vez, todos los datos faltantes de la X_2 son reemplazados por valores obtenidos de la distribución predictiva posterior de X_2 . Este proceso se repite para todas las variables con datos faltantes. A todo este proceso se lo llama ciclo. Todo un ciclo produce una base de datos imputados. Todo este procedimiento se repite m veces para generar m cantidad de bases imputadas (White *et al*, 2011). Luego, una vez que las m bases están completas se realizan los análisis de interés final (por ejemplo, estimación de modelos de regresión logística) de manera individual en cada m base y finalmente las estimaciones independientes se combinan para obtener una única estimación final. Para obtener estas estimaciones finales se utilizan las reglas de Rubin

(1976,1987). Entonces, el parámetro de interés (coeficiente de regresión) se calcula como la media de los parámetros obtenidos en cada base de datos imputada. Para calcular los errores estándar combinados primero se obtiene la media de los errores estándar cuadrados de cada m base de datos (varianza intra-imputación), luego se calcula la varianza de los m parámetros estimados entre las bases (varianza entre-imputaciones) y finalmente se combinan estas dos medidas de variabilidad a través de la siguiente fórmula, donde b_k es el coeficiente de regresión de k de las M bases y S_k es su error estándar estimado. La media de b_k es \bar{b} ; y el error estándar se obtiene mediante la expresión

$$EE = \sqrt{\frac{1}{m} \sum S_k^2 + \left(1 + \frac{1}{m}\right) \left(\frac{1}{m-1}\right) \sum_k (b_k - \bar{b})^2}.$$

Esta sería la raíz cuadrada de la media de las varianzas de las m bases más la varianza de los coeficientes estimados multiplicados por $(1 + 1/m)$ bases (Allison, 2000). Los resultados combinados que se obtienen en el tercer paso del proceso de imputación incorporan la variabilidad dentro de cada m base de datos imputada (incertidumbre sobre los resultados de una base imputada) y la variabilidad entre las distintas imputaciones (reflejando la incertidumbre debido a información faltante).

2.1.4.2. Sesgo de información por error de clasificación

El error de clasificación puede llegar a distorsionar severamente la magnitud de una asociación entre un factor de riesgo y una enfermedad. Este error puede ocurrir en la variable de exposición, el *outcome* o alguna covariable (Lash *et al*, 2009). En los estudios caso-control conducidos por el GEACC, se planteó la necesidad de indagar acerca del posible error de clasificación de la exposición. Esto surge debido a las características de diseño de dichos estudios, ya que puede existir clasificación errónea de los sujetos en relación a la exposición, derivado del sesgo de memoria o *recall bias*. En estos estudios, se indaga sobre hábitos de años anteriores al momento de la entrevista o diagnóstico y, la habilidad para recordar la exposición en el pasado puede depender de ser sujeto caso con enfermedad o sin enfermedad. En general, los casos pueden tender a declarar en mayor medida la exposición, en comparación con los controles, éstos tratan de identificar algunos eventos inusuales que hayan ocurrido, si la enfermedad que presentan se debe a algo que la persona hizo, si fueron responsables, etc. Este error de clasificación puede ser también no diferencial, esto es, el

grado de error de la exposición se presenta en casos y en controles en la misma medida (Lash *et al*, 2009).

La información sobre la exposición en los estudios caso-control del GEACC se recopila a través de frecuencias de consumo alimentario, utilizadas generalmente para evaluar ingestas medias de largos períodos de tiempo. Actualmente, el campo de la epidemiología nutricional se ve limitado por la falta de estándares de referencia (o *Gold Standard*) disponibles para validar la información alimentaria obtenida de diferentes instrumentos. Sólo algunas mediciones de referencia ideales están disponibles, por ejemplo, para el consumo de energía, de proteínas o de potasio. Sin embargo, por razones prácticas, se suelen utilizar mediciones de referencia no ideales como las que surgen de instrumentos denominados los recordatorios de 24 horas o los registros alimentarios (Lachat *et al*, 2016). En la presente tesis, que intenta conocer la calidad del esquema de clasificación de los sujetos expuestos y los no expuestos derivado de la frecuencia de consumo alimentario, se lo comparó con el recordatorio de 24 horas, instrumento considerado como el *Gold Standard*.

2.1.4.2.1. Una propuesta para analizar el error de clasificación de las frecuencias de consumo alimentario

A continuación se propone un esquema para modelar los consumos obtenidos a partir de recordatorios de 24 horas (R24H) y obtener a partir de estos, la clasificación *Gold Standard* que servirá para comparar la calidad con la que clasifica la alimentación de los sujetos estudiados a través de la frecuencia de consumo alimentaria (FFQ).

En los análisis de sesgos realizados hasta el momento la variable de exposición utilizada fueron los patrones alimentarios emergentes obtenidos de los estudios caso-control del GEACC. Debido a que esta variable es una medida integral de exposición que combina información de numerosas variables de naturaleza alimentaria, se decide trabajar con las variables de los grupos de alimentos con mayores cargas factoriales obtenidos del primer patrón identificado (carnes procesadas y productos de pastelería), derivado de los R24H, para poder realizar los análisis de error de clasificación en este estudio de validación.

La mayoría de los estudios de validación han realizado los análisis en base a los consumos de nutrientes, sin embargo, es posible realizar las comparaciones a nivel de alimentos individuales o grupos de alimentos (Willett, 1998). Estas comparaciones pueden

llegar a ser muy útiles para enfocar la atención en ciertos ítems alimentarios que tienen desempeño pobre, es decir, que la recolección de la información de estos alimentos a través del instrumento de recolección de información utilizado puede no ser precisa (en este caso la FFQ), con el fin de sugerir áreas específicas para mejorar o modificar en próximas aplicaciones del mismo.

El procedimiento de recolección de información alimentaria, sobre la misma unidad o sujeto a través de los R24H produce una serie de medidas repetidas sobre cada unidad, que se caracterizan por estar correlacionadas. Entonces, para cada sujeto existen varios consumos diarios alimentarios registrados durante un período de tiempo. Para obtener un único consumo representativo de todos los días de consumo registrados que consideren la correlación entre las medidas repetidas, se construyeron modelos lineales generalizados mixtos para los dos grupos de alimentos mencionados (Rabe-Hesketh 2012). Estos modelos se conocen con el nombre de modelos lineales mixtos, debido a que incorporan tanto factores fijos, como aleatorios. Estos diseños mixtos, comprenden por un lado, los diseños multinivel o jerárquicos; y por otro, los modelos longitudinales o de medidas repetidas. La característica distintiva de los segundos, como se mencionó anteriormente, es la medición repetida a lo largo del tiempo en cada individuo o sujeto de estudio. Por tanto, se puede considerar la estructura como mixta, con las observaciones (repetidas) agrupadas dentro de cada individuo y el tiempo se puede considerar como una variable explicativa más dentro de cada grupo (Sáez, 2001). La expresión general para este tipo de modelos es la siguiente:

$$\text{Log}(\mu_{ij}) = \alpha + \varepsilon_{ij},$$

donde:

- μ_{ij} es el valor medio del consumo del grupo de alimento de interés en uno de los 12 momentos registrados ($i = 1, \dots, 12$) de una de las 57 mujeres o sujetos estudiados ($j = 1, \dots, 57$).
- α es el valor general del intercepto (efecto fijo) en el origen para todas las mujeres estudiadas, esto es, la media general.
- ε_{ij} es el error del predictor lineal representado por un intercepto aleatorio para el consumo del grupo de alimento, tal que $\varepsilon_{ij} \sim N(0, \sigma^2)$ (Seoane, 2014).

En los modelos también se incluyó el tiempo como variable independiente. A partir de los valores predictivos de estos modelos (con media y con media más efecto tiempo) se obtienen las modas y las medias de los consumos de los grupos de alimentos. Luego de realizar pruebas diagnósticas gráficas, se decide trabajar con la moda del consumo de cada

grupo de alimentos sin cuantificar el efecto del tiempo, ya que resultaron ser más homogéneas. Finalmente, se calcula la media del consumo a partir de las modas de los consumos diarios de los sujetos estudiados. Entonces, se obtiene así para cada sujeto, el consumo medio por grupo de alimento.

Si el investigador, de alguna manera, pudiera conocer el verdadero valor de categoría de exposición, se podría evaluar la calidad del sistema de clasificación utilizado y evaluar si clasifica correctamente a sujetos expuestos de los no expuestos. Para ello, la clasificación de la exposición entre expuestos y no expuestos de los consumos de grupos de alimentos derivados del *Gold Standard* se compararon con la clasificación de la FFQ que recabó información durante el mismo período en el que se levantaron los R24H. Del estudio de validación se obtuvieron los parámetros para realizar los análisis de sensibilidad específicos para el error de clasificación de la exposición y evaluar la calidad del sistema de clasificación de la FFQ de los estudios caso-control.

2.2. Materiales

Del objetivo general de esta tesis de doctorado deriva el interés por analizar la calidad de los estudios caso-control del GEACC. A continuación se describen los diseños y características principales de los estudios caso-control de cáncer de mama y colorrectal en los cuales se llevaron a cabo los análisis de sensibilidad. Además, se describe el estudio de validación interno realizado en cáncer de mama.

2.2.1. Estudios caso control para cáncer en Córdoba

2.2.1.1. Diseño de los estudios

El presente trabajo se enmarca en el proyecto sobre “Epidemiología Ambiental del Cáncer en Córdoba”, como se mencionó anteriormente. Una de las líneas de estudio sobre alimentación y cáncer lleva a cabo estudios de tipo caso-control en el territorio provincial.

Específicamente, para la presente tesis, fueron indagadas mujeres con cáncer de mama (CM) y sujetos con cáncer colorrectal (CCR) que conformaron los grupos de casos para los respectivos estudios de tipo caso-control. El grupo de casos del estudio de caso-control de CM estuvo conformado por 318 mujeres mayores de 18 años de edad con diagnóstico histopatológicamente confirmado de CM (CIE-10 C50) (OMS, 1992) reciente (menos de tres años), residentes en la provincia de Córdoba en el período 2008-2015. En el grupo casos del

estudio caso-control de CCR se estudiaron 161 sujetos casos mayores de 18 años de edad con diagnóstico histopatológicamente confirmado de CCR (CIE-10 C18-20) (OMS, 1992) residentes en la provincia de Córdoba en el período 2010-2016. Como criterios de inclusión para la selección de los sujetos controles en ambos estudios, se consideró que los individuos procedan del mismo departamento de residencia que los casos, que presenten el mismo sexo y edad similar a estos (± 5 años), que no padezcan o hayan padecido alguna enfermedad oncológica, ni otras asociadas a los tipos de tumores en cuestión, y que no presenten hábitos alimentarios particulares (por costumbres, salud o creencias religiosas). Luego, fueron incluidos 526 y 329 controles para el estudio de CM y CCR respectivamente. Además, se establecieron como criterios de exclusión, para ambos grupos casos y controles, ser menor de 18 años, presentar una neoplasia benigna o alguna patología asociada a los cánceres en estudio, presentar hábitos alimentarios particulares (como ser vegetariano o adventista), y presentar como tumor primario otra localización tumoral diferente de la estudiada.

La construcción de la muestra tuvo en cuenta la necesidad de los test estadísticos para trabajar con una potencia de por lo menos 80%: seleccionando aproximadamente 2 controles por cada caso.

Los sujetos con la patología fueron identificados con la colaboración de profesionales de la salud (clínicos, oncólogos, ginecólogos, anatomopatólogos) pertenecientes a los hospitales públicos e instituciones de salud privadas de Córdoba Capital y de localidades del interior de la provincia (siendo las principales Hernando, Jesús María, La Falda, Río Cuarto, Río Tercero, Villa Carlos Paz, Villa del Rosario, Villa María, entre otros). Por otra parte, los controles fueron seleccionados aleatoriamente de la población de cada localidad y en función de los criterios de inclusión antes descriptos. Los sujetos fueron invitados a participar en el presente estudio y dieron su conformidad, luego de una adecuada información de la naturaleza del estudio (consentimiento informado) (Anexo 1), según las Pautas Éticas Internacionales para la Investigación y Experimentación Biomédica en Seres Humanos del Consejo de Organizaciones Internacionales de las Ciencias Médicas (CIOMS/OMS/2002, y sus modificaciones). Es importante destacar que este estudio se enmarca dentro del estudio marco del equipo de investigación que fue sometido a evaluación y aprobado por el Comité de Ética del Hospital Nacional de Clínicas de Córdoba (CIEIS-HNC) (REPIS N°044/10) (Anexo 2).

Los sujetos participantes fueron encuestados por la tesista y encuestadores entrenados, utilizando un cuestionario semiestructurado que indaga, por un lado, características antropométricas, sociodemográficas, de estilos de vida, y también sobre la

historia médica personal y familiar, entre otras. En una segunda parte, para obtener información precisa sobre los hábitos alimentarios de los sujetos se recoge la frecuencia de consumo alimentaria donde se recaba información sobre 127 ítems alimentarios, las formas de cocción de los mismos y las porciones consumidas. Este cuestionario validado para su utilización en estudios epidemiológicos en la población de Córdoba (Navarro *et al*, 2001) (Anexo 3), se acompaña con un atlas de alimentos fotográfico (Navarro *et al*, 2000, 2007), para facilitar la cuantificación de las porciones (pequeña, mediana o grande). La conversión de las porciones de alimentos reportados a gramos o centímetros cúbicos de alimento consumido por día, así como la posterior cuantificación de la ingesta diaria en términos de calorías, macro y micronutrientes, fue realizada empleando el software de análisis nutricional Nutrio 2.0 (Peyrano *et al*, 1998). Todos los encuestados fueron indagados sobre sus prácticas alimentarias en el pasado: 5 años previos al momento del diagnóstico en el grupo casos y 5 años previos al momento de la entrevista, en los controles.

2.2.1.2. Variables de estudio

Los patrones alimentarios, representados por sus scores de adherencia en cada sujeto se definieron como covariables, como variable respuesta (outcome) la presencia de CM o CCR (de naturaleza dicotómica). A continuación se presentan las definiciones operacionales de las mismas, así como de las variables intervinientes consideradas y demás variables de ajuste utilizadas en los modelos y metodologías empleadas. Asimismo, se definieron otras variables, utilizadas en la fase exploratoria de los datos muestrales, las cuales se presentan a continuación.

La variable respuesta de los estudios caso-control fue:

- Presencia de cáncer de mama/ presencia de cáncer colorrectal

Esta variable es de naturaleza dicotómica (si/no) y se define como la presencia de adenocarcinoma mamario o adenocarcinoma de colon o recto histológicamente confirmado de manera reciente (no más de tres años previos al momento de realización de la encuesta).

Las covariables de exposición principales fueron:

- Patrones Alimentarios

Se define como el conjunto de productos que un individuo, familia o grupo de familias consumen de manera ordinaria según un promedio habitual de frecuencia estimado en por lo menos una vez al mes. En el campo de la Epidemiología Nutricional, se utilizan distintos

métodos estadísticos de reducción o agregación de componentes para detectar ese conjunto o grupos de alimentos (Kac *et al*, 2007). En este caso, se aplicó el Análisis Factorial de Componentes Principales.

Se utilizaron grupos alimentarios representativos de la dieta típica de la población local como variables intermedias para la construcción de patrones. Para el estudio caso-control sobre cáncer de mama, un total de 20 grupos alimentarios fueron empleados en el análisis de identificación de patrones alimentarios, llevado a cabo en estudios anteriores (Tumas *et al*, 2014). Los grupos alimentarios fueron los siguientes: cereales refinados, cereales integrales, productos de pastelería, legumbres, vegetales no amiláceos, vegetales amiláceos, frutas, leche y yogur, quesos blandos, quesos duros, aceites y mayonesa, manteca y crema, azúcares y dulces, confituras, bebidas alcohólicas, bebidas analcohólicas, huevos, carnes procesadas, carnes grasas y carnes magras (Apéndice 1).

A partir de tales grupos de alimentos se identificaron patrones alimentarios característicos, mediante un Análisis Factorial de Componentes Principales (Johnson y Wichern, 1992). De los 3 factores (patrones) identificados, se utilizó como covariable principal el primer factor denominado patrón dietario Tradicional, caracterizado por elevadas cargas de carnes grasas, productos de pastelería y aceites y mayonesa.

Para el estudio caso-control sobre cáncer colorrectal, 12 grupos alimentarios fueron empleados en el análisis de identificación de patrones alimentarios realizado por Pou *et al* (2012). Los grupos fueron los siguientes: cereales y derivados, vegetales amiláceos, frutas y vegetales no amiláceos, lácteos, carnes procesadas, carnes rojas (vacuna y porcina), pescado y pollo, huevos, azúcares y confituras, bebidas azucaradas, vino, grasas y aceites. En este trabajo se utilizó el patrón Cono Sur (principal patrón identificado) como covariable, caracterizado por un elevado consumo de carnes rojas, vino y vegetales amiláceos.

Luego, para cada sujeto en cada estudio caso-control se calcularon los coeficientes de puntuación o escores para el factor o patrón identificado mediante método de regresión. Estos escores indican el grado en que la dieta del sujeto se ajusta a cada patrón identificado (Edefonti *et al*, 2008). Estas nuevas variables construidas, el patrón Tradicional y el patrón Cono Sur, fueron las principales covariables utilizadas en los estudio caso-control de CM y CCR, respectivamente.

-Sexo: femenino/masculino,

-Edad: el tiempo de vida a la fecha de la entrevista (en años),

-Nivel socioeconómico: esta variable se construyó indagando sobre la ocupación y el nivel educativo alcanzado por el principal sostén del hogar, la cantidad de aportantes del hogar, la posesión de algunos bienes y servicios (autos, computadora, internet, tarjeta de débito), así como la cobertura de salud. A cada uno de los diferentes ítems le fue asignado un valor y a partir de ello se obtuvo un puntaje total final que permitió incluir al hogar del encuestado en alguna de las seis clases establecidas (Asociación Argentina de Marketing, 2002) (Anexo 4). Esta variable fue utilizada como categórica marginal, baja inferior, baja superior, media típica, media alta y alta o en dos baja: marginal, baja inferior, baja superior, y media-alta: media típica, media alta y alta.

Las variables que se incluyeron en la construcción del nivel socioeconómico se describen brevemente con ocho modalidades:

- Nivel educativo: máximo nivel de instrucción alcanzado por el entrevistado (o en el caso de no ser el principal sostén del hogar, el máximo nivel educativo de este último). En este estudio fue categorizada como sigue: sin estudios, primaria incompleta, primaria completa, secundario incompleto, secundario completo, estudios terciarios o universitarios (incompletos o completos).
- Ocupación: Tiene que ver con la inserción en el sistema productivo del principal sostén del hogar. Para poder clasificar a las actividades laborales de los sujetos se tienen en cuenta varias dimensiones, entre ellas, la condición de actividad, la modalidad de inserción laboral, la jerarquía ocupacional, la dedicación horaria, etc.
- Cantidad de aportantes del hogar: cantidad de sujetos que aportan económicamente en el hogar del entrevistado. Las categorías utilizadas fueron: un aportante, dos o tres aportantes o más de tres aportantes.
- Autos: cantidad de automóviles que poseen en el hogar. Categorizada en: ninguno, uno, dos, tres o más.
- Computadora: variable dicotómico (Si/No) que refiere si poseen computadoras en el hogar del entrevistado.
- Internet: variable dicotómico (Si/No) que refiere si poseen servicio de internet en el hogar del entrevistado.
- Tarjeta de débito: variable dicotómico (Si/No) que refiere si el entrevistado posee tarjeta de débito.
- Obra social: variable dicotómica (Si/No) que refiere a si el entrevistado posee obra social o no.

-Presencia de antecedentes familiares de CCR: Se consideró que un individuo presente este antecedente familiar cuando declara que al menos un familiar de primer grado (padre, madre o hermanos) padece o padeció CCR.

-Hábito de fumar: o consumo de tabaco. Para su valoración se tuvo en cuenta el hecho de que los sujetos fumen al momento de realizar la encuesta o que hayan fumado alguna vez (fumador) versus que nunca hayan fumado (no fumador).

-Actividad Física: se consideró actividad física a cualquier movimiento corporal producido por los músculos esqueléticos que exija gasto de energía. (OMS, 2017 *b*; Shephard y Balady, 1999). Se utilizó una versión adaptada del Cuestionario Internacional de Actividad Física (The International Physical Activity Questionnaire [IPAQ], 2005) a fin de conocer el tipo de actividad efectuada cinco años previos al diagnóstico de la enfermedad en los casos y cinco años anteriores a la realización del cuestionario en los controles. Para obtener el nivel de actividad física, se utilizó como criterio los equivalentes metabólicos (METs) semanales, calculados a partir de la sumatoria de los METs de cada tipo de actividad. Conforme lo establecido en el IPAQ, los METs se obtienen a partir de la multiplicación de los minutos de actividad por un coeficiente de 3,3 para la caminata, 4 para la actividad moderada y 8 para la actividad vigorosa (IPAQ, 2005). Para este estudio se consideró a la variable como continua y categorizada de la siguiente manera: sedentaria (menos de 600 METs semanales) y activa (600 o más METs semanales) (Anexo 5).

-Índice de masa corporal (IMC): como indicador del estado nutricional de un individuo. Se calcula dividiendo el peso corporal (kg) por la talla elevada al cuadrado (m^2). Para su cálculo se usaron medidas de peso y talla declaradas, considerando el peso habitual. Esta variable fue continua (kg/m^2), y también categórica, conforme la clasificación del IMC establecida por la OMS: bajo peso-normopeso (menos de $24,9 kg/m^2$), preobesidad (25 a $29,9 kg/m^2$) y obesidad (más de $25 kg/m^2$) (OMS, 2006).

-Estatus ginecológico: se refiere a si las mujeres se encuentran en estado premenopáusico o postmenopáusico. Para los casos se consideró si las mujeres menstruaban o no al momento del diagnóstico de la enfermedad, y para los controles al momento de la realización de la encuesta. Esta diferenciación responde a que muchas veces el tratamiento oncológico recibido por las mujeres con cáncer de mama les induce la menopausia.

-Hijos: hace referencia a si las mujeres tuvieron hijos biológicos o no (nulíparas). Esta variable se definió como dicotómica: No (nulípara)/ Si (tuvo uno o más hijos).

-Lactancia materna: puede entenderse a la lactancia materna como una práctica sociocultural, y también como una variable de naturaleza biológica, vinculada a la reproducción humana. A los fines de este estudio se consideró si las mujeres no amamantaron a sus hijos o lo hicieron por menos de 6 meses; si lo hicieron durante 6 meses o más (variable dicotómica cuyas categorías: 0-5 meses /6 o más meses).

-Edad de la menarca: edad de la mujer en la que tuvo su primera menstruación. En este estudio se consideró a esta variable como continua (en años) y dicotómica (<12 años/ ≥12 años).

2.2.2. Estudio de validación para cáncer de mama en Córdoba

2.2.2.1. Diseño del estudio

En esta etapa se utilizó otro de los estudios del GEACC, uno epidemiológico de tipo prospectivo, que también fue sometido a evaluación y aprobado por el Comité de Ética del Hospital Nacional de Clínicas de Córdoba (CIEIS-HNC) (REPIS N°159/13) (Anexo 6). De éste se deriva, para la presente tesis, específicamente un estudio longitudinal, en el que se invitaron a participar a mujeres que fueron incluidas en el estudio caso-control de CM. Veinte mujeres casos y 37 controles fueron seguidas durante un año, evaluándolas cada 3 meses. En cada encuentro se aplicó un recordatorio de 24 horas para recabar información sobre la alimentación de 3 días (de semana y fin de semana). Así, para cada mujer participante se obtuvieron hasta 12 recordatorios de 24 horas (en el caso de que hubiese participado en todas las instancias). Los recordatorios de 24 horas fueron conducidos por personas capacitadas a tal fin, solicitando a las mujeres entrevistadas informar sobre todos los alimentos y bebidas que habían consumido durante el día anterior, facilitando la cuantificación de los consumos con el atlas fotográfico de alimentos anteriormente mencionado. La posterior cuantificación de la ingesta diaria en términos de calorías, macro y micronutrientes, fue realizada empleando el mismo software de análisis nutricional Nutrio 2.0 (Peyrano *et al*, 1998). Al finalizar el año de seguimiento, se aplicó una frecuencia de consumo alimentario obteniendo información sobre la alimentación habitual del último año transcurrido, mismo tiempo en el que se realizaron los recordatorios de 24 horas.

2.2.2.2. Variables de estudio

En este estudio de validación específico, la información alimentaria es recogida a través de dos instrumentos. La alimentación de las mujeres se definió como covariable o de exposición principal. Los grupos de alimentos identificados en el primer factor (Tumas *et al*, 2014) mediante el análisis de componentes principales (carnes procesadas y productos de pastelería), fueron las variables alimentarias de exposición utilizadas en los subsiguientes análisis. El resto de las variables de interés, utilizadas con fines descriptivos presentan las mismas definiciones operacionales y ya han sido descriptas. A continuación, se definen las variables alimentarias de exposición utilizadas en este estudio longitudinal según el instrumento utilizado para recolectar la información.

Los grupos de Alimentos consumidos a partir de recordatorios de 24 horas se describen como sigue:

En este estudio longitudinal de 57 mujeres, se decidió trabajar con los grupos de alimentos con más cargas factoriales identificados en el primer patrón emergente. De los 20 grupos de alimentos mencionados anteriormente, se decidió agrupar aún más los grupos de alimentos consumidos para obtener estimaciones más eficientes en el Análisis Factorial de Componentes Principales. Así, quedaron conformados 13 grupos de alimentos mencionados a continuación: cereales y vegetales amiláceos; productos de pastelería; frutas y vegetales no amiláceos; lácteos (leche, yogur, quesos)¹; aceites y mayonesa; manteca y crema; azúcares, dulces y confituras; bebidas alcohólicas; bebidas analcohólicas; huevos; carnes procesadas; carnes grasas; carnes magras.

Los grupos de alimentos consumidos definidos como de exposición, con cargas factoriales elevadas en el primer patrón identificado, fueron las carnes procesadas (bondiola, jamón cocido, jamón crudo, mortadela, panceta, queso de cerdo, salame de milán, salamín, salchichón, chorizo, morcilla, salchicha parrillera, salchicha de viena) y los productos de pastelería (criollos, facturas, galletitas dulces, pan con grasa, torta frita, tortas dulces).

¹ Dado que todos los alimentos que constituyen este grupo contienen los mismos nutrientes, pero en muy diferente concentración, se convirtió la ingesta diaria de leche y yogur a *equivalentes de gramos de queso* de la siguiente manera: 100g de leche/yogur= 30g de equivalentes de queso. Esta última medida fue usada en la sumatoria final del grupo a los fines de equiparar el volumen de ingesta de la leche y el yogur al volumen de ingesta de los quesos.

Como se mencionó anteriormente, utilizando modelos lineales generalizados mixtos se obtuvieron los consumos modales de los grupos de alimentos para cada sujeto utilizados en los análisis, como fuera explicado en la sección 2.1.4.2.1. *Una propuesta para analizar el error de clasificación de las frecuencias de consumo alimentario.*

Los grupos de alimentos construidos a partir de la frecuencia de consumo alimentario fueron carnes procesadas y productos de pastelería, los mismos grupos derivados de los R24H. La frecuencia de consumo alimentario se aplicó para conocer el consumo alimentario habitual del último año, durante el mismo período en el que se recabo información a través de los R24H.

2.3. Análisis estadísticos

A continuación se detallará cómo se llevaron a cabo los análisis estadísticos en ambos estudios al abordar cada sesgo estudiado. Para los mismos se utilizó el software estadístico Stata 14 (StataCorp, 2015).

2.3.1. Confusión

Para el estudio de caso-control de cáncer de mama, se aplicó el método de Mantel-Haenszel para probar el efecto confundente de la variable edad (Benz, 2008), en la relación entre la exposición Patrón Tradicional y la patología. Luego, en el análisis *stepwise* se indagó el posible efecto confundente de las siguientes variables: edad, nivel socioeconómico, IMC, actividad física, estatus ginecológico, hijos, lactancia materna y edad de la menarca. Estas variables también se incluyeron en el análisis combinado, donde se representaron gráficamente y se visualizaron las estimaciones de riesgo de todas las posibles combinaciones del conjunto de variables confundentes incluidas en el análisis. La cantidad de combinaciones observadas dependieron de la cantidad y tipo de variables incluidas. En cuanto al análisis para evaluar el posible efecto de una variable no observada en las estimaciones de riesgo, se realizaron simulaciones de variabilidad en diferentes escenarios hipotéticos respecto a la variable glucemia (no registrada) asignando diferentes coeficientes de asociación a saber: 0,20; 0,30; 0,40 y 0,50, respecto a una variable registrada en la base e incluida en el modelo de análisis (IMC). La evidencia científica sugiere que esta variable puede llegar a confundir la asociación entre la exposición y la patología estudiada (Lee *et al*, 2017; Chen *et al*, 2016; Barclay *et al*, 2008; Dong y Qin, 2011).

Para el estudio caso-control de cáncer colorrectal se estudió el efecto confundente de la variable Nivel socioeconómico (Doubeni *et al*, 2012) utilizando el método Matel-Haenszel, en la asociación entre Patrón Cono Sur y la enfermedad. Edad, sexo, nivel socioeconómico, IMC, actividad física y antecedentes de primer grado de cáncer colorrectal, fueron las variables que se incluyeron en los análisis *stepwise* y combinado. Para el análisis de simulación de variabilidad para una variable no registrada, se decidió utilizar la variable aumento de peso durante la vida adulta, la cual no está registrada en el presente estudio y existe evidencia de su asociación al CCR (Keum *et al*, 2015; Chen *et al*, 2015; Song *et al*, 2015 *b*). Se hipotetizan diferentes coeficientes de asociación: 0,20; 0,30; 0,40 y 0,50 respecto una variable registrada en la base (IMC).

2.3.2. Sesgo de Selección

Este sesgo fue estudiado utilizando como variables de exposición el Patrón Tradicional y el Cono Sur, respectivamente en los estudios de cáncer de mama y colorrectal. Los correspondientes análisis de sensibilidad determinísticos se realizaron utilizando distribuciones *a priori* de tipo trapezoidal, imponiendo distribuciones diferenciales a los grupos estudiados (según condición caso-control, expuesto-no expuesto). Los escenarios planteados al realizar los análisis fueron cuatro, uno no diferencial y tres diferenciales. En la Tabla 2.2 se presentan las probabilidades de participación trapezoidal asignadas *a priori* para el análisis determinístico de sesgo de selección para ambos estudios caso-control. En el primer caso, de tipo no diferencial, las probabilidades de participación de los cuatro grupos a comparar (casos expuestos, casos no expuestos, controles expuestos y controles no expuestos) fueron similares (probabilidades trapezoidales *a priori*: 0,7; 0,75; 0,85; 0,9). Luego, un primer y un segundo escenario diferencial presentaron mayores probabilidades de participación del grupo de los casos expuestos en comparación con el resto de los grupos. En el primer escenario diferencial planteado la participación del resto de los grupos fue baja (0,35; 0,4; 0,45; 0,5); y en el segundo caso fue moderada (0,55; 0,6; 0,65; 0,7) en comparación con el grupo de casos expuestos (0,7; 0,75; 0,85; 0,9). Se planteó un tercer posible escenario de participación diferencial donde los casos expuestos presentaron menor participación que el resto de los grupos, situación inversa a la ocurrida a lo planteado en el primer escenario.

Tabla 2.2. Probabilidades de participación trapezoidal asignadas *a priori* para el análisis determinístico de sesgo de selección para los estudios caso-control de cáncer de mama y colorrectal, según grupos de estudio.

	Grupos en estudio			
	Casos expuestos	Casos no expuestos	Controles expuestos	Controles no expuestos
Escenario no diferencial	0,7;0,75;	0,7;0,75;	0,7;0,75;	0,7;0,75;
	0,85; 0,9	0,85; 0,9	0,85; 0,9	0,85; 0,9
Escenarios diferenciales	0,7;0,75;	0,35;0,4;	0,35;0,4;	0,35;0,4;
	0,85; 0,9	0,45; 0,5	0,45; 0,5	0,45; 0,5
Escenario 1	0,7;0,75;	0,55;0,6;	0,55;0,6;	0,55;0,6;
	0,85; 0,9	0,65; 0,7	0,65; 0,7	0,65; 0,7
Escenario 2	0,35;0,4;	0,7;0,75;	0,7;0,75;	0,7;0,75;
	0,45; 0,5	0,85; 0,9	0,85; 0,9	0,85; 0,9
Escenario 3				

Por otro lado, se asignó *a priori* una distribución de probabilidad al factor de sesgo de selección. Se asumió una distribución log normal con media 0 y desvío estándar 0,21, que incluye el 95% de probabilidad *a priori* del factor de sesgo. En todos los análisis se presentan los percentiles 2,5; 50 y 97,5 de las distribuciones de probabilidades de las estimaciones de riesgo, que se utilizan como análogos de las estimaciones puntuales (la mediana) y los límites simulados del intervalo de confianza del 95% (percentiles 2,5 y 97,5).

2.3.3. Sesgo de Información

2.3.3.1. Sesgo de información derivado de datos faltantes

La metodología de imputación múltiple de datos faltantes en covariables se aplicó en ambos estudios caso-control. En ambos estudios, se llevaron a cabo modelos de regresión logística múltiple, siendo el outcome la presencia de la patología (cáncer de mama o colorrectal) y la exposición, el patrón dietario Tradicional y Cono Sur, respectivamente. En el siguiente capítulo se presentan los resultados derivados del análisis de datos completos y el análisis de imputación múltiple. Para este último análisis, se generaron 20 bases de datos imputadas (Sterne *et al*, 2009) y se realizó la imputación múltiple de datos en los casos donde

existía más del 10% de datos faltantes (Bennett, 2001). Para cada una de las variables imputadas se construyeron modelos de imputación considerando todas las variables incluidas en los análisis posteriores (regresión logística) y variables predictiva de los valores faltantes, en el caso que correspondiera. El modelo final seleccionado fue el más adecuado en base al resto de los modelos de imputación y al incremento medio de la varianza relativa (IVR) obtenido para cada uno (Acock, 2014). Además, se incluyeron gráficos que comparan las distribuciones de los valores imputados con los valores observados en el caso de imputación de la variable continua actividad física (Eddings y Marchenko, 2012).

En el caso del estudio de cáncer de mama, las covariables imputadas fueron Actividad Física y Nivel Socioeconómico. Estas variables tuvieron importantes porcentajes de datos faltantes debido principalmente a que fueron incluidas luego de haberse iniciado el estudio (Tabla 2.3, 2.4 y Apéndice 2). La Actividad Física fue tratada como variable continua, medida en equivalentes metabólicos semanales (METs). En cuanto al nivel socioeconómico, como se explicó en apartados anteriores se construye en base a 8 ítems (Asociación Argentina de Marketing, 2002). Cuando una de estas variables presenta datos faltante, el Nivel Socioeconómico también. Por esto, primero se imputaron estas variables y luego se realizó el cálculo final del Nivel Socioeconómico. Esta estrategia evita la pérdida de información y es la recomendada en estos casos (Eekhout *et al*, 2014). Seis de estas 8 variables fueron imputadas por tener más de 10% de datos faltantes: cantidad de aportantes en el hogar, autos, computadora, internet, tarjeta de débito. En total, en éste estudio se imputaron 7 variables.

Las Tablas 2.3 y 2.4 presentan las distribuciones de frecuencias de las variables de interés y la distribución de los datos faltantes. Trescientos dieciocho casos de cáncer de mama y 526 controles fueron incluidos en este estudio. Aproximadamente la mitad de las mujeres tuvieron más de 60 años. Más de un tercio presentaron elevada adherencia al Patrón Tradicional identificado en la población femenina y más de la mitad presentaron preobesidad u obesidad (51,3%). Respecto a la actividad física, 21,6% de las mujeres declaró ser sedentaria, aunque cabe señalar que el porcentaje de datos faltantes fue elevado en esta variable (29,1%). En relación al nivel educativo, la mitad de las mujeres no terminó la escuela secundaria, y un 29,5% realizó estudios superiores (universitario o terciario, completos o no) (Tabla 2.3).

En cuanto a las variables ginecológicas, cerca del 80% de las mujeres tuvo hijos, 61,7% amamantó, 69,8% declaró ser menopáusica al momento del diagnóstico (o entrevista los controles) y un 17,4% declararon haber tenido menos de 12 años al momento de la menarca.

Respecto a las variables socioeconómicas, alrededor de un 68% declaró tener obra social y un 24,3% fueron datos faltantes. Aproximadamente la mitad de las mujeres entrevistadas tenían computadora, internet y tarjeta de débito y un 43% no tenía auto. Todas estas variables presentaron un 10% de datos faltantes (Tabla 2.4, información sobre el principal sostén del hogar se presenta en Apéndice 2).

En general, la mayoría de las distribuciones de las variables con datos faltantes difieren entre las categorías de las variables incluidas en el análisis, aunque no todas estas diferencias son estadísticamente significativas. Por ejemplo, las mujeres sedentarias presentaron 52,7% de datos faltantes en Obra Social, comparado con el 10,7% de *missing* en la categoría superior de Actividad Física. Las distribuciones también fueron estadísticamente significativas respecto a las categorías de las variables Nivel de Educación, Hijos y Lactancia Materna. También se observan patrones similares de datos faltantes en el resto de variables socioeconómicas. Los porcentajes de datos faltantes parecen declinar a mayor nivel educativo de las participantes. Cuando se considera el Patrón Tradicional, los porcentajes de *missing* se elevan a mayor adherencia al patrón. Los patrones de datos faltantes parecen presentar un mecanismo de tipo MAR (es decir, de manera aleatoria).

Tabla 2.3. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas del resultado, la exposición y otras covariables, estudio caso-control de cáncer de mama (CM), Córdoba, Argentina 2008-2015.

	n	%	% de observaciones con datos faltantes						
Total	44	00							
			Actividad Física	Obra social	N° de proveedores	Computadora	Internet	Tarjeta Débito	Autos
Presencia CM									
No	526	62.3	29.7	26.05	11.22	11.22	11.22	11.2	11.2
Si	318	37.7	28.3	21.38	8.81	8.81	8.81	8.8	9.4
Patrón Tradicional									
Tecil 1	245	29.0	31.0	24.1	6.9*	6.9*	6.9*	6.9*	7.3*
Tercil 2	277	32.8	28.5	24.2	10.5	10.1	10.1	10.1	10.1
Tercil 3	322	38.2	28.4	24.5	12.7	13.0	13.0	13.0	13.3
Edad									
<45 años	137	16.2	26.3	21.9	9.5	9.5	9.5	9.5	10.2
45-60 años	307	36.4	24.4	26.4	11.4	11.4	11.4	11.4	11.4
>60 años	400	47.4	33.7	23.5	9.7	9.7	9.7	9.75	10

IMC									
<25kg/mt ²	400	47.4	29	27.2	8.5*	8.7*	8.7*	8.75*	8.75*
25-30kg/mt ²	268	31.7	27.9	22.0	9.3	9.3	9.3	9.33	9.33
>30kg/mt ²	165	19.6	31.5	22.4	15.8	15.1	15.1	15.1	16.4
Desconocido	11	1.3	27.3	0	18.2	18.2	18.2	18.2	18.2
Actividad Física									
Sedentario	182	21.6	—	52.7*	4.9*	4.4*	4.4*	4.4*	4.9*
Moderado	228	27.0	—	28.1	8.8	8.8	8.8	8.8	8.8
Vigorosa	188	22.3	—	10.6	23.9	24.5	24.5	24.5	25
Desconocido	246	29.1	—	10.2	5.3	5.3	5.3	5.3	5.3
Educación									
Sin Estudios	5	0.6	80*	0*	20*	20*	20*	20*	20*
Primaria Incompleta	63	7.5	33.3	0	17.6	15.9	15.9	15.9	17.9
Primaria Completa	281	33.3	22.1	46.6	9.2	9.2	9.2	9.2	9.2
Secundaria Incompleta	79	9.4	34.2	0	8.9	10.1	10.1	10.1	10.1
Secundaria Completa	145	17.2	21.4	19.3	10.3	10.3	10.3	10.3	10.3
Educación Superior	249	29.5	36.9	13.6	6.0	6.0	6.0	6.0	6.4
Desconocido	22	2.6	40.9	54.5	54.5	54.5	54.5	54.5	54.5

*Diferencias estadísticamente significativas (p<0,05).

**Variables sombreadas con más de 10% de missing.

Tabla 2.4. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas de variables ginecológicas y socioeconómicas, estudio caso-control de cáncer de mama, Córdoba, Argentina 2008-2015

	n	%	% de observaciones con datos faltantes						
Total	844	100							
			Act. Física	Obra Social	N° de proveedores	Computadora	Internet	Tarjeta Débito	Auto
Hijos									
No	120	14.2	32.5*	24.2*	10.83	10.83	10.83	10.83	11.67
Si	666	78.9	29.9	26.3	10.36	10.36	10.36	10.36	10.51
Desconocido	58	6.9	13.8	1.7	8.62	8.62	8.62	8.62	8.62

Lactancia Materna

No	266	31.5	31.2	21.8*	7.52	7.89	7.89	7.89	8.27
Si	521	61.7	29.37	28.2	11.71	11.52	11.52	11.52	11.71
Desconocido	57	6.8	17.54	0	10.53	10.53	10.53	10.53	10.53

Estado**ginecológico**

No	201	23.8	25.37*	24.9	9.45	9.45	9.45	9.45	9.45
Si	589	69.8	31.92	26.3	11.38	11.38	11.38	11.38	11.54
Desconocido	54	6.4	12.96	0	1.85	1.85	1.85	1.85	3.7

Edad de la Menarca

<12 edad	147	17.4	29.93	22.4	11.56	11.56	11.56	11.56	11.56*
>=12 edad	676	80.1	28.99	25	9.62	9.62	9.62	9.62	9.76
Desconocido	21	2.5	28.57	14.3	23.81	23.81	23.81	23.81	28.57

Obra social

No	68	8.1	33.82*	–	14.71*	14.7*	14.7*	14.7*	14.7*
Si	571	67.6	34.68	–	11.38	11.4	11.4	11.4	11.4
Desconocido	205	24.3	12.2	–	5.85	5.8	5.8	5.8	5.8

N° de proveedores

Uno	364	43.1	34.07*	23.1*	0	0*	0*	0*	0.5*
Dos o tres	379	44.9	27.97	28.0	0	0.3	0.3	0.3	0.3
Más de tres	14	1.7	21.43	21.4	0	0	0	0	0
Desconocido	87	10.3	14.94	13.8	100	98.8	98.8	98.8	98.8

Computadora

No	293	34.7	27.99*	32.8*	0*	–	0	0	0.3*
Si	464	55.0	32.54	20.9	0.22	–	0	0	0.2
Desconocido	87	10.3	14.94	13.8	98.85	–	100	100	100

Internet

No	346	41	27.46*	32.7*	0.29*	0	–	0	0.6*
Si	411	48.7	33.58	19.5	0	0	–	0	0
Desconocido	87	10.3	14.94	13.8	98.85	100	–	100	100

Tarjeta de**débito**

No	354	41.9	28.81*	26.0*	0.28*	0	0	–	0.6*
Si	403	47.8	32.51	25.1	0	0	0	–	0
Desconocido	87	10.3	14.94	13.8	98.85	100	100	–	100

Autos

Ninguno	361	42.8	35.18*	19.7*	0.3*	0	0	0	–
---------	-----	------	--------	-------	------	---	---	---	---

Uno	332	39.3	26.51	30.4	0	0	0	0	–
Dos	58	6.9	29.31	36.2	0	0	0	0	–
Tres o más	4	0.5	25	0	0	0	0	0	–
Desconocido	89	10.5	14.6	13.5	96.6	97.7	97.7	97.7	–

*Diferencias estadísticamente significativas ($p < 0,05$).

**Variables sombreadas con más de 10% de missing.

El estudio de cáncer de mama, presentaba una base con aproximadamente un 10% de datos faltantes en determinadas variables, no así el estudio caso-control de cáncer colorrectal, como se describe a continuación.

En el caso del estudio caso-control de cáncer colorrectal, la variable imputada fue el Nivel Socioeconómico. Similar a lo ocurrido en el estudio de cáncer de mama, ésta variable fue incluida luego de iniciado el estudio, por lo que presentaba elevados porcentajes de datos faltantes (Tabla 2.5, 2.6 y Apéndice 3). De las 8 variables o ítems necesarios para construir el Nivel Socioeconómico, se imputaron las que tuvieron más del 10% de datos faltantes, en este caso fueron 6, a saber, cantidad de aportantes en el hogar, autos, computadora, internet, tarjeta de débito, obra social. En este caso, las variables que se imputaron presentaban elevados porcentajes de *missing* (40% aproximadamente), como se presenta en las Tablas 2.5 y 2.6. Para llevar a cabo los análisis de imputación múltiple de datos se decide trabajar con la mitad de la información incompleta. Así, a través de la extracción de una muestra aleatoria de éstos, del 50%, la base con sujetos con datos completos sumados a los de la muestra con *missing* quedó constituida en tanto 391 sujetos (125 casos y 266 controles).

En las Tablas 2.5 y 2.6 se presentan las distribuciones de frecuencias de las variables de estudio y los porcentajes de datos faltantes. Ciento sesenta y un casos de cáncer colorrectal y 329 controles pertenecían al estudio. La mayoría de los sujetos tuvieron más de 60 años y un poco más de la mitad fueron hombres (53,1%). Más de un tercio presentaron elevada adherencia al Patrón Cono Sur y más de la mitad presentaron preobesidad u obesidad (62,9%). Respecto a la actividad física, 24,5% de las personas declaró ser sedentaria (Tabla 2.5).

En relación al nivel educativo, la mayoría (58,8%) no había concluido la escuela secundaria. Respecto a las variables socioeconómicas, alrededor de un 86,5% declaró tener obra social. Aproximadamente un tercio de los sujetos tenían computadora, internet y tarjeta de débito y un 43% no tenía auto. Todas estas variables presentaron un 40% de datos

faltantes (Tabla 2.6, información sobre el principal sostén del hogar se presenta en Apéndice 3).

Respecto a las distribuciones de los datos faltantes según categorías de las variables incluidas en los análisis, éstas presentaron diferencias en la mayoría de los casos, aunque no todas significativas. Por ejemplo, la distribución de datos faltantes de fue diferentes entre categorías de Nivel de Actividad Física. A mayor IMC, mayor fue el porcentaje de datos faltantes en las variables con *missing*. También se observa en la variable nivel de educación que en la categoría de estudios superiores el porcentaje de *missing* fue menor que el resto de las categorías. Similar a lo ocurrido en el estudio de cáncer de mama, los patrones de datos faltantes estarían indicando que el mecanismo por el cual se manifiestan los *missing* en esta base de datos no sería de tipo MNAR (es decir, que los datos faltan de manera aleatoria).

Tabla 2.5. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas del resultado, la exposición y otras covariables, estudio caso-control de cáncer colorrectal (CCR), Córdoba, Argentina 2010-2016.

	n	%	% de observaciones con datos faltantes				
Total	490	100					
Presencia CCR			N° de proveedores	Computadora	Internet	Tarjeta Débito	Autos
No	329	67,1	39,21	40,73	39,51	40,73	40,12
Si	161	32,9	39,75	39,75	39,75	39,75	39,75
Patrón Cono Sur							
Tecil 1	144	29,5	38,89	38,89	38,89	38,89	38,89
Tercil 2	162	33,1	41,98	44,4	41,98	44,4	43,21
Tercil 3	184	37,4	37,7	38,25	38,25	38,25	38,25
Sexo							
Femenino	230	46,9	39,57	40,43	39,57	40,43	40,0
Masculino	260	53,1	39,23	40,38	39,62	40,38	40,0
Edad							
<45 años	49	10,0	30,61	32,65	30,61	32,65	32,65
45-60 años	117	23,9	41,17	41,03	40,17	41,03	41,03
>60 años	324	66,1	40,43	41,36	40,74	41,36	40,74
IMC							
<25kg/mt ²	180	36,7	33,33	33,89	33,33	33,89	33,89
25-30kg/mt ²	193	39,4	36,27	37,31	36,27	37,31	36,79
>30kg/mt ²	115	23,5	54,78	56,52	55,65	56,52	55,65
Desconocido	2	0,4	0,0	0,0	0,0	0,0	0,0
Antecedentes CCR							
No	456	93,1	39,25	40,13	39,47	40,13	39,91
Si	34	6,9	41,18	44,12	41,18	44,12	41,18
Actividad Física							

Sedentario	120	24,5	15,83	16,67	15,83	16,67	16,67
Moderado	237	48,4	63,29	64,14	63,29	64,14	63,29
Vigorosa	97	19,8	23,71	25,77	24,74	25,77	24,74
Desconocido	36	7,3	2,78	2,78	2,78	2,78	5,56

Tabla 2.6. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas de variables socioeconómicas, estudio caso-control de cáncer de mama, Córdoba, Argentina 2010-2016.

	n	%	% de observaciones con datos faltantes				
Total	490	100					
Educación			N° de proveedores	Computadora	Internet	Tarjeta Débito	Auto
Sin Estudios	4	0,8	75,0	75,0	75,0	75,0	75,0
Primaria Incompleta	8	13,9	39,71	41,18	41,18	41,18	39,71
Primaria Completa	28	26,1	46,09	46,88	46,09	46,88	46,09
Secundaria Incompleta	8	18,0	40,91	42,05	40,91	42,05	43,18
Secundaria Completa	2	18,8	40,30	42,39	41,30	42,39	42,39
Educación Superior	03	21,0	23,30	24,27	23,30	24,27	23,30
Desconocido	22	1,4	85,71	85,71	85,71	85,71	85,71
Obra social							
No	6	11,5	51,79	53,57	51,79	53,57	55,36
Si	24	86,5	36,79	37,74	37,03	37,74	37,03
Desconocido	10	2,0	80,0	80,0	80,0	80,0	80,0
N° de proveedores							
Uno	33	27,1	–	2,26	0	2,26	0,75
Dos o tres	61	32,9	–	1,24	0,62	1,24	1,24
Más de tres		0,6	–	0	0	0	0
Desconocido	193	39,4	–	100	100	100	100
Computadora							
No	23	25,1	0	–	0	0	0,81
Si	69	34,5	0	–	0	0	0
Desconocido	198	40,4	97,47	–	100	100	98,48
Internet							
No	48	30,2	0	0	–	0	0,68
Si	48	30,2	0	2,70	–	2,7	1,35
Desconocido	194	39,6	99,48	100	–	100	99,48
Tarjeta de débito							

No	56	31,8	0	0	0	–	0,64
Si	36	27,8	0	0	0	–	0
Desconocido	198	40,4	97,47	100	100	–	98,48
Autos							
Ninguno	38	42.8	0	0	0	0	–
Uno	44	39.3	0	2,08	0,69	2,08	–
Dos	2	29,4	0	0	0	0	–
Tres o más		0,0	0	0	0	0	–
Desconocido	196	40,0	98,47	99,49	98,47	99,49	–

*Diferencias estadísticamente significativas ($p < 0,05$).

**Variables sombreadas con más de 10% de missing.

2.3.3.2 Sesgo de información por error de clasificación

El sesgo de información derivado del error de clasificación se evaluó mediante el estudio de validación interno en cáncer de mama. Las variables de exposición evaluadas fueron el consumo de carnes procesadas y el consumo de productos de pastelería, identificadas con mayores cargas del primer factor en el Análisis Factorial.

Para realizar las correspondientes comparaciones de los sistemas de clasificación derivados de los R24H (*Gold Standard*) y FFQ las variables de exposición (consumo de carnes procesadas y el consumo de productos de pastelería de los R24H) fueron categorizadas en base a terciles de consumo. Se consideró expuesto al sujeto que tenía un consumo superior al primer tercil de consumo del grupo control (es decir, consumo alto-medio) y no expuesto cuando el consumo es menor a este (consumo bajo).

Se llevaron a cabo los análisis de sensibilidad para obtener los parámetros de sensibilidad, especificidad, valores predictivos positivos y negativos (Apéndice 4). Posteriormente, en base a éstos parámetros obtenidos del estudio de validación, se realizaron correcciones por error de clasificación no diferencial y diferencial. Por último, se realizaron simulaciones de diferentes escenarios donde los parámetros adoptan valores distintos para probar posibles situaciones donde exista error de clasificación diferencial entre los grupos de estudio a comparar, es decir, entre casos y controles.

CAPÍTULO 3. RESULTADOS

Este apartado presenta los resultados obtenidos luego de aplicarse los análisis ya descriptos para afrontar posibles errores sistemáticos en los estudios caso-control conducidos por el GEACC, en cáncer de mama primero y, colorrectal luego. En una primera instancia, se presentan los resultados derivados del análisis del sesgo por fenómeno de confusión empleando el método de Mantel-Haenszel, seguido por el método *stepwise*, el combinado y, por último, el análisis para variables confundentes no observadas propuesto. Luego, se presentan resultados derivados del análisis del sesgo de selección. En último lugar, se exponen los resultados sobre los análisis de sesgo de información por falta de datos utilizando el método de imputación múltiple de datos y, finalmente, los resultados derivados del análisis de error de clasificación de la exposición.

3.1. Análisis de sesgos por fenómeno de confusión

Estudio caso-control de cáncer de mama

La Tabla 3.1 presenta las estimaciones de riesgo entre la adherencia al patrón dietario Tradicional y la presencia de cáncer de mama. La estimación cruda indica mayor riesgo de ocurrencia de la patología a mayor adherencia al patrón Tradicional. Las estimaciones de riesgo obtenidas en los diferentes estratos de la variable edad propuesta como posible confundente, muestran dirección similar al OR global. No existe evidencia de que la variable edad está confundiendo la asociación entre la exposición de interés y la patología.

Tabla 3.1. Estimación de las medidas de asociación (ORs) y sus intervalos de confianza (95%) entre la adherencia al patrón dietario Tradicional (PT) y cáncer mama. Análisis crudo y estratificado por método Mantel-Haenszel. Estudio caso-control, provincia de Córdoba. Año 2008-2015.

Análisis crudo		Controles (n)	Casos (n)	OR (IC95%)	Valor p
Adherencia al PT					
Baja-Media		351	171	1	
Alta		175	147	1,72(1,296-2,293)	<0,001
Totales		526	318		
Análisis estratificado por Edad					
Edad	Adherencia al PT	Controles (n)	Casos (n)	OR (IC95%)	Valor p
<45años	Baja-Media	49	29	1	
	Alta	34	25	1,24(0,620-2,488)	0,539
≥45años	Baja-Media	302	142	1	
	Alta	141	122	1,84(1,339-2,527)	0,001
Totales		526	318		
Análisis ponderado por Mantel-Haenszel (ajustado por Edad)					
Adherencia al PT				OR(IC)	Valor p
Baja-Media				1	
Alta				1,71(1,289-2,292)	0,002

Los resultados obtenidos a partir del método *Stepwise* se presentan en la Tabla 3.2. Se incluyen todas las posibles confundentes para cáncer de mama: edad, nivel socioeconómico, IMC, actividad física, hijos, estatus ginecológico y lactancia materna. La tabla presenta las estimaciones de riesgo y los porcentajes de cambio de los mismos, incluyendo aditivamente, una por una a dichas variables. El orden queda determinado según la magnitud de modificación de la estimación de cada variable incorporada, incluyendo primero la variable con mayor efecto hasta la de menor impacto en los OR. En este caso, las variables lactancia materna e hijos fueron las de mayor porcentaje de cambio ejercido en las estimaciones. La primera, aumentando la magnitud del OR un 2% y la segunda disminuyéndolo. La Figura 3.1 demuestra de manera gráfica la situación explicada.

Tabla 3.2. Estimación de las medidas de asociación (ORs) y sus intervalos de confianza (95%) entre la adherencia al patrón dietario Tradicional y cáncer de mama. Análisis crudo y ajustado por posibles confundentes por método *Stepwise*. Estudio caso-control, provincia de Córdoba. Año 2008-2015.

Variables Incorporadas	Odds Ratio	IC 95%	% Cambio
Crudo	1,14	0,689-1,913	-
Lactancia Materna	1,17	0,706-1,971	2,67%
Hijos	1,15	0,687-1,929	-2,39%
Nivel socioeconómico	1,17	0,698 -1,968	1,81%
Edad de la menarca	1,5	0,684-1,938	-1,75%
IMC	1,16	0,692-1,970	1,44%
Actividades Física	1,17	0,697-1,985	0,69%
Estado ginecológico	1,16	0,691-1,974	-0,68%
Edad	1,16	0,689-1,973	-0,19%

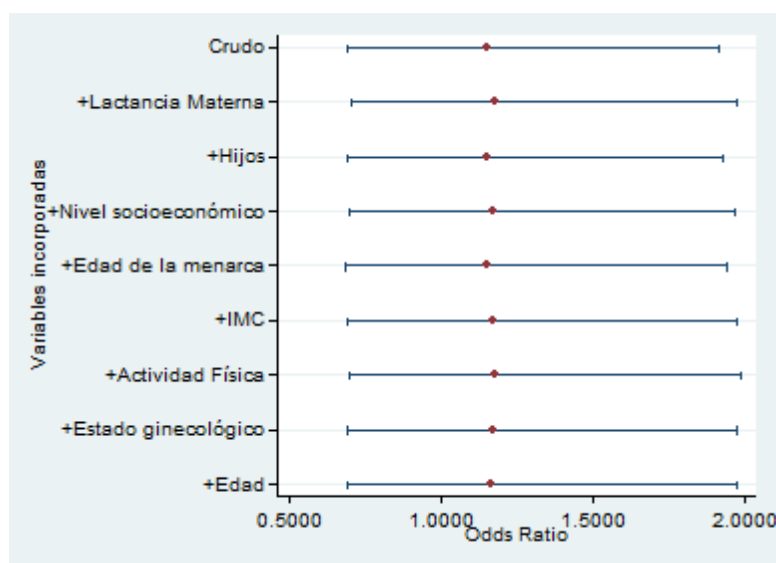


Figura 3.1. Estimación de las medidas de asociación (ORs) entre la adherencia al patrón dietario Tradicional y cáncer de mama. Análisis crudo y ajustado por posibles confundentes por método *Stepwise*. Estudio caso-control, provincia de Córdoba. Año 2008-2015.

Respecto al análisis combinado, la Figura 3.2 muestra todas las posibles estimaciones de riesgo con sus respectivos valores p que surgen al combinar en el modelo de regresión logística todas las variables confundentes. Se realizaron 512 set de combinaciones entre confundentes por lo que se graficaron igual cantidad de OR, incluido el OR crudo o estimación inicial del modelo. Este gráfico permite observar que la mayoría de las estimaciones de riesgo

calculadas fueron significativas ($p < 0,05$), los modelos que tuvieron mejor desempeño fueron aquellos de menor valor de los estadísticos BIC (Criterio de Información Bayesiana) y AIC (Criterio de Información de Akaike), situándose en el extremo superior izquierdo del gráfico, con mayores valores de OR y menores valores p .

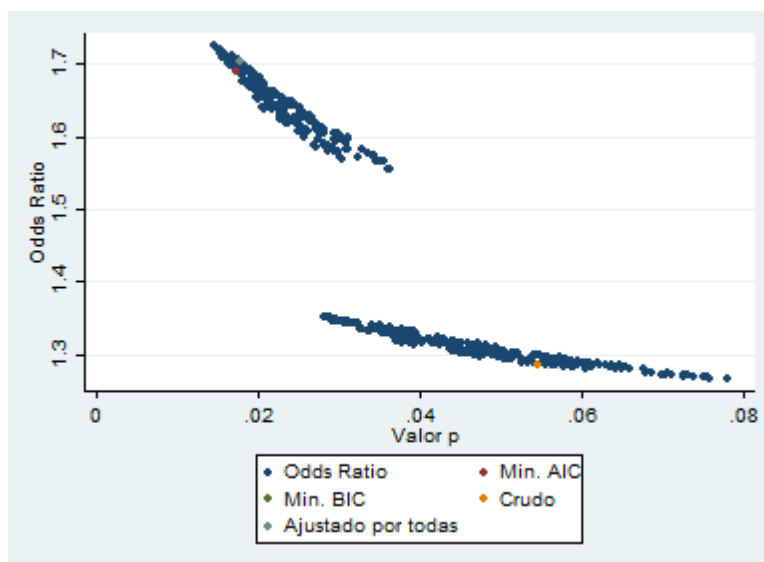


Figura 3.2. Diagrama de dispersión entre las medidas de asociación (ORs) y los valores p crudos y ajustados. Análisis crudo y ajustado por posibles confundentes por método combinado. Estudio caso-control de cáncer de mama, provincia de Córdoba. Año 2008-2015.

En cuanto al análisis planteado para evaluar el efecto de una variable confundente no registrada, se puede observar en la Tabla 3.3 las estimaciones de riesgo y los valores p obtenidos en los diferentes escenarios planteados, según los diferentes desvíos estándar (0, 1, 2 y 3) y coeficientes de asociación con la variable IMC (0,20; 0,30; 0,40 y 0,50) impuestos *a priori*. Se observa que todos los posibles escenarios con diferente variabilidad y coeficientes de asociación con la variable registrada son estadísticamente significativos. Los OR estimados aumentan en magnitud a medida que aumenta la variabilidad impuesta y el coeficiente de asociación, llegando a un aumento máximo aproximado del 43% del OR. En la Figura 3.3 se presentan gráficamente las estimaciones de riesgo obtenidas en cada situación propuesta. En el primer gráfico de línea (A), se observa el modelo básico, donde no hay heterogeneidad en la variable no observada, es decir, no se propuso variabilidad, por lo que las estimaciones de riesgo se mantuvieron constantes cuando se modificaron los coeficientes de asociación. En el resto de los casos, donde se asignaron desvíos estándar de 1, 2 o 3 (gráficos B, C y D

respectivamente) se observa que los ORs aumentan conforme la variabilidad impuesta y a mayor coeficiente de asociación propuesto, en prácticamente todas las situaciones.

Tabla 3.3. Estimación de las medidas de asociación (ORs) entre la adherencia al patrón dietario Tradicional (continuo) y cáncer de mama. Simulación de diferentes escenarios de riesgo con asignación de variabilidad *a priori* de la variable confundente no observada. Estudio caso-control, provincia de Córdoba. Año 2008-2015.

	Asoc. Variable Observada (IMC)	Odds Ratio	Valor p
DE=0	0.2	1,31	0,048
	0.3	1,31	0,048
	0.4	1,31	0,048
	0.5	1,31	0,048
DE=1	0.2	1,40	0,038
	0.3	1,40	0,034
	0.4	1,41	0,031
	0.5	1,41	0,027
DE=2	0.2	1,59	0,034
	0.3	1,60	0,029
	0.4	1,41	0,031
	0.5	1,61	0,020
DE=3	0.2	1,85	0,033
	0.3	1,87	0,027
	0.4	1,88	0,022
	0.5	1,88	0,017

DE: desvío estándar

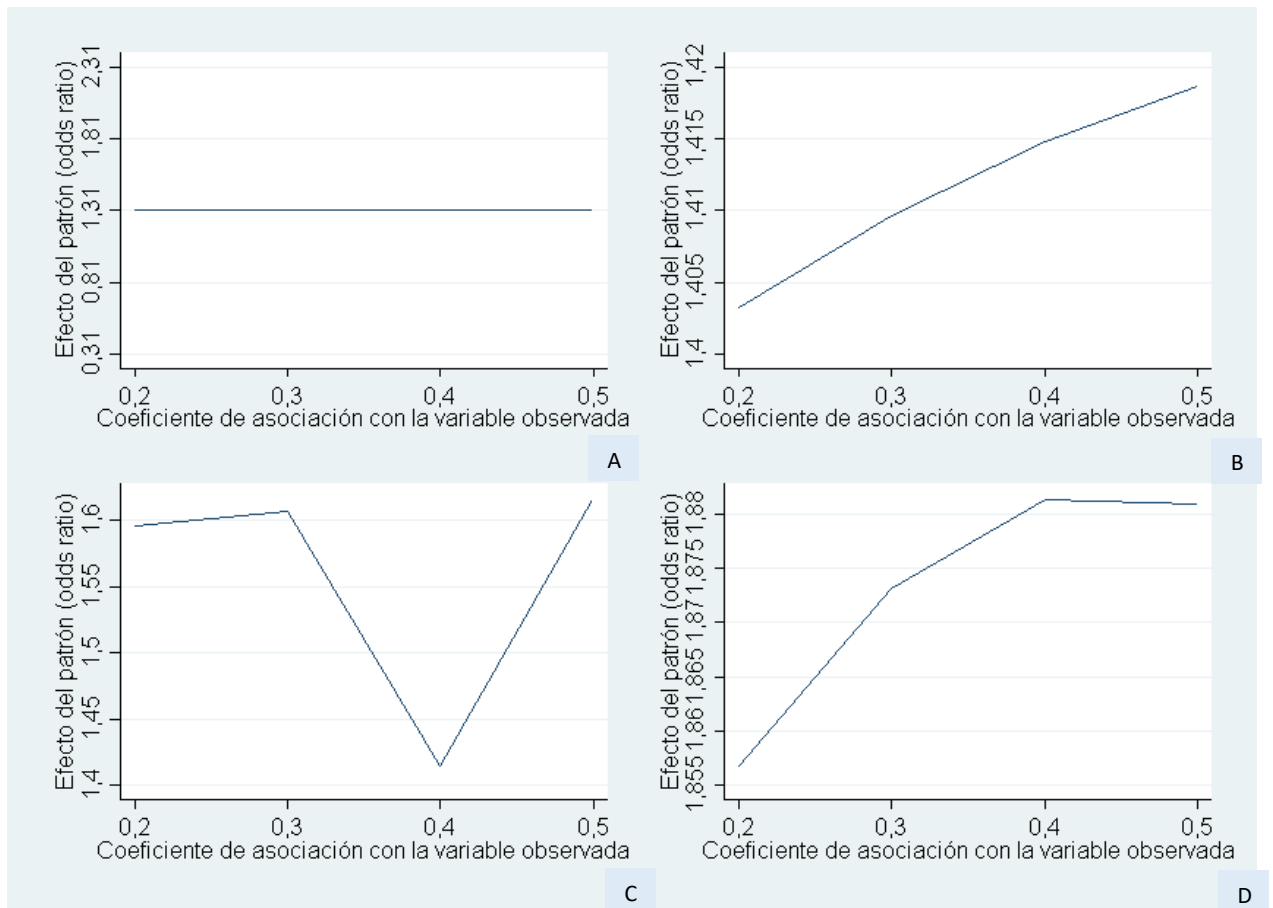


Figura 3.3. Tendencias entre el efecto del patrón dietario Tradicional y el correspondiente riesgo de presentar cáncer de mama. Estimación de las medidas de asociación (ORs) entre la adherencia al patrón dietario Tradicional (continuo) y cáncer de mama. Simulación de diferentes escenarios (A: Desvío estándar=0; B: Desvío estándar=1; C: Desvío estándar=2; D: Desvío estándar=3) de riesgo con asignación de variabilidad *a priori* de la variable confundente no observada. Estudio caso-control, provincia de Córdoba. Año 2008-2015.

Estudio caso-control de cáncer colorrectal

El análisis exploratorio que se presenta a continuación muestra los métodos propuestos para evaluar el fenómeno de confusión en el estudio caso-control de cáncer colorrectal en la provincia de Córdoba. En la Tabla 3.4 se presentan las estimaciones de riesgo entre la adherencia al patrón dietario Cono Sur y la ocurrencia del cáncer colorrectal. Se observa la estimación cruda de la asociación que estaría indicando un mayor riesgo de presentar la patología a mayor adherencia al patrón dietario Cono Sur. En cuanto al análisis estratificado por la confundente propuesta, el Nivel socioeconómico (NSE), las estimaciones de riesgo para cada NSE presentan similar dirección, siendo el OR del NSE medio-alto de

mayor magnitud que el del nivel bajo. Respecto al NSE bajo, se observó un OR de 1,41 indicando un 40% más de chance de presentar la patología de las personas con mayor adherencia al patrón Cono Sur respecto a los sujetos de menor adherencia. Por su parte, los sujetos pertenecientes al NSE medio-alto con una adherencia alta a este patrón, presentaron un OR de 2,31 respecto a los de menor adherencia al patrón. Además, las estimaciones obtenidas mediante el método de Mantel- Haenszel comparadas con los análisis crudos difieren en un 10% aproximadamente (OR crudo=1,71; OR ajustado=1,89). Este análisis permitió observar el efecto confundente de la variable NSE en la asociación entre Patrón Cono Sur y cáncer colorrectal.

Tabla 3.4. Estimación de las medidas de asociación (ORs) y sus intervalos de confianza (95%) entre la adherencia al patrón dietario Cono Sur (PCS) y cáncer colorrectal. Análisis crudo y estratificado por método Mantel-Haenszel. Estudio caso-control, provincia de Córdoba. Año 2010-2016.

Análisis crudo		Controles (n)	Casos (n)	OR (IC95%)	Valor p
Adherencia al PCS					
Baja-Media		220	87	1	
Alta		109	74	1,71(1,167-2,524)	0,006
Totales		329	161		
Análisis estratificado por NSE					
NSE	Adherencia al PCS	Controles (n)	Casos (n)	OR (IC95%)	Valor p
Bajo	Baja-Media	34	24	1	
	Alta	21	21	1,41(0,632-3,172)	0,394
Medio-Alto	Baja-Media	95	26	1	
	Alta	41	26	2,31(1,187-4,522)	0,011
Desconocido [†]	Baja-Media	91	37		
	Alta	47	27		
Totales		329	161		
Análisis ponderado por Mantel-Haenszel (ajustado por NSE)					
Adherencia al PCS				OR(IC)	Valor p
Baja-Media				1	
Alta				1,89(1,135-3,156)	0,012

[†]Por ser desconocido este dato, no se reporta este OR.

En la Tabla 3.5 se presentan los resultados del método *Stepwise*, el cual parte de un modelo de regresión logística múltiple en donde se incluyen todas las posibles confundentes

de interés: edad, sexo, nivel socioeconómico, IMC, actividad física y antecedentes primarios de cáncer colorrectal. La tabla presenta las estimaciones de riesgo y los porcentajes de cambio de los mismos, incluyendo de manera aditiva una a una las variables mencionadas. El orden de incorporación de cada variable al modelo se determinada según la magnitud del porcentaje de cambio ejercido en la estimación por cada variable incorporada. El orden fue descendente, presentándose primero la variable con mayor efecto en el OR hasta la de menor impacto. En este caso, el sexo y el NSE son las que mayor porcentaje de cambio ejercen en las estimaciones. En la Tabla 3.5 se observa que al incluir la variable sexo en el modelo crudo se provoca una modificación del OR de aproximadamente un 34% y al incorporar el NSE se modificó un 10%. La representación gráfica de la modificación de efecto se presenta en la Figura 3.4.

Tabla 3.5. Estimación de las medidas de asociación (ORs) y sus intervalos de confianza (95%) entre la adherencia al patrón dietario Cono Sur y cáncer colorrectal. Análisis crudo y ajustado por posibles confundentes por método *Stepwise*. Estudio caso-control, provincia de Córdoba. Año 2010-2016.

Variables Incorporadas	Odds Ratio	IC 95%	% Cambio
Crudo	1,72	1,017-2,918	-
Sexo	2,30	1,208-4,399	33,78%
Nivel socioeconómico	2,08	1,077-4,030	-9,61%
Antec. 1º CCR	2,00	1,031-3,911	-3,61%
IMC	1,94	0,993-3,821	-3,01%
Actividades Física	1,90	0,968-3,750	-2,16%
Edad	1,89	0,960-3,745	-0,49%

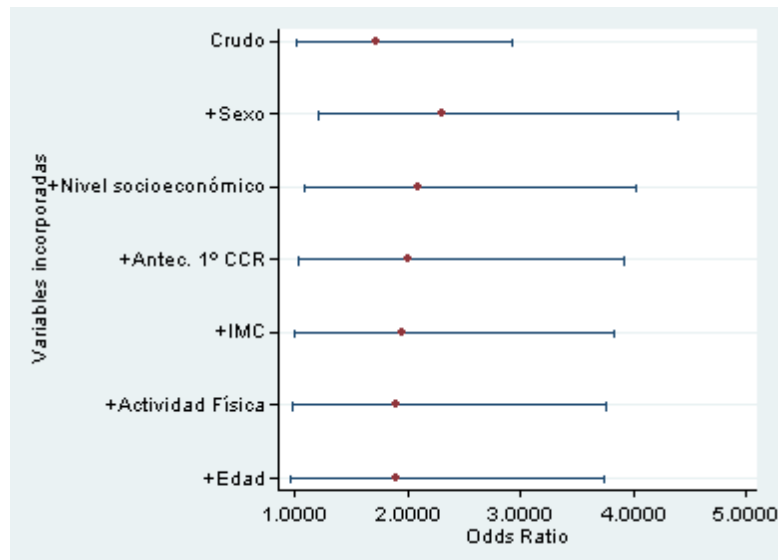


Figura 3.4. Estimación de las medidas de asociación (ORs) entre la adherencia al patrón dietario Cono Sur y cáncer colorrectal. Análisis crudo y ajustado por posibles confundentes por método *Stepwise*. Estudio caso-control, provincia de Córdoba. Año 2010-2016.

En cuanto al análisis combinado, la Figura 3.5 presenta gráficamente las estimaciones de riesgo puntuales derivadas del modelo de regresión logística al combinar todas las variables confundentes del modelo, junto con los respectivos valores p. En este caso, se combinaron en 64 set de confundentes por lo que se graficaron 64 OR junto con el OR crudo inicial del modelo, observándose que la magnitud del mismo se modifica de manera modesta y que todos los modelos fueron significativos. La estimación puntual de OR donde se ajustó por todas las variables incluidas en el modelo, se encuentra en el extremo superior izquierdo del gráfico, cercana a la estimación puntual con mejor desempeño según el estadístico AIC.

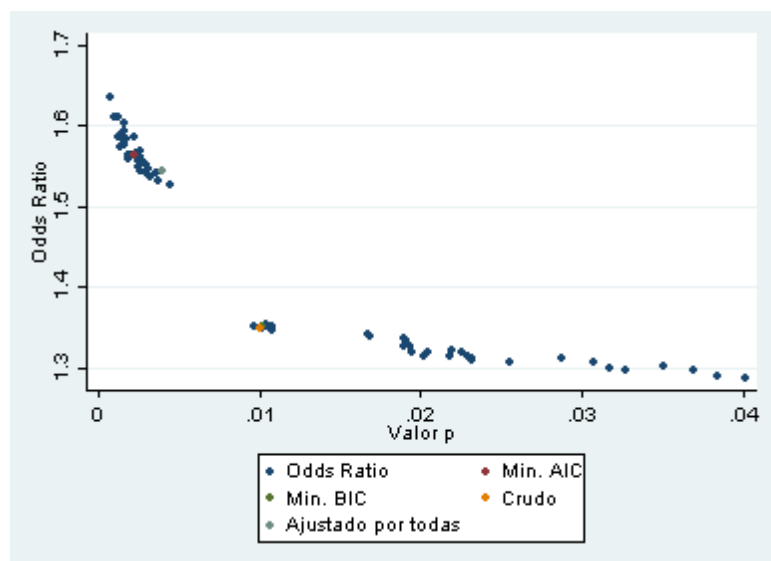


Figura 3.5. Diagrama de dispersión entre las medidas de asociación (ORs) y los valores p crudos y ajustados. Análisis crudo y ajustado por posibles confundentes por método combinado. Estudio caso-control de cáncer colorrectal, provincia de Córdoba. Año 2010-2016.

La Tabla 3.6 presenta los resultados del análisis de la variable confundente no observada. Se obtuvieron las estimaciones de riesgo y los valores p, según los diferentes desvíos estándar (0, 1, 2 y 3) y coeficientes de asociación con la variable IMC (0,20; 0,30; 0,40 y 0,50) impuestos *a priori*. Se observa que todos los posibles escenarios con diferente variabilidad y coeficientes de asociación con la variable registrada son estadísticamente significativos, similar a lo ocurrido en el caso del análisis en el estudio de cáncer de mama. Los OR estimados van aumentando en magnitud a medida que aumenta la variabilidad impuesta y el coeficiente de asociación. La modificación máxima observada del OR en los escenarios planteados alcanzó un 70% de diferencia comparado con el OR planteado sin variabilidad (OR=1,44). En la Figura 3.6 se presentan gráficamente las estimaciones de riesgo obtenidas en cada situación propuesta. El modelo básico se presenta en el primer gráfico de línea (A), allí se propone una variabilidad nula para la confundente no observada; por ello, las estimaciones de riesgo no presentan modificaciones al variar los coeficientes de asociación. En los gráficos B, C y D se presentan las estimaciones de riesgo obtenidas al imponer desvíos estándar de 1, 2 y 3 respectivamente a la variable no observada. En todos los casos los valores de OR son superiores al modelo básico, observándose, en la mayoría de los casos, mayor riesgo a mayor coeficiente de asociación con la variable registrada.

Tabla 3.6. Estimación de las medidas de asociación (ORs) entre la adherencia al patrón dietario Cono Sur (continuo) y cáncer colorrectal. Simulación de diferentes escenarios de riesgo con asignación de variabilidad *a priori* de la variable confundente no observada. Estudio caso-control, provincia de Córdoba. Año 2010-2016.

	Asoc. Variable Observada (IMC)	Odds Ratio	Valor p
DE=0	0.2	1,44	0,01
	0.3	1,44	0,01
	0.4	1,44	0,01
	0.5	1,44	0,01

DE=1	0.2	1,58	0,006
	0.3	1,60	0,005
	0.4	1,62	0,004
	0.5	1,64	0,003
DE=2	0.2	1,88	0,005
	0.3	1,91	0,003
	0.4	1,62	0,004
	0.5	1,98	0,001
DE=3	0.2	2,30	0,004
	0.3	2,37	0,002
	0.4	2,42	0,001
	0.5	2,47	0,001

DE: desvío estándar

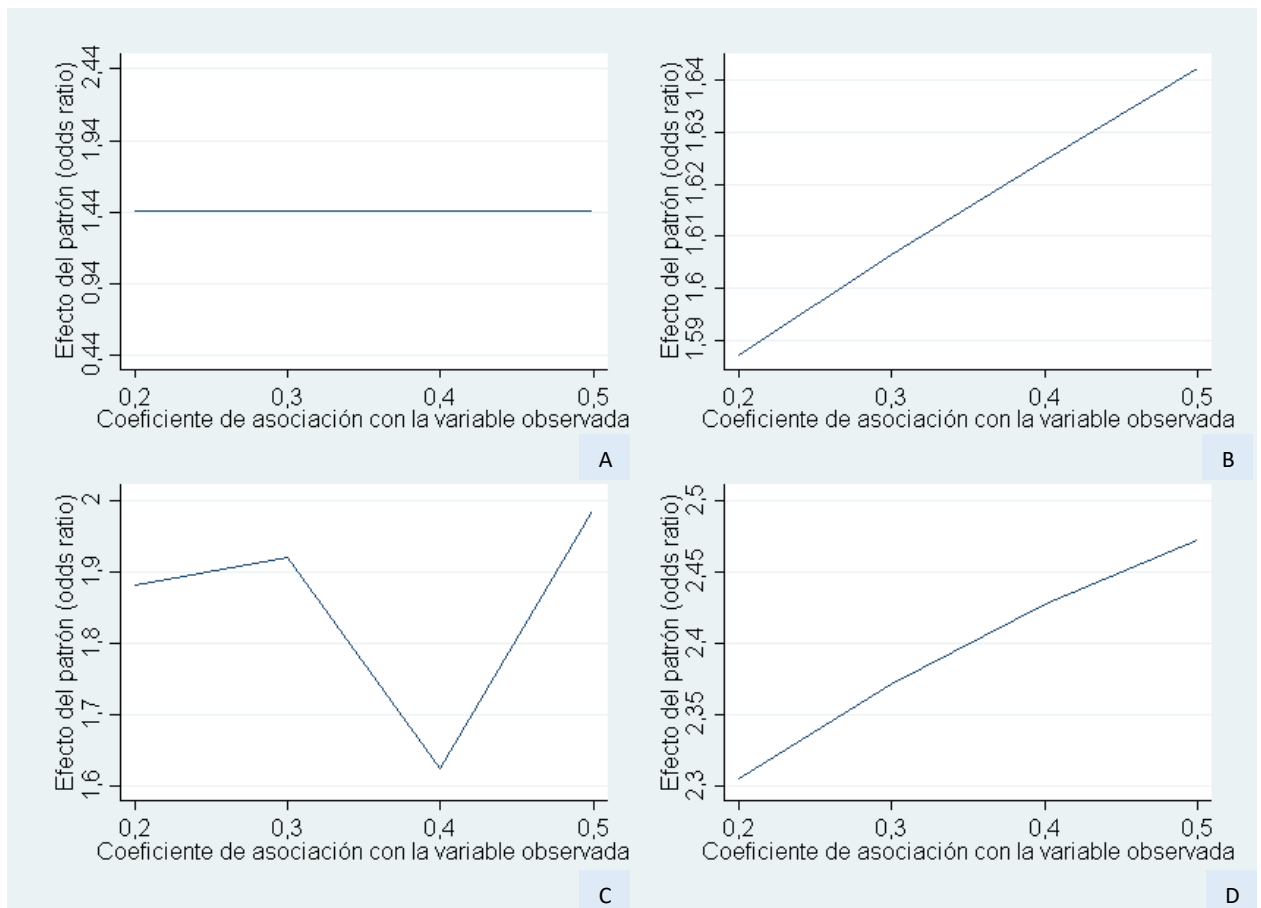


Figura 3.6. Tendencias entre el efecto del patrón dietario Cono Sur y el correspondiente riesgo de presentar cáncer colorrectal. Estimación de las medidas de asociación (ORs) entre la adherencia al patrón dietario Cono Sur (continuo) y cáncer colorrectal. Simulación de diferentes escenarios (A: Desvío estándar=0; B: Desvío estándar=1; C: Desvío estándar=2; D: Desvío estándar=3) de riesgo con asignación de variabilidad *a priori* de la variable confundente no observada. Estudio caso-control, provincia de Córdoba. Año 2010-2016.

3.2. Análisis de sesgo de selección

Estudio caso-control de cáncer de mama

A continuación se presentan en la Tabla 3.7 los resultados del análisis de sensibilidad por el sesgo de selección a través de métodos determinístico y probabilístico. Se puede observar que al plantearse un análisis determinístico no diferencial, es decir, donde los grupos en estudio (caso expuestos, casos no expuesto, controles expuesto y controles no expuestos) tienen las mismas probabilidades de participar en el estudio, las estimaciones de riesgo obtenidas son iguales a los análisis que no considerarían este posible sesgo (convencionales), siendo los límites del rango intercuartílico (RIC) más extremos y por ende el RIC más amplio. Al plantear escenarios con distribuciones de probabilidades *a priori* diferenciales, se observa que en los escenarios 1 y 2, donde las probabilidades de participación asignadas fueron mayores para el grupo de casos expuestos en comparación con el resto, las estimaciones puntuales de los ORs fueron menores a las convencionales (ver Tabla 2.2 en Métodos). Lo contrario ocurrió en el escenario 3 propuesto donde las menores probabilidades de participación se asignaron a los casos expuestos. En cuanto al análisis probabilístico, la estimación puntual del OR ajustado por el sesgo de selección fue prácticamente el mismo que el análisis convencional (OR=1,71), pero el cociente entre los límites del rango simulado incluyendo el error sistemático y aleatorio (OR=2,89) fue 63% mayor que el convencional (OR=1,77).

Tabla 3.7. Estimación de las medidas de asociación (ORs) entre la adherencia al patrón dietario Tradicional y cáncer de mama. Análisis determinístico y probabilístico del sesgo de selección. Estudio caso-control, provincia de Córdoba. Año 2008-2015.

		Error	Percentiles			Ratio
			2,5	50	97,5	97,5/2,5
		Convencional	1,30	1,72	2,29	1,77
<hr/>						
Análisis determinístico						
<hr/>						
Escenario no diferencial	Error sistemático	1,39	1,72	2,13	1,53	
	Aleatorio y sistemático	1,22	1,72	2,47	2,02	
Escenarios diferenciales						
Escenario 1	Error sistemático	0,70	0,91	1,19	1,71	
	Aleatorio y sistemático	0,63	0,91	1,34	2,14	

Escenario 2	Error sistemático	1,10	1,35	1,63	1,49
	Aleatorio y sistemático	0,96	1,35	1,91	1,99
Escenario 3	Error sistemático	2,74	3,45	4,31	1,57
	Aleatorio y sistemático	2,74	3,45	5,00	2,06
Análisis probabilístico					
	Error sistemático	1,15	1,71	2,65	2,30
	Aleatorio y sistemático	1,03	1,74	2,97	2,89

Estudio caso-control de cáncer colorrectal

En la Tabla 3.8 se presentan los diferentes escenarios planteados para evaluar el posible sesgo de selección del estudio caso-control de cáncer colorrectal. Se presentan los percentiles de las distribuciones de las estimaciones de riesgo calculadas y el ratio entre los límites del intervalo obtenido. En cuanto al análisis determinístico, los resultados obtenidos presentan similitudes con los conseguidos en el estudio de cáncer de mama. Al postular escenarios donde el grupo casos expuestos presentó mayor probabilidad de participación, menor fue la estimación de riesgo puntual obtenida (escenarios 1 y 2). En cambio, cuando la probabilidad de participación fue menor para el grupo de casos expuestos, la estimación puntual de riesgo fue del doble a la obtenida en los análisis convencionales (escenario 3). Respecto al análisis probabilístico, la estimación puntual del OR ajustado por el sesgo de selección fue similar al obtenido en el análisis convencional (OR=1,73), aunque el ratio entre los percentiles 75 y 25 simulados, incluyendo el error sistemático y aleatorio (OR=3,24) fue 50% mayor que el convencional (OR=2,16).

Tabla 3.8. Estimación de las medidas de asociación (ORs) entre la adherencia al patrón dietario Cono Sur y cáncer colorrectal. Análisis determinístico y probabilístico del sesgo de selección. Estudio caso-control, provincia de Córdoba. Año 2010-2016.

		Error	Percentiles		Ratio	
			2,5	50	97,5	97,5/2,5
		Convencional	1,17	1,72	2,52	2,16
Análisis determinístico						
Escenario no diferencial	Error sistemático	1,39	1,72	2,12	1,53	
	Aleatorio y sistemático	1,13	1,72	2,68	2,37	
Escenarios diferenciales						

Escenario 1	Error sistemático	0,70	0,91	1,19	1,71
	Aleatorio y sistemático	0,58	0,91	1,45	2,50
Escenario 2	Error sistemático	1,10	1,34	1,63	1,49
	Aleatorio y sistemático	0,88	1,34	2,09	2,37
Escenario 3	Error sistemático	2,73	3,43	4,29	1,57
	Aleatorio y sistemático	2,24	3,44	5,47	2,44
Análisis probabilístico					
	Error sistemático	1,14	1,70	2,63	2,30
	Aleatorio y sistemático	0,96	1,73	3,12	3,24

3.3. Análisis de sesgo de información

Como se mencionó, el sesgo de información derivado de datos faltantes siguiendo un patrón de datos aleatorio, es decir MAR (Missing At Random), es el asumido en el presente trabajo. A continuación se presenta este análisis mediante la metodología de Imputación Múltiple de datos por ecuaciones encadenadas en los estudios caso-control. Se presentan luego los resultados derivados del análisis de sesgo de información por error de clasificación aplicado al estudio prospectivo de cáncer de mama específicamente.

3.3.1. Análisis de sesgo de información por datos faltantes

Estudio caso-control de cáncer de mama

La Tabla 3.9 muestra las estimaciones de efecto (OR e intervalos de confianza) de las variables del modelo de regresión logística derivados los análisis de casos completos y de imputación múltiple. El análisis de datos completos se aplicó en aproximadamente un tercio de los sujetos estudiados, porque de la base total sólo 265 sujetos presentaban datos completos para todas las variables incluidas en el análisis. Un efecto promotor significativo del patrón dietario Tradicional se observó en los dos enfoques. Sin embargo, sólo se identificó efecto de otras covariables como el IMC y la lactancia materna cuando se aplicó el método de imputación múltiple. Si bien la incertidumbre asociada al proceso de imputación es considerada en las estimaciones, se obtienen intervalos de confianza más precisos luego de aplicar este análisis. La medida diagnóstica de imputación (IVR, incremento de la varianza relativo) del modelo final de riesgo mostró un valor igual a 0,07, indicando que la variabilidad de la muestra estimada para este conjunto de variables fue sólo un 7% mayor de aquella que se hubiese observado en el caso que las covariables hubiesen presentado datos completos.

Los factores que tuvieron mayor influencia en el incremento del error estándar, y por lo tanto en el IVR, fueron la variable actividad física y las categorías más bajas de nivel socioeconómico (imputadas). La Figura 3.7 muestra las distribuciones de los valores observados e imputados para la variable actividad física de dos ejemplos de bases imputadas (2 y 17), que ilustran el comportamiento del modelo de imputación. Para el resto de las bases imputadas las distribuciones observadas e imputadas resultaron similares entre sí (gráficos no presentados).

Tabla 3.9. Estimación de las medidas de asociación (ORs) entre la adherencia al patrón dietario Tradicional y cáncer de mama. Análisis de casos completos y de imputación múltiple de datos. Estudio caso-control, provincia de Córdoba. Año 2008-2015.

	Análisis de casos completos (n=265; 31,4%)			Imputación múltiple (n=703; 83,3%)			% Incremento del error estándar
	Odds Ratio	95% CI	p value	Odds Ratio	95% CI	p value	
Patrón Tradicional	1,33	1,015-1,755	0,039	1,40	1,184-1,657	<0,001	0,14
IMC	1,03	0,972-1,106	0,262	1,03	1,004-1,067	0,026	0,19
Lactancia Materna	0,58	0,319-1,080	0,087	0,54	0,365-0,813	0,003	0,13
NSE							
Baja Inferior	0,43	0,104-1,838	0,260	1,14	0,461-2,857	0,766	12,8
Baja Superior	0,98	0,332-2,949	0,985	1,32	0,633-2,773	0,454	7,19
Media típico	1,09	0,364-3,291	0,872	1,88	0,863-4,096	0,112	6,07
Media alto	1,49	0,536-4,144	0,443	1,57	0,771-3,194	0,213	6,81
Alto	1,26	0,430-3,691	0,673	1,39	0,674-2,886	0,369	5,94
Estatus ginecológico	1,20	0,559-2,595	0,633	1,50	0,922-2,457	0,102	0,14
Actividad Física	0,99	0,999-1,000	0,694	1,00	0,999-1,000	0,510	19,32
Edad	0,99	0,968-1,030	0,961	0,99	0,977-1,012	0,571	0,64
Edad de la menarca	0,93	0,778-1,119	0,458	1,02	0,921-1,139	0,653	0,14
Hijos	1,54	0,648-3,663	0,328	1,57	0,913-2,709	0,102	0,17

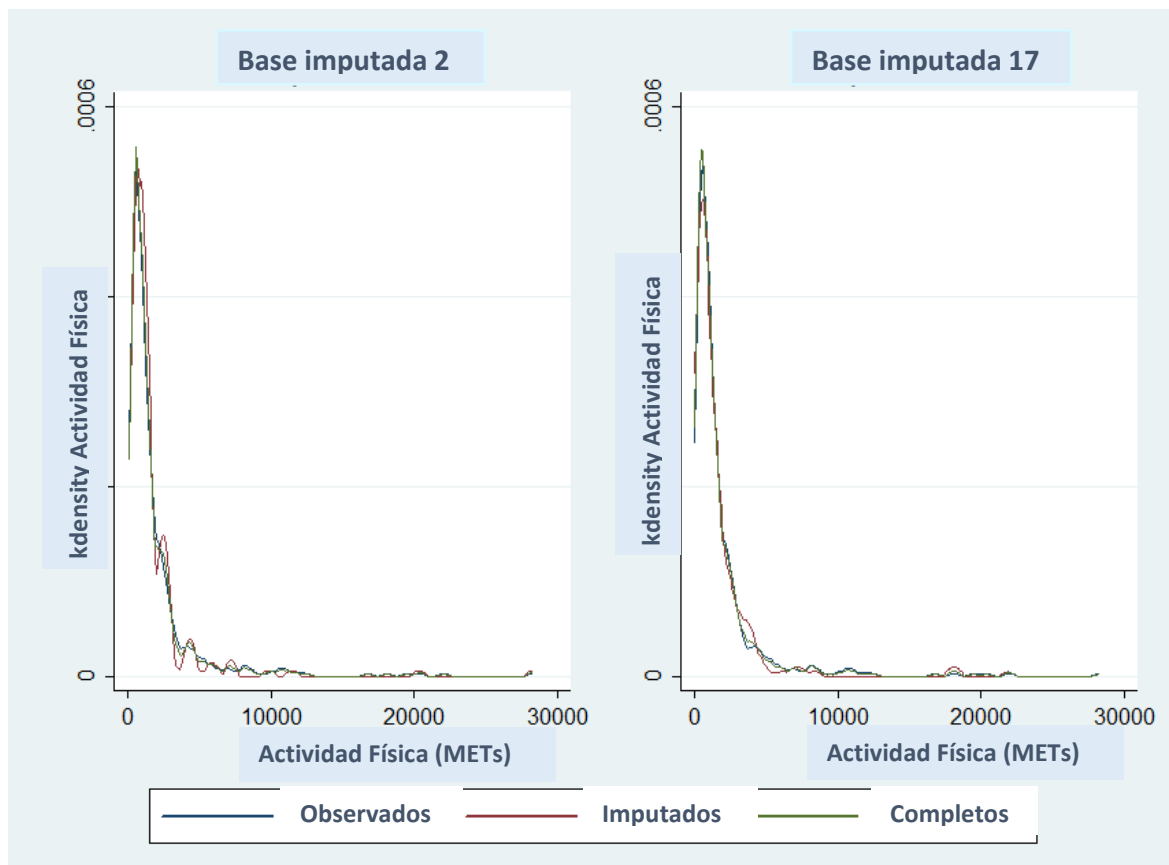


Figura 3.7. Gráficos diagnósticos de las distribuciones de valores observados e imputados de actividad física para las bases 2 y 17. Estudio caso-control de cáncer de mama, provincia de Córdoba. Año 2008-2015.

Estudio caso-control de cáncer colorrectal

A continuación se presentan en la Tabla 3.10 los resultados del análisis de sesgo de información por falta de datos aplicado al estudio de cáncer colorrectal, mostrando las estimaciones de efecto derivadas del análisis de datos completos y de Imputación Múltiple por ecuaciones encadenadas. El análisis de datos completos se aplicó en aproximadamente la mitad de los sujetos estudiados. En términos generales, los valores de OR estimados para las variables se atenúan cuando la Imputación Múltiple es usada. Los dos enfoques metodológicos mostraron un efecto promotor significativo del patrón dietario Cono Sur, además de asociaciones significativas de las variables sexo, IMC, antecedentes primarios de cáncer colorrectal y ciertas categorías de nivel socioeconómico. En general, se obtuvieron intervalos de confianza más precisos, aun considerando la incertidumbre asociada al proceso de imputación (como sucedió en el estudio de cáncer de mama). El modelo final de riesgo mostro un valor de 0,11 en la medida diagnóstica IVR, indicando que la variabilidad de la

muestra estimada para este conjunto de variables fue de un 11% mayor del que se hubiese observado en el caso de que las covariables hubiesen presentado datos completos. Se observa en la tabla que las categorías más bajas de nivel socioeconómico fueron las que mayor influencia tuvieron en el % de incremento del error estándar.

Tabla 3.10. Estimación de las medidas de asociación (ORs) entre la adherencia al patrón dietario Cono Sur y cáncer colorrectal. Análisis de casos completos y de imputación múltiple de datos. Estudio caso-control, provincia de Córdoba. Año 2010-2016.

	Análisis de casos completos (n=252; 51,4%)			Imputación múltiple (n=346; 70,6%)			% Incremento del error estándar
	Odds Ratio	95% CI	p value	Odds Ratio	95% CI	p value	
Patrón Cono Sur	1,56	1,164-2,116	0,003	1,44	1,115-1,863	0,005	0,46
Edad	0,99	0,977-1,020	0,918	0,99	0,976-1,014	0,624	1,25
Sexo	2,36	1,158-4,846	0,018	2,17	1,184-3,99	0,012	0,8
NSE							
Bajo	0,54	0,144-2,059	0,372	0,68	0,198-2,347	0,543	10,64
Bajo-Superior	0,66	0,201-2,217	0,511	0,73	0,234-2,272	0,586	13,12
Medio	0,27	0,077-0,998	0,050	0,31	0,095-1,064	0,063	13,51
Medio alto	0,30	0,097-0,985	0,047	0,33	0,116-0,977	0,045	9,35
Alto	0,21	0,063-0,722	0,013	0,21	0,069-0,679	0,009	9,27
IMC	1,07	1,000-1,146	0,049	1,05	0,994-1,113	0,075	1,45
Actividad Física	1,00	0,999-1,000	0,583	1,00	0,999-1,000	0,236	0,43
Antecedentes de primer grado de CCR	3,57	1,251-10,218	0,017	5,12	2,025-12,94	0,001	2,83

3.3.2. Análisis de sesgo de información por error de clasificación en la exposición

Estudio prospectivo de cáncer de mama

A continuación se presentan resultados descriptivos de las mujeres incluidas en el estudio prospectivo y, luego, los resultados derivados del análisis de sensibilidad por error de clasificación de la exposición.

Las Tablas 3.11 y 3.12 presentan una caracterización del grupo de mujeres estudiadas según presencia y ausencia de la patología. La media de edad de las 57 mujeres que se siguieron a lo largo de un año fue 51,29(13,6) años. Un 37,5 % de mujeres presentó sobrepeso, un 26,3% era sedentaria y un 42,11% presentaba el hábito de fumar (fumaba al inicio del estudio o fumó en el pasado). La mayoría había tenido hijos (63%) y de éstas,

prácticamente todas habían amamantado. Un 56% de las mujeres declaró ser menopáusica al momento del inicio del estudio. Las distribuciones fueron similares para ambos grupos de estudio en gran parte de las categorías de las variables de interés. El 84% de las mujeres respondió a 7 o más recordatorios de 24 horas R24H (Tabla 12).

Tabla 3.11. Distribución de frecuencias de variables de interés según presencia o ausencia de cáncer de mama al inicio del subestudio prospectivo 2015-2016.

	Cáncer de mama		Total n(%)
	Ausencia n(%)	Presencia n(%)	
Edad*	50,94(13,70)	51,95(13,75)	51,29(13,60)
IMC			
<25kg/mt ²	24(64,86)	11(57,89)	35(62,5)
25-30kg/mt ²	8(21,62)	4(21,05)	12(21,43)
>30kg/mt ²	5(13,51)	4(21,05)	9(16,07)
Actividad Física			
Sedentario	10(27,03)	5(25)	15(26,32)
Activo	27(72,97)	15(75)	42(73,68)
Fuma/Fumaba			
No	21(56,76)	12(60)	33(57,89)
Si	16(43,24)	8(40)	24(42,11)
Estado civil			
Soltera	13(35,14)	10(50)	23(40,35)
Casada	11(28,73)	6(30)	17(29,82)
Divorciada	5(13,51)	3(15)	8(14,04)
Viuda	8(21,62)	1(5)	9(15,79)
Hijos			
No	15(40,54)	6(30)	21(36,84)
Si	22(59,46)	14(70)	36(63,16)
Lactancia Materna			
No	3(8,11)	1(5)	4(7,02)
Si	19(51,35)	13(65)	32(56,14)
No corresponde	15(40,54)	6(30)	21(36,84)
Menopausia			
No	19(51,35)	6(30)	25(43,86)
Si	18(48,65)	14(70)	32(56,14)

* \bar{X} (Desvío Estándar).

Tabla 3.12. Distribución de frecuencias de cantidad de recordatorios de 24 horas aplicados según presencia o ausencia de cáncer de mama en las mujeres del subestudio 2015-2016.

Cáncer de mama				Frecuencia Acumulada (%)
Cantidad de R24H	Ausencia	Presencia	Total n (%)	
1	0	4	4 (7)	7
2	0	0	0 (0)	7
3	0	0	0 (0)	7
4	1	0	1 (1,8)	8,8
5	3	0	3 (5,3)	14,1
6	1	0	1 (1,8)	15,9
7	8	2	10 (17,5)	33,4
8	4	0	4 (7)	40,4
9	0	1	1 (1,8)	42,2
10	12	9	21 (36,8)	79
12	8	4	12 (21)	100

En las Tablas 3.13 y 3.14 se presentan las distribuciones de los sujetos según las variables de exposición consumo de carnes procesadas y productos de pastelería, respectivamente. Al inicio de la tabla se presentan las distribuciones del total de sujetos en estudio, según consumo alto-medio vs. consumo bajo. Luego, lo mismo para los grupos controles y casos. En la siguiente tabla (Tabla 3.15), se observan las medidas de error de clasificación, la sensibilidad, especificidad, valores predictivos positivos y negativos. Para la exposición de consumo de carnes procesadas, las mediciones de sensibilidad y especificidad fueron de 0,7 y 0,71 respectivamente del total de sujetos estudiados. Respecto al consumo de productos de pastelería, la sensibilidad fue superior (0,81) y la especificidad fue inferior a las mencionadas (0,6). Los valores predictivos obtenidos fueron de 0,63 a 0,77. Al analizar ambos grupos de estudio por separado, se observa que las mediciones utilizadas para evaluar la calidad de sistema de clasificación de la exposición de la frecuencia de consumo alimentario varió desde 0,5 (especificidad para casos en consumo de carnes procesadas) a 0,86 (sensibilidad para casos en consumo de productos de pastelería). La sensibilidad fue superior en los grupos casos en comparación con los controles (principalmente en relación a los productos de pastelería, Tabla 3.15). Este desempeño se discutirá en la próxima sección.

Tabla 3.13. Distribución de sujetos en estudio según consumo de carnes procesadas (dicotómica) utilizando el R24H como *Gold Standard* y cálculo de medidas de error de clasificación, subestudio 2015-2016

		R24H			
		Consumo de carnes procesadas	Alto-Medio	Bajo	Totales
Todos	Alto-Medio		21	6	27
	Bajo		9	15	24
	Totales		30	21	51
FFQ	Alto-Medio		16	2	18
	Controles	Bajo	8	11	19
	Totales		24	13	37
Casos	Alto-Medio		5	4	9
	Bajo		1	4	5
	Totales		6	8	14

Tabla 3.14. Distribución de sujetos en estudio según consumo de productos de pastelería (dicotómica) utilizando el R24H como *Gold Standard* y cálculo de medidas de error de clasificación, subestudio 2015-2016.

		R24H			
		Consumo de productos de pastelería	Alto-Medio	Bajo	Totales
Todos	Alto-Medio		25	8	33
	Bajo		6	12	18
	Totales		31	20	51
FFQ	Alto-Medio		19	4	23
	Controles	Bajo	5	9	14
	Totales		24	13	37
Casos	Alto-Medio		6	4	10
	Bajo		1	3	4
	Totales		7	7	14

Tabla 3.15. Sensibilidad (S), Especificidad (E) y valores predictivos positivos (VPP) y negativos (VPN) de la frecuencia de consumo alimentaria para evaluar el consumo de carnes procesadas y productos de pastelería.

Grupos de alimentos	Totales				Casos		Controles	
	S	E	VPP	VPN	S	E	S	E
Carnes procesadas	0,70	0,71	0,77	0,63	0,83	0,50	0,66	0,84
Productos de Pastelería	0,81	0,60	0,75	0,66	0,86	0,42	0,79	0,69

A continuación se presentan los análisis realizados para obtener la información que se hubiese observado si el instrumento para clasificar a los sujetos (frecuencia de consumo alimentario) hubiese clasificado correctamente a todos los sujetos. En una primera instancia, se presentan los resultados asumiendo error de clasificación no diferencial. Los valores asignados impuestos fueron similares a los obtenidos en cálculos de sensibilidad y especificidad del total de los sujetos estudiados (Tabla 3.15). Luego, se asume error diferencial basándose en las mediciones de sensibilidad y especificidad obtenidas para casos y controles (Tabla 3.15). Finalmente, se presentan diferentes escenarios donde se asignan diferentes valores de sensibilidad y especificidad a los dos grupos de estudio, teniendo en cuenta las posibles distribuciones de estos parámetros en base a los análisis previos. Se presentan en un principio los resultados derivados del análisis de la exposición del consumo de carnes procesadas, y luego, de la exposición de productos de pastelería.

En la Tabla 3.16, se impone una sensibilidad y especificidad no diferencial de 0,70 para el caso de la exposición de consumo de carnes procesadas. A la izquierda de la tabla se presenta la clasificación de sujetos por frecuencia de consumo alimentaria observada en el estudio y a la derecha los datos corregidos, es decir, la distribución de los sujetos que se hubiese observado si el esquema de clasificación clasificara correctamente a los sujetos según las mediciones de sensibilidad y especificidad. En este caso, se observa que los casos fueron los que se clasificarían con menor adherencia que los controles. La estimación de riesgo cruda (OR) derivada de los datos observados fue de 1,96, en tanto que la estimación de riesgo derivada de los datos corregidos fue mayor (más de tres veces) a la obtenida de los datos observados. Esto indica que el OR observado estaría sesgado hacia el valor nulo. En la Tabla 3.17, se presentan los resultados derivados de analizar el error de clasificación corrigiendo por los valores de sensibilidad y especificidad derivados de la Tabla 3.15. En esta situación se

obtiene un OR con una dirección opuesta al OR crudo. Esto hace suponer que la sensibilidad y especificidad obtenida para casos y controles presenta tanta amplitud (0,5-0,84) que modifican la dirección de la estimación de riesgo. Esta amplitud en el rango de mediciones de sensibilidad y especificidad será discutida opinando acerca del tamaño muestral. En la siguiente tabla (Tabla 3.18) se presentan estimaciones de riesgo derivadas de diferentes escenarios donde se asignan distintas sensibilidades y especificidades a casos y controles. Se observa que los ORs obtenidos son superiores a 1. A medida que la sensibilidad y especificidad de controles aumenta, las estimaciones de riesgo se alejan del valor nulo. Ocurre lo contrario respecto al grupo casos. Los ORs se acercan al valor nulo a medida que la sensibilidad y especificidad aumentan. Se observa en este caso que el instrumento utilizado asumiría una sensibilidad y especificidad que oscila entre 0,75 y 0,60 en casos, mientras que en controles, la sensibilidad estaría entre 0,65-0,7 y la especificidad entre 0,75 y 0,80, dado que el OR observado (1,9) se encuentra en ese rango.

Tabla 3.16. Corrección de error de clasificación de consumo de carnes procesadas en estudio de cáncer de mama asumiendo error de clasificación no diferencial, subestudio 2015-2016.

	Datos observados		Datos corregidos	
	Consumo de Carnes Procesadas		Consumo de Carnes Procesadas	
S=0,70; E=0,70	Alto-Medio	Bajo	Alto-Medio	Bajo
Casos	9	5	12	2
Controles	18	19	17,3	19,8
OR	1,9		6,87	

Tabla 3.17. Corrección de error de clasificación de consumo de carnes procesadas en estudio de cáncer de mama asumiendo error de clasificación diferencial, subestudio 2015-2016.

	Datos observados		Datos corregidos	
	Consumo de Carnes Procesadas		Consumo de Carnes Procesadas	
	Alto-Medio	Bajo	Alto-Medio	Bajo
Casos (SE=0,83; SP=0,5)	9	5	6,1	7,9
Controles(SE=0,66; SP=0,84)	18	19	24,2	12,8
OR	1,9		0,41	

Tabla 3.18. Odds ratios ajustados del análisis de carnes procesadas y riesgo de presentar cáncer de mama según diferentes combinaciones sensibilidad y especificidad en casos y controles asumiendo error de clasificación diferencial, subestudio 2015-2016.

		Controles				
		S	0,60	0,65	0,70	0,75
		E	0,70	0,75	0,80	0,85
Casos	S	E				
	0,70	0,60	2,59	2,94	3,17	3,33
	0,75	0,65	1,66	1,89	2,04	2,14
	0,80	0,70	1,33	1,51	1,63	1,71
	0,85	0,75	1,15	1,31	1,41	1,49

Respecto al análisis sobre la exposición a consumo de productos de pastelería, en una primera instancia se impuso una sensibilidad y especificidad no diferencial de 0,80 y 0,6 respectivamente (Tabla 3.15). En la Tabla 3.19 se presentan las correcciones realizadas derivadas del error de clasificación no diferencial en este grupo de alimentos, mostrando a la derecha de la tabla los valores esperados de no existir este error. La estimación de riesgo cruda (OR) derivada de los datos observados fue de 1,5 y el OR derivada de los datos corregidos fue aproximadamente dos veces mayor (2,95). Esto indica que el OR observado estaría sesgado hacia el valor nulo. En la Tabla 3.20, se presentan los resultados derivados de analizar el error de clasificación corrigiendo por los valores de sensibilidad y especificidad derivados de los análisis anteriores según casos y controles (Tabla 3.15). Así se obtiene un OR con una dirección opuesta al OR crudo, situación similar a la ocurrida en el caso de las carnes procesadas. A continuación se plantean diferentes escenarios donde se asignan distintas sensibilidades y especificidades a los casos y controles y las estimaciones de riesgo derivadas. Se observa que en algunas situaciones los ORs obtenidos son superiores el OR crudo (1,5) y en otras son inferiores incluso se modifica la dirección del OR en una situación. A medida que la sensibilidad y especificidad de casos disminuye y estas medidas aumentan en controles, las estimaciones de riesgo se alejan del valor nulo (Tabla 3.21). El instrumento utilizado asumiría una sensibilidad y especificidad que oscila entre 0,75-0,8 y 0,6-0,7 respectivamente en casos; mientras que en controles, la sensibilidad estaría entre 0,65-0,75 y la especificidad entre 0,60-0,70, dado que el OR observado (1,5) se encuentra en esos rangos.

Se sintetizan a continuación los resultados obtenidos. Respecto al fenómeno de confusión, los análisis realizados en el estudio de cáncer de mama indicaron que no existió efecto confusor de las variables edad, lactancia materna, hijos, nivel socioeconómico, edad de la menarca, IMC, actividad física y estado ginecológico. En cambio, en el estudio caso-control de cáncer colorrectal, se observaron efectos confundentes de las variables sexo y nivel socioeconómico (mostrando modificaciones en las estimaciones de riesgo en un 34 y 10%, respectivamente). Los análisis para evaluar el posible efecto confusor de variables no medidas mostraron modificaciones despreciables, existiendo apenas un posible efecto de glucemia y aumento de peso durante la vida adulta (variables no registradas y propuestas en esta tesis) para cáncer de mama y colorrectal respectivamente. Los cambios en las estimaciones, entre un 40 y 70%, ocurrieron al imponer gran variabilidad y correlación con la variable medida (Índice de Masa Corporal) asociada a la variable no medida. En cuanto al sesgo de selección, las diferencias respecto a las estimaciones convencionales fueron pequeñas, afectando, en parte, la magnitud de las asociaciones y la precisión de la estimación por intervalo, no así la dirección de las estimaciones. Situación similar ocurrió en ambos estudios. Respecto al sesgo de información por datos faltantes en el estudio de cáncer de mama, un 32% de mujeres presentó información completa y un 83% de la base fue considerada al aplicar el método de Imputación Múltiple. Ambos enfoques mostraron efecto promotor del patrón Tradicional (datos completos OR: 1,33; IC 95%:1,015-1,755; con imputación OR: 1,4; IC 95%: 1,184-1,657). No obstante, los efectos de otras variables, como Índice de Masa Corporal (OR: 1,03; IC 95%: 1,004-1,067) y la práctica de lactancia materna (OR: 0,54; IC 95%: 0,365-0,813) solo se evidenció bajo el esquema de Imputación Múltiple. Por su parte, el estudio de cáncer colorrectal, presentó un 51% de los sujetos con información completa que se incluyeron en el análisis de datos completos y, considerando un 20% más de sujetos en estudio al realizar la imputación (71%). Un efecto promotor significativo del patrón dietario Cono Sur se observó en ambos enfoques (datos completos OR: 1,56; IC 95%:1,164-2,116; con imputación OR: 1,44; IC 95%: 1,115-1,863). Además, se observaron asociaciones significativas respecto a las variables sexo, IMC, antecedentes primarios de CCR y nivel socioeconómico. Los modelos de imputación propuestos en ambos estudios mostraron un desempeño satisfactorio. En relación al sesgo de información por error de clasificación de la exposición, se destaca que se obtuvo una clasificación de la exposición aceptable para los grupos de consumo de carnes procesadas y productos de pastelería analizados, con valores de sensibilidad, de especificidad y valores predictivos positivos y negativos entre 63-86%, 42-84%, y 63-77%, respectivamente.

CAPÍTULO 4. DISCUSIÓN

En el presente trabajo se construyó un marco metodológico para valorar cuantitativamente la calidad de los estudios caso-control, estudiando exhaustivamente los errores sistemáticos presentes en los estudios epidemiológicos en cáncer de mama y colorrectal, conducidos por el Grupo de Epidemiología Ambiental del Cáncer en Córdoba. Los principales resultados mostraron que, en el estudio de cáncer de mama no se observaron importantes efectos confundentes de las variables analizadas, aunque sí podría existir, en determinadas situaciones, un posible efecto confusor de la variable propuesta no medida (glucemia). Respecto al estudio de cáncer colorrectal, fueron mostrados los efectos confundentes de las variables sexo y nivel socioeconómico y, además, se verificó un posible efecto confusor posible de la variable no registrada (aumento de peso durante la vida adulta). En consideración al sesgo de selección, los análisis realizados en ambos estudios escasos indicios de que este sesgo esté presente y afecte estimaciones derivadas de los mismos. En cuanto al sesgo de información derivado de datos faltantes, la metodología de Imputación Múltiple de datos aplicada, mostró beneficios al considerar mayor información en los análisis, mejorando las estimaciones de riesgo obtenidas e identificando efectos de covariables. Respecto al error de clasificación derivado del estudio de validación de cáncer de mama, los resultados mostraron valores aceptables para la clasificación de la exposición de los sujetos para las dos variables de exposición propuestas (grupo de alimentos de carnes procesadas y productos de pastelería). Estos resultados confirman la hipótesis planteada, ya que el diseño del análisis de sesgo sistemático y su cuantificación permitió obtener estimaciones de riesgo que difieren de las estimaciones clásicas, respecto a la consideración de la incertidumbre en las estimaciones. Las magnitudes de las estimaciones de riesgo fueron, en gran medida, atenuadas, luego de considerar los sesgos estudiados.

La presente tesis contribuye a mejorar la validez de las estimaciones de riesgo derivadas de estudios epidemiológicos sobre enfermedades crónicas, como el cáncer, en la provincia de Córdoba, Argentina. Esta enfermedad es la segunda causa de muerte en el mundo y en Argentina, precedida sólo por las enfermedades cardiovasculares (OMS, 2017a; Dirección de Estadísticas e Información de Salud, 2013). Se conoce que cerca del 70% de las muertes por cáncer se registran en países de ingresos medios y bajos (OMS, 2017a). La evidencia sugiere que las personas que presentan un nivel socioeconómico bajo serán las que

aportarán de manera desproporcionada al número de casos nuevos de cáncer (Doubeni *et al*, 2012), lo que coloca a la Argentina en una situación preocupante. Aunque el país ha sido clasificado por The World Bank (The World Bank Group, 2017) como un país de ingresos alto-medio, según datos oficiales, aproximadamente un tercio de la población argentina se encuentra en condición de pobreza (Instituto Nacional de Estadística y Censos, 2017). Un estudio poblacional llevado a cabo en Córdoba, (Aballay *et al*, 2016) reportó que el 51% de los sujetos pertenecían al nivel socioeconómico bajo. Respecto a la incidencia y mortalidad a nivel mundial por sitio tumoral, el cáncer de mama es el cáncer más diagnosticado en las mujeres y la principal causa de muerte por cáncer, con más de medio millón de muertes relacionadas a la patología estimadas en 2012 (Ferlay *et al*, 2012). Específicamente, en América Central y América Latina la mortalidad por este cáncer ha ido aumentando en las últimas décadas. Los cambios en los estilos de vida derivados de las tendencias demográficas y epidemiológicas en un paralelo con el continuo crecimiento económico de la región, junto con la falta de acciones, de concientización pública y la escasa capacidad de desarrollo de los sistemas de salud, entre otros factores, sugieren que las cifras seguirán aumentando aún por encima de las proyecciones realizadas para esta región (Di Sibio *et al*, 2016). En cuanto al cáncer colorrectal, registros internacionales muestran también a éste cáncer como uno de los más incidentes a nivel mundial, con aproximadamente 1,4 millones de casos nuevos y 700000 muertes ocurridas en 2012. El cáncer colorrectal es el tercer cáncer más diagnosticado y la cuarta causa de muerte asociada a cáncer en ambos sexos combinados (Ferlay *et al*, 2012). A pesar de que las tasas de incidencia y mortalidad estandarizadas por edad de éste cáncer son superiores en países desarrollados en comparación a los países menos desarrollados, los primeros parecieran haber estabilizado o declinado sus tasas, mientras que países en desarrollo como Argentina han aumentado estos indicadores para cáncer colorrectal, llegando a igualar a aquellos de países con mayores niveles de desarrollo humano (Sierra y Forman, 2016).

La alimentación se reconoce asociada al cáncer de mama y al colorrectal. Los primeros estudios epidemiológicos sobre el tema comenzaron a publicarse en la segunda mitad del siglo pasado aproximadamente, cuando el interés de la epidemiología nutricional giró en torno a las enfermedades no transmisibles más prevalentes en las poblaciones desarrolladas (como el cáncer y las enfermedades cardiovasculares). Estos cambios de intereses y prioridades de la epidemiología nutricional coincidieron también con los inicios de una nueva

etapa en el proceso denominado transición nutricional. Estas poblaciones con niveles de desarrollo más alto, comenzaron a mostrar cambios en los patrones de consumo alimentario y de bebidas, como también en los hábitos de actividad e inactividad física, con las consiguientes modificaciones en la composición corporal y la emergencia de enfermedades relacionadas a la alimentación (Popking, 2015). Este momento estuvo caracterizado por un aumento en el consumo de grasas, azúcares simples, alimentos procesados, bebidas azucaradas y una disminución de los niveles de actividad física total y un aumento de tiempo dedicado a actividades sedentarias (Popkin 2002, 2006). En los países menos desarrollados, también comenzaron a observarse cambios hacia esta nueva etapa de la transición nutricional (asociada a la emergencia de enfermedades no transmisibles), que sucedieron de una manera más acelerada, con características particulares asociadas al contexto socioeconómico. En América Latina, como también en otras regiones de bajos y medianos ingresos, se encontraban problemas de mal nutrición por deficiencia y por exceso en el mismo hogar, incluso en el mismo sujeto (Popking, 2002). El desafío de los investigadores abocados al estudio de la relación entre alimentación y cáncer se complejizó aún más frente a estos nuevos escenarios de transiciones epidemiológicas y nutricionales, reflejándose en las metodologías aplicadas en el área. Tradicionalmente, la epidemiología nutricional se había dedicado a estudiar el efecto de nutrientes o de alimentos específicos sobre los procesos de salud-enfermedad. La evidencia sugeriría que nutrientes y alimentos actuarían de manera sinérgica en el riesgo de ocurrencia de enfermedades crónicas. Entonces, surgió la propuesta de utilizar el enfoque de patrones alimentarios como enfoque superador (Kac *et al*, 2007) y, aporta elementos al análisis de la red causal entre alimentación y enfermedades no transmisibles emergentes en los últimos tiempos.

En Argentina y en Córdoba, precisamente, este enfoque fue utilizado por Pou *et al* (2014), basándose en una muestra de controles, en representación de la población general e identificando que existen cuatro patrones alimentarios, que caracterizan la dieta habitual de la población cordobesa. Dos de estos patrones, los que captaron mayor variabilidad, el patrón dietario Tradicional y el otro denominado Cono Sur, fueron las variables de exposición adoptadas en los estudios de cáncer analizados en el presente trabajo. Respecto al patrón dietario Tradicional, identificado para cáncer de mama y caracterizado por elevadas cargas de carnes grasas, productos de pastelería y aceites y mayonesa, estaría asociado a ciertos hábitos alimentarios de mujeres ya reportados en la literatura asociados a la ocurrencia del

cáncer de mama mediante diferentes mecanismos. Entre ellos, el alto consumo de lípidos e hidratos de carbono se relacionarían con los niveles circulantes de hormonas sexuales y/o factores de crecimiento (Renehan *et al*, 2015; Amadou *et al*, 2014; Lajous *et al*, 2005); el consumo de carnes grasas, también promovería un aumento del riesgo de presentar la patología a través de su elevado contenido de lípidos, como así también la producción de aminas heterocíclicas (Reartes *et al*, 2016; Ronco *et al*, 2010) entre otros mecanismos. Además, el bajo consumo de fibra dietética y vitaminas antioxidantes a mayor adherencia a este patrón alimentario podría relacionarse, en parte, a la ocurrencia del cáncer de mama (Karimi *et al*, 2014). En cuanto al patrón Cono Sur, caracterizado por un elevado consumo de carnes rojas, vino y vegetales amiláceos, existen también diversos mecanismos propuestos por la literatura que relacionan dietas con características “occidentales” como este patrón, a un mayor riesgo de presentar cáncer colorrectal (Dermadi *et al*, 2017). Se destacan algunos de estos mecanismos a continuación. Las carnes rojas contienen hierro Hem, el cual promueve la formación de compuestos N-nitrosos potencialmente carcinógenos, como también otros compuestos citotóxicos derivados de la peroxidación de los lípidos (Vieira *et al*, 2017; Zhou *et al*, 2017). Respecto al consumo de etanol y el riesgo de presentar cáncer colorrectal, varios mecanismos plausibles son propuestos, entre los que se menciona el efecto del acetaldehído, metabolito del etanol, como carcinógeno en humanos. Además, el alcohol actúa como solvente facilitando el ingreso de otros compuestos carcinógenos a las células. También ha sido reportado que el etanol interviene en el metabolismo del retinol, el cual afecta el crecimiento y diferenciación celular y la apoptosis (Choi *et al*, 2017; Vieira *et al*, 2017). Un mayor consumo de vegetales amiláceos (papa, batata, choclo) en detrimento del grupo no amiláceo, también fue asociado a un mayor riesgo de presentar la patología (Makarem *et al*, 2015; Nagle *et al*, 2015). Los mecanismos involucrados podrían estar asociados principalmente a los efectos anti-cancerígenos de la fibra dietética, entre los cuales se menciona la formación de ácidos grasos de cadena corta por fermentación de las bacterias colónicas, la reducción de la producción de ácidos biliares secundarios, la reducción del tiempo de tránsito intestinal y de la resistencia a la insulina (Murphy *et al*, 2012).

La calidad de los estudios epidemiológicos, por lo general, no es indagada por los investigadores en ciencias de la salud. Es poco frecuente encontrar trabajos que reportan análisis de sensibilidad y muestran estudios sobre los sesgos presentes en estudios, tanto en los observacionales como en los experimentales. Tampoco informan detalladamente la

metodología utilizada y cómo la aplicaron en el caso de haberlo hecho (Zhang *et al*, 2017; Agha *et al*, 2016; Kahan *et al*, 2015; Eekhout *et al*, 2012; Groenwold *et al*, 2008; Klebanoff and Cole, 2008; Lee *et al*, 2007). Los sesgos, como se describió anteriormente, son errores sistemáticos que afectan las estimaciones de efecto e influyen en la validez de los mismos. A pesar de que todos los estudios son susceptibles a éstos, sus efectos suelen subestimarse. Existen recomendaciones sobre cómo enfrentar y reportar estos errores o sesgos en estudios observacionales en general (von Elm *et al*, 2007) y nutricionales específicamente (Lachat *et al*, 2016) y, sugerencias sobre cómo aplicar y reportar metodologías particulares (Sterne *et al*, 2009; Klebanoff y Cole, 2008), sin embargo, los trabajos científicos que publican análisis de sesgos son escasos. Esto puede deberse, en parte, a que los investigadores evitan este tipo de análisis por falta de conocimiento metodológico a ese respecto y, en algunos casos, no aplican las metodologías adecuadas para enfrentar el problema específico que se les presenta. Además, en ciertas situaciones, los investigadores no dan cuenta del efecto que genera el hecho de no controlar el sesgo. Una revisión sistemática de estudios observacionales en revistas de medicina y epidemiología demostró que la calidad del reporte sobre el fenómeno de confusión es pobre (Groenwold *et al*, 2008). Otra revisión, en este caso de estudios experimentales, concluyó que un análisis sobre el sesgo de selección no pudo llevarse a cabo debido a la escasa información publicada respecto a este sesgo (Kahan *et al*, 2015). Según Karahalios *et al* (2013) la gran mayoría de los estudios de cohorte presentan análisis sesgados por no considerar los datos faltantes en los mismos o aplicar metodologías reconocidas como sesgadas. La falta de datos en estos estudios deriva, en parte, de la dificultad de retener a los participantes en el mismo. En esta revisión, se reporta que la mayoría de las investigaciones consideradas excluyen a los participantes con datos faltantes, realizando así el análisis sobre los casos completos (en un 66% de los estudios) y sólo un 7% usando Imputación Múltiple o métodos bayesianos, metodologías reconocidas como no sesgadas. Estas publicaciones dan cuenta de la situación actual sobre la temática abordada y los esfuerzos que los investigadores deben realizar para mejorar la calidad de sus publicaciones.

Respecto al fenómeno de confusión, éste debe considerarse tanto en la etapa del diseño de los estudios caso-control, como en el momento de los análisis. El *matching* sirve para reducir el efecto confundente de ciertas variables (como la edad y el sexo) en el diseño y que ambos grupos de estudio, casos y controles, sean factibles de analizar conjuntamente. La literatura recomienda luego incluir, en los modelos de regresión, las variables por las cuales

se realizó el *matching* para evitar otro posible sesgo. El objetivo de tener un grupo control en este tipo de estudios es el de proveer una estimación de la distribución de la exposición en la población de referencia. Si los controles son seleccionados en base a un factor para coincidir con los casos, es decir, la variable por la que se realiza el *matching*, y ésta se correlaciona con la exposición, luego, la frecuencia de exposición cruda en los controles estará distorsionada en dirección similar a la de los casos (Rothman y Greenland, 1998). En los estudios caso-control analizados en el presente trabajo, las variables *matching* también fueron incluidas como covariables en los modelos de regresión múltiple junto con otras covariables o posibles confundentes.

Si bien la relación entre factores alimentarios y cáncer de mama y cáncer colorrectal es ampliamente reportada (WCRF/AICR, 2012), son escasos los trabajos que detallan las metodologías aplicadas y las variables consideradas como posibles confundentes en los análisis, más aún en el campo de la epidemiología nutricional del cáncer (Jamshidinaeini *et al*, 2016; Becaria Coquet *et al*, 2014; Bingham *et al*, 2005). Algunos estudios incluyeron como variables de ajuste las mismas variables confundentes utilizadas en el presente trabajo, aunque sin profundizar sobre las metodologías (Shin *et al*, 2017; Go *et al*, 2016; Klurfeld *et al*, 2015). El modelo de regresión es una metodologías muy utilizada en estudios observacionales para controlar por potenciales confundentes, incorporándolos como factores de riesgo explicativos (Hammer *et al*, 2009). En esta tesis, dos de las metodologías propuestas para enfrentar el fenómeno de confusión de variables observadas se basaron en modelos de regresión. En estos análisis, junto con el método de Mantel-Haenszel, se analizó la magnitud del efecto confundente de las variables sexo y nivel socioeconómico en el estudio de cáncer colorrectal. Además de la modificación de las estimaciones observadas por estos métodos, la evidencia científica sugiere que las variables propuestas como confundentes cumplan con los criterios para considerarlas como tales. Esto es, la variable confundente debe estar asociada a la exposición, asociada de manera “causal” a la patología y no debe ser una variable intermedia entre la exposición y el resultado (Szklo y Nieto, 2007). Numerosas investigaciones reportaron asociaciones entre el sexo y la enfermedad en estudio (Kim *et al*, 2015; Murphy *et al*, 2011). A pesar de las diferencias entre sitios tumorales que existen entre hombres y mujeres, la mayoría de los estudios no considera las especificidades por sexo en el diseño del estudio y la interpretación de los resultados. Un estudio de revisión sugiere introducir algunas innovaciones respecto a la temática, para reducir la morbilidad y mortalidad del cáncer colorrectal en mujeres principalmente, ya que éstas presentan mayor mortalidad y menores

tasas de supervivencia a 5 años que los hombres (Kim *et al*, 2015). El impacto de las desigualdades socioeconómicas en el contexto de diferentes patologías, no sólo las oncológicas, ha sido estudiado en las últimas décadas (Salgado-Barreira *et al*, 2014). Específicamente, respecto al cáncer colorrectal, estudios sugieren una asociación entre esta patología y el nivel socioeconómico (Manse y Bauerfeind, 2014; Doubeni *et al*, 2012; Aarts *et al*, 2010). Cabe destacar, que en algunos casos la variable nivel socioeconómico suele incluirse en las investigaciones como un indicador compuesto construido a partir de dos o más variables; y, en otros casos, el abordaje de los condicionantes socioeconómicos se realiza utilizando variables de manera independiente. Esta variable de naturaleza multidimensional, además de influir en la asociación entre la exposición y la patología por su posible efecto confundente, impacta en las estimaciones, ya que por su naturaleza es susceptible de presentar datos faltantes en alguna de las variables que la componen. De hecho, éste fue el caso en ambos estudios caso-control analizados en la presente tesis, donde la variable nivel socioeconómico fue construida, como se explicó anteriormente, a partir de 8 variables relacionadas con el nivel de educación, el trabajo, el tipo de atención médica, los bienes y servicios declarados por el principal sostén del hogar. Ésta ya ha sido asociada a la ocurrencia de cáncer colorrectal en la provincia de Córdoba (Becaria Coquet *et al*, 2014) y, en un estudio poblacional llevado a cabo en la capital de la provincia, también ha vinculado esta variable a otros condicionantes de salud, como el sobrepeso (Aballay *et al*, 2016). Otro estudio en esta misma ciudad, consideró a las variables ingresos mensuales, gastos mensuales reportados y al nivel de educación de manera independiente, para evaluar el nivel socioeconómico de los sujetos y su asociación con otra enfermedad no transmisible, la diabetes tipo 2 (Elgart *et al*, 2014). Grigsby *et al* (2016), también consideraron un conjunto de variables aisladas para estudiar el nivel socioeconómico y su asociación con la enfermedad pulmonar obstructiva crónica en cinco países (entre ellos Argentina), sin construir o utilizar un índice o score multidimensional, como el caso de la presente tesis. Deberá evaluarse en cada situación, la posible presencia de sesgo de confusión si correspondiere y, además, considerar la posibilidad de existencia de sesgos de información por falta de datos, especialmente en el caso de una variable multidimensional o la presencia de error de clasificación conceptual en el caso de la definición de la variable nivel socioeconómico a través de variables independientes.

Respecto al análisis de posibles confundentes no registradas, existen metodologías para analizar posibles escenarios sobre este sesgo (Orsini *et al*, 2008; Schneeweiss, 2006; McCandless *et al*, 2008; Groenwold *et al*, 2009; Luiz y Cabral Borges, 2010). Sin embargo, la

mayoría de los investigadores dedican apenas una mínima mención en la discusión sobre esta temática, sin llevar a cabo análisis cuantitativo alguno. La metodología propuesta en la presente tesis es relativamente sencilla para avanzar y profundizar respecto a lo que ofrece el mero análisis cualitativo y más divulgado en la literatura y, explora exhaustivamente este problema potencial, permitiendo estimar posibles resultados, variando un amplio conjunto de escenarios respecto a la confundente no observada. En ambos casos propuestos en el presente trabajo (para el estudio de cáncer de mama y colorrectal), las variables no observadas, glucemia y aumento de peso durante la vida adulta, no impactaron de manera importante en la mayoría de los escenarios propuestos. Sin embargo, podrían haber influido sustancialmente en las estimaciones de riesgo si efectivamente éstas presentaran una gran variabilidad y, además, estuvieran muy correlacionadas con el IMC, variable registrada en la base y asociada en ambos estudios caso-control. Muchas variables biológicas satisfacen los supuestos de distribución normal en su comportamiento estocástico. Esta distribución, indica altas probabilidades de obtener observaciones cercanas al centro y bajas probabilidades a medida que se aleja del mismo (McDonald, 2014; Ghasemi y Zahediasl, 2012). Así, sería poco probable que surgiera una modificación de la estimación de la mediana entre un 40 y un 70% como se observó en los análisis, ya que las variables antropométricas descritas en estas poblaciones suelen presentar una distribución simétrica y con desvíos estándar pequeños.

El sesgo de selección, como se ha mencionado, se reconoce como una posible amenaza a la validez de estimaciones derivadas no sólo de estudios comparativos, como los de tipo caso-control; sino también de estudios como los de cohorte, cuando por ejemplo, existen pérdidas en el seguimiento (Howe *et al*, 2016), o en estudios experimentales cuando existe sesgo por asignación al tratamiento (La Caze, 2013). En la presente tesis, los análisis propuestos para enfrentar este posible sesgo fueron de naturaleza determinística y probabilística. En el caso de que existiera sesgo de selección no diferencial, en la primera propuesta mencionada (determinística), la estimación de riesgo convencional sería conservadora. Es menester, destacar que en este tipo de estudios, la preocupación se asocia con la existencia de un sesgo diferencial entre los grupos a comparar (Gordis, 2009; Vrijheid *et al*, 2006). Debido a que el proceso de selección de casos y controles puede ser diferente en estos estudios, no siempre se tiene la certeza de que las condiciones de intercambiabilidad se mantienen (Geneletti *et al*, 2009) y por lo tanto, que este proceso de selección afecte la probabilidad de participación de manera diferente entre los grupos de estudio. Que el sesgo sea diferencial significa que la probabilidad de participación es diferente entre los grupos, es

decir, entre casos expuesto, casos no expuesto, sanos expuestos o no expuestos. En general, en los estudios caso-control, es más probable que ocurra una participación de esta modalidad, diferencial, donde el grupo de casos expuestos tenga mayor participación que el resto de los grupos en estudio. En los escenarios donde se planteó esta situación, las estimaciones se acercaron al valor nulo. Esto ocurrió en ambos estudios, el de cáncer de mama y colorrectal. Estas situaciones son las que presentan más chance de ocurrir, ya que los casos expuestos podrían estar más motivados o tener más interés en participar en estudios sobre hábitos modificables o estilos de vida y salud. Si éste fuera el caso, las estimaciones de riesgo a observar en ambos estudios sobre la exposición y su asociación con la patología se acercarían más al valor nulo a mayor diferencia entre las probabilidades de participación asignadas entre los casos expuestos y el resto de los grupos. La evidencia sugiere que las mujeres toman decisiones sobre alimentación más saludables que los hombres, lo que podría estar asociado en parte a que ellas se involucran más en el control del peso y también a que sus creencias sobre el efecto de la alimentación en la salud son mayores que los hombres (Wardle *et al*, 2004). En el caso del estudio de cáncer de mama, como la población de mujeres adultas es reconocida por su interés en la alimentación y la salud (Ek, 2015), es probable que en el caso que existieran diferencias entre los grupos no sean importantes, generando menor impacto en las estimaciones de riesgo. Respecto al estudio de cáncer colorrectal, la magnitud y dirección del impacto de este sesgo podría estar influenciado por la distribución de exposición entre hombres y mujeres. Es decir, la probabilidad de participación del grupo casos expuestos comparado con el resto de los grupos podría diferir en mayor medida si hubiese en este grupo mayor porcentaje de hombres expuestos en la población objeto de estudio que mujeres. Respecto al análisis probabilístico, las estimaciones puntuales obtenidas fueron similares a las convencionales considerando la incertidumbre del sesgo en las estimaciones.

Existen otras metodologías, descritas en la bibliografía, para explorar el sesgo de selección que pueden aportar elementos a los modelos causales que subyacen al problema de investigación de interés como son los diagramas causales, conocidos como Gráficos Acíclicos Dirigidos (en inglés *directed acyclic graphs* o DAGs) (Hernán y Robins, 2018). Estos diagramas causales codifican conocimientos expertos cualitativos, o supuestos, sobre la estructura causal de un problema, y por lo tanto, de los determinantes causales de los sesgos. Esta herramienta es muy útil conceptualmente y aporta a la inferencia causal del problema estudiado, aunque debe combinarse con enfoques cuantitativos para estimar esos efectos

causales. Varios estudios epidemiológicos han aplicado esta metodología para identificar este posible sesgo y proponer explicaciones biológicas plausibles y estimaciones válidas (Hernán *et al*, 2004; Geneletti *et al*, 2009). Los diagramas causales también sirven para proponer cuestiones menos obvias sobre resultados aparentemente sorprendentes o que no conciden con la evidencia y el consenso científico, como por ejemplo respecto a la paradoja del peso al nacer (Hernández-Díaz *et al*, 2008) o la paradoja de la obesidad (Banack *et al*, 2015). Este enfoque también se utiliza para estudiar el fenómeno de confusión (Hernán *et al*, 2002).

Los datos faltantes son un problema frecuente en los estudios epidemiológicos, ya que impactan en los análisis estadísticos y sus inferencias. Si bien aplicar un análisis al subconjunto de datos completos, es decir, a la parte balanceada de la información, es mucho más simple, las estimaciones obtenidas pueden estar sesgadas si los participantes con datos faltantes son omitidos de los análisis (Hernandez *et al*, 2017). Excluir a las observaciones que presentan datos faltantes también ignora la posibilidad de que existan diferencias sistemáticas entre los casos completos y los casos incompletos. Por consiguiente, las inferencias derivadas podrían no aplicarse a la población de referencia, especialmente cuando el número de casos completos es pequeño (National Research Council, 2010). El presente estudio analizó la fiabilidad que los estudios caso-control de cáncer de mama y colorrectal tienen para identificar factores de riesgo utilizando Imputación Múltiple. La Imputación Múltiple implica reemplazar los valores faltantes con más de un valor plausible extraídos de las distribuciones predictivas, condicionadas a la información observada o completa. Como resultado se obtienen múltiples bases de datos imputadas que pueden luego ser analizadas de manera independiente por métodos convencionales y los resultados son combinados por las reglas de Rubin (1987). Cuando el tamaño de la muestra no es muy grande y/o sólo un bajo porcentaje de los sujetos se incluyen en el análisis de casos completos, el problema de los datos faltantes debe ser considerado. Al comparar los resultados obtenidos de un análisis de datos completos con los obtenidos luego de aplicar Imputación Múltiple de datos, se obtienen diferencias en los p-valores (Ibrahim *et al*, 2012). Asimismo, en algunos casos la Imputación Múltiple puede también ser beneficiosa para la estimación de los coeficientes de covariables relativamente completas en presencia de otras covariables incompletas (White y Carlin, 2010). En este caso, los efectos de la exposición podrían resultar no confiables, además, del hecho que el tiempo y los recursos invertidos en recolección de información no compensarían, ya que ciertas observaciones se eliminan al momento del análisis.

En el presente trabajo no se utilizó Imputación Múltiple para estimar cada valor faltante usando valores simulados, sino una muestra aleatoria de valores faltantes. Este proceso obtiene inferencias estadísticas válidas cuando los mecanismos seleccionados son acordes a las bases de datos (Molenberghs y Kenward, 2007). Respecto a los mecanismos de patrones de datos faltantes ampliamente descritos en la literatura, se descartó la posibilidad que los *missing* tuvieran un patrón MCAR, en función de análisis exploratorio llevado a cabo previamente. Además, es muy poco probable que suceda este mecanismo en estudios epidemiológicos donde se registran muchas variables que pueden llegar a explicar en cierta medida los patrones de datos faltantes en otras (Soley-Bori, 2013). Tampoco se asumió un patrón MNAR para los datos faltantes, por no poder probarse este supuesto. Toda la exploración empírica realizada antes brindó indicios sobre la posibilidad que los datos faltantes se presentaran como aleatorios (MAR). Así, este trabajo asumió que la información faltaba siguiendo un patrón aleatorio (MAR), es decir, que para la variable X, la probabilidad de que una observación estuviera faltando depende sólo de los valores observados de otras variables, no de los valores no observados de X. El supuesto MAR no puede ser verificado, ya que justamente los valores faltantes no pueden observarse. No obstante, la medida diagnóstica de Incremento de la Varianza Relativo, calculada luego de llevar a cabo la imputación, indicó una buena performance del enfoque de modelación utilizado.

En América Latina, pocos son los estudios que han aplicado Imputación Múltiple para resolver el problema de datos faltantes (Benjet *et al*, 2008; Camargos *et al*, 2011; Fries *et al*, 2013; Nunes *et al*, 2009; Rubinstein *et al*, 2010). Respecto a los estudios sobre epidemiología nutricional de cáncer en la región, ninguno ha propuesto el método de Imputación Múltiple de datos para tratar con este posible sesgo derivado de información faltante. En Argentina, sólo un estudio relacionado a enfermedades cardiovasculares aplicó metodologías de imputación para tratar datos faltantes (Rubinstein *et al*, 2010). A pesar de que en los últimos años la metodología de Imputación Múltiple de datos se ha utilizado en la región, no se han encontrado trabajos que utilicen esta metodología en estudios sobre epidemiología del cáncer en América Latina. Es más, de los estudios de la región que utilizaron Imputación Múltiple ninguno de ellos presentó información relacionada a la calidad de los modelos de imputación propuestos. Los valores relativamente bajos de los Incrementos de la Varianza Relativos declarados en este trabajo son una estimación de la media de incremento relativo de la varianza de las estimaciones provocadas por los datos faltantes considerados. Idealmente estas estimaciones debieran ser próximas a cero (Acock, 2014). Se recomiendan

acciones tendientes a fortalecer la calidad de los estudios en la región, especialmente en el territorio del Cono Sur. En la región, por lo general, los estudios epidemiológicos tienen tamaños no muy grandes, por lo que debería evaluarse la posibilidad de que datos faltantes puedan estar influenciando los resultados.

En el presente trabajo se evaluó el error de clasificación a través del desempeño de la Frecuencia de Consumo Alimentario comparando el consumo de dos grupos de alimentos obtenidos de este instrumento con el derivado de recordatorios de 24 horas recolectados en el período de un año. Para estimar el consumo alimentario habitual se necesitan varios recordatorios de 24 horas, y esta cantidad depende de la variabilidad intra-sujeto de consumo del nutriente de interés (o nutrientes) y del nivel de precisión que se desea obtener de ese consumo habitual. Se reconoce que el número de días utilizados como parámetro a comparar es un factor que puede afectar la validez (Macedo-Ojeda *et al*, 2013). Willet (1998) ya recomendaba entre 3 y 10 recordatorios para obtener una caracterización general del consumo habitual de energía y macronutrientes; mayor cantidad para estimar consumo de nutrientes con mayor variabilidad entre día y día, como el colesterol por ejemplo. Estudios actuales también sugieren 3 réplicas para estimar la energía y nutrientes con buena precisión (Shamah-Levy *et al*, 2016; Ma *et al*, 2009). Para estimar el consumo habitual de nutrientes específicamente en poblaciones de mujeres adultas, la bibliografía sugiere entre 3 y 8 recordatorios de 24 horas dependiendo del nutriente de interés (Pereira *et al*, 2010). En el presente estudio, a la gran mayoría de las mujeres respondió a 7 o más recordatorios de 24 horas para obtener estimaciones precisas y contemplar esta variabilidad intra-sujeto; superando la media recomendada para este grupo poblacional.

Para la obtención de una única medida de consumo alimentario derivado de los múltiples recordatorios de 24 horas, se utilizaron modelos lineales generalizados mixtos. Éstos presentan ventajas para el cálculo de consumos habituales, cuando la distribución tiende a ser asimétrica o presentar valores extremos de consumos (Corrente y Fumes, 2016). Los resultados del análisis de sensibilidad sugieren valores aceptables para la recolección de información alimentaria de estos grupos de alimentos estudiados en la población femenina de Córdoba. La sensibilidad y especificidad para evaluar el consumo de los grupos de alimentos carnes procesadas y productos de pastelería son aceptables, aunque la especificidad es mayor en el caso del primer grupo de alimentos mencionado. Esto significa que el instrumento frecuencia de consumo alimentario utilizado tiene una mayor capacidad de identificar a los sujetos con bajo consumo “verdadero” (según el *Gold Standard*) de carnes

procesadas que de productos de pastelería. Estos valores son consistentes con otros encontrados en la literatura, especialmente respecto a la sensibilidad donde se reportan valores entre 70-86% (Mohammadifard *et al*, 2011; Clover *et al*, 2007); y en algunos trabajos, los valores de especificidad son menores a los de sensibilidad obtenidos en la presente tesis (Clover *et al*, 2007). En cuanto a la especificidad, relativamente baja, obtenida especialmente para el consumo de productos de pastelería, un aspecto a considerar es que las mujeres participantes tal vez hayan sub-declarado el consumo de alimentos pertenecientes a este grupo en la frecuencia de consumo alimentario. Un estudio de validación llevado a cabo en población adulta en México, obtuvo valores bajos de coeficientes de Pearson para un grupo de alimentos denominado “alimentos dulces”. Los grupos estudiados en ese trabajo podría considerarse como análogo al grupo de productos de pastelería (Macedo-Ojeda *et al*, 2013) Algunos factores que podrían explicar esto pueden estar asociados a que las mujeres suelen tener un mayor interés sobre alimentación y su relación con los procesos de salud-enfermedad (Ek, 2015). En el grupo de casos estudiados se observaron valores de especificidad especialmente bajos, lo que permite suponer que un mayor conocimiento podría haber conllevado a una sub-declaración del consumo al reconocer su potencial efecto en la salud. También deberían considerarse variables como la edad y el IMC en el análisis. Paalanen *et al* (2006) validaron un cuestionario de frecuencia de consumo alimentario con registros alimentarios y sugieren que los consumos reportados por el grupo de mujeres mayores a 50 años presentaron mayor correlación entre los instrumentos comparados y que el grupo de sujetos con IMC elevado tuvieron valores de correlaciones más bajos. En la presente tesis, aproximadamente un 40% de las mujeres presentó sobrepeso con una media de edad de 51 años.

Sería ideal en un estudio de validación obtener valores elevados de sensibilidad y especificidad para un instrumento, aunque en varias ocasiones obtener valores elevados de uno compromete al otro parámetro (Gordis, 2009). Cabe destacar que los análisis de validación del presente estudio se realizaron con grupos de alimentos que no se analizan tan frecuentemente como otros grupos (por ejemplo, frutas y verduras, o calcio como nutriente), como los trabajos citados anteriormente (Mohammadifard *et al*, 2011; Clover *et al*, 2007). No se encontraron trabajos que analizaran los grupos de alimentos de carnes procesadas y de productos de pastelería o equivalentes, con la misma metodología y/o poblaciones similares, aunque se señala existen estudios sobre validación de frecuencias de consumo alimentario, respecto a grupos de alimentos, aplicando otras metodologías, como análisis de correlaciones

de Spearman, Pearson o de correlación de interclases (Steinemann *et al*, 2017; Martínez *et al*, 2013; Shrestha *et al*, 2017). En general, se observó una validez aceptable para los grupos de alimentos analizados similares a los del presente estudio, principalmente para los grupos de carnes consumidas.

Los valores predictivos positivos y negativos obtenidos en el estudio de validación indicaron que el cuestionario de frecuencia de consumo alimentario es capaz de identificar el 75-77% de los sujetos que tuvieron un consumo elevado de carnes procesadas o de productos de pastelería, es decir, considerados expuestos; y, puede identificar entre un 63 a 66% de los sujetos no expuestos (consumo bajo de estos grupos de alimentos). Esto indicaría que el instrumento tiene mayor habilidad para clasificar como expuesto a alguien realmente expuesto que a uno no expuesto, siendo éste realmente un sujeto no expuesto. Rangos similares de valores predictivos se obtuvieron en el estudio de Mohammadifard y otros colaboradores (2011). Este estudio se realizó en una ciudad de Irán, en el cual se incluyeron 123 adultos sanos, evaluándose la reproducibilidad y validez de un cuestionario de frecuencia de consumo alimentario sobre el consumo de los grupos de alimentos frutas y verduras específicamente. Otro estudio llevado a cabo en Puerto Rico (Palacios *et al*, 2012), donde validaron un cuestionario de frecuencia de consumo, incluyeron a 93 adultos para evaluar el consumo del nutriente calcio, encontrando valores predictivos positivos de 86% y negativos de 45%.

Algunas de las limitaciones identificadas en el presente trabajo se mencionan a continuación. Una de ellas es la falta de información respecto la clasificación del tipo de tumor según receptores hormonales para el estudio de cáncer de mama y la falta de discriminación entre sitios tumorales para el de cáncer colorrectal. Otra limitación es el tamaño del estudio, lo que hace imperativo utilizar la mayor cantidad de información posible controlando por el sesgo de información derivado de datos faltantes. En el estudio de cáncer de mama se demuestra que el patrón dietario Tradicional, el IMC y la lactancia materna están asociados a la ocurrencia de cáncer de mama en la población femenina Argentina cuando se aplica la metodología de Imputación Múltiple de datos. En el estudio de cáncer colorrectal, mejoran las estimaciones de riesgo luego de aplicar esta metodología. Así, se demuestran los beneficios de aplicar este método en bases de datos de epidemiología del cáncer con elevados porcentajes de datos faltantes en ciertas covariables. En la región de América del Sur, los tamaños de los estudios epidemiológicos caso control sobre cáncer varían mucho,

siendo en la mayoría de éstos una limitación. Algunos estudios en Chile o Colombia incluyeron desde 300-350 sujetos (Martinez *et al*, 2008; Pereira *et al*, 2012), un estudio llevado a cabo en Paraguay y otro en Brasil, enrolaron cada uno 500 sujetos aproximadamente (Felden y Figueiredo, 2011; Rolón *et al*, 1995). En Argentina, algunos estudios en Córdoba, incluyeron 120 sujetos y otros casi 1000 sujetos (Becaria Coquet *et al*, 2016; Pou *et al*, 2014; Steinmaus *et al*, 2010). Esfuerzos para publicar estudios con tamaños mayores se realizan en la región, entre las posibles estrategias utilizadas, se cita el incluir varios estudios similares provenientes de distintas ciudades o países en el estudio, alcanzando tamaños de aproximadamente 4 mil (incluyendo Brasil, Argentina, Paraguay, Uruguay, Cuba) (Lubin *et al*, 2014; Szymańska *et al*, 2011); o estudiar varios sitios tumorales en conjunto llegando a 13 mil sujetos por ejemplo en una investigación de origen uruguayo (De Stefani *et al*, 2004). Además, pocas veces se declara si el total de los sujetos del estudio son los que finalmente se consideran en los análisis de riesgo, o de hecho se descartan algunos por falta de datos o desbalanceamiento.

Otra debilidad del presente trabajo es no haber profundizado en la aplicación de la metodología de diagramas causales para abordar el sesgo de selección. Este método hubiera aportado a una mayor comprensión de las relaciones causales entre las variables de interés y posiblemente brindado mayores herramientas al momento de explicar posibles escenarios, mediante las metodologías cuantitativas, como las utilizadas en la presente tesis.

El estudio de validación utilizó como método *Gold Standard* el recordatorio de 24 horas, instrumento muy usado en la práctica, aunque no es considerado el ideal para referencia (Lachat *et al*, 2016), ya que en el ámbito de la epidemiología nutricional no existe un método perfecto para evaluar la alimentación. Por esto, en esta área, los estudios de validación deben utilizar un *Gold Standard* que sea un método superior al instrumento a evaluar. Teniendo esto presente, se decidió aplicar múltiples recordatorios de 24 horas con el objetivo de obtener con mayor precisión información sobre la dieta habitual del último año de las mujeres participantes. Esto es una característica metodológica a destacar de este estudio, ya que algunos sujetos respondieron hasta 12 recordatorios, en un intento de captar más genuinamente el hábito de la alimentación en esta muestra. Cuando se comparan dos instrumentos considerados “no perfectos” para evaluar la alimentación es importante garantizar que los posibles errores derivados de ambos instrumentos sean lo más independiente posibles. Esto evitaría obtener estimaciones sesgadas de elevada validez que harían suponer buena calidad del instrumento. Por ello, los encuestadores fueron entrenados

y realizaron la aplicación de ambos instrumentos utilizando cuestionarios validados. Además, este proceso fue acompañado de un atlas fotográfico para la cuantificación de porciones, facilitando la estimación de las mismas y evitando errores de interpretación de los encuestados y/o dificultades al momento de recordar (Ferrari, 2013). En este aspecto específico de la validación, una limitación del presente estudio fue el tamaño muestral (n=57). La bibliografía recomienda usar estudios con un mínimo de 110 sujetos para estudios de validación (Willet, 1998). No obstante, otros autores también han trabajado con tamaños similares al de este estudio. Steinman *et al* (2017), condujeron un estudio de validación en Alemania, donde incluyeron 56 sujetos adultos sanos y compararon la frecuencia de consumo alimentario con registros alimentarios de 4 días (Steinmann *et al*, 2017). Otro estudio, en México (Macedo-Ojeda *et al*, 2013), incluyó 97 adultos.

En el contexto socioeconómico particular de Argentina, dadas las condiciones epidemiológicas y demográficas, junto con los cambios nutricionales que se identifican en su población y de características particulares, se torna indispensable continuar realizando investigaciones sobre epidemiología nutricional del cáncer en la región. Este marco analítico sobre análisis de sesgos específicos, como los presentados en este trabajo, pueden aplicarse a estudios que consideren a la exposición alimentaria como factor de riesgo, trabajando con otros patrones o grupos de alimentos específicos contruidos particularmente para el estudio de ciertos sitios tumorales. También se puede aplicar a nutrientes específicos, por ejemplo para cáncer de recto o colón independientemente, o trasladar estas metodologías a investigaciones sobre otras patologías crónicas prevalentes en la región, como las enfermedades cardiovasculares, diabetes, obesidad, Chagas, entre otras. Aportar a la red causal de este conjunto de patologías llevando a cabo estudios de calidad metodológica debería ser un objetivo transversal a todas las investigaciones que se conducen en el país y el resto de la región, obteniendo como resultado derivado recomendaciones alimentarias-nutricionales que se traduzcan en acciones preventivas específicas para estos grupos poblacionales.

Conclusiones

Los estudios sobre epidemiología nutricional y cáncer son muy relevantes, ya que esta patología se relaciona con los hábitos alimentarios y representa una importante carga de enfermedad en la región, y en el mundo. El presente estudio permitió establecer un marco metodológico para cuantificar errores sistemáticos en los estudios de tipo caso-control de cáncer de mama y colorrectal en la provincia de Córdoba. Los aportes de este trabajo de investigación radican en mejoras en la validez de los resultados derivados de estos estudios respecto al rol de la alimentación en la etiología de esta enfermedad y, de esta manera, contribuyen a la calidad de los mismos y, por ende fortalecen e incorporan conocimiento sobre epidemiología nutricional y cáncer a nivel local.

En términos generales, los hallazgos derivados de la presente tesis generan un enfoque original respecto a la calidad de las metodologías aplicadas al diseño y análisis de los estudios caso-control en cáncer, no solamente en cáncer de mama y colorrectal, coincidiendo en parte con resultados de otros estudios. Se aporta conocimiento acerca de las posibles hipótesis sobre los mecanismos etiopatogénicos de estas enfermedades. Entre los hallazgos más importantes, se pueden mencionar:

- Las metodologías propuestas para enfrentar el fenómeno de confusión para variables observadas no condujeron a identificar efectos importantes de confusión en relación a las variables edad, lactancia materna, hijos, nivel socioeconómico, edad de la menarca, IMC, actividad física, estado ginecológico (confundentes propuestas) en el estudio de cáncer de mama. En cambio, en el estudio de cáncer colorrectal, se identificaron efectos confundentes de las variables sexo y nivel socioeconómico. Esto aporta conocimientos en lo que respecta al diseño de futuros estudios en la región, alertando sobre la consideración de estas variables para posibles análisis por estratos, entre otros.

- El efecto confundente de variables no medidas propuestas, a saber: glucemia y aumento de peso durante la vida adulta para los estudios de cáncer de mama y colorrectal respectivamente; no mostró cambios sustantivos en las estimaciones de riesgo. No obstante, se manifestó un posible efecto confusor al imponer un escenario extremo de una importante variabilidad y elevada correlación con la variable registrada (covariable IMC).

- En cuanto al sesgo de selección, la metodología propuesta permitió plantear diferentes escenarios posibles. De las situaciones más probables, las diferencias encontradas con las estimaciones convencionales fueron mínimas, afectando en parte la magnitud de las asociaciones y la precisión de los rangos intercuartílicos, no así la dirección de las estimaciones. Situaciones similares se dieron en ambos estudios caso-control.

- La utilización de la metodología de Imputación Múltiple de datos para afrontar el sesgo de información por falta de datos en ambos estudios proveyó varios beneficios. En cáncer de mama, se mostró que el patrón dietario Tradicional, el IMC y la lactancia materna están asociados a la ocurrencia de la patología en la población femenina de Córdoba. Ya en el estudio de cáncer colorrectal, a pesar de no haber identificado nuevos efectos en covariables luego de aplicar ésta metodología, los efectos de las variables identificados a saber: patrón dietario Cono Sur, sexo, IMC, nivel socioeconómico y antecedentes de primer grado de cáncer colorrectal, fueron más precisos, se atenuaron y consideraron mayor información del estudio, luego del uso de Imputación Múltiple.

- Para cáncer de mama, el error de clasificación de la exposición se afrontó mediante una metodología sencilla y los resultados obtenidos indicaron una clasificación de la exposición para los grupos de consumo de carnes procesadas y productos de pastelería aceptable por parte del instrumento de recolección de información alimentario-nutricional utilizado en estos estudios (frecuencia de consumo alimentario).

En el campo de la Epidemiología Nutricional y la salud pública es de gran importancia poder traducir los hallazgos en recomendaciones. Se espera que el conocimiento derivado de esta tesis permita elevar la calidad de los estudios observacionales llevados a cabo en la provincia y permitan el diseño de recomendaciones alimentarias, que orienten a las políticas públicas de salud y nutrición, con el fin de mejorar las condiciones de salud de las poblaciones a través de la alimentación, hábito éste modificable. Se anhela también que estos hallazgos puedan considerarse en pro de la calidad de los estudios caso-control en particular y, observacionales en general, provenientes de la región. También, que estos nuevos conocimientos junto con el diseño de estudios que incorporen desde el inicio un análisis de

sensibilidad, sean considerados y valorados por la comunidad científica, y para futuros estudios de revisión internacionales, de los cuales se desprenden lineamientos y recomendaciones para las poblaciones.

Finalmente, del análisis de la presente tesis se indican posibles temáticas de interés o líneas a profundizar en futuras investigaciones a abordar con el fin de obtener estimaciones válidas en los estudios sobre epidemiología nutricional del cáncer y aportar a la red causal de esta enfermedad. Entre ellas:

- Aplicar protocolo de análisis de errores sistemáticos a los estudios caso-control de cáncer de próstata y tumores de vías urinarias del GEACC.

- Diseñar y llevar a cabo estudios sobre errores sistemáticos específicos para el estudio de cohorte de mujeres con cáncer de mama y para los estudios poblacionales con diferentes grupos poblacionales (adultos y niños escolares) del equipo GEACC.

- Profundizar en los análisis del sesgo de selección mediante diagramas causales y enriquecer lo trabajado mediante metodología cuantitativa para su análisis conjunto.

- Ahondar sobre el error de clasificación en la exposición de los grupos de alimentos representantes de los patrones alimentarios identificados en la población de la provincia de Córdoba mediante la metodología propuesta y otras.

CAPÍTULO 5. BIBLIOGRAFIA

Aarts MJ, Lemmens VE, Louwman MW, Kunst AE, Coebergh JW. Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. *Eur J Cancer*. 2010; 46(15):2681-95. doi: 10.1016/j.ejca.2010.04.026.

Aballay LR, Eynard AR, Díaz Mdel P, Navarro A, Muñoz SE. Overweight and obesity: a review of their relationship to metabolic syndrome, cardiovascular disease, and cancer in South America. *Nutr Rev*. 2013; 71(3):168-79. doi: 10.1111/j.1753-4887.2012.00533.x.

Aballay LR, Osella AR, De La Quintana AG, Diaz MDP. Nutritional profile and obesity: results from a random-sample population-based study in Córdoba, Argentina. *Eur J Nutr*. 2016; 55(2):675-685. doi: 10.1007/s00394-015-0887-0.

Abid Z, Cross AJ, Sinha R. Meat, dairy, and cancer. *Am J Clin Nutr*. 2014; 100(1):386S-93S. doi: 10.3945/ajcn.113.071597.

Acock AC. Working with Missing Values. *J Marriage Fam*. 2005; 67(4): 1012–1028.

Acock AC. Working with missing values-multiple imputation. In *A Gentle Introduction to Stata* (Fourth edition.). College Station, Texas, USA: Stata Press, 2014.

Agha RA, Lee SY, Jeong KJ, Fowler AJ, Orgill DP. Reporting Quality of Observational Studies in Plastic Surgery Needs Improvement: A Systematic Review. *Ann Plast Surg*. 2016; 76(5):585-9. doi: 10.1097/SAP.0000000000000419.

Agogo GO, van der Voet H, van 't Veer P, Ferrari P, Muller DC, Sánchez-Cantalejo E, Bamia C, Braaten T, Knüppel S, Johansson I, van Eeuwijk FA, Boshuizen HC. A method for sensitivity analysis to assess the effects of measurement error in multiple exposure variables using external validation data. *BMC Med Res Methodol*. 2016 Oct 13;16(1):139.

Alejos B, Castelló A. Métodos de Imputación simple y múltiple. Instituto de Salud Carlos III. Ministerio de Economía y Competitividad, 2015.

Alfaro R, Fuenzalida M. Imputación Múltiple en Encuestas Microeconómicas. Cuadernos de Economía, vol. 46, 2009, p.273-288.

Allison PD. Chapter 4: Missing Data. Chapter 4: Missing Data. En: The SAGE Handbook of Quantitative Methods in Psychology. Editores: Millsap RE, Maydeu-Olivares A.SAGE Publications. 2009. doi: <http://dx.doi.org/10.4135/9780857020994.n4>

Allison PD. Multiple Imputation for Missing Data: A cautionary tale. Socio. Meth. Res. 2000; 28(3):301-309.

Amadou A, Torres-Mejía G, Hainaut P, Romieu I. Breast cancer in Latin America: global burden, patterns, and risk factors. Salud pública Mex. 2014; 56(5):547-54.

American Cancer Society. Acerca del cáncer de colon y recto. 2016. [Internet]. [Consultado el 6/Mar/2017]. Disponible en:<https://www.cancer.org/es/cancer/cancer-de-colon-o-recto/acerca/que-es-cancer-de-colon-o-recto.html#referencias>

American Cancer Society. Acerca del cáncer de seno. 2016. [Internet]. [Consultado el 6/Mar/2017]. Disponible en:<https://www.cancer.org/es/cancer/cancer-de-seno/acerca/que-es-el-cancer-de-seno.html>

Archer E, Hand GA, Blair SN. Validity of U.S. nutritional surveillance: National Health and Nutrition Examination Survey caloric energy intake data, 1971-2010. PLoS One. 2013; 8(10):e76632. doi: 10.1371/journal.pone.0076632.

Asociación Argentina de Marketing (AAM) y Cámara de Control de Audiencia (CCMA). Índice de Nivel Socioeconómico 2002. [Internet]. [Consultado el 6/Mar/2017]. Disponible en: <http://www.iprofesional.com/adjuntos/documentos/18/0001836.pdf>.

Aune D, Chan DS, Greenwood DC, Vieira AR, Rosenblatt DA, Vieira R, Norat T. Dietary fiber and breast cancer risk: a systematic review and meta-analysis of prospective studies. *Ann Oncol.* 2012; 23(6):1394-402. doi: 10.1093/annonc/mdr589.

Aune D, Lau R, Chan DS, Vieira R, Greenwood DC, Kampman E, Norat T. Dairy products and colorectal cancer risk: a systematic review and meta-analysis of cohort studies. *Ann Oncol.* 2012; 23(1):37-45. doi: 10.1093/annonc/mdr269.

Aykan NF. Red Meat and Colorectal Cancer. *Oncol Rev.* 2015 ; 9(1):288. doi: 10.4081/oncol.2015.288.

Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res.* 2011; 20(1):40-9. doi: 10.1002/mpr.329.

Babaknejad N1, Sayehmiri F, Sayehmiri K, Rahimifar P, Bahrami S, Delpesheh A, Hemati F, Alizadeh S. The relationship between selenium levels and breast cancer: a systematic review and meta-analysis. *Biol Trace Elem Res.* 2014; 159(1-3):1-7. doi: 10.1007/s12011-014-9998-3.

Baena R, Salinas P. Diet and colorectal cancer. *Maturitas.* 2015; 80(3):258-64. doi: 10.1016/j.maturitas.2014.12.017.

Banack HR, Kaufman JS. Does selection bias explain the obesity paradox among individuals with cardiovascular disease? *Ann Epidemiol.* 2015; 25(5):342-9. doi: 10.1016/j.annepidem.2015.02.008.

Barclay AW, Petocz P, McMillan-Price J, Flood VM, Prvan T, Mitchell P, Brand-Miller JC. Glycemic index, glycemic load, and chronic disease risk--a meta-analysis of observational studies. *Am J Clin Nutr.* 2008; 87(3):627-37.

Benjet C, Borges G, Medina-Mora ME. DSM-IV personality disorders in Mexico: results from a general population survey. *Rev Bras Psiquiatr.* 2008; 30, 227–34.

Bennett DA. How can I deal with missing data in my study? *Aust N Z J Public Health.* 2001; 25, 464–69.

Benz CC. Impact of aging on the biology of breast cancer. *Crit Rev Oncol Hematol*. 2008; 66(1):65-74.

Bingham SA, Norat T, Moskal A, Ferrari P, Slimani N, Clavel-Chapelon F, Kesse E, Nieters A, Boeing H, Tjønneland A, Overvad K, Martinez C, Dorronsoro M, González CA, Ardanaz E, Navarro C, Quirós JR, Key TJ, Day NE, Trichopoulou A, Naska A, Krogh V, Tumino R, Palli D, Panico S, Vineis P, Bueno-de-Mesquita HB, Ocké MC, Peeters PH, Berglund G, Hallmans G, Lund E, Skeie G, Kaaks R, Riboli E. Is the association with fiber from foods in colorectal cancer confounded by folate intake?. *Cancer Epidemiol Biomarkers Prev*. 2005; 14(6):1552-6.

Breslow NE, Day NE. *Statistical Methods in Cancer Research. Vol. I. The analysis of case-control studies*. IARC Scientific Publications No. 32. Switzerland, Lyon: IARC; 1980. 338 p.

Boffetta P. Epidemiology of environmental and occupational cancer. *Oncogene*. 2004; 23(38):6392-403.

Bosma-den Boer MM, van Wetten ML, Pruimboom L. Chronic inflammatory diseases are stimulated by current lifestyle: how diet, stress levels and medication prevent our body from recovering. *Nutr Metab (Lond)*. 2012; 9(1):32. doi: 10.1186/1743-7075-9-32.

Boyle T, Fritschi L, Tabatabaei SM, Ringwald K, Heyworth JS. Smoking, alcohol, diabetes, obesity, socioeconomic status, and the risk of colorectal cancer in a population-based case-control study. *Cancer Causes Control*. 2014; 25(12):1659-68. doi: 10.1007/s10552-014-0470-7.

Buis, Maarten L. Chapter 7, The consequences of unobserved heterogeneity in a sequential logit model, En: Buis, Maarten L. *Inequality of Educational Outcome and Inequality of Educational Opportunity in the Netherlands during the 20th Century*. PhD thesis. 2010. Disponible en: http://www.maartenbuis.nl/dissertation/chap_7.pdf

Camargos VP1, César CC, Caiaffa WT, Xavier CC, Proietti FA. Imputação múltipla e análise de casos completos em modelos de regressão logística: uma avaliação prática do impacto das perdas em covariáveis. *Cad Saude Publica*. 2011; 27, 2299–313.

Cerda, Jaime, Vera, Claudio, & Rada, Gabriel. Odds ratio: aspectos teóricos y prácticos. *Revista médica de Chile*. 2013; 141(10): 1329-35. <https://dx.doi.org/10.4067/S0034-98872013001000014>.

Chajès V, Romieu I. Nutrition and breast cancer. *Maturitas*. 2014; 77(1):7-11. doi: 10.1016/j.maturitas.2013.10.004.

Chen Q, Wang J, Yang J, Jin Z, Shi W, Qin Y, Yu F, He J. Association between adult weight gain and colorectal cancer: a dose-response meta-analysis of observational studies. *Int J Cancer*. 2015; 136(12):2880-9. doi: 10.1002/ijc.29331.

Chen Y, Wen YY, Li ZR, Luo DL, Zhang XH. The molecular mechanisms between metabolic syndrome and breast cancer. *Biochem Biophys Res Commun*. 2016;471(4):391-5. doi: 10.1016/j.bbrc.2016.02.034.

Chiavarini M, Minelli L, Fabiani R. Garlic consumption and colorectal cancer risk in man: a systematic review and meta-analysis. *Public Health Nutr*. 2016; 19(2):308-17. doi: 10.1017/S1368980015001263.

Choi YJ, Myung SK, Lee JH. Light Alcohol Drinking and Risk of Cancer: A Meta-analysis of Cohort Studies. *Cancer Res Treat*. 2017. doi: 10.4143/crt.2017.094.

Civetta, MT, Civetta, JD. Carcinogénesis. *Salud Pública de Méx*, 2011; 53(5): 405-414.

Colditz GA, Sellers TA, Trapido E. Epidemiology - identifying the causes and preventability of cancer? *Nat Rev Cancer*. 2006;6(1):75-83.

Cooke AA, Connaughton RM, Lyons CL, McMorrow AM, Roche HM. Fatty acids and chronic low grade inflammation associated with obesity and the metabolic syndrome. *Eur J Pharmacol*. 2016; 15; 785:207-14. doi: 10.1016/j.ejphar.2016.04.021.

Corrente JE y Fumes G. Use of Asymmetric Models to Adjust the Vitamin Intake Distribution Data for Older People. *Health*, 2016;8: 887-893.

Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. 1959. *Int J Epidemiol*. 2009 Oct;38(5):1175-91. doi: 10.1093/ije/dyp289.

Crosara Teixeira M, Braghiroli MI, Sabbaga J, Hoff PM. Primary prevention of colorectal cancer: myth or reality? *World J Gastroenterol*. 2014; 20(41):15060-9. doi: 10.3748/wjg.v20.i41.15060.

De La Quintana AG, Pou SA, Forte CA, Aballay LR. Ingesta alimentaria, estado nutricional y estilos de vida pro-inflamatorios en una cohorte de mujeres con cáncer de mama de la ciudad de Córdoba, Argentina. *Revista de Nutrición de la CONFELANYD*. 2014; 4: 49-63.

Di Sibio A, Abriata G, Forman D, Sierra MS. Female breast cancer in Central and South America. *Cancer Epidemiol*. 2016; 44 (1):S110-S120. doi: 10.1016/j.canep.2016.08.010.

Diaz MP, Corrente JE, Osella AR, Muñoz SE, Aballay LR. Modelling spatial distribution of cancer incidence in Cordoba, Argentina. *Applied Cancer Research* 2010;30(2)245-252.

Dermadi D, Valo S, Ollila S, Soliymani R, Sipari N, Pussila M, Sarantaus L, Linden J, Baumann M, Nyström M. Western Diet Deregulates Bile Acid Homeostasis, Cell Proliferation, and Tumorigenesis in Colon. *Cancer Res*. 2017. doi: 10.1158/0008-5472.CAN-16-2860.

Dirección de Estadísticas e Información de Salud. Estadísticas Vitales. Información Básica. Año 2013 (No. 5(57)). Ministerio de Salud. Presidencia de la Nación. [Internet]. Disponible en: <http://www.deis.msal.gov.ar/wp-content/uploads/2016/01/Serie5Nro58.pdf>

Doll, R.-(1967) *The prevention of cancer: pointers from epidemiology*. Inglaterra: Nuffield Provincial Hospitals Trust.

Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst*. 1981; 66(6):1191-308.

Dong JY, Qin LQ. Dietary glycemic index, glycemic load, and risk of breast cancer: meta-analysis of prospective cohort studies. *Breast Cancer Res Treat*. 2011; 126(2):287-94. doi: 10.1007/s10549-011-1343-3.

Draper NR, Smith H. Applied Regression Analysis. 3er ed. Estados Unidos: John Wiley & Sons, Inc.; 1998. p. 736.

Doubeni CA, Laiyemo AO, Major JM, Schootman M, Lian M, Park Y, Graubard BI, Hollenbeck AR, Sinha R. Socioeconomic status and the risk of colorectal cancer: an analysis of more than a half million adults in the National Institutes of Health-AARP Diet and Health Study. *Cancer*. 2012; 118(14):3636-44. doi: 10.1002/cncr.26677.

Eddings W, Marchenko Y. Diagnostics for multiple imputation in Stata. *Stata J*. 2012; 12, 353–67.

Edefonti V, Decarli A, La Vecchia C, Bosetti C, Randi G, Franceschi S, Dal Maso L, Ferraroni M. Nutrient dietary patterns and the risk of breast and ovarian cancers. *Int J Cancer* 2008; 122(3):609-13. doi: 10.1002/ijc.23064.

Eekhout I, de Boer RM, Twisk JW, de Vet HC, Heymans MW. Missing data: a systematic review of how they are reported and handled. *Epidemiology*. 2012; 23(5):729-32. doi: 10.1097/EDE.0b013e3182576cdb.

Eekhout I, de Vet HC, Twisk JW, Brand JP, de Boer MR, Heymans MW. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *J Clin Epidemiol*. 2014; 67(3):335-42. doi: 10.1016/j.jclinepi.2013.09.009.

Ek S. Gender differences in health information behaviour: a Finnish population-based survey. *Health Promot Int*. 2015; 30(3):736-45. doi: 10.1093/heapro/dat063.

Elgart JF, Caporale JE, Asteazarán S, De La Fuente JL, Camilluci C, Brown JB, González CD, Gagliardino JJ. Association between socioeconomic status, type 2 diabetes and its chronic complications in Argentina. *Diabetes Res Clin Pract*. 2014; 104(2):241-7. doi: 10.1016/j.diabres.2014.02.010.

Felden JB., Figueiredo AC. Distribution of body fat and breast cancer: a case-control study in the South of Brazil. *Cien Saude Colet*. 2011; 16(5):2425-33.

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. [Internet]. Lyon, France: Centro Internacional de Investigaciones sobre el Cáncer; 2013. . Disponible en: <http://globocan.iarc.fr>

Ferrante D, Linetzky B, Konfino J, King A, Virgolini M, Laspiur S. Encuesta Nacional de factores de riesgo 2009: evolución de la epidemia de enfermedades crónicas no transmisibles en Argentina. Estudio de corte transversal. *Rev Argent Salud Pública* 2011; 2:34-41.

Ferrari, MA. Estimación de la Ingesta por Recordatorio de 24 Horas. *Diaeta*. 2013; 31(143), 20-25.

Fries L, Grogan-Kaylor A, Bares C, Han Y, Delva J. Gender differences in predictors of self-reported physical aggression: Exploring theoretically relevant dimensions among adolescents from Santiago, Chile. *Int Perspect Psychol*. 2013; 2, 255–68.

Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2005; 34(6):1370-6.

Fung AE, Palanki R, Bakri SJ, Depperschmidt E, Gibson A. Applying the CONSORT and STROBE statements to evaluate the reporting quality of neovascular age-related macular degeneration studies. *Ophthalmology*. 2009; 116(2):286-96. doi: 10.1016/j.ophtha.2008.09.014.

Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab*. 2012 Spring;10(2):486-9. doi: 10.5812/ijem.3505.

Geneletti S, Richardson S, Best N. Adjusting for selection bias in retrospective, case-control studies. *Biostatistics*. 2009; 10(1):17-31. doi: 10.1093/biostatistics/kxn010.

Go Y, Chung M, Park Y. Dietary Patterns for Women With Triple-negative Breast Cancer and Dense Breasts. *Nutr Cancer*. 2016;68(8):1281-1288.

Gordis L. Chapter 5: Assessing the validity and reliability of diagnostic and screening tests. En: *Epidemiology*. Estados Unidos: 4ta ed. Saunders Elsevier.2009.p85-108.

Gordis L. *Epidemiology*. 4ta Ed. Saunders Elsevier. Estados Unidos, 2009. 375 p.

Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol*. 1996; 25(6):1107-16.

Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol*. 1986;15(3):413-9.

Grigsby M, Siddharthan T, Chowdhury MA, Siddiquee A, Rubinstein A, Sobrino E, Miranda JJ, Bernabe-Ortiz A, Alam D, Checkley W. Socioeconomic status and COPD among low- and middle-income countries. *Int J Chron Obstruct Pulmon Dis*. 2016; 11:2497-2507.

Groenwold RH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies. *J Clin Epidemiol*. 2009;62(1):22-8. doi: 10.1016/j.jclinepi.2008.02.011.

Groenwold RH, Van Deursen AM, Hoes AW, Hak E. Poor quality of reporting confounding bias in observational intervention studies: a systematic review. *Ann Epidemiol*. 2008; 18(10):746-51. doi: 10.1016/j.annepidem.2008.05.007.

Hammer GP, du Prel JB, Blettner M. Avoiding bias in observational studies: part 8 in a series of articles on evaluation of scientific publications. *Dtsch Arztebl Int*. 2009; 106(41):664-8. doi: 10.3238/arztebl.2009.0664.

Hammerling U, Bergman Laurila J, Grafström R, Ilbäck NG. Consumption of Red/Processed Meat and Colorectal Carcinoma: Possible Mechanisms Underlying the Significant Association. *Crit Rev Food Sci Nutr*. 2016; 56(4):614-34. doi: 10.1080/10408398.2014.972498.

Harvie M, Howell A, Evans DG. Can diet and lifestyle prevent breast cancer: what is the evidence? *Am Soc Clin Oncol Educ Book*. 2015:e66-73. doi: 10.14694/EdBook_AM.2015.35.e66.

Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004; 15(5):615-25.

Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 2002; 155(2):176-84.

Hernán MA, Robins JM. Causal Inference. Estados Unidos, Boca Ratón: Chapman & Hall/CRC, forthcoming. 2018. p.130.

Hernández-Avila M, Garrido F, Salazar-Martínez E. Sesgos en estudios epidemiológicos. Salud pública de Méx. 2000; 42(5).

Hernández-Díaz S, Wilcox AJ, Schisterman EF, Hernán MA. From causal diagrams to birth weight-specific curves of infant mortality. Eur J Epidemiol. 2008; 23(3):163-6. doi: 10.1007/s10654-007-9220-4.

Hernández G, Moriña D, Navarro A. Imputación de valores ausentes en salud pública: conceptos generales y aplicación en variables dicotómicas. Gac Sanit. 2017;31(4):342-5. doi: 10.1016/j.gaceta.2017.01.001

Hotamisligil GS. Inflammation and metabolic disorders. Nature. 2006; 444(7121):860-7.

Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ Jr. Selection Bias Due to Loss to Follow Up in Cohort Studies. Epidemiology. 2016; 27(1):91-7. doi: 10.1097/EDE.0000000000000409.

Ibrahim, JG, Chu H, Chen MH. Missing Data in Clinical Studies: Issues and Methods. J Clin Oncol. 2012; 30, 3297–303.

Instituto Nacional de Estadística y Censos. Informes Técnicos. Vol. 1, nº 53. Condiciones de vida. Vol. 1, nº 4. Incidencia de la pobreza y la indigencia en 31 aglomerados urbanos. Segundo semestre de 2016. ISSN 2545-6768. Argentina: Instituto Nacional de Estadística y Censos; 2017. 13 p.

Barker B, Degenhardt L. Accidental drug-induced deaths in Australia 1997- 2001. Sydney (Australia): University of New South Wales, National Drug and Alcohol Research Centre; 2003. 46 p.

Jamshidinaeini Y, Akbari ME, Abdollahi M, Ajami M, Davoodi SH. Vitamin D Status and Risk of Breast Cancer in Iranian Women: A Case-Control Study. J Am Coll Nutr. 2016; 35(7):639-646.

Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis. 3ra. ed. Estados Unidos, Nueva York: Prentice-Hall. 1992.

Kac G, Sichieri R, Petrucci Gigante D. *Epidemiología Nutricional*. Rio de Janeiro: Ed. Fiocruz/Atheneu; 2007. 579 p.

Kahan BC, Rehal S, Cro S. Risk of selection bias in randomised trials. *Trials*. 2015; 16:405. doi: 10.1186/s13063-015-0920-x.

Kamangar F. Confounding variables in epidemiologic studies: basics and beyond. *Arch Iran Med*. 2012; 15(8):508-16. doi: 012158/AIM.0014.

Karimi Z, Jessri M, Houshiar-Rad A, Mirzaei HR, Rashidkhani B. Dietary patterns and breast cancer risk among women. *Public Health Nutr*. 2014; 17(5):1098-106. doi: 10.1017/S1368980013001018.

Keum N, Greenwood DC, Lee DH, Kim R, Aune D, Ju W, Hu FB, Giovannucci EL. Adult weight gain and adiposity-related cancers: a dose-response meta-analysis of prospective observational studies. *J Natl Cancer Inst*. 2015; 107(2). doi: 10.1093/jnci/djv088.

Kim SE, Paik HY, Yoon H, Lee JE, Kim N, Sung MK. Sex- and gender-specific disparities in colorectal cancer risk. *World J Gastroenterol*. 2015; 21(17):5167-75. doi: 10.3748/wjg.v21.i17.5167.

Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. *Am J Epidemiol*. 2008; 168(4):355-7. doi: 10.1093/aje/kwn071.

Klurfeld DM. Research gaps in evaluating the relationship of meat and health. *Meat Sci*. 2015; 109:86-95. doi: 10.1016/j.meatsci.2015.05.022.

La Caze A. Why randomized interventional studies. *J Med Philos*. 2013;38(4):352-68. doi: 10.1093/jmp/jht028.

Lachat C, Hawwash D, Ocké MC, Berg C, Forsum E, Hörnell A, Larsson C, Sonestedt E, Wirfält E, Åkesson A, Kolsteren P, Byrnes G, De Keyzer W, Van Camp J, Cade JE, Slimani N, Cevallos M, Egger M, Huybrechts I. Strengthening the Reporting of Observational Studies in Epidemiology-Nutritional Epidemiology (STROBE-nut): An Extension of the STROBE Statement. *PLoS Med*. 2016; 13(6):e1002036. doi: 10.1371/journal.pmed.1002036.

Lajous M, Willett W, Lazcano-Ponce E, Sanchez-Zamorano LM, Hernandez-Avila M, Romieu I. Glycemic load, glycemic index, and the risk of breast cancer among Mexican women. *Cancer Causes Control*. 2005; 16(10):1165-9.

Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology*. 2003; 14(4):451-8.

Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic studies. Springer, *Statistics for Biology and Health*. New York; 2009, p. 192.

Lash TL, Silliman RA. A sensitivity analysis to separate bias due to confounding from bias due to predicting misclassification by a variable that does both. *Epidemiology*. 2000; 11(5):544-9.

Last JM. *A dictionary of Epidemiology*. 4ta ed. Estados Unidos: Oxford University Press. 2001.

Lazcano-Ponce E, Salazar-Martínez E, Hernández-Ávila M. Estudios epidemiológicos de casos y controles. Fundamento teórico, variantes y aplicaciones. *Salud pública de Méx*, 2001; 43(2).

Lee W, Bindman J, Ford T, Glozier N, Moran P, Stewart R, Hotopf M. Bias in psychiatric case-control studies: literature survey. *Br J Psychiatry*. 2007; 190:204-9.

Lee JA, Yoo JE, Park HS. Metabolic syndrome and incidence of breast cancer in middle-aged Korean women: a nationwide cohort study. *Breast Cancer Res Treat*. 2017; 162(2):389-393. doi: 10.1007/s10549-017-4131-x.

Libby P. Inflammatory mechanisms: the molecular basis of inflammation and disease. *Nutr Rev*. 2007; 65(12 Pt 2):S140-6.

Linnet MS. Evolution of cancer epidemiology. *Epidemiol Rev*. 2000; 22(1):35-56.

Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 2da ed. Estados Unidos:Wiley. 2002. p.389.

Liu Y, Nguyen N, Colditz GA. Links between alcohol consumption and breast cancer: a look at the evidence. *Womens Health (Lond)*. 2015; 11(1):65-77. doi: 10.2217/whe.14.62.

Lubin JH, De Stefani E, Abnet CC, Acosta G, Boffetta P, Victora C, Graubard BI, Muñoz N, Deneo-Pellegrini H, Franceschi S, Castellsagué X, Ronco AL, Dawsey SM. Maté drinking and esophageal squamous cell carcinoma in South America: pooled results from two large multicenter case-control studies. *Cancer Epidemiol Biomarkers Prev.* 2014; 23(1):107-16. doi: 10.1158/1055-9965.EPI-13-0796.

Luiz RR, Cabral Borges MD. Sensitivity analysis for an unmeasured confounder: a review of two independent methods. *Rev. bras. Epidemiol.* 2010; 13(2): 188-198. <http://dx.doi.org/10.1590/S1415-790X2010000200002>.

Ma Y, Olendzki BC, Pagoto SL, Hurley TG, Magner RP, Ockene IS, Schneider KL, Merriam PA, Hébert JR. Number of 24-hour diet recalls needed to estimate energy intake. *Ann Epidemiol.* 2009; 19(8):553-9. doi: 10.1016/j.annepidem.2009.04.010.

Macedo-Ojeda G, Vizmanos-Lamotte B, Márquez-Sandoval YF, Rodríguez-Rocha NP, López-Uriarte PJ, Fernández-Ballart JD. Validation of a semi-quantitative food frequency questionnaire to assess food groups and nutrient intake. *Nutr Hosp.* 2013; 28(6):2212-20. doi: 10.3305/nutr.hosp.v28in06.6887.

Maclure M, Hankinson S. Analysis of selection bias in a case-control study of renal adenocarcinoma. *Epidem* 1990; 1(6), 441-47.

Makarem N, Lin Y, Bandera EV, Jacques PF, Parekh N. Concordance with World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) guidelines for cancer prevention and obesity-related cancer risk in the Framingham Offspring cohort (1991-2008). *Cancer Causes Control.* 2015;26(2):277-86. doi: 10.1007/s10552-014-0509-9.

Manser CN, Bauerfeind P. Impact of socioeconomic status on incidence, mortality, and survival of colorectal cancer patients: a systematic review. *Gastrointest Endosc.* 2014; 80(1):42-60.e9. doi: 10.1016/j.gie.2014.03.011.

Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959;22(4):719-48.

Martín de Civetta MT, Civetta JD. Carcinogénesis. *Salud pública Méx.* 2011; 53(5): 405-414.

Martinez MF, Philippi ST, Estima C, Leal G. Validity and reproducibility of a food frequency questionnaire to assess food group intake in adolescents. *Cad Saude Publica.* 2013; 29(9):1795-804.

Martínez T, Hernández GA, Rojas CA. Diet and Its Association with Preneoplastic Lesions and Gastric Cancer in a High-Risk Area for Gastric Cancer in Colombia I, 2000-2006. *Rev Colomb Cancerol.* 2008; 12:74–88.

McCandless LC, Gustafson P, Levy AR. A sensitivity analysis using information about measured confounders yielded improved uncertainty assessments for unmeasured confounding. *J Clin Epidemiol.* 2008; 61(3):247-55. doi: 10.1016/j.jclinepi.2007.05.006.

McDonald, J.H. *Handbook of Biological Statistics.* 3er ed. Sparky House Publishing. Estados Unidos: Baltimore, Maryland. 2014. p.133-136.

Mezei G, Kheifets L. Selection bias and its implications for case-control studies: a case study of magnetic field exposure and childhood leukaemia. *Int J Epidemiol.* 2006; 35(2):397-406.

Michels KB. Nutritional epidemiology—past, present, future. *Int J Epidemiol.* 2003; 32 (4): 486-488. doi: 10.1093/ije/dyg216.

Mohammadifard N, Omidvar N, Houshiarrad A, Neyestani T, Naderi GA, Soleymani B. Validity and reproducibility of a food frequency questionnaire for assessment of fruit and vegetable intake in Iranian adults. *J Res Med Sci.* 2011; 16(10): 1286–1297.

Molenberghs G, Kenward MG. *Missing Data in Clinical Studies.* Estados Unidos: John Wiley and Sons, Ltd. 2007. p.526.

Mourouti N, Kontogianni MD, Papavagelis C, Panagiotakos DB. Diet and breast cancer: a systematic review. *Int J Food Sci Nutr.* 2015; 66(1):1-42. doi: 10.3109/09637486.2014.950207.

Murphy G, Devesa SS, Cross AJ, Inskip PD, McGlynn KA, Cook MB. Sex disparities in colorectal cancer incidence by anatomic subsite, race and age. *Int J Cancer.* 2011; 128(7):1668-75. doi: 10.1002/ijc.25481.

Murphy N, Norat T, Ferrari P, Jenab M, Bueno-de-Mesquita B, Skeie G, Dahm CC, Overvad K, Olsen A, Tjønneland A, Clavel-Chapelon F, Boutron-Ruault MC, Racine A, Kaaks R, Teucher B, Boeing H, Bergmann MM, Trichopoulou A, Trichopoulos D, Lagiou P, Palli D, Pala V, Panico S, Tumino R, Vineis P, Siersema P, van Duijnhoven F, Peeters PH, Hjartaker A, Engeset D, González CA, Sánchez MJ, Dorransoro M, Navarro C, Ardanaz E, Quirós JR, Sonestedt E, Ericson U, Nilsson L, Palmqvist R, Khaw KT, Wareham N, Key TJ, Crowe FL, Fedirko V, Wark PA, Chuang SC, Riboli E. Dietary fibre intake and risks of cancers of the colon and rectum in the European prospective investigation into cancer and nutrition (EPIC). *PLoS One*. 2012; 7(6):e39361. doi: 10.1371/journal.pone.0039361.

Nagle CM, Wilson LF, Hughes MC, Ibiebele TI, Miura K, Bain CJ1, Whiteman DC, Webb PM. Cancers in Australia in 2010 attributable to inadequate consumption of fruit, non-starchy vegetables and dietary fibre. *Aust N Z J Public Health*. 2015; 39(5):422-8. doi: 10.1111/1753-6405.12449.

Nasef NA, Mehta S, Ferguson LR. Susceptibility to chronic inflammation: an update. *Arch Toxicol*. 2017; 91(3):1131-1141. doi: 10.1007/s00204-016-1914-5.

National Research Council (NRC). The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. 2010. [Internet]. [Consultado el 2/Feb/2017]. Disponible en: <http://www.cytel.com/hs-fs/hub/1670/file-2411099288-pdf/Pdf/MissingDataNationalAcademyof-Medicine.2010.pdf>.

Navarro A, Cristaldo PE, Andreatta MM, Díaz MP, Muñoz SE, Lantieri MJ, et al. Atlas de Alimentos. Córdoba: Editorial Universidad Nacional de Córdoba; 2007. 86 p.

Navarro A, Cristaldo PE, Díaz MP, Eynard AR. Atlas fotográfico para cuantificar el consumo de alimentos y nutrientes en estudios nutricionales epidemiológicos en Córdoba, Argentina. *Rev Fac Cienc Méd Córdoba* 2000; 57(1): 67-74.

Navarro A, Díaz MP, Muñoz SE, Lantieri MJ, Eynard AR. Characterization of meat consumption and risk of colorectal cancer in Cordoba, Argentina. *Nutrition*. 2003; 19(1):7-10.

Navarro A, Osella AR, Guerra V, Muñoz SE, Lantieri MJ, Eynard AR. Reproducibility and Validity of a Food-Frequency Questionnaire in Assessing Dietary Intakes and Food Habits in Epidemiological Cancer Studies in Argentina. *J ExpClin Cancer Res* 2001; 20(3):203-208.

Niclis C, Díaz MP, La Vecchia C. Breast cancer mortality trends and patterns in Córdoba, Argentina in the period 1986-2006. *Eur J Cancer Prev.* 2010;19(2):94-9. doi: 10.1097/CEJ.0b013e328333fb52.

Niclis C, Román MD, Osella AR, Eynard AR, Díaz MP. Traditional Dietary Pattern Increases Risk of Prostate Cancer in Argentina: Results of a Multilevel Modeling and Bias Analysis from a Case-Control Study. *J Cancer Epidemiol.* 2015; (2):1-10. doi: 10.1155/2015/179562.

Nunes LN, Klück MM, Fachel JMG. Uso da imputação múltipla de dados faltantes: uma simulação utilizando dados epidemiológicos Multiple imputations for missing data: a simulation with epidemiological data. *Cad Saude Pública.* 2009; **25**, 268–78.

Organización Mundial de la Salud (OMS). International Statistical Classification of Diseases and Related Health Problems, 10th Revision. Vol 2. Ginebra, OMS, 1992. [Internet]. [Consultado el 5/Feb/2017]. Disponible en: <http://ais.paho.org/classifications/Chapters/pdf/Volume2.pdf>

Organización Mundial de la Salud (OMS). Cáncer. 2017. [internet]. [Consultado el 5/Feb/2017]. Disponible en: <http://www.who.int/mediacentre/factsheets/fs297/es/>

Organización Mundial de la Salud (OMS). Actividad Física. 2017. [internet]. [Consultado el 6/Mar/2017]. Disponible en: <http://www.who.int/dietphysicalactivity/pa/es/>

Organización Mundial de la Salud (OMS). Enfermedades no transmisibles. 2017. [Internet]. [Consultado el 6/Jun/2017]. Disponible: <http://www.who.int/mediacentre/factsheets/fs355/es/>.

Organización Panamericana de la Salud (OPS). Mejoremos la Salud a todas las edades. Un manual para el cambio de comportamiento. 2005. [Internet]. [Consultado el 6/Jun/2017]. Disponible en: <http://www.paho.org/spanish/dd/pub/mejorar-lu.pdf>.

Paalanen L, Männistö S, Virtanen MJ, Knekt P, Räsänen L, Montonen J, Pietinen P. Validity of a food frequency questionnaire varied by age and body mass index. *J Clin Epidemiol.* 2006; 59(9):994-1001.

Palacios C, Segarra A, Trak M, Colón I. Reproducibility and validity of a food frequency questionnaire to estimate calcium intake in Puerto Ricans. *Arch Latinoam Nutr.* 2012;62(3):205-12.

Peyrano M, Gigena J, Muñoz SE, Lantieri M, Eynard AR, Navarro A. A computer software system for the analysis of Dietary data in cancer epidemiological research 17th International Cancer Congress, Monduzzi Editore; 1998. p. 381-384.

Pereira RA, Araujo MC, Lopes Tde S, Yokoo EM. How many 24-hour recalls or food records are required to estimate usual energy and nutrient intake? *Cad Saude Publica.* 2010; 26(11):2101-11.

Popkin BM. The shift in stages of the nutrition transition in the developing world differs from past experiences! *Public Health Nutr.* 2002; 5(1A):205-14.

Popkin BM. Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. *Am J Clin Nutr.* 2006;84(2):289-298.

Popkin BM. Nutrition Transition and the Global Diabetes Epidemic. *Curr Diab Rep.* 2015; 15(9):64. doi: 10.1007/s11892-015-0631-4.

Potter JD. Nutritional epidemiology--there's life in the old dog yet! *Cancer Epidemiol Biomarkers Prev.* 2015; 24(2):323-30. doi: 10.1158/1055-9965.EPI-14-1327.

Pou SA, Niclis C, Aballay LR, Tumas N, Román MD, Muñoz SE, Becaria Coquet J, Díaz M del P. Cancer and its association with dietary patterns in Córdoba (Argentina). *Nutr Hosp.* 2014; 29(3):618-28. doi: 10.3305/nh.2014.29.3.7192.

Pou SA, Díaz M del P, Osella AR. Applying multilevel model to the relationship of dietary patterns and colorectal cancer: an ongoing case-control study in Córdoba, Argentina. *Eur J Nutr.* 2012; 51(6):755-64. doi: 10.1007/s00394-011-0255-7.

Reartes GA, Di Paola Naranjo RD, Eynard AR, Muñoz SE. Cooking methods and the formation of PhIP (2-Amino, 1-methyl, 6-phenylimidazo[4,5-b] pyridine) in the crust of the habitually consumed meat in Argentina. *Food Chem Toxicol.* 2016; 92:88-93. doi: 10.1016/j.fct.2016.03.026.

Renehan AG, Zwahlen M, Egger M. Adiposity and cancer risk: new mechanistic insights from epidemiology. *Nat Rev Cancer*. 2015; 15(8):484-98. doi: 10.1038/nrc3967.

Registro Provincial de Tumores de Córdoba (RPTC). Informe sobre cáncer en la Provincia de Córdoba 2004-2009. Ministerio de Salud de la Provincia de Córdoba. Editorial de la Provincia de Córdoba. Córdoba, 2013.

Rolón PA, Castellsagué X, Benz M, Muñoz N. Hot and cold mate drinking and esophageal cancer in Paraguay. *Cancer Epidemiol Biomarkers Prev*. 1995; 4(6):595-605.

Román MD, Niclis C, Tumas N, Díaz Mdel P, Osella AR, Muñoz SE. Tobacco smoking patterns and differential food effects on prostate and breast cancers among smokers and nonsmokers in Córdoba, Argentina. *Eur J Cancer Prev*. 2014; 23(4):310-8. doi: 10.1097/CEJ.0000000000000044.

Robey RB, Weisz J, Kuemmerle NB, Salzberg AC, Berg A, Brown DG, Kubik L, Palorini R, Al-Mulla F, Al-Temaimi R, Colacci A, Mondello C, Raju J, Woodrick J, Scovassi AI, Singh N, Vaccari M, Roy R, Forte S, Memeo L, Salem HK, Amedei A, Hamid RA, Williams GP, Lowe L, Meyer J, Martin FL, Bisson WH, Chiaradonna F, Ryan EP. Metabolic reprogramming and dysregulated metabolism: cause, consequence and/or enabler of environmental carcinogenesis? *Carcinogenesis*. 2015; 36 (1):S203-31. doi: 10.1093/carcin/bgv037.

Ronco AL, De Stefani E, Deneo-Pellegrini H, Boffetta P, Aune D, Silva C, Landó G, Luaces ME, Acosta G, Mendilaharsu M. Dietary patterns and risk of ductal carcinoma of the breast: a factor analysis in Uruguay. *Asian Pac J Cancer Prev*. 2010; 11(5):1187-93.

Rosenbaum PR. Sensitivity Analysis in Observational Studies. En: *Encyclopedia of Statistics in Behavioral Science*. Editorial John Wiley & Sons, Ltd, 2005. Volumen 4, p. 1809–1814.

Rossi RE, Pericleous M, Mandair D, Whyand T, Caplin ME. The role of dietary factors in prevention and progression of breast cancer. *Anticancer Res*. 2014; 34(12):6861-75.

Rothman, K.J., Greenland, S. *Modern Epidemiology*, 2nd Edition. Philadelphia, PA: Lippincott, Williams & Wilkins; 1998. p.738.

Rothman KJ, Greenland S. Matching. En: Modern Epidemiology. Rothman KJ, Greenland S. 2da ed. Philadelphia: Lippincott-Raven Publishers, 1998. p. 147-162.

Royo Bordonada MA, Moreno JD. Método Epidemiológico: Instituto de Salud Carlos III - Ministerio de Ciencia e Innovación. Escuela Nacional de Sanidad (ENS); 2009. 226 p.

Rubin DB. Inference and missing data. Biometrika. 1976; 63, 581–92.

Rubin, DB. Multiple Imputation for Nonresponse in Surveys. J. Wiley & Sons, New York.1987.

Rubinstein A, Colantonio L, Bardach A, et al. .Estimación de la carga de las enfermedades cardiovasculares atribuible a factores de riesgo modificables en Argentina. Rev Panam Salud Publica.2010; 27.

Sachin J, Preethi SG, Aleyamma M. Assessment of Confounding and Interaction using the Mantel-Haenszel Risk Estimation Method. Asian Pacific J Cancer Prev.2008; 9, 323-326.

Saez M. El problema de las medidas repetidas. Análisis longitudinal en epidemiología. Gac Sanit. 2001;15(4):347-352.

Salgado-Barreira A, Estany-Gestal A, Figueiras A. Efecto del nivel socioeconómico sobre la mortalidad en áreas urbanas: revisión crítica y sistemática. Cad. Saúde Pública. 2014; 30(8):1609-1621. doi: 10.1590/0102-311X00152513

Satija A, Yu E, Willett WC, Hu FB. Understanding nutritional epidemiology and its role in policy. AdvNutr. 2015; 6(1):5-18. doi: 10.3945/an.114.007492.

Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf. 2006, 15(5):291–303.

Schoenfeld JD, Ioannidis JP. Is everything we eat associated with cancer? A systematic cookbook review. Am J Clin Nutr. 2013; 97(1):127-34. doi: 10.3945/ajcn.112.047142.

Scoccianti C, Lauby-Secretan B, Bello PY, Chajes V, Romieu I. Female breast cancer and alcohol consumption: a review of the literature. *Am J Prev Med.* 2014; 46(3-1):S16-25. doi: 10.1016/j.amepre.2013.10.031.

Seoane J. ¿Modelos mixtos (lineales)? Una introducción para el usuario temeroso. *Etología.*2014; 24:1-23.

Shamah-Levy T, Rodríguez-Ramírez S, Gaona-Pineda EB, Cuevas-Nasu L, Carriquiry AL, Rivera JA. Three 24-Hour Recalls in Comparison with One Improve the Estimates of Energy and Nutrient Intakes in an Urban Mexican Population. *J Nutr.* 2016; 146(5):1043-50. doi: 10.3945/jn.115.219683.

Shephard, RJ, Balady, GJ. Exercise as cardiovascular therapy. *Circulation,* 99, 963-972. doi:10.1161/01.CIR.99.7.963

Shin S, Saito E, Sawada N, Ishihara J, Takachi R, Nanri A, Shimazu T, Yamaji T, Iwasaki M, Sasazuki S, Inoue M, Tsugane S; JPHC Study Group. Dietary patterns and colorectal cancer risk in middle-aged adults: A large population-based prospective cohort study. *Clin Nutr.* 2017; S0261-5614(17)30154-1. doi: 10.1016/j.clnu.2017.04.015.

Shrestha A, Koju RP, Beresford SAA, Chan KCG, Connell FA, Karmacharya BM, Shrestha P, Fitzpatrick AL. Reproducibility and relative validity of food group intake in a food frequency questionnaire developed for Nepalese diet. *Int J Food Sci Nutr.* 2017; 68(5):605-612. doi: 10.1080/09637486.2016.1268099.

Sierra MS, Forman D. Burden of colorectal cancer in Central and South America. *Cancer Epidemiol.* 2016; 44 (1):S74-S81. doi: 10.1016/j.canep.2016.03.010.

Skrondal A, Rabe-Hesketh S. Generalized latent variable modeling: Multilevel, longitudinal and structural equation models. Boca Raton, FL: Chapman& Hall; 2004. 528 p.

Soley-Bori M. Dealing with missing data: Key assumptions and methods for applied analysis. Technical Report No. 4. Boston University School of Public Health. Department of Health Policy and Management. 2013. 20 p.

Song M, Garrett WS, Chan AT. Nutrients, foods, and colorectal cancer prevention. *Gastroenterology*. 2015; 148(6):1244-60.e16. doi: 10.1053/j.gastro.2014.12.035.

Song M, Hu FB, Spiegelman D, Chan AT, Wu K, Ogino S, Fuchs CS, Willett WC, Giovannucci EL. Adulthood weight change and risk of colorectal cancer in the Nurses' Health Study and Health Professionals Follow-up Study. *Cancer Prev Res (Phila)*. 2015; 8(7):620-7. doi: 10.1158/1940-6207.CAPR-15-0061.

StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

Steinemann N, Grize L, Zieseimer K, Kauf P, Probst-Hensch N, Brombach C. Relative validation of a food frequency questionnaire to estimate food intake in an adult population. *Food Nutr Res*. 2017; 61(1):1305193. doi: 10.1080/16546628.2017.1305193.

Steinmaus C, Yuan Y, Kalman D, Rey OA, Skibola CF, Dauphine D, Basu A, Porter KE, Hubbard A, Bates MN, Smith MT, Smith AH. Individual differences in arsenic metabolism and lung cancer in a case-control study in Cordoba, Argentina. *Toxicol Appl Pharmacol*. 2010; 247(2):138-45. doi: 10.1016/j.taap.2010.06.006.

Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009; 338: b2393. doi: 10.1136/bmj.b2393

Stewart BW, Wild CP, editors. *World cancer report 2014*. Lyon: International Agency for Research on Cancer; 2014. p.630.

Szklo M, Nieto FJ. *Epidemiology beyond the basics*. 2da ed. Estados Unidos: John and Barlett Publishers; 2007.p.489.

Szymańska K, Hung RJ, Wünsch-Filho V, Eluf-Neto J, Curado MP, Koifman S, Matos E, Menezes A, Fernandez L, Daudt AW, Boffetta P, Brennan P. Alcohol and tobacco, and the risk of cancers of the

upper aerodigestive tract in Latin America: a case-control study. *Cancer Causes Control*. 2011; 22(7):1037-46. doi: 10.1007/s10552-011-9779-7.

The International Physical Activity Questionnaire. International Physical Activity Questionnaire-Short Format. 2005. [Internet]. [Consultado el 6/Mar/2017]. Disponible en: <https://sites.google.com/site/theipaq>

The World Bank Group. World Bank Country and Lending Groups. 2017. [Internet]. [Consultado el 22/Ago/2017]. Disponible en: <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519>

Tumas N, Niclis C, Aballay LR, Osella AR, Díaz M del P. Traditional dietary pattern of South America is linked to breast cancer: an ongoing case-control study in Argentina. *Eur J Nutr*. 2014;53(2):557-66. doi: 10.1007/s00394-013-0564-0.

Vera-Ramirez L, Ramirez-Tortosa MC, Sanchez-Rovira P, Ramirez-Tortosa CL, Granados-Principal S, Lorente JA, Quiles JL. Impact of diet on breast cancer risk: a review of experimental and observational studies. *Crit Rev Food Sci Nutr*. 2013; 53(1):49-75. doi: 10.1080/10408398.2010.521600.

Vieira AR, Abar L, Chan D, Vingeliene S, Polemiti E, Stevens C, Greenwood D, Norat T. Foods and beverages and colorectal cancer risk: a systematic review and meta-analysis of cohort studies, an update of the evidence of the WCRF-AICR Continuous Update Project. *Ann Oncol*. 2017. doi: 10.1093/annonc/mdx171.

von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Prev Med*. 2007; 45(4):247-51.

Vrijheid M, Deltour I, Krewski D, Sanchez M, Cardis E. The effects of recall errors and of selection bias in epidemiologic studies of mobile phone use and cancer risk. *J Expo Sci Environ Epidemiol*. 2006; 16(4):371-84.

Wang, Z. Two postestimation commands for assessing confounding effects in epidemiological studies. *The Stata Journal*. 2007; 7(2):183–196.

Wardle J, Haase AM, Steptoe A, Nillapun M, Jonwutiwes K, Bellisle F. Gender differences in food choice: the contribution of health beliefs and dieting. *Ann Behav Med.* 2004; 27(2):107-16.

White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat Med.* 2010; 29, 2920–31.

White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011; 30(4):377-99. doi: 10.1002/sim.4067.

Willet W (1998). *Nutritional Epidemiology*. 2da ed. New York: Oxford University Press. 514 p.

Weisberg HI. *Bias and Causation: models and judgment for valid comparisons*. 1a ed. Estados Unidos: John Wiley & Sons, Inc.; 2010. p.348.

World Cancer Research Fund / American Institute for Cancer Research (WCRF/AICR). *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective*. Washington DC: AICR, 2007. p.517.

World Cancer Research Fund/American Institute for Cancer Research: *Colorectal Cancer 2011 Report*. Food, nutrition, physical activity and the prevention of colorectal cancer [Internet]. [Consultado el 6/Jun/2017]. Disponible en: <http://www.wcrf.org/sites/default/files/Colorectal-Cancer-2011-Report.pdf>

World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR). *Food, nutrition, physical activity and the prevention of cancer: A global perspective*. 2012. [Internet]. [Consultado el 6/Jun/2017]. Disponible en: <http://www.dietandcancerreport.org/index.php>

World Health Organization (WHO). *Global Database on Body Mass Index. BMI classification* [internet] 2006. [Consultado el 10/Mar/2017] Disponible en: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html.

Zhang Y, Flórez ID, Colunga Lozano LE, Bakar Aloweni FA, Kennedy SA, Li A, Craigie S, Zhang S, Agarwal A, Lopes LC, Devji T, Wiercioch W, Riva JJ, Wang M, Jin X, Fei Y, Alexander P, Morgano GP, Zhang Y, Carrasco-Labra A, Kahale LA, Akl EA, Schünemann HJ, Thabane L, Guyatt G. A systematic

survey on reporting and methods for handling missing participant data for continuous outcomes in randomized controlled trials: a systematic survey. *J Clin Epidemiol.* 2017; S0895-4356(17)30573-5. doi: 10.1016/j.jclinepi.2017.05.017.

Zheng J, Zhou Y, Li S, Zhang P, Zhou T, Xu DP, Li HB. Effects and Mechanisms of Fruit and Vegetable Juices on Cardiovascular Diseases. *Int J Mol Sci.* 2017 Mar 4;18(3). pii: E555. doi: 10.3390/ijms18030555.

Zhou L, Zahid M, Anwar MM, Pennington KL, Cohen SM, Wisecarver JL, Shostrom V, Mirvish SS. Suggestive evidence for the induction of colonic aberrant crypts in mice fed sodium nitrite. *Nutr Cancer.* 2016; 68(1):105-12. doi: 10.1080/01635581.2016.1102298.

CAPÍTULO 6. APÉNDICES Y ANEXOS

APÉNDICES

APÉNDICE 1. Grupos de alimentos utilizados para la construcción de los patrones alimentarios de la población de Córdoba.

APÉNDICE 2. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas del principal sostén del hogar, estudio caso-control de cáncer de mama, Córdoba, Argentina 2008-2015.

APÉNDICE 3. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas del principal sostén del hogar, estudio caso-control de cáncer colorrectal, Córdoba, Argentina 2010-2016.

APÉNDICE 4. Parámetros para el análisis del error de clasificación

ANEXOS

ANEXO 1. Consentimiento Informado estudios caso-control de cáncer en Córdoba.

ANEXO 2. Registro Provincial de Investigaciones en Salud (RePIS N°044/10)

ANEXO 3. Instrumento Encuesta Alimentaria

ANEXO 4. Categorización de la variable Nivel Socioeconómico

ANEXO 5. Categorización de la variable Actividad Física

ANEXO 6. Registro Provincial de Investigaciones en Salud (RePIS N°159/13)

ANEXO 7. Publicaciones científicas realizadas

APÉNDICE 1. Grupos de alimentos utilizados para la construcción de los patrones alimentarios de la población de Córdoba.

Tabla a. Grupos de alimentos utilizados para la construcción de los patrones alimentarios de la población de Córdoba, mediante Análisis Factorial de Componentes Principales.

Grupos alimentarios	Alimentos
Cereales refinados	Pastas, granos no integrales, galletitas saladas, pan.
Cereales integrales	Granos integrales, pan de salvado, salvado, galletitas integrales.
Productos de pastelería	Criollos, facturas, galletitas dulces, pan con grasa, torta frita, tortas dulces.
Legumbres	Lentejas, soja, arvejas, porotos, garbanzos.
Vegetales no amiláceos	Acelga, achicoria, ajo, alcaucil, apio, arvejas frescas, berenjena, berro, brócoli, calabaza, cebolla, coliflor, chaucha, espárrago, espinaca, lechuga, pepino, pimienta, rabanito, repollo, tomate, zanahoria, zapallito, zapallo.
Vegetales amiláceos	Papa, batata, choclo.
Frutas	Ananá, banana, ciruela, durazno, damasco, frutilla, kiwi, limón, mandarina, manzana, melón, naranja, pera, pomelo, sandía, uva.
Leche y yogur	Leche descremada, leche entera, yogur entero, yogur descremado.
Quesos blandos	Queso mantecoso, ricota.
Quesos duros	Queso de rallar, queso rallado, queso semiduro.
Aceites y mayonesa	Mayonesa, aceite mezcla, aceite de uva, aceite de soja, aceite de oliva, aceite de maíz, aceite de girasol.
Manteca y crema	Manteca y crema de leche.
Azúcares y dulces	Azúcar, caramelos, golosinas, miel, mermeladas.
Confituras	Chocolate, cacao, dulce de leche, helado, mantecol.
Bebidas alcohólicas	Aperitivos, bebidas espirituosas, cerveza, ginebra, grapa, vino, whisky.
Bebidas analcohólicas	Gaseosas, amargo serrano, jugos.
Huevos	Clara, yema, huevo duro, huevo frito.
Carnes procesadas	Bondiola, jamón cocido, jamón crudo, mortadela, panceta, queso de cerdo, salame de milán, salami, salchichón, chorizo, morcilla, salchicha parrillera, salchicha de viena.
Carnes grasas	Carne vacuna y de cerdo grasa –incluye asado y achuras como mollejas, lengua, etc.- y pollo con piel en sus diferentes formas de cocción.
Carnes magras	Carne vacuna y de cerdo magra y pollo sin piel en sus diferentes formas de cocción.

APÉNDICE 2. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas del principal sostén del hogar, estudio caso-control de cáncer de mama, Córdoba, Argentina 2008-2015.

Tabla b. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas del principal sostén del hogar, estudio caso-control de cáncer de mama, Córdoba, Argentina 2008-2015.

	n	%	% de observaciones con datos faltantes						
Total	844	100							
Occupación del principal sostén del hogar			AF	OS	Nº de prov.	Compu	Internet	TD	Autos
Dueño o alta dirección >50 empleados	0	0	0	0	0	0	0	0	0
Dueño o alta dirección 6-50 empleados	14	1.7	21.4	14.3	0	0	0	0	0
Dueño o alta dirección 1-5 empleados	19	2.2	42.1	15.8	21.0	21.0	21.0	21.0	21.0
Profesional Independiente	89	10.5	30.3	14.6	4.5	4.5	4.5	4.5	4.5
Técnicos independientes y en relación de dependencia	100	11.9	32	21	13	12	12	12	13
Comerciante sin personal, artesano, encargados y supervisores	94	11.1	31.9	20.2	21.3	21.3	21.3	21.3	21.3
Empleado sin jerarquía	259	30.7	32.0	16.2	6.6	6.6	6.6	6.6	6.6
Obrero calificado	12	1.4	25	66.7	8.3	8.3	8.3	8.3	8.3
Trabajador autónomo u obrero no calificado	136	16.1	26.5	25	13.2	12.5	12.5	12.5	13.2
Desocupado o trabajador no formal	1	0.1	0	100	0	0	0	0	0

(changas)

Jubilado o pensionado	97	11.5	17.5	51.5	0	0	0	0	0
Desconocido	23	2.7	30.4	52.2	52.2	52.2	52.2	52.2	52.2

Referencias: AF: actividad física; OS:obra social; N° de prov.:N° de proveedores en el hogar;

Compu: computadora; TD: tarjeta de débito.

APÉNDICE 3. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas del principal sostén del hogar, estudio caso-control de cáncer colorrectal, Córdoba, Argentina 2010-2016.

Tabla c. Sujetos y datos faltantes: Distribuciones de frecuencias absolutas y relativas del principal sostén del hogar, estudio caso-control de cáncer colorrectal, Córdoba, Argentina 2010-2016.

	n	%	% de observaciones con datos faltantes				
Total	490	100					
Occupación del principal sostén del hogar			Nº de proveedores	Compu	Internet	TD	Autos
Dueño o alta dirección >50 empleados	0	0	0	0	0	0	0
Dueño o alta dirección 6-50 empleados	15	3,06	40,0	40,0	40,0	40,0	40,0
Dueño o alta dirección 1-5 empleados	29	5,92	24,14	24,14	24,14	24,14	24,14
Profesional Independiente	65	13,27	20,0	20,0	20,0	20,0	20,0
Técnicos independientes y en relación de dependencia	19	3,88	47,37	47,37	47,37	47,37	47,37
Comerciante sin personal, artesano, encargados y supervisores	56	11,43	25,0	26,79	25,0	26,79	26,79
Empleado sin jerarquía	141	28,78	39,72	41,13	39,72	41,13	41,13
Obrero calificado	62	12,65	48,39	50	50,0	50,0	48,39
Trabajador autónomo u obrero no calificado	90	18,37	55,56	56,67	55,56	56,67	55,56
Desocupado o trabajador no formal (changas)	5	1,02	20,0	20,0	20,0	20,0	20,0
Jubilado o pensionado	0	0	0	0	0	0	0

Desconocido	8	1,63	87,50	87,50	87,50	87,50	87,5
-------------	---	------	-------	-------	-------	-------	------

APÉNDICE 4. Parámetros para el análisis del error de clasificación

Tabla d. Nomenclatura de ecuaciones para calcular los parámetros de error de clasificación

	Verdaderamente expuesto	Verdaderamente no expuesto
Clasificado como expuesto	A	B
Clasificado como no expuesto	C	D

A continuación se presentan las ecuaciones para calcular las mediciones de clasificación:

$$\text{Sensibilidad (S)} = \frac{A}{A+C};$$

$$\text{Especificidad (E)} = \frac{D}{D+B};$$

$$\text{Valor Predictivo Positivo (VPP)} = \frac{A}{A+B};$$

$$\text{Valor Predictivo Negativo (VPN)} = \frac{D}{C+D}.$$

En la siguiente tabla se presentan las ecuaciones para obtener los verdaderos valores esperados, es decir, corregidos, a partir de los datos observados con error de clasificación. Se utilizan las letras minúsculas para designar valores observados y mayúsculas para designar valores esperados o verdaderos. E_1 y D_1 representan a los grupos expuestos y casos, respectivamente, y E_0 y D_0 a los no expuestos y controles, respectivamente (Lash *et al*, 2009).

Tabla d. Ecuaciones para calcular valores verdaderos esperados a partir de los datos observados con error de clasificación de la exposición.

	Observado		Corregido	
	E_1	E_0	E_1	E_0
D_+	a	b	$[a - D_{+total}(1 - SP_{D_+})] / [SE_{D_+} - (1 - SP_{D_+})]$	$D_{+total} - A$
D_-	c	d	$[c - D_{-total}(1 - SP_{D_-})] / [SE_{D_-} - (1 - SP_{D_-})]$	$D_{-total} - C$
Total	a+c	b+d	A+C	B+D

*A+C es el número total de sujetos expuestos corregidos y B+D es el número total de sujetos no expuestos.

ANEXOS

ANEXO 1. Consentimiento Informado estudios caso-control de cáncer en Córdoba.



**UNIVERSIDAD NACIONAL DE CÓRDOBA
FACULTAD DE CIENCIAS MÉDICAS
ESCUELA DE NUTRICIÓN**

Se explica al individuo que se realizará un estudio sobre la relación entre la alimentación y la presencia de cáncer en la Provincia de Córdoba, para lo cual se entrevistará a un grupo de personas con la patología y a otro de personas sanas, sobre datos generales (como nacionalidad, sexo, edad, hábito de fumar, etc) y particularmente, sobre hábitos alimentarios desde cinco años atrás a la fecha.

A posterior se le solicita, siempre que sea mayor de edad, su Declaración Voluntaria y Consentimiento para que se realice una encuesta.

A continuación el sujeto declara:

Se me ha dado a conocer que este estudio se hace por primera vez a nivel provincial, y que se realiza para identificar posibles factores de riesgo de origen alimentario que se relacionen con la aparición de cáncer.

Se me ha dado a conocer que la finalidad de este estudio es servir para el diseño de políticas públicas de salud, como pueden ser campañas de educación alimentaria-nutricional, para la prevención del cáncer, redundando en beneficio de la comunidad.

Como también hago constar que conozco detalladamente el tema: Se me informó que los factores alimentarios pueden jugar un rol importante en la prevención de las enfermedades crónicas, como lo es el cáncer. Dentro de los distintos tipos de tumores, se comenzó a estudiar en este caso el cáncer colo-rectal, próstata, mama y tumores de urotelio (localizados en las vías urinarias), los cuales se considera que pueden estar asociados al consumo frecuente de distintos componentes de los alimentos. Estas asociaciones pueden estudiarse en la etapa de análisis de los datos recolectados en la encuesta mediante técnicas matemático-estadísticas como los "modelos multilevel".

Entiendo que no se recibe remuneración por parte de los individuos estudiados de los investigadores, realizándose por ambas partes con acuerdo y libertad, y que puedo abandonar el estudio cuando lo desee.

Acepto la realización de una encuesta para que se arribe a una caracterización del patrón alimentario, la cual implica solamente responder a las preguntas del encuestador.

Se me ha informado que mi participación en este estudio no implica ningún tipo de riesgo para mi salud.

Se me ha informado y entiendo cómo los datos de estudio de mi persona serán mantenidos en confidencialidad.

Se me ha informado y entiendo que puedo suspender mi participación en el estudio, en el momento que mi persona así lo desee, con total libertad.

Conozco la importancia de este estudio y su proyección hacia la prevención si fuese posible, por lo que me complace que una vez concluido, los datos resultantes puedan ser difundidos y publicados para ayudar a la humanidad, en especial los más desfavorecidos.

*Inicialmente en 2007.



**UNIVERSIDAD NACIONAL DE CÓRDOBA
FACULTAD DE CIENCIAS MÉDICAS
ESCUELA DE NUTRICIÓN**

ESTUDIO: "EPIDEMIOLOGÍA AMBIENTAL DEL CÁNCER EN CÓRDOBA: ESTUDIO DE FACTORES DE RIESGO NUTRICIONALES Y TÓXICOS, E IDENTIFICACIÓN DE BIOMARCADORES DE EXPOSICIÓN, DE EFECTO Y DE DIAGNÓSTICO TEMPRANO."
Estudio epidemiológico - Córdoba - Argentina.

Una vez escuchado y leído la explicación de este estudio digo en consentimiento voluntario:

1.- Confirmando que he leído y comprendido la hoja suplementaria de información para el encuestado* con fecha Julio 2007 para el estudio mencionado, y que he tenido oportunidad de formular preguntas.

Si
No

Iniciales:

2.- Entiendo que mi participación es voluntaria y que tengo libertad de retirarme del mismo cuando lo desee.

Si
No

Iniciales:

3.- Se me ha informado y entiendo cómo los datos de estudio de mi persona serán mantenidos en confidencialidad.

Si
No

Iniciales:

4.- Acepto participar en el estudio mencionado.

Si
No

Iniciales:

INDIVIDUO ENCUESTADO

FIRMA:.....

ACLARACIÓN: DNI:

FECHA: HORA:

TESTIGO INDEPENDIENTE

FIRMA:.....

ACLARACIÓN: DNI:

FECHA: HORA:

PERSONA QUE OBTUVO EL CONSENTIMIENTO

FIRMA:.....

ACLARACIÓN: DNI:

FECHA: HORA:

DATOS DEL INVESTIGADOR

FIRMA:.....


ACLARACIÓN: DNI:

FECHA: HORA:

*Inicialmente en 2007.

ANEXO 2. Registro Provincial de Investigaciones en Salud (RePIS N°044/10)

Formulario_Inscripción_ENMIENDAS_V_3_2010




REGISTRO PROVINCIAL DE INVESTIGACIÓN EN SALUD (RePIS)

REGISTRO DE ENMIENDAS

Hospital Nacional de Clínicas

Registro N° 044/10



Gobierno de Córdoba
Ministero de Salud

N° DE RePIS INVESTIGACIÓN ORIGINAL	N° DE INSCRIPCIÓN DE LA ENMIENDA

I. INVESTIGACIÓN

Título de la Investigación y versión: **EPIDEMIOLOGÍA AMBIENTAL DEL CÁNCER EN CÓRDOBA: ESTUDIO DE FACTORES DE RIESGO NUTRICIONALES Y TÓXICOS E IDENTIFICACIÓN DE BIOMARCADORES DE EXPOSICIÓN, DE EFECTO Y DE DIAGNÓSTICO TEMPRANO.**

II. INVESTIGADOR

Nombre Investigador Principal: Díaz, María del Pilar N° RePIS Investigador: 388

III. COMITÉ INSTITUCIONAL DE ÉTICA DE LA INVESTIGACIÓN EN SALUD DONDE SE PRESENTÓ EL ESTUDIO

Nombre: Comité Institucional de Ética del Hospital Nacional de Clínicas (UNC)

IV. ENMIENDA AL PROTOCOLO

Versión de Enmienda: Primera (1) Fecha de la Enmienda: 18-06-2012


Tipo de Enmienda	<input checked="" type="checkbox"/> Administrativa	<input type="checkbox"/> Seguridad	Tamaño muestra	<input type="checkbox"/> Diseño del ensayo	<input type="checkbox"/> Otra
Alcance de la Enmienda	<input checked="" type="checkbox"/> Al centro		A centros de Argentina	<input type="checkbox"/> Internacional	
Resultado Evaluación CIEIS	<input type="checkbox"/> Aprobada		Aprobada con recomendación	<input type="checkbox"/> Rechazada	

V. ENMIENDA AL CONSENTIMIENTO INFORMADO

Versión de Enmienda: _____ Fecha de la Enmienda: _____

Tipo de Enmienda	<input type="checkbox"/> Administrativa	<input type="checkbox"/> Seguridad	Tamaño muestra	<input type="checkbox"/> Diseño del ensayo	<input type="checkbox"/> Otra
Alcance de la Enmienda	<input type="checkbox"/> Al centro		A centros de Argentina	<input type="checkbox"/> Internacional	
Resultado Evaluación CIEIS	<input type="checkbox"/> Aprobada		Aprobada con recomendación	<input type="checkbox"/> Rechazada	

La información de este formulario tiene carácter de declaración jurada por lo que no se aceptarán inscripciones manuales ni enmiendas.



Firma Investigador

18-06-2012

Fecha

DR. MARIA DEL PILAR DIAZ
PROFESORA TITULAR
ESCUELA DE NUTRICION
FCN° UNC

VI. PAGO DE TASA RETRIBUTIVA

Tasa de Registro: _____ N° de Ticket: _____ Fecha de Pago: ____/____/20__

VII. PÓLIZA DE SEGURO

Compañía Aseguradora: _____

Fecha de vencimiento (deber tener vigencia mayor a 15 días al momento de presentar este formulario): ____/____/20__

VIII. EN CASO DE OBSERVACIONES DEL CIEIS A LA ENMIENDA AL PROTOCOLO


Con observación - especificar: _____


Sin observación

IX. EN CASO DE RECOMENDACIONES DEL CIEIS A LA ENMIENDA AL CONSENTIMIENTO INFORMADO

Con observación - especificar: _____

Sin observación





Prof. Dra. LIDIA L. MENTRULL
Firma Responsable del Comité de Ética

18/06/12

Sello Fecha

CIEIS-HNC



Gobierno de Córdoba
Ministerio de Salud

REGISTRO PROVINCIAL DE INVESTIGACIÓN EN SALUD (REPIS)



Hospital Nacional de Clínicas

Registro de Investigaciones en Salud

COMITE DE ETICA

Hospital Nacional de Clínicas
Registro N° 0144 / 10

Deberán ser registradas todas aquellas investigaciones en seres humanos que se realicen en el ámbito de la Provincia de Córdoba, de carácter experimental u observacional, que impliquen o no nuevos métodos de prevención, diagnóstico, tratamiento y/o rehabilitación.

I. PATROCINANTE

Identificación del Patrocinante de la Investigación ante Financiador: Agencia Nacional de Promoción Científica y Tecnológica

Carácter: Público Privado Otro No corresponde

Domicilio Legal: No corresponde

II. TITULO

Título de la Investigación incluyendo el N° de encuesta si la tiene: Exposiciones ambientales y cáncer en Córdoba: Estudio de la relación dieta-cáncer y construcción de una escala de exposición a contaminantes.

Número: ninguno Nombre abreviado: SEECA (Study of Environmental Epidemiology of Cancer)

Especialidad Vinculada: Epidemiología-Nutrición-Toxicología-Geoquímica

Nombre de la Droga en Estudio: No corresponde

Consentimiento Informado (Versión y Fecha): Cáncer de colon y vías urinarias, versión 1 junio 2007, cáncer de próstata, versión 1 diciembre 2008 y versión 1 febrero 2009, cáncer de mama versión 1 noviembre 2009

III. INVESTIGADOR

Nombre Investigador Responsable: Dra. Sonia Edith Muñoz

Domicilio: Uriteco 1040 Cumbres de Villa Allende

Teléfono/s: 0351-4334020/21(laboral) Mail: smunoz@cmefcm.uncor.edu

Nro de Inscripción: 388 / 371

IV. TIPO DE INVESTIGACIÓN (marcar con cruz lo que corresponda)

Descriptiva	<input type="checkbox"/> Descriptivo simple	Observacional	<input type="checkbox"/> de Cohortes	Experimental	<input type="checkbox"/> Ensayo clínico no farmacológico.
	<input type="checkbox"/> Descriptivo correlacional		<input checked="" type="checkbox"/> de Casos y Controles		<input type="checkbox"/> Ensayo clínico farmacológico con grupo control con placebo.
	<input type="checkbox"/> Descriptivo longitudinal				<input type="checkbox"/> Ensayo clínico farmacológico con grupo control con otro fármaco.
FASE DEL ESTUDIO: No corresponde		Otros (Describe): Observacional	<input type="checkbox"/> Ensayo clínico farmacológico en grupo control		

V. CARACTERÍSTICAS DEL ESTUDIO (marcar lo que corresponda, pueden ser varias marcas)

- Investigación en Genética humana
- Investigación en Reproducción Humana
- Estudios Farmacológicos en Fase I o II
- Estudios Clínicos con Vacunas (cualquier Fase)
- Investigaciones consideradas de alto riesgo
- Nuevos procedimientos aún no descritos y/o validados en la literatura (sólo en Fase IV)
- Investigación con grupos vulnerables, niños, embarazadas, ancianos, pacientes preoperatorios, discapacitados, poblaciones indígenas, prisioneros, etc.
- Investigaciones coordinadas desde el extranjero o con participación extranjera
- Estudios que incluyen envío de material biológico al extranjero

<input type="checkbox"/>	Estudios multicéntricos	<input type="checkbox"/>	Local	<input type="checkbox"/>	Nacional	<input type="checkbox"/>	Internacional
<input checked="" type="checkbox"/>	Otra: Estudios de tipo caso-control						

VI. LUGAR DE REALIZACIÓN	
Ciudad	Córdoba
Establecimiento:	Instituto de Biología Celular, Facultad de Ciencias Médicas – Universidad Nacional de Córdoba.
Domicilio:	Enrique Barros, esq. Enf. Gordillo, Ciudad Universitaria
Ámbito	Público <input checked="" type="checkbox"/> Privado <input type="checkbox"/> Otros (Especificar) -

VII. DURACIÓN	
Duración de la investigación (tiempo estimado desde el inicio a la terminación)	Periodo mínimo de 5 años

VIII. AUTORIDAD RESPONSABLE DEL ESTABLECIMIENTO	
Apellido y Nombre	Dr. Humberto Cabral
Cargo	Director del Instituto de Biología Celular- Facultad de Ciencias Médicas, Universidad Nacional de Córdoba.

IX. SEGURO DE DAÑOS (Específico para ensayos clínicos)	
Compañía Aseguradora	No corresponde
Domicilio:	No corresponde
Ciudad:	No corresponde
País:	No corresponde
Fecha de vencimiento:	No corresponde

X. COMITÉ INSTITUCIONAL DE ÉTICA DE LA INVESTIGACIÓN EN SALUD DONDE SE PRESENTÓ EL ESTUDIO	
Nombre	Comité de Ética del Hospital Nacional de Clínicas
Institución:	Hospital Nacional de Clínicas
Domicilio Legal	Santa Rosa 1569 Córdoba
Coordinador	Dra. Montrull Hilda L.

La información de este formulario tiene carácter de declaración jurada por lo que no se aceptarán inscripciones manuales ni enmiendas realizadas al mismo.

	Dr. SONIA E. MUÑOZ FAC. CIENCIAS MÉDICAS UNIVERSIDAD NACIONAL DE CÓRDOBA INVESTIGADORA ADJUNTA CONICET	12 May 2010
Firma Investigador	Sello	Fecha

XI. EN CASO DE OBSERVACIONES DEL CIEIS (si en la evaluación del protocolo se solicitó enmienda o condicionamientos, etc.)	
Resultado Evaluación	Solicitud de enmienda (especifique)
	Condicionamiento (especifique)
	Otra (especifique)

XII. INFORME CIEIS	
Resultado Evaluación CIEIS	<input checked="" type="checkbox"/> Aprobada
	<input type="checkbox"/> Aprobada con recomendación
	<input type="checkbox"/> Rechazada

	Dr. HILDA L. MONTRULL Coordinadora del Comité de Ética CIEIS-HAC	Prof. Dra. HILDA L. MONTRULL Coordinadora del Comité de Ética CIEIS-HAC	20/5/2010
Firma Autoridad Responsable del CIEIS	Sello	Fecha	

XIII. EN CASO DE OBSERVACIONES DEL CIEIS (si en la evaluación del protocolo se solicita enmienda o condicionamiento, etc.)	
Resultado Evaluación	<input type="checkbox"/> Solicitud de enmienda (especifique)
	<input type="checkbox"/> Condicionamiento (especifique)
	<input type="checkbox"/> Otros (especifique)
XIV. INFORME CONSEJO DE EVALUACIÓN ÉTICA DE INVESTIGACIÓN EN SALUD COEIS (usado retrospectiva)	
Resultado Evaluación	<input type="checkbox"/> Aprobado (Fase I, Fase II, Reproducción Humana, Nuevos procedimientos no farmacológicos)
	<input type="checkbox"/> Aprobado con recomendación
	<input type="checkbox"/> Rechazado

Firma Autoridad/ Responsable

Firma Autoridad/ Responsable

Firma Autoridad/ Responsable

Fecha

Versión 4 - Agosto07

Instrucciones para promotores/ investigadores y CIEIS:

- El investigador debe presentar ante el CIEIS este formulario por triplicado debidamente fechado, firmado y sellado.
- Este formulario debe ser llenado en computadora.
- El CIEIS presentará ante el RePIS el formulario de inscripción, debidamente llenado, fechado y firmado con la resolución del CIEIS.
- Para los casos de estudios previstos en el Art. 18° de la Resolución Ministerial 00622/07 el CIEIS presentará el formulario firmado y sellado ante el COEIS, acompañado de toda la documentación prevista en el Anexo III, y el Informe del CIEIS con detalle de su decisión.
- Para los estudios previstos en el Art. 19° de la presente resolución presentará el formulario firmado y sellado ante el COEIS acompañado de: Resumen del Protocolo, Consentimiento Informado y el Informe del CIEIS con detalle de su decisión.
- En caso de haber solicitado enmienda o condicionamiento deberá especificar los mismos y registrar el protocolo según el resultado de los mismos (aprobado o rechazado)
- El CIEIS deberá informar en un plazo de 10 días hábiles el inicio, finalización, suspensión y/o cancelación del protocolo registrado con el formulario correspondiente.

ENC.	2
------	---

21- Domicilio actual- Calle y número

Barrio

Teléfono

Localidad

22-Residencia/s anterior/es:

¿Vivió en otro departamento de la Provincia de Córdoba anteriormente? SI NO

¿Cuál?..... ¿Cuánto tiempo?..... (en años)

23- Agua de beber habitual o para infusiones:

¿Tomaba agua de red? SI NO ¿Durante cuánto tiempo?.....

¿Tomaba agua de pozo? SI NO ¿Durante cuánto tiempo?

¿Le agregaba gotas de lavandina al agua? SI NO ¿Durante cuantos años?

Cantidad de líquidos: _____ cm³

24- Ocupación (principal sostén del hogar):

a) ¿Trabaja actualmente? SI NO Jubilado/Pensionado

b) ¿Cuántas ocupaciones tiene? 1 empleo 2 o más

c) ¿En qué trabaja o trabajaba?

Ocupación principal _____ Años de antigüedad

Ocupación secundaria _____ Años de antigüedad

d) ¿Tiene empleados a su cargo? SI NO

¿Cuántos? De 1 a 5 más de 5

e) Si el encuestado/a no es el principal sostén del hogar: ¿Trabaja? SI NO Jubilado/Pensionado

f) ¿En qué trabaja o trabajaba?

Ocupación principal _____ Años de antigüedad

g) Trabajo o trabaja en industrias de colorantes, pinturas, textil, plástico, caucho, cuero, herbicidas, automotor, químicos, carbón?

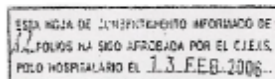
SI NO ¿Cuál?..... ¿Cuántos años?.....

¿Cuántas horas/día?.....

¿Trabajaba al aire libre? SI NO ¿Cuántos años?..... ¿Cuántas horas/día?.....

¿Qué tarea realizaba?.....

¿Le realizaba examen físico pre-ocupacional? SI NO



Prof. Dr. HUGO O. VILARRODHA
COORDINADOR

ENC.		3
------	--	---

h) ¿Tiene obra social? SI NO

Situación Ocupacional A B C D E
(Marcar con un círculo)

25-Nivel Socioeconómico.

Cantidad de aportantes en el hogar 1 2 a 3 4 o más

Conexión a internet: SI NO

Computadora: SI NO

Tarjeta de débito: SI NO

Cantidad de autos (menos de 15 años de antigüedad):.....

26-Situación Educativa: ¿Qué estudios ha realizado? (marcar con una X)

- | | | | |
|------------------------------|--------------------------|-----------------------|--------------------------|
| a) Sin instrucción | <input type="checkbox"/> | <u>Estrato social</u> | |
| b) Primaria incompleta | <input type="checkbox"/> | Alto | <input type="checkbox"/> |
| c) Primaria Completa | <input type="checkbox"/> | Medio/Alto | <input type="checkbox"/> |
| d) Secundaria Incompleta | <input type="checkbox"/> | Medio | <input type="checkbox"/> |
| e) Secundaria Completa | <input type="checkbox"/> | Bajo | <input type="checkbox"/> |
| f) Terciaria o Universitaria | <input type="checkbox"/> | Carenciado | <input type="checkbox"/> |

27- Considerando la actividad física que realizaba en su trabajo y en su tiempo libre ¿Cómo la clasificaría?

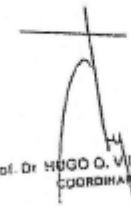
L	M	I	¿Cuál/es?.....
---	---	---	----------------

28- ¿Tuvo o ha tenido alguna de estas enfermedades?

- | | NO | SI |
|---|--------------------------|--------------------------|
| • Diabetes | <input type="checkbox"/> | <input type="checkbox"/> |
| • Pólipos intestinales | <input type="checkbox"/> | <input type="checkbox"/> |
| • Enfermedades del hígado (Hepatitis, cirrosis) | <input type="checkbox"/> | <input type="checkbox"/> |
| • Gota | <input type="checkbox"/> | <input type="checkbox"/> |
| • Cálculos de vesícula biliar (y otros problemas de vesícula) | <input type="checkbox"/> | <input type="checkbox"/> |
| • Tumores benignos | <input type="checkbox"/> | <input type="checkbox"/> |
| • Tumores malignos | <input type="checkbox"/> | <input type="checkbox"/> |
| • Úlcera estomacal y/o duodenal | <input type="checkbox"/> | <input type="checkbox"/> |
| • Grasas y/o colesterol elevado en sangre | <input type="checkbox"/> | <input type="checkbox"/> |

Localización.....
Mes..... Año.....
Diagnóstico.....

ESTA HOJA DE CONSENTIMIENTO APROBADO DE
LA POLIO HA SIDO APROBADA POR EL C.J.E.L.S.
POLO HOSPITALARIO EL 13.FEB.2008..


Prof. Dr. HUGO O. VILARRODONA
COORDINADOR

ENC.		4
------	--	---

	NO	SI
• Enfermedades de la piel	<input type="checkbox"/>	<input type="checkbox"/>
• Enfermedad de Chagas	<input type="checkbox"/>	<input type="checkbox"/>
• Hipertensión crónica	<input type="checkbox"/>	<input type="checkbox"/>
• Insuficiencia renal crónica	<input type="checkbox"/>	<input type="checkbox"/>
• Asma crónico	<input type="checkbox"/>	<input type="checkbox"/>
• Lastimaduras crónicas que no curan (Fistulas, osteomielitis crónica)	<input type="checkbox"/>	<input type="checkbox"/>
Enfermedades del aparato urinario		
*Actualmente, ¿Orina con sangre?	<input type="checkbox"/>	<input type="checkbox"/>
*¿Cistitis o ardor al orinar?	<input type="checkbox"/>	<input type="checkbox"/>
*¿Cuántas veces en su vida? _____ ¿y al año? _____		
*Litiasis (cálculos)?	<input type="checkbox"/>	<input type="checkbox"/>
*¿Infecciones crónicas de vías urinarias?	<input type="checkbox"/>	<input type="checkbox"/>
¿Ha ido al odontólogo el último año?	<input type="checkbox"/>	<input type="checkbox"/>
*¿Le encontró manchas blancas?	<input type="checkbox"/>	<input type="checkbox"/>
*¿Usa dentadura postiza?	<input type="checkbox"/>	<input type="checkbox"/>
¿Cuántos años hace que usa dentadura postiza? _____		
¿Tiene o ha tenido dificultad para tragar?	<input type="checkbox"/>	<input type="checkbox"/>
*¿Desde hace cuantos meses? _____		
¿Otras enfermedades?	<input type="checkbox"/>	<input type="checkbox"/>
¿Cuál?.....		

29- ¿Algún pariente cercano (abuelo, padre, madre, hermanos, hijos) ha tenido tumores malignos o cáncer?
 SI NO NO SE

En caso de que la respuesta sea afirmativa: ¿Quién? _____
 ¿De qué tipo? _____

30- ¿Fuma Ud.? SI NO ¿Cuántos años hace que fuma?
 ¿Qué fuma? C P O ¿Con filtro? SI NO
 ¿Cuántos fuma por día? Rubios _____ Negros _____
 ¿Inhala el humo? SI NO

ESTA HOJA DE RESULTADOS FINANCIADO POR EL C.I.E.S. PUEDEN HA SIDO APROBADA POR EL C.I.E.S. PUEDEN HOSPITALARIO EL 13 FEB 2006

Prof. Dr. HUGO O. VILARRODONA
 COORDINADOR 124

ENC.		5
------	--	---

¿Fumaba Ud.? SI NO ¿Cuántos años fumo?

¿Qué fumaba? C P O

¿Cuántos fumaba por día? Rubios _____ Negros _____

¿Inhala el humo? SI NO

¿Cuántas veces por día fumaba pipa?

¿Cuántos gramos tenía el paquete? ___ ¿Cuántos días le duraba el paquete? ___

Cuando fuma o fumaba,

*¿Acostumbra/ba a tener la pipa o el cigarrillo siempre en los labios? SI NO

*Habitualmente ¿Retiene/retenía el humo en la boca? SI NO

PREGUNTAS DIRIGIDAS A MUJERES

-Edad menarca:.....

-Menopausia: SI NO Edad:.....

-Nº embarazos a término:

-Nº embarazos interrumpidos:

Antes del 3º mes	3 a 6 meses	Después del 6º mes
------------------	-------------	--------------------

-Hijos: SI NO ¿Cuántos?..... Edad al nacer el primer hijo:.....

-Amamantó: SI NO Tiempo en meses:.....

-Anticonceptivos orales: SI NO Tiempo de consumo:.....

-Terapia de reemplazo hormonal: SI NO Tiempo de consumo:

-Otros tratamientos hormonales: SI NO Tipo:.....

Motivo:..... Tiempo de consumo:

-Mamografías: SI NO Cantidad a la fecha: ¿Consumo Tamoxifeno actualmente? SI NO

-Radiología prolongada: SI NO

-¿Conoce el método de autoexamen mamario? SI NO ¿Lo practica? SI NO

31- Hace 5 años atrás o más... Ud.

A) ¿Estaba acostumbrado/a a tomar laxantes o purgantes habitualmente? SI NO

B) ¿Estaba acostumbrado/a a tomar analgésicos y antipiréticos (antifebriles) habitualmente?

SI NO ¿Qué marca?.....

Prof. Dr. HUGO O. VILARRODONA
COORDINADOR

ESTA HOJA DE CONSENTIMIENTO INFORMADO DE
LA FOLIOVA HA SIDO APROBADA POR EL C.I.E.S.
PUEDO HOSPITALIZADO EL 13.FEB.2006..

- C) ¿Tomaba habitualmente suplementos vitamínicos y/o minerales? SI NO
 ¿Qué marca?..... ¿Cuántos años?.....
- D) ¿Usaba habitualmente edulcorantes artificiales? SI NO
 ¿Qué marca?..... ¿Cuántos años?.....
- E) ¿Tomaba habitualmente infusiones muy calientes? SI NO o ¿muy frías? SI NO
- F) ¿Sigue o seguía alguna alimentación especial por razones de salud, religiosas u otros motivos?
 SI NO
 Si la respuesta es afirmativa, indicar el motivo: _____
 ¿Durante cuantos años? _____

F. 1.) ¿Podría decirnos 2 de sus comidas preferidas?

1- Nombre _____
 Principales ingredientes y forma de cocción _____

 ¿Cuántas veces a la semana la consumía? _____

2- Nombre _____
 Principales ingredientes y forma de cocción _____

 ¿Cuántas veces a la semana la consumía? _____

F. 2.) ¿Cuántas comidas realizaba al día? (marcar con una X)

- | | | | |
|---------------|--------------------------|-----------|--------------------------|
| -Desayuno | <input type="checkbox"/> | -Merienda | <input type="checkbox"/> |
| -Media mañana | <input type="checkbox"/> | -Cena | <input type="checkbox"/> |
| -Almuerzo | <input type="checkbox"/> | -Otras | <input type="checkbox"/> |

CUESTIONARIO INTERNACIONAL DE ACTIVIDAD FISICA (Versión Adaptada)

Piense en todas las actividades intensas que usted realizó hace... años atrás o más. Las actividades intensas se refieren a aquellas que implican un esfuerzo físico intenso y que lo hacen respirar mucho más intensamente que lo normal. Piense solo en aquellas actividades físicas que realizaba durante por lo menos 10 minutos seguidos.

1. Hace... años atrás o más, ¿Cuántas veces por semana realizaba actividades físicas intensas tales como levantar pesos pesados, cavar, hacer ejercicios aeróbicos o andar rápido en bicicleta?
días por semana
 Ninguna actividad física intensa Vaya a la pregunta 3


 Prof. Dr. HUGO O. VILARRODONA
 COORDINADOR

ESTE INDICADOR DE DESPENTENIMIENTO INFORMADO DE
 LA POLIPOSIS HA SIDO APROBADA POR EL CIEHS,
 POLO HOSPITALARIO EL 13 FEB 2006.

ENC.		7
------	--	---

2. Habitualmente, ¿Cuánto tiempo en total dedicaba a una actividad física intensa en uno de esos días?
-horas por día
minutos por día
 No sabe/No está seguro

Piense en todas las actividades moderadas que usted realizó hace... años atrás o más. Las actividades moderadas son aquellas que requieren un esfuerzo físico moderado que lo hacen respirar algo más intensamente que lo normal. Piense solo en aquellas actividades físicas que realizaba durante por lo menos 10 minutos seguidos.

3. Hace... años atrás o más, ¿Cuántas veces por semana realizaba actividades físicas moderadas tales como transportar pesos livianos, andar en bicicleta a velocidad regular o jugar dobles de tenis? No incluya caminar
-días por semana
 Ninguna actividad física moderada *Vaya a la pregunta 5*
4. Habitualmente, ¿Cuánto tiempo en total dedicaba a una actividad física moderada en uno de esos días?
-horas por día
minutos por día
 No sabe/No está seguro

Piense en el tiempo que usted dedico a caminar hace... años atrás o más. Esto incluye caminar en el trabajo o en la casa, para trasladarse de un lugar a otro o cualquier otra caminata que usted podría hacer solamente para la recreación, el ejercicio o el ocio.

5. ¿Cuántas veces por semana realizaba caminata de por lo menos 10 minutos seguidos?
-días por semana
 Ninguna caminata *Vaya a la pregunta 7*
6. Habitualmente, ¿Cuánto tiempo en total dedicaba a caminar en uno de esos días?
-horas por día
minutos por día
 No sabe/No está seguro

La última pregunta es acerca del tiempo que paso usted sentado, habitualmente, durante los días hábiles hace... años atrás o más. Esto incluye el tiempo dedicado al trabajo, en la casa, en una clase, y durante el tiempo libre. Puede incluir el tiempo que paso sentado ante un escritorio, visitando amigos, leyendo, viajando en ómnibus, o sentado o recostado mirando televisión.

7. Hace... años atrás o mas ¿Cuánto tiempo solía pasar sentado durante un día hábil?
-horas por día
minutos por día
 No sabe/No está seguro

¿Cuántas horas diarias dormía habitualmente hace 5 años atrás? _____

¿Cuántas horas diarias duerme actualmente? _____

Prof. Dr. HUGO O. VILARRODONA
 COORDINADOR

ESTE REGISTRO DE CONSENTIMIENTO INFORMADO DE
 LA FOLIOVA HA SIDO APROBADA POR EL C.I.E.I.S.
 DEL HOSPITALARIO EL 13.FEB.2006.

ENCUESTA ALIMENTARIA

8

COD.	¿Que alimentos consume habitualmente?.. ALIMENTOS Tipos de cocción	N	¿Con que frecuencia?..			¿Que cantidad?..			Observaciones
			Dias/ Mes	Dias/ Sem.	Veces /Día	P	M	G	
L908	LECHE DESCREMADA								
L554	LECHE ENTERA								
L832	YOGUR ENTERO								
L243	YOGUR DESCREMADO								
L768	QUESO MANTECOSOS								
L557	QUESO SEMI-DURO (Tipo Senda)								
L135	QUESO DE RALLAR								
L380	QUESO RALLADO								
L578	RICOTTA								
H500	HUEVO FRITO								
H380	HUEVO ENTERO								
C672	CARNE MAGRA HERVIDA								
C698	CARNE MAGRA PLANCHA C/C								
C052	CARNE MAGRA PLANCHA S/C								
C406	CARNE MAGRA PARRILLA C/C								
C804	CARNE MAGRA PARRILLA S/C								
C731	CARNE MAGRA HORNO C/C								
C269	CARNE MAGRA HORNO S/C								
C002	CARNE FRITA C/C								
C096	CARNE FRITA S/C								
C402	CARNE GRASA HERVIDA								
C851	CARNE GRASA PLANCHA C/C								
C554	CARNE GRASA PLANCHA S/C								
C908	CARNE GRASA PARRILLA C/C								
C832	CARNE GRASA PARRILLA S/C								
C089	CARNE GRASA HORNO C/C								
C803	CARNE GRASA HORNO S/C								
C001	CARNE GRASA FRITA C/C								
C067	CARNE GRASA FRITA S/C								
C822	PUCHERO HERVIO C/ CARACU								
C336	PUCHERO HERVIO S/ CARACU								
C161	ASADO DE TIRA-COSTILLA/PARRILLA C/C								
C621	ASADO DE TIRA-COSTILLA/PARRILLA S/C								
C192	MOLLEJA PARRILLA C/C								
C927	MOLLEJA PARRILLA S/C								
C284	LENGUA HERVIDA								
C886	CHINCHULIN PARRILLA C/C								
C282	CHINCHULIN PARRILLA S/C								
C881	SESOS HERVIDOS								

Prof. Dr. HUGO O. VILARRODONA
COORDINADOR

ESTA HOJA DE RESULTADOS INFORMADO DE
LA FOLIOS HA SIDO APROBADA POR EL C.I.E.A.S.
POLO HOSPITALARIO EL 13.FEB.2006..

COD.	ALIMENTOS Tipos de cocción	N	Dias/ Mes	Dias/ Sem.	Veces /Dia	P	M	G	Observaciones
C243	CHORIZO HERVIDO (GUISO)								
C254	CHORIZO PARRILA C/C								
C472	CHORIZO PARRILA S/C								
C768	MORCILLA								
C557	SALCHICHA PARRILERA HERVIDA								
C835	SALCHICHA PARRILERA PARRILLA C/C								
C566	SALCHICHA PARRILERA PARRILLA S/C								
C135	SALCHICHA DE VIENA HERVIDA								
C607	CERDO HERVIDO								
C075	CERDO PLANCHA C/C								
C993	CERDO PLANCHA S/C								
C174	CERDO PARRILLA C/C								
C788	CERDO PARRILLA S/C								
C262	CERDO HORNO C/C								
C654	CERDO HORNO S/C								
C618	GALLINA C/PIEL HERVIDA								
C350	GALLINA S/PIEL HERVIDA								
C396	POLLO C/PIEL HERVIDO								
C518	POLLO S/PIEL HERVIDO								
C854	POLLO C/PIEL PLANCHA C/C								
C500	POLLO C/PIEL PLANCHA S/C								
C353	POLLO S/PIEL PLANCHA C/C								
C465	POLLO S/PIEL PLANCHA S/C								
C359	POLLO C/PIEL PARRILLA C/C								
C218	POLLO C/PIEL PARRILLA S/C								
C659	POLLO S/PIEL PARRILLA C/C								
C403	POLLO S/PIEL PARRILLA S/C								
C144	POLLO C/PIEL HORNO C/C								
C663	POLLO C/PIEL HORNO S/C								
C616	POLLO S/PIEL HORNO C/C								
C732	POLLO S/PIEL HORNO S/C								
C865	POLLO C/PIEL FRITO C/C								
C219	POLLO C/PIEL FRITO								
C573	POLLO S/PIEL FRITO C/C								
C946	POLLO S/PIEL FRITO S/C								
P339	PESCADO GRASO PLANCHA C/C								
P338	PESCADO GRASO PLANCHA S/C								
P944	PESCADO GRASO PARRILLA C/C								
P503	PESCADO GRASO PARRILLA S/C								
P004	PESCADO GRASO FRITO C/C								
P314	PESCADO GRASO FRITO S/C								

Prof. Dr. HUGO O. VILARRODONA
COORDINADOR

SE HA HECHO DE CONFORMIDAD INFORMADO DE
QUE FOLIOS HA SIDO APROBADA POR EL C.I.E.S.
POLO HOSPITALARIO EL 13.FEB.2006.

COD.	ALIMENTOS Tipos de cocción	N	Días/ Mes	Días/ Sem.	Veces /Dia	P	M	G	Observaciones
P388	PESCADO MAGRO HERVIDO								
P669	PESCADO MAGRO PLANCHA C/C								
P469	PESCADO MAGRO PLANCHA S/C								
P319	PESCADO MAGRO PARRILLA C/C								
P939	PESCADO MAGRO PARRILLA S/C								
P568	PESCADO MAGRO HORNO C/C								
P399	PESCADO MAGRO HORNO S/C								
P005	PESCADO MAGRO FRITO C/C								
P153	PESCADO MAGRO FRITO S/C								
P724	ATUN Y CABALLA								
P383	SARDINA EN ACEITE								
C017	BONDIOLA								
C380	SALAMIN								
C578	JAMON CRUDO								
C388	JAMON COCIDO-PALETA								
C105	MORTADELA								
C724	SALCHICHON								
C383	SALAME MILAN								
C306	PANCETA-TOCINO FRITO C/C								
C379	PANCETA-TOCINO HERVIDO (guiso)								
C242	QUESO DE CERDO								
V698	ACELGA OTRA COCCION								
V118	ACHICORIA								
V318	AJO COCIDO								
V145	AJO CRUDO								
V822	ALCAUCIL OTRA COCCION								
V282	APIO CRUDO								
V426	ARVEJAS FRESCAS O EN LATA								
V406	BATATA OTRA COCCION								
V804	BERENJENA OTRA COCCION								
V889	BERRO CRUDO								
V017	BROCOLI OTRA COCCION								
V231	CALABAZA O CALABACIN OTRA COCCION								
V927	CEBOLLA CRUDA								
V562	CEBOLLA FRITA								
V972	CEBOLLA OTRA COCCION								
V953	CHAUCHA OTRA COCCION								
V629	CHOCLO OTRA COCCION								
V242	COLIFLOR OTRA COCCION								
V243	ESPARRAGO OTRA COCCION								
V862	ESPINACA OTRA COCCION								
V580	LECHUGA								

Prof. Dr. HUGO O. VILARRODHA
COORDINADOR

ESTA HOJA DE DESPESIERO INFORMADO DE
LA POLISA HA SIDO APROBADA POR EL C.I.E.A.S.
POLO HOSPITALARIO EL 13.FEB.2006..

COD.	ALIMENTOS Tipos de cocción	N	Días/ Mes	Días/ Sem.	Veces /Día	P	M	G	Observaciones
V016	PAPA FRITA								
V078	PAPA OTRA COCCION								
V404	PEPINO CRUDO								
V947	PIMIENTO CRUDO								
V638	PIMIENTO OTRA COCCION								
V089	RABANITO CRUDO								
V573	REMOLACHA OTRA COCCION								
V803	REPOLLO CRUDO								
V748	REPOLLO OTRA COCCION								
V041	TOMATE CRUDO								
V746	TOMATE OTRA COCCION								
V938	ZANAHORIA CRUDA								
V310	ZANAHORIA OTRA COCCIO								
V192	ZAPALLITO OTRA COCCION								
V269	ZAPALLO OTRA COCCION								
F211	ANANA (crudo)								
F926	ANANA CONSERVA								
F337	BANANA (cruda)								
F254	CIRUELA (cruda)								
F605	CIRUELA OTRA COCCION								
F472	DAMASCO (crudo)								
F432	DURAZNO (crudo)								
F273	DURAZNO OTRA COCCION								
F665	FRUTILLA (cruda)								
F077	KIWI (crudo)								
F208	LIMON (crudo)								
F675	MANDARINA (cruda)								
F434	MANZANA (cruda)								
F735	MANZANA OTRA COCCION								
F242	MELON (crudo)								
F867	NARANJA (cruda)								
F594	PERA (cruda)								
F556	PERA OTRA COCCION								
F377	POMELO (crudo)								
F453	SANDIA (cruda)								
F804	UVA (cruda)								
N371	MANI								
N375	NUEZ								
N711	ALMENDRA								

Prof. Dr. HUGO O. VILARRODONA
COORDINADOR

ESTA NGIA DE LICENCIAMIENTO INFORMADO DE
LA FOLIOS HA SIDO APROBADA POR EL C.I.E.S.
POLO HOSPITALARIO EL 13.EEB.2006..

COD.	ALIMENTOS Tipos de cocción	N	Días/ Mes	Días/ Sem.	Veces /Dia	P	M	G	Observaciones
T379	PAN FRANCES								
T631	PAN NEGRO								
T613	CRIOLOS								
T301	PAN CON GRASA								
T300	TORTA FRITA								
T297	TORTA								
T303	FACTURAS								
T811	GALLETITAS DULCES								
T892	GALLETITAS SALADAS								
T655	GALLETITAS INTEGRALES								
T920	GRANOS NO INTEGRALES								
T393	GRANOS INTEGRALES								
T188	PASTAS								
D802	LENTEJA								
D951	SOJA								
D231	POROTO-GARBANZO-ARVEJA								
G262	ACEITE DE GIRASOL (crudo)								
G654	ACEITE DE MAIZ (crudo)								
G338	ACEITE DE OLIVA (crudo)								
G788	ACEITE DE UVA (crudo)								
G993	ACEITE MEZCLA (crudo)								
G188	ACEITE PATITO (crudo)								
G075	CREMA DE LECHE (cruda)								
G908	GRASA DE CERDO (cruda)								
G832	GRASA DE VACA (cruda)								
G557	MANTECA (cruda)								
G244	MARGARINA (cruda)								
G607	MAYONESA								
I803	CAFE								
I426	MATE BOMBILLA								
I940	MATE COCIDO								
I089	TE								
M803	CACAO								
A889	AZUCAR								
A916	DULCE DE LECHE								
A242	MERMELADAS								
A089	MIEL								

Prof. Dr. HÉGO O. V. LARRODONA
COORDINADOR

ESTA HOJA DE CONSENTIMIENTO INFORMADO DE
LA FOLIO HA SIDO APROBADA POR EL C.I.E.I.S.
POZO HOSPITALANO EL 13.FEB.2006..

13

COD.	ALIMENTOS Tipos de cocción	N	Dias/ Mes	Dias/ Sem.	Veces /Dia	P	M	G	Observaciones
M052	CARAMELOS								
M940	CHOCOLATES								
M698	GOLOSINAS (ALFAJOR-TURRON)								
M554	HELADOS								
M426	MANTECOL								
B468	APERITIVOS								
B803	BEBIDAS ESPIRITUOSAS								
B406	CERVEZA								
B044	GINEBRA								
B600	GRAPPA								
B804	VINO								
B703	WHISKY								
B851	AMARGO SERRANO								
B731	GASEOSAS								
B269	JUGOS PARA DILUIR AL 20%								

Describir otro/s alimento/s o bebida/s de consumo habitual que no figure/n en el listado (ACLARAR):

COD.	ALIMENTOS Tipos de cocción	N	Dias/ Mes	Dias/ Sem.	Veces /Dia	P	M	G	Observaciones

Sub-total

Total


 Prof. Dr. HUGO O. VILARRODONA
 COORDINADOR

ESPANOLA DE CONSENTIMIENTO INFORMADO DE
 AL. FOLIOS HA SIDO AFRODEIDA POR LA C.I.E.S.
 POLO HOSPITALARIO EL 1.3.EEB. 2006.

ANEXO 4. Categorización de la variable Nivel Socioeconómico

Índice del nivel socioeconómico 2002. Asociación Argentina de Marketing – Cámara de Control y Medición de Audiencias. 2003. Argentina.

Se calcula por sumatoria de puntos de las variables seleccionadas de interés para la construcción del mismo. Para construir los estratos sociales se tuvieron en cuenta las siguientes variables: cantidad de personas que aportan en el hogar, posesión de bienes y servicios, tipo de atención médica y nivel educativo del principal sostén del hogar.

Las categorías son las siguientes:

1. Nivel económico social alto
2. Nivel económico social medio alto
3. Nivel económico social medio típico
4. Nivel económico social bajo superior
5. Nivel económico social bajo inferior
6. Marginal.

OCUPACION PRINCIPAL SOSTÉN DEL HOGAR	PUNTAJE
Dueño o alta dirección de empresas (más de 50 empleados)	32 ptos.
Dueño o alta dirección de empresas (6 a 50 empleados)	28 ptos.
Dueño o alta dirección de empresas (1 a 5 empleados)	22 ptos.
Profesionales independientes	16 ptos.
Técnico independientes y en relación de dependencia	12 ptos.
Comerciantes sin personal, artesanos, encargados y supervisores	10 ptos.
Empleados sin jerarquía	7 p
Supervisores	10 p
Obreros calificados o especializados	6 p
Trabajos autónomos u obreros no calificados	4 p
Ocupación sin inserción en act. económica genuina (changas)	2 p
Pasivos e inactivos (jubilados)	4 p
Desocupados	2 p

Cantidad de aportantes en el hogar:

- 4 ó + → 9 p
- 2 a 3 → 7 p
- 1 → 1 p

Nivel educativo del PSH:

- Univ. completo o Postgrado.....13 p
- Univ. inc. o Terciario o Secundario completo.....4 p
- Secundario inc., primario completo o inc.....0 p

Posesión de Bienes y Servicios:

- Conexión a internet en el hogar.....8 p
- Computadora en el hogar.....6 p
- Tarjeta de débito.....5 p
- Cantidad de autos (de menos de 15 años de antigüedad)
 - 2 ó +.....22 p
 - 1.....11 p

Tipo de atención médica:

- Privada o con obra social o medicina prepaga....5 p
- Hospital público0 p

Puntuaciones:

Código	CLASE	PUNTAJE
1	Alta	55 a 100
2	Media alta	37 a 54
3	Media típica	29 a 36
4	Baja superior	17 a 28
5	Baja inferior	12 a 16
6	Marginal	0 a 11

ANEXO 5. Categorización de la variable Actividad Física

Categorización de la variable Nivel de Actividad Física según IPAQ

Cálculo de METs

Caminata METs minutos/semana: $3,3 * \text{minutos caminados} * \text{días}$.

Actividad Moderada METs minutos/semana: $4 * \text{minutos de actividad física de intensidad moderada} * \text{días}$.

Actividad Física Alta METs minutos/semana: $8 * \text{minutos de actividad física de intensidad vigorosa} * \text{días}$.

METs Totales = Caminata METs minutos/semana + Actividad Moderada METs minutos/semana + Actividad Física Alta METs minutos/semana.

Categoría 1 Bajo

Este es el nivel más bajo de actividad física. Los individuos que no cumplen con las Categorías 2 o 3 son considerados dentro de éste nivel de actividad física.

Categoría 2 Moderado

El nivel de actividad física del individuo se clasifica como moderado cuando cumple con los siguientes criterios:

- a) 3 o más días de actividad física de intensidad vigorosa de por lo menos 20 minutos por día.
ó
- b) 5 o más días de actividad física de intensidad moderada y/o caminar por lo menos 30 minutos por día.
ó
- c) 5 o más días de alguna combinación de caminata, de actividad física de intensidad moderada o vigorosa, llegando a cumplir con un mínimo total de actividad física de por lo menos 600 METs minutos/semana.

Categoría 3 Alto

Cuando se cumplan con los siguientes criterios, el nivel de actividad física de un individuo será clasificado como alto.

a) Un mínimo de 3 días de actividad física de intensidad vigorosa llegando a cumplir un mínimo total de 1500 METs minutos/semana.

ó

b) 7 días de combinación de caminata, de actividad física de intensidad moderada o vigorosa, llegando a cumplir con un mínimo total de actividad física de por lo menos 3000 METs minutos/semana.

ANEXO 6. Registro Provincial de Investigaciones en Salud (RePIS N°159/13)

Formulario Inscripción INVESTIGACIONES ClEIS V. 1. 2011



Gobierno de Córdoba
Ministerio de Salud

REGISTRO PROVINCIAL DE INVESTIGACIÓN EN SALUD (RePIS)
REGISTRO DE INVESTIGACIONES NO PATROCINADAS

HNG
Hospital Nacional de Clínicas

COMITE DE ETICA
Hospital Nacional de Clínicas

Registro N° 159/13


N° DE REPIS
DE INVESTIGACIONES NO
PATROCINADAS
REGISTRADAS EN ClEIS

I. PATROCINANTE (Universidad, Fundación, Institución que otorga beca. Corresponde marcar NO, si es independiente de autogestión)			
Posee Patrocinante:	SI <input type="checkbox"/> NO <input checked="" type="checkbox"/>		
Identificación:			
Carácter:	Público <input type="checkbox"/> Privado <input type="checkbox"/> Otro <input type="checkbox"/>		
II. TÍTULO			
Título de la Investigación	Estudio prospectivo sobre el efecto del sobrepeso y obesidad en sujetos con cáncer de mama en la ciudad de Córdoba.		
Especialidad Vinculada	Nutrición		
Objetivo principal	Determinar el efecto del sobrepeso y la obesidad en el pronóstico de pacientes con cáncer de mama, diagnosticados en fase temprana, en Córdoba (Argentina), considerando factores alimentarios asociados.		
Consentimiento Informado (Versión y Fecha)			
III. INVESTIGADOR			
Nombre Investigador Principal	Ana Gabriela De La Quintana		
IV. EQUIPO DE INVESTIGADORES			
Nombres			
V. TIPO DE INVESTIGACIÓN (marcar con cruz lo que corresponda)			
Descriptivo	<input type="checkbox"/> Descriptivo simple	Observacional	<input checked="" type="checkbox"/> de Cohortes
	<input type="checkbox"/> Descriptivo correlacional		<input type="checkbox"/> de Casos y Controles
	<input type="checkbox"/> Descriptivo longitudinal		
Otros (Describir):			
VI. CARACTERÍSTICAS DEL ESTUDIO (marcar lo que correspondo, pueden ser varias marcas)			
<input type="checkbox"/>	Investigación en Genética humana		
<input type="checkbox"/>	Investigación en Reproducción Humana		
<input type="checkbox"/>	Estudios Farmacológicos en Fase I o II		
<input type="checkbox"/>	Estudios Clínicos con Vacunas (cualquier Fase)		
<input type="checkbox"/>	Investigaciones consideradas de alto riesgo		
<input type="checkbox"/>	Nuevos procedimientos aún no descriptos y/o validados en la literatura (aún en Fase IV)		

<input checked="" type="checkbox"/>	Investigación con grupos vulnerables, (niños, embarazadas, ancianos, pacientes psiquiátricos, discapacitados, poblaciones marginales, prisioneros, etc.)
<input type="checkbox"/>	Estudios que incluyen investigación en farmacogenómica
VII. VINCULACIÓN INTERNACIONAL	
<input type="checkbox"/>	Investigaciones coordinadas desde el extranjero o con participación extranjera
<input type="checkbox"/>	Estudios que incluyen envío de material biológico al extranjero

VIII. ALCANCE DEL ESTUDIO			
<input type="checkbox"/>	Estudios multicéntricos	<input checked="" type="checkbox"/>	Local
<input type="checkbox"/>		<input type="checkbox"/>	Nacional
<input type="checkbox"/>	Otros:	<input type="checkbox"/>	Internacional

IX. DURACIÓN	
Duración de la investigación (tiempo estimado desde el inicio a la terminación)	4 años

X. LUGAR DE REALIZACIÓN		N° DE REPIS DEL CENTRO
Establecimiento/s:	Cátedra de Estadística y Bioestadística, Escuela de Nutrición, Facultad de Ciencias Médicas, UNC	
Ciudad:	Córdoba	
Domicilio:	Enrique Barrios s/n. Ciudad Universitaria	
Ámbito:	Público <input checked="" type="checkbox"/> Privado <input type="checkbox"/> Otros (Especificar):	
XI. AUTORIDAD RESPONSABLE DEL ESTABLECIMIENTO		Firma y sello de la autoridad - Fecha
Apellido y Nombre	Perovic Nilda Raquel	 DRA. NILDA RAQUEL PEROVIC DIRECTORA ESCUELA DE NUTRICION F.C.M. - U.N.C.
Cargo	Directora	

XII. PÓLIZA DE SEGURO	
Compañía Aseguradora	
Domicilio:	
Ciudad:	
País:	
Fecha de vencimiento (deber tener vigencia mayor a 15 días al momento de presentar este formulario) ____ / ____ / 20__	
Póliza presenta renovación automática:	<input type="checkbox"/> Si <input type="checkbox"/> No Período de duración de cada Renovación: _____

XIII. COMITE INSTITUCIONAL DE ETICA DE LA INVESTIGACION EN SALUD DONDE SE PRESENTO EL ESTUDIO	
Nombre	CIEIS
Institución	Hospital Nacional de Clínicas, Universidad Nacional de Córdoba
Coordinador	Dra. Hilda Montrull

La información de este formulario tiene carácter de declaración jurada por lo que no se aceptarán inscripciones manuales ni enmiendas realizadas al mismo.


 Firma Investigador

 Sello

 Fecha 30/05/13

XIV. INFORME CIEIS	
Resultado Evaluación CIEIS	<input type="checkbox"/> Condicionada <input checked="" type="checkbox"/> Aprobada <input type="checkbox"/> Rechazada

[Firma manuscrita]
 Firma Autoridad Responsable
 del CIEIS

Prof. Dra. HILDA L. MONTRULL *13 de Oct. 13*
 Coordinadora del Comité de Bioética
 del BELLVU Fecha

XV. EVALUACIÓN DE LA INSTITUCIÓN ACADÉMICA (Comité de tesis, Comité Capacitación y Docencia, otros)					
Tiene Evaluación:	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Si	<input type="checkbox"/>	No
Organismo evaluador	<input type="checkbox"/>	Comité de tesis universitaria o Entidad deontológica (Consejo de Médicos)			
	<input type="checkbox"/>	Comité de Capacitación y docencia			
	<input type="checkbox"/>	Otro			
	<input type="checkbox"/>				

XVI. EVALUACIÓN DE COEIS					
Se Solicita Evaluación:	<input type="checkbox"/>	<input type="checkbox"/>	Si	<input checked="" type="checkbox"/>	No
Razón:	<input type="checkbox"/>	La investigación posee características nombradas en el Art. 21- Ley 9694/09 (Art. 18 Res. 22/07)			
	<input type="checkbox"/>	La investigación posee características nombradas en el Art.22- Ley 9694/09 (Art. 19 Res. 22/07)			
	<input type="checkbox"/>	Ambos			

XVII. INFORME CONSEJO DE EVALUACIÓN ÉTICA DE INVESTIGACIÓN EN SALUD COEIS (cuando corresponda)		
Resultado Evaluación	<input type="checkbox"/>	Aprobado (Fase I, Fase II, Reproducción Humana, Nuevos procedimientos no farmacológicos)
	<input type="checkbox"/>	Aprobado con recomendación
	<input type="checkbox"/>	Condicionada y aprobada
	<input type="checkbox"/>	Condicionada y rechazada
	<input type="checkbox"/>	Rechazada
	<input type="checkbox"/>	Rechazada por no cumplir condicionamiento

Firma Autoridad/ Responsable

Firma Autoridad/ Responsable

Firma Autoridad/ Responsable

Fecha

Instrucciones para promotores/ Investigadores y CIEIS:

- El Investigador debe presentar ante el CIEIS este formulario por duplicado debidamente fechado, firmado y sellado.
- Este formulario debe ser llenado en computadora.
- El CIEIS deberá informar el inicio, finalización, suspensión y/o cancelación del protocolo registrado en el informe trimestral, formato electrónico, enviado al RePIS.
- El CIEIS presentará ante el RePIS el formulario de inscripción, debidamente llenado, fechado y firmado con la resolución del CIEIS, en caso de encuadrarse dentro del Art. 21 y Art. 22 de la Ley 9694/09.

HNC

Hospital Hospital de Clínicas

COMITÉ DE ÉTICA

Hospital Hospital de Clínicas

Registro: 159/13

HNC 100 años



Comité de ética

1. Datos y características del estudio

1	Título de la investigación	Estudio prospectivo sobre el efecto del sobrepeso y obesidad en sujetos con cáncer de mama en la ciudad de Córdoba.
2	Institución donde se llevara a cabo el estudio	Escuela de Nutrición. FCM - UNC
3	Autoridad responsable de la institución o Director de tesis (según corresponda)	Dra. Nilda Raquel Perovic
4	Investigador responsable	De La Quintana Ana Gabriela Lic. en Nutrición
5	Tipo de investigación	De cohortes

2. Resolución final

El CEHNC, en reunión plenaria, ha decidido aprobar el protocolo ad referendum de la reunión plenaria del 13 de junio de 2013 según consta en el Libro de Actas 1, Nº 125. La investigación cumple con las disposiciones y regulaciones provinciales y nacionales que salvaguardan los derechos de los participantes en investigación clínica.

Se comunica al investigador principal deberá notificar a este Comité:

- o El inicio de la ejecución del estudio dentro de las 72 horas (mediante el formulario correspondiente).
- o La suspensión o cancelación del estudio de manera inmediata
- o El reporte de avance anual (según lo disponga este comité) donde incluya el estado de progreso del protocolo y cualquier otra situación que modifique el curso de la investigación y/o revista mayor seguridad para los participantes
- o Las enmiendas al protocolo de manera oportuna y para su aprobación



Firma del Presidente / Coordinador

ANEXO 7. Publicaciones científicas

Becaria Coquet J, Muñoz SE, Díaz MP. A framework to address potential bias in colorectal cancer: its implementation on a nutritional epidemiologic study in Argentina. *Indian Journal of Applied Research* 2017; 7(12). En prensa.

Becaria Coquet J, Tumas N, Osella AR, Tanzi M, Franco I, Díaz MP. Breast Cancer and Modifiable Lifestyle Factors in Argentinean Women: Addressing Missing Data in a Case-Control Study. *Asian Pac J Cancer Prev* 2016; 17 (10), 4567-4575.

Becaria Coquet J, Juárez SM, Flores MA, Pou SA, Aballay LR, Díaz MP. Identificación de factores de confusión en el estudio de la relación cáncer-dieta. *Rev. Chi. Salud Pública. Rev Chil Salud Pública* 2014; 18(2): 161-172.