UNIVERSIDAD NACIONAL DE COLOMBIA

Estimación de los Costos Directos de los Eventos Reportables en Pacientes Hospitalizados en el Hospital Universitario de Neiva en 2006, usando la Puntuación de Propensión para disminuir los Sesgos de Selección

Marlio Charry Barrios

Universidad Nacional de Colombia
Facultad de Medicina
Instituto de Investigaciones Clínicas
Bogotá D.C., Colombia
2014

Estimación de los Costos Directos de los Eventos Reportables en Pacientes Hospitalizados en el Hospital Universitario de Neiva en 2006, usando la Puntuación de Propensión para disminuir los Sesgos de Selección

Marlio Charry Barrios
Código: 597861

Tesis presentada como requisito parcial para optar al grado de:
Magister en Epidemiología Clínica

Director:

M.D. MSc. Oscar Guevara.

Codirector:

QF. MSc. PhD (c) Jorge Diaz.

Línea de Investigación:
Seguridad del paciente y mejoramiento continuo de Calidad
Grupo de Investigación:
Evaluación de Tecnologías y Políticas en Salud

Universidad Nacional de Colombia
Facultad de Medicina
Instituto de Investigaciones Clínicas
Bogotá D.C., Colombia
2014

A Jesús David, mi obra consentida quien hoy sigue mis pasos

A Patricia, mi apoyo permanente

A mi madre, mi padre (qepd) y mis hermanos

A mis compañeros de maestría, colegas y pacientes

A Dios, mi Ser Superior que me ha permitido ser lo que soy

Agradecimientos

A todos los docentes de la Maestría en Epidemiología Clínica de la Universidad Nacional de Colombia por todas sus enseñanzas y hacerme un mejor profesional.

A todas las personas del Hospital Universitario de Neiva que hicieron posible esta tesis

Un agradecimiento muy especial al Dr. Oscar Guevara y al professor Jorge Diaz por la dirección de la tesis y al Dr. Hernando Gaitán Duarte por su ánimo permanente y disposición a sacar adelante este proyecto

Resumen

Introducción: En Colombia, el Ministerio de Salud y Protección Social está desarrollando la Política de Seguridad del Paciente cuyo objetivo es prevenir la ocurrencia de eventos reportables (ER). En Colombia se ha encontrado una incidencia acumulada de 4,6% de ER en pacientes hospitalizados, pero no hay estudios de análisis de costos directos de eventos reportables (ER) en general en pacientes hospitalizados en instituciones públicas de alta complejidad y utilizando la puntuación de propensión para disminuir los sesgos de selección.

Objetivo: Determinar si existen diferencias en los costos directos totales de hospitalización entre los pacientes con ER y sin ER en el Hospital Universitario de Neiva en Colombia.

Metodología: Cohorte retrospectiva durante el año 2006 y análisis de costos directos de pacientes con y sin ER hospitalizados mínimo 12 horas en una institución pública general de alta complejidad en Neiva, Colombia. Se excluyeron los pacientes con alteraciones psiquiátricas. La muestra final fue de 124 pacientes, 58 (56,31% de los 103 individuos que presentaron ER) con ER y 66 sin ER. Fuente de información: Base de datos del estudio de Gaitán[1] y colaboradores y facturas de las hospitalizaciones. Procedimientos: Determinación de los factores de riesgo de los individuos asociados con los ERs y los principales insumos y servicios utilizados responsables de los costos directos. Cálculo de la puntuación de propensión de los individuos con ER y sin ER según los factores de riesgo -variables independientes - pertinentes, emparejamiento de los grupos de expuestos y no expuestos de acuerdo a la puntuación de propensión para disminuir el sesgo de selección. Estimación de la diferencia de costo por ER, expresado en unidades monetarias, calculando el ATT (*average treatment on the treated)* con la puntuación de propensión. Variable dependiente: ER y el costo total institucional de la hospitalización.

Resultados: De los factores de riesgo medidos, la edad, el puntaje de comorbilidad, los días de estancia y el servicio de hospitalización fueron los que mostraron diferencia significativa entre los dos grupos, revelando que los individuos con ER tienen mayor edad (11 años) y puntaje de comorbilidad, más días de estancia (10 días), y son

hospitalizados en los servicios de medicina interna, cirugía general y gineco-obstetricia. La estimación de la puntuación de propensión utilizó un modelo probit y el poder del modelo indicó una excelente capacidad discriminatoria para los individuos de tener o no un ER. El emparejamiento se realizó mediante un algoritmo *kernel.* Después del emparejamiento, la distribución de las puntuaciones de propensión en ambos grupos no fue completamente coincidente y se consideraron no balanceadas en su totalidad; sin embargo si produjo una reducción importante del sesgo de selección entre los grupos. El estudio mediante la aplicación de la puntuación de propensión estimó una diferencia en el costo promedio total institucional en cada paciente de $22'154.423,1. Esta cantidad representa un incremento de 3.72 veces en el costo promedio de los pacientes con ER. Como información adicional, sin emparejar se estimó el costo institucional para las diferentes categorías de los recursos utilizados encontrando diferencias significativas (p=0.0000) en la mayoría de las categorías de los recursos utilizados, siendo mayor en los individuos que tuvieron ER. Estas diferencias podrían ser explicadas por sesgos de selección.

Conclusiones: La metodología de puntuación de propensión es útil para el análisis de estudios observacionales incluidos los estudios de cohortes y análisis de costo al disminuir el sesgo de selección y calculando el ATT (*average treatment on the treated)*. El estudio logró el objetivo de mostrar que si existen diferencias de los costos directos totales de hospitalización entre los pacientes con y sin ERs. El costo fue mayor en los pacientes que presentaron ERs. Este resultado fue consistente teniendo en cuenta que se mostró de tres formas diferentes y en el mismo sentido. También justifica en el Hospital Universitario de Neiva la implementación de intervenciones que disminuyan la ocurrencia de los ERs.

Palabras Claves: Costos directos - hospitalización - eventos reportables - puntuación de propensión – sesgo de selección

Abstract

Introduction: In Colombia, the Ministry of Health and Social Protection is developing the Patient Safety Policy which aims to prevent the occurrence of reportable events (RE). In Colombia we have found a cumulative incidence of 4.6% of ER in hospitalized patients, but no studies analyzing direct costs of reportable events (RE) in general in patients hospitalized in public institutions of higher complexity and using propensity score to reduce selection bias.

Objective:To determine whether there are differences in the total direct costs of hospitalization among patients without RE or with RE at University Hospital of Neiva in Colombia.

Methods: Retrospective cohort study in 2006 and analysis of direct costs of patients with and without RE hospitalized at least 12 hours in a general public institution of higher complexity in Neiva, Colombia. Patients with psychiatric disorders were excluded. The final sample included 124 patients, 58 (56.31% of the 103 individuals who presented RE) RE and 66 without RE. Source:

Database Gaitán study [1] and collaborators and hospitalization bills. Procedures: Identification of risk factors for individuals associated with the REs and the main inputs and services used responsible for direct costs. Calculation of propensity score individuals without ER or with RE according –variables independent risk factors - relevant, matching groups of exposed and unexposed according to the propensity score to reduce selection bias. Estimating the cost difference for RE, expressed in monetary units, calculating the ATT (average treatment on the treated) with the propensity score. Dependent variable:
RE and total institutional cost of hospitalization.

Results: Of the risk factors measured, age, comorbidity score, length of hospital
stay and hospital service were those that showed significant difference between
the two groups, revealing that individuals with RE are older (11 years) and comorbidity score, more days of stay (10 days), and are hospitalized in the
internal medicine, general surgery and obstetrics and gynecology. The estimation of the

propensity score used a probit model and the power of model indicated excellent discriminatory ability for individuals to have or not an RE.

Matching is performed using a kernel algorithm. After pairing, the distribution of propensity scores in both groups was not entirely consistent and considered unbalanced in its entirety; however there was a significant reduction of selection bias between groups. The study by applying the propensity score estimated a difference in the average total cost institutional for each patient $22'154.423,1. This amount represents an increase of 3.72 times the average cost of patients with RE.

As additional information, the institutional cost unmatched for different categories of resources spent finding significant differences (p = 0.0000) in most of the categories of resources used, being higher in individuals who had RE was estimated. These differences could be explained by selection bias.

Conclusions: The propensity score methodology is useful for the analysis of observational studies included cohort studies and cost analysis to reduce selection bias and calculating the ATT (average treatment on the treated).

The study achieved the goal of showing that if there are differences in the total direct costs of hospitalization among patients with and without REs. The cost was higher in patients who had REs. This result was consistent considering that showed three different forms and in the same direction. It also justifies the Hospital Universitario de Neiva the implementation of interventions that reduce
the occurrence of REs.

Keywords: Direct costs - hospitalization- reportables events - propensity score - selection bias

Contenido

Pág.

Resumen ..... **IX**

Lista de figuras ..... **XV**

Lista de tablas ..... **XVI**

Introducción **1**

1. Planteamiento del problema **3**

2. Justificación **5**

3. Pregunta de investigación **7**

4. Marco teórico **9**
4.1 Factores de riesgo para eventos reportables 9
4.2 Desenlaces en los eventos reportables 11
4.3 Costos directos totales de hospitalización 12
4.4 Técnicas de costeo 13
4.5. Puntuación de propensión o puntaje de predisposición o probabilidad de participación (pscore): 13
4.5.1 Definición: 13
4.5.2 Usos del índice de propensión 16
4.5.3. Procedimiento para la aplicación de la Puntuación de Propensión: 18

5. Objetivos **33**
5.1 Objetivo general 33
5.2 Objetivos específicos 33

6. Metodología **35**
6.1. Diseño 35
6.2. Población a estudio y lugar: 36
6.3. Tamaño muestral: 36
6.4. Identificación de los recursos utilizados: 37
6.5. Técnica de costeo de los recursos utilizados: 38
6.6 Procedimiento 38
6.6.1 Verificación preliminar de los costos antes del emparejamiento. 38
6.6.2 Análisis 39
6.6.3. Análisis estadístico 40
6.6.4. Consideraciones éticas 40

7. Resultados 41
7.1 Población original en Neiva: 41
7.2. Selección de variables 44
7.4. Región de soporte común 45
7.5 Revisión de balance 47
7.6. Estimación del efecto de la exposición y el error estándar 50
7.7. Análisis de sensibilidad 51
7.8. Costos institucionales según los recursos utilizados 52

8. Discusión 55
8.1. Resultados claves 55
8.2 Limitaciones 58
8.3. Interpretación 60
8.4 Generalizabilidad 63

9. Conclusiones 65

Bibliografía 67

Lista de gráficas

Pág.

Grafica 1. Procedimientos para la implementación de la Puntuación de Propensión 19
Grafica 2. Diseño de la cohorte retrospectiva 37
Grafica 3. Curva ROC del modelo predictivo de pscore 45
Grafica 4. Distribución de la Puntuación de Propensión Región de Soporte Común 47
Grafica 5. Puntuación de propensión por estado de exposición zonas de emparejamiento 48
Grafica 6. Proporción de sesgos y su reducción antes y de después del emparejamiento de las covariables 50

Lista de tablas

Pág.

Tabla.1. Análisis preliminar de costos por eventos reportables 42
Tabla.2.Factores de riesgo asociados a eventos reportables 43
Tabla 3. Covariables incluidas para la estimación de la puntuación de propensión 44
Tabla.4. Límites de Región de Soporte Común 46
Tabla 5. Puntuaciones de Propensión en soporte común 47
Tabla 6. Proporción de sesgos y su reducción antes y de después del emparejamiento de las covariables 49
Tabla 7.Diferencia del costo total institucional con algoritmo *kernel* 51
Tabla.8. Análisis de Sensibilidad por método de límites de Rosenbaum 51
Tabla 9. Costos institucionales discriminados por recursos utilizados 53

Introducción

Una de los principios del ejercicio médico se resume en la expresión más antigua “Primum non nocere”, atribuida a Hipócrates y puede tomar varias formas como, "Primero no hacer daño", "Sobre todo no hacer daño", "Ante todo no hacer daño", "Primero que nada no dañar", "Antes que nada no dañar"[2]. El Juramento Hipocrático promulga “Aplicaré mis tratamientos para beneficio de los enfermos, según mi capacidad y buen juicio, y me abstendré de hacerles daño o injusticia”[3]. La 55ª Asamblea Mundial de la Salud en 2002 hizo un llamado a los Estados miembros para que prestarán la mayor atención posible al problema de la seguridad del paciente[4]. Actualmente la Organización Mundial de la Salud (OMS), esta incentivando en el mundo la investigación sobre la seguridad del paciente como un componente de la calidad en la atención en salud, debido a que el cuidado no seguro es una causa importante de morbilidad y mortalidad aumentando los costos de la atención en salud[5]. La OMS define los eventos reportables (ER) como lesiones debidas al cuidado médico[6]; el Ministerio de Salud de Nueva Zelandia los define como una situación que termina en daño no intencional al paciente, que ocurre por causa o con ocasión del servicio y es descubierta después de que éste ingresa a la institución[7]. En Colombia, el Ministerio de Salud y Protección Social está desarrollando en este mismo sentido la Política de Seguridad del Paciente cuyo objetivo es prevenir la ocurrencia de eventos reportables[8]. Gaitán y colaboradores en su estudio de ERs en pacientes hospitalizados, encontraron una incidencia acumulada de 4,6 % (IC 95% 4,1-5,1) durante la hospitalización; prevenible en 61% de los pacientes (189 de 310), con discapacidad permanente en 1,3%, una mortalidad asociada a los ERs de 6,4% (20 de 310) y un aumento de la hospitalización con una mediana de 2 días y 1072 días en total[1].

Estudios diferentes han evaluado los costos económicos de los ERs; para el Sistema de Salud de Nueva York se estimaron mediante simulación en US$161 millones de 1989[9], para el gobierno de Australia representó más de AU$900 millones en 1995[10] y en los estados de Utah y Colorado, más de US$348 millones[11], Brown en Nueva Zelandia

encontró un costo promedio de NZ$10.264 por paciente y para el sistema médico representó NZ$870 millones, de los cuales NZ$590 millones correspondían a ERs prevenibles, además encontró un aumento de la estancia de 9 días y en los eventos prevenibles fue de 10 días. Estos resultados sugieren que más del 30% del presupuesto de los hospitales públicos se gastan en tratar ERs; que los pacientes de mayor edad, los neonatos y los que tienen comorbilidades moderadamente serias tienden a tener ERs más costosos y también que recursos importantes pueden ahorrarse eliminando los ERs prevenibles[12]. En Colombia, Tribiño analizando las reacciones adversas a medicamentos detectó 268 reacciones adversas en 208 de 836 pacientes, y encontró que su atención originó unos costos directos de COL$ 93,633,422 (USD$ 35,014.92) a COL $122,155,406 (USD$ 45,680.94) utilizando la conversión a dólares norteamericanos teniendo en cuenta la tasa representativa del mercado a julio 6 de 2004 (Col $2.674,1 = US$1) y concluyen que debido al impacto negativo de las reacciones adversas en el bienestar de los pacientes, los recursos que se emplean en su atención y la proporción importante de reacciones adversas prevenibles, se hace necesario la implementación de programas operativos de farmacovigilancia en los hospitales[13].

Por lo anterior es importante poder determinar cuál es la diferencia entre los costos directos totales institucionales de hospitalización de los pacientes con y sin ER, en nuestras instituciones de alta complejidad, los factores de riesgo asociados a la ocurrencia de los ERs, y los insumos y servicios atribuidos a los eventos reportables responsables de los costos directos. En una cohorte retrospectiva, mediante un análisis de costo, con el objetivo de determinar esta diferencia, se tomó una muestra aleatoria de los pacientes que participaron en el estudio de Gaitán y colaboradores[1] en el Hospital Universitario de Neiva durante el año 2006 y se formaron 2 grupos; uno que presentó ER y otro que no presentó ER y se emparejaron teniendo en cuenta su puntuación de propensión (probabilidad de participación o índice de propensión o propensity score en inglés) para disminuir el sesgo de selección. Los resultados de este estudio permiten a las directivas del Hospital Universitario de Neiva tomar la decisión de implementar intervenciones para aumentar la calidad de la atención haciéndola más segura y menos costosa.

1. Planteamiento del problema

Gaitán y colaboradores[1] encontraron en Colombia una incidencia acumulada de eventos reportables de 4,6 % (IC 95% 4,1-5,1) durante la hospitalización. Pinzón en 2011 encontró que 24,8 % de los eventos reportados eran prevenibles, con mortalidad de 1,1 % y costos directos totales asociados con su atención entre $ 33´620.346 (US$ 16.687) y $ 37´754.856 (US$ 18.739). Los factores asociados más frecuentemente con el hecho de ser prevenibles fueron las interacciones farmacológicas y las dosis o frecuencias inadecuadas de administración.[14]

El Hospital Universitario de Neiva actualmente está en proceso de acreditación; como parte de este proceso ha implementado una política de seguridad en la atención, sin que hasta el momento se hayan logrado los objetivos trazados. La ocurrencia de eventos reportables en pacientes hospitalizados puede tener diferentes desenlaces relacionados con el paciente como son la mortalidad, la morbilidad, así como los relacionados con la institución como son las estancias prologadas y el aumento de los costos de atención. Al día de hoy no se conocen con precisión cuales son los costos directos derivados de la ocurrencia de ERs que reflejen el costo de la asistencia sin calidad. Este estudio pretende determinar la diferencia que pueda existir en los costos directos de hospitalización entre los pacientes que presentan ERs y los que no presentan este tipo de eventos.

La información obtenida de este estudio, le facilita a las directivas del Hospital Universitario justificar la implementación de intervenciones que busquen disminuir la ocurrencia de los ERs, aumenten la seguridad en la atención de los usuarios, reduzcan los costos de la atención para el Sistema de Salud y ayude a lograr la acreditación de la Institución.

2. Justificación

En Colombia no hay muchos estudios que hayan analizados los costos directos totales institucionales de los ERs ocurridos en hospitales públicos de alta complejidad, que nos permita dimensionar la carga económica de este problema de salud.

En la literatura se han descrito factores de riesgo diferentes para ER relacionados al paciente o a la institución, los cuales se sabe son un predictor importante sobre los resultados en salud incluidos los costos de la atención. En nuestro país no conocemos muy bien la asociación de estos factores y su impacto en las consecuencias derivadas de los ER como son la morbilidad, la discapacidad, la mortalidad y los costos directos.

Por lo tanto el propósito de este trabajo fue determinar si existen diferencias y su su posible magnitud entre los costos directos totales de hospitalización de los pacientes con ER y los costos directos totales de hospitalización de los pacientes sin ER, en nuestra institución pública de alta complejidad, que nos permita hacia el futuro cercano realizar evaluaciones económicas completas como son los estudios de costo-efectividad de intervenciones tendientes a disminuir la ocurrencia de los ERs en nuestros hospitales.
Poder conocer la relación de costo-efectividad de las diferentes intervenciones permite implementar una mejor política de seguridad del paciente en nuestro país con un aumento en la calidad en la atención de los usuarios en las instituciones hospitalarias de alta complejidad

3. Pregunta de investigación

¿Los costos directos totales de hospitalización de los pacientes con ER [Ptes ER(+)] son diferentes a los costos directos totales de hospitalización de los pacientes sin ER [Ptes ER(-)]?

Hipótesis del estudio:

Hipótesis Nula:

Los costos directos totales de hospitalización de los pacientes con ER no son diferentes a los costos directos totales de hospitalización de los pacientes sin ER

Ho: Costos totales de Ptes. ER(+) = Costos totales de Ptes. ER(-)

Hipótesis Alterna:

Los costos directos totales de hospitalización de los pacientes con ER son diferentes a los costos directos totales de hospitalización de los pacientes sin ER

Ha: Costos totales de Ptes. ER(+) ≠ Costos totales de Ptes. ER(-)

4. Marco teórico

4.1 Factores de riesgo para eventos reportables

Los esfuerzos serios para manejar la calidad necesitan estrategias multinivel, de varios componentes coordinados, que manejen el ambiente organizacional y la cultura que afecta como se proporciona el cuidado. Incluye factores externos, e internos como el entrenamiento del personal clínico, procedimientos organizacionales y mecanismos para coordinar la atención y prevenir los errores, la implementación de mejores prácticas, el uso efectivo de tecnologías de información e incentivos apropiados; los externos incluye mecanismos de financiación y reembolso, disposiciones regulatorias que protejan el acceso y los derechos de los pacientes en situaciones de vulnerabilidad y contratos de desempeño. La integración del sistema es difícil, y siempre incompleta, pero el movimiento hacia este objetivo es una estrategia esencial objetiva[15].En estudios diferentes se han encontrado factores de riesgo relacionados al paciente como son la edad, el sexo, la raza, el nivel socioeconómico, el puntaje de comorbilidades, la escolaridad, la severidad de la enfermedad, el tipo de seguridad social, área de procedencia, órgano o sistema comprometido, signos vitales, automedicación; y los relacionados al hospital como son las deficiencias institucionales, el número y tipo de los medicamentos, la continuidad del control, el tipo de atención, el origen de la remisión, los días de estancia y el servicio de hospitalización.

En los Estados Unidos los errores médicos resultan en 44.000 a 98.000 muertes innecesarias cada año y 1 millón de exceso de errores. Estos errores ocurren con frecuencia cuando los médicos son inexpertos y se implementan procedimientos nuevos. Las edades extremas, los casos complejos, las urgencias y una estancia prolongada están asociados con más errores[16]. Andrews et. al. encontraron que la probabilidad de experimentar un ER aumentó un 6% por cada día de estancia en el hospital[17]. Collins estudio los factores de riesgo para estancia prolongada después de cirugía electiva mayor, teniendo en cuenta que la duración de la estancia es un desenlace importante

como indicador de consumo de recursos. Determinar los factores que prolongan la estancia ayuda a disminuir los costos y mejorar la atención en salud. Estancia prolongada se definió como una estancia mayor que o igual al percentil 75 (en días) para cada operación; en este estudio encontró que la estancia prolongada está asociada a factores preoperatorios, intraoperatorios y postoperatorios; las variables preoperatorios asociadas con estancia prolongada fueron el estado funcional, la clasificación de la Sociedad Americana de Anestesiología y la edad; siendo los ERs posoperatorios junto con los procesos intraoperatorios los de mayor riesgo para prolongar la estancia. Una forma de disminuir los costos es minimizando los eventos posoperatorios y mejorando los procesos intraoperatorios de la atención[18]. Kane-Gill investigó la asociación de ERs y los medicamentos de alto costo o alta frecuencia de uso en cuidados intensivos, sin encontrar que la frecuencia, severidad y evitabilidad de los ERs en cuidados intensivos estuviera asociado con el costo o la frecuencia de uso de los medicamentos[19].

Los eventos asociados con la hospitalización como los ERs por medicamentos, caídas, infecciones nosocomiales, úlceras de presión, y complicaciones quirúrgicas y perioperatorias, son más comunes en personas mayores de 65 años y pueden ser más severas y más frecuentemente prevenibles; este riesgo mayor va desde un aumento de 2,2 veces para las complicaciones perioperatorias a un aumento de 10 veces para las caídas. La causa mayor de ese aumento del riesgo parece ser la disminución de la reserva fisiológica de los ancianos; sin embargo la edad sola es un predictor menos importante de ERs que las comorbilidades y el estado funcional. Intervenciones diferentes sobre estos factores han demostrado ser efectivas como por ejemplo los programas de prevención de caídas[20]. Se ha encontrado que barreras del lenguaje y las discapacidades que afectan la comunicación disminuyen la calidad de la atención en salud. Bartlett encontró que pacientes con ERs prevenibles tenían más probablemente un problema de comunicación que aquellos que no tenían tales eventos (odds ratio [OR] 3.00; IC 95% 1.43–6.27)[21]. Wong demostró en pacientes de unidades de cuidado agudo para ancianos una disminución funcional posterior a la hospitalización con pérdida de la independencia, reingresos hospitalarios y muerte. Una prueba basal anormal de marcha, está asociada con estancia prolongada y resultados reportables en el futuro; encontró que el riesgo de un evento adverso se incrementó en un 4% por cada día adicional en el hospital (Hazard Ratio (HR) no ajustado 1,04, IC del 95%: 1,02 a 1,06; p<0,0001)[22]. Los

diferentes factores de riesgo presentes se consideran en este estudio como las variables independientes o explicativas para la ocurrencia de un ER.

4.2 Desenlaces en los eventos reportables

Los ERs pueden presentar clases diferentes de desenlaces, pero los más importantes son los relacionados con la morbilidad, discapacidad, mortalidad y costos. En hospitales británicos el 10.8% de los pacientes experimentaron un ER, con una tasa total de ERs de 11.7% cuando se incluyeron múltiples eventos reportables. La mitad de ellos se consideraron evitables con una atención estándar, la tercera parte llevó a discapacidad moderada o mayor o a la muerte, por eso son una fuente importante de daño al paciente y un consumo grande de recursos para el Sistema Nacional de Salud[23]. Wolff confirmó en un hospital rural en el servicio de urgencias ERs en 250 pacientes (1.24% de los 20050 consultantes en un periodo de dos años); 81 (32.4%) se consideraron severos y 169 (67.6%) de severidad menor. En un período de dos años el número de ERs cayó de 84 (3.26% de todos los consultantes) a 12 (0.48% de todos los consultantes) (reducción del riesgo relativo 85.3% (IC 95%, 62.7% a 100%)) después de implementar actividades de mejoría de la calidad[24]. Cho encontró que el aumento de 1 hora de atención por enfermeras certificadas por paciente día se asoció con una disminución de 8.9% en la probabilidad de neumonía, así como un aumento del 10% en la proporción de enfermeras certificadas se asoció con una disminución del 9.5% en la probabilidad de neumonía. La ocurrencia de cada ER se asoció con una estancia prolongada significativa y el aumento de los costos médicos. Los pacientes que tenían neumonía, infección de heridas o sepsis tenían una probabilidad mayor de muerte durante la hospitalización. Concluye que se necesitan sistemas de atención que incluyan un personal de enfermería apropiado que disminuyan los ERs y sus consecuencias[25]. Zhan encontró un exceso de estancia, costo y mortalidad atribuibles a lesiones médicas como lesiones neonatales, sepsis postoperatoria, dehiscencia de suturas. A pesar de que estas lesiones tienen un impacto altamente variable, significan una amenaza significante para los pacientes y costo para la sociedad[26]. Encontramos que en el caso de la ocurrencia de ER los días de estancia están asociados tanto con la exposición – probabilidad de ocurrir un ER – como con el desenlace costos, comportándose entonces como un factor de confusión el cual debe ser controlado y en este caso se puede utilizar el emparejamiento.

En este estudio se tuvieron en cuenta los ERs - una variable binaria - como un desenlace intermedio para discriminar los grupos a comparar y su consecuencia los costos directos institucionales de la hospitalización – una variable continua - como desenlace final o variable dependiente. Este tipo de variable cuando se trata de costo usualmente no tienen una distribución normal cuando el tamaño de muestra es mediano o pequeño, por lo tanto los métodos como *t* de Student y regresión lineal basados en una distribución normal de los promedios no están indicados en estos casos[27].

4.3 Costos directos totales de hospitalización

Rigby encontró en Australia que el costo de doce tipo de lesiones iatrogénicas prevenibles es significante (US$0.25 millones) y explica el 2 a 3% del presupuesto anual de un hospital australiano típico de 120 camas[28]. Ehsani encontró que los ERs están asociados con unos costos significantes al estimar los costos de los ERs intrahospitalarios utilizando modelamiento con regresión lineal y ajustándolos para edad y comorbilidad; los desenlaces principales fueron costo de cada paciente, estancia y mortalidad. Los pacientes con ERs tenían 10 días más de estancia y un aumento de la mortalidad intrahospitalaria 7 veces más que aquellos sin complicaciones. El costo promedio sin ER fue de $2.181 (IC95% $2.171– $2.190) y el costo promedio con ER fue de $14.027 (IC95% $13.865–$14.187). Después de ajustar para edad y comorbilidad la presencia de un ER sumo $6826 al costo de cada episodio admitido. El costo total de los ERs fue de $460.311 millones, representando 15.7% del gasto total sobre los costos directos del hospital, o un 18.6% adicional del presupuesto total del hospital[29].

Cuando se pagan reclamaciones que no incluyen errores, los pagos son significativamente más bajos en promedio que las reclamaciones que incluyen errores ($313,205 vs. $521,560, *p*=0.004). Las reclamaciones con errores explican el 78% de los costos administrativos totales. Los costos fijos de litigios de mala práctica son exorbitantes[30]. Marchetti estudió los costos y beneficios de un sistema de vigilancia de alerta temprana para pacientes hospitalizados; encontró que los costos directos esperados por paciente, para caídas y paros cardiopulmonares fueron de $191.73 por hospitalización y los beneficios económicos estimados asociados con el sistema de vigilancia documentan un costo de $14.95 por día para el sistema[31]. En un análisis sobre

el costo de los ERs, Mello encontró relacionados a lesiones un costo promedio de US$2.013 y relacionados a lesiones negligentes costos de US$1.246 por paciente dado de alta. Sin embargo el hospital pagó solamente el 22% de esos costos[32]. Pappas concluye que los sistemas de costos en los hospitales son útiles para determinar el costo de los ERs y pueden ayudar a determinar la cantidad de personal de enfermería necesario para evitar costos por ERs[33].

4.4 Técnicas de costeo

La técnica de microcosteo es el enfoque más valido para la estimación de los costos en pacientes hospitalizados, pero es costoso y puede limitar la muestra de pacientes o instituciones. La estimación de los costos basado en protocolos varia ampliamente en su validez dependiendo de la fuente de los datos. Cuando la precisión y la resolución son objetivos importantes, los enfoques de costeo clínico proporcionan el estimado de costos intrahospitalarios más validos a unos costos de datos razonables. Cuando la validez externa es importante o donde se desea estandarizar los costos hospitalarios, el uso de costos ponderados nacionales publicados puede ser preferido[34]. El método de costo por día fue perfeccionado por Jacobs y Bachynsky[34, 35] con el enfoque de 'Hotel y Auxiliares', donde los costos incurridos en la sala son asignados por día, y otros servicios específicos (denominados 'auxiliares') son contados a nivel del paciente. Se asigna un costo unitario por auxiliar, y el producto de los costos auxiliares de un paciente se suman a sus gastos de hotel por día para producir un costo total a nivel del paciente. Esto mejora la precisión del cálculo de costos por día, pero necesita una recolección considerable de datos primarios. Este análisis de costo de acuerdo a la perspectiva del hospital utiliza un método de costo por día perfeccionado con el enfoque de “Hotel y auxiliares” usando los precios de compra.

4.5. Puntuación de propensión o puntaje de predisposición o probabilidad de participación (pscore):

4.5.1 Definición:

La puntuación de propensión es la probabilidad condicional de exposición a un tratamiento dadas las covariables observadas. En un estudio de cohortes, los sujetos tratados y controles emparejados o estratificados en una variable única – la puntuación

de propensión- tiende a balancear todas las otras covariables observadas, sin embargo, a diferencia de la asignación aleatoria de los tratamientos en los ensayos clínicos, la puntuación de propensión no puede balancear las covariables no observadas[36].

Cuando los pacientes tratados son comparados a los controles, los resultados diferentes pueden reflejar los efectos causados por el tratamiento o las diferencias en el pronóstico antes del tratamiento. Cuando no se usa la asignación aleatoria – esto es, en un estudio observacional- los grupos tratados y de control pueden diferir en el pronóstico, y los resultados diferentes pueden no ser efecto del tratamiento. Los ajustes analíticos para sesgos evidentes son de dos tipos. 1) los que se fijan en la relación entre covariables pronosticas o independientes y los resultados y 2) los que se fijan en la relación entre las covariables pronosticas y la asignación de los pacientes a tratamiento o control. La puntuación de propensión es la posibilidad de recibir el tratamiento más que el control para un paciente con ciertas covariables pronosticas observadas. Un ajuste o corrección usando puntuaciones de propensión intenta anular el problema creado por las posibilidades diferentes de recibir el tratamiento. Si el resultado es raro pero el tratamiento es común, puede haber pocos datos con los cuales estimar la relación entre el resultado y las covariables pronosticas, pero bastantes datos con los cuales estimar la relación entre la asignación del tratamiento y las covariables pronosticas, esto es, para estimar la puntuación de propensión, pueden ser utilizados de otras maneras. Una aplicación forma parejas coincidentes de un paciente tratado y un control con puntuaciones de propensión similares. Otra usa un modelo de regresión logística para predecir el tratamiento. Las puntuaciones de propensión eliminan sesgos evidentes pero hace poco o nada para manejar sesgos ocultos debido a diferencias no observadas o registradas entre los pacientes tratados o controles antes del tratamiento. El problema de los sesgos ocultos no pueden ser eliminados de un estudio observacional aunque se puede reducir su magnitud[37].

La puntuación de propensión para el individuo *i* se define como:

$$\Pr(T_i=1|\ X_i = x_i),$$

donde T_i=1 si el individuo recibe el tratamiento y T_i=0 si recibe el control. Se asume, que condicionado a un grupo de variables explicatorias X_i, el T_i es independiente. Esta

probabilidad puede estimarse fácilmente usando una regresión *logit* o una *probit*, siendo la *logit* la más utilizada. Una vez estimada se necesita predecir la propensión individual de ser asignado al control o al tratamiento. Esto facilita la creación de un asignación casi-aleatorizada, en que dos individuos – uno que recibe el tratamiento y otro el control - teniendo puntuaciones de propensión muy similares, pueden ser vistos como si hubieran sido aleatorizados. La metodología de la puntuación de propensión transmite toda la información contenida en el grupo de covariables incluidas en el modelo de predicción en una variable única, la cual puede entonces ser usada como una variable en un modelo de regresión, crear parejas de individuos similares con la misma puntuación de propensión y por lo tanto la misma distribución de covariables, y crear estratos de igual tamaño – usualmente definidos por los quintiles de la distribución de la puntuación de propensión - dentro de los cuales la estimación del efecto promedio del tratamiento (ATT por sus siglas en ingles, *average treatment on the treated*) puede realizarse.[38]

El cálculo de la puntuación de propensión es un método para controlar las diferencias entre los grupos de expuestos y no expuestos. Tiene limitaciones cuando el emparejamiento es incompleto o hay datos insuficientes en la puntuación de propensión calculada[39].

La puntuación de propensión para un individuo es la probabilidad de ser tratado en forma condicional según (o únicamente sobre la base de) los valores de las covariables del individuo. La puntuación de propensión puede ser usada si existe independencia condicional entre la asignación del tratamiento y los resultados potenciales (conocido como la asignación del tratamiento fuertemente ignorable). En la práctica, el éxito de las puntuaciones de propensión después de su uso se juzga si se alcanza el balance entre los grupos de tratamiento según los valores de las covariables. Una ventaja de las puntuaciones de propensión es que si 2 sujetos son encontrados, un sujeto en el grupo tratado y un sujeto en el control, con la misma puntuación de propensión, entonces uno podría imaginar que esos 2 individuos fueron asignados “aleatoriamente” a cada grupo en el sentido de tener la misma probabilidad de ser tratado o ser control. Al construir el modelo de la puntuación de propensión, únicamente las covariables presentes antes del tratamiento deben ser incluidas.

$$p(X) = Pr\ (D=1|\ X_1, X_2..., Xn) = E(D|X)$$

Donde “D” es el tratamiento o exposición analizada (0 o 1), y X_1…Xn, las covariables que predicen su utilización[40].

4.5.2 Usos del índice de propensión:

Las tres técnicas más comunes que usan la puntuación de propensión son el emparejamiento, la estratificación y el ajuste con regresión[41].

El emparejamiento es una técnica común utilizada para seleccionar individuos control que son emparejados con individuos expuestos sobre la base de covariables que el investigador considera necesitan ser controladas. El emparejamiento con puntuación de propensión resuelve este problema al permitir al investigador controlar al inicio muchas covariables simultáneamente por emparejamiento de una sola variable, la puntuación de propensión. Para evaluar el éxito del emparejamiento con puntuación de propensión, una técnica común es comparar las variables en el grupo de individuos expuestos y de control, antes y después del emparejamiento. En muchos escenarios, el emparejamiento con puntuación de propensión puede también ser muy costo-efectivo. En el proceso de emparejamiento, Hall agrupó los sujetos 1 a 1 con puntuaciones de propensión ± 0.01[42]. Cuando realizamos procedimientos de emparejamiento están disponibles un número grande de técnicas posibles, sin embargo todavía no existen guías coherentes para seleccionar la aplicación más apropiada. El propósito principal de un procedimiento de emparejamiento es disminuir los sesgos de selección aumentando el equilibrio entre el grupo de tratamiento y el de control. Baser para revisar el equilibrio propone los siguientes enfoques, consistentes de cinco pasos cuantificables:1) Usar la prueba-t para dos muestras para comparar los promedios de los grupos del tratamiento y de control para cada variable explicatoria; 2) Comparar la diferencia de medias como un porcentaje del promedio de las desviaciones estándar; 3) Comparar el porcentaje de disminución de los sesgos en los promedios de las variables explicadoras antes y después del emparejamiento; 4) Comparar las densidades estimadas para tratamiento y control para las variables explicadoras; y 5) Comparar la densidad estimada de las puntuaciones de propensión de los individuos control con las de los individuos tratados. Investigó siete técnicas diferentes de emparejamiento y su desempeño respecto a los 5 pasos propuestos y comparó los efectos promedio del tratamiento estimados con análisis multivariado con los estimados con el emparejamiento con las puntuaciones de propensión. Se hizo análisis de emparejamiento del vecino más cercano de siete

maneras: reemplazo, emparejamiento 2 a 1, emparejamiento de Mahalanobis, emparejamiento de Mahalanobis con pinza, emparejamiento kernel, emparejamiento radial y el método de estratificación. Encontró que la elección del emparejamiento tenía efectos significantes sobre los resultados. El emparejamiento de Mahalanobis con pinza dio los mejores resultados de acuerdo a todos los cinco criterios[43].

Estratificación: La estratificación se usa comúnmente también en estudios observacionales para controlar diferencias sistemáticas entre los grupos control y expuestos[41].

Ajuste con regresión (Covarianza): La puntuación de propensión también se puede utilizar en ajuste con regresión – covarianza - . En el ajuste con regresión, el efecto del tratamiento es estimado ajustando para el impacto de las covariables iniciales en un modelo de regresión. Otro enfoque para ajustar con regresión es usar un conjunto grande de covariables iniciales, para estimar la puntuación de propensión y entonces usar un subgrupo de esas covariables y la puntuación de propensión en el ajuste con regresión. Un interrogante que puede surgir cuando se usa el ajuste con regresión con puntuación de propensión es si cualquier ganancia resulta del uso de la puntuación de propensión más que del desempeño de un ajuste con regresión con todas las covariables usadas para estimar la puntuación de propensión utilizada en el modelo. Uno puede combinar las dos primeras técnicas, estratificación y ajuste con regresión, estratificando primero los datos según la puntuación de propensión y luego ajustando con regresión con un subgrupo de covariables importantes dentro de cada estrato. La puntuación de propensión no es la única herramienta que puede ser utilizada para el análisis de datos de estudios observacionales; más bien, debe ser vista como una herramienta adicional disponible para los investigadores que tratan de estimar los efectos de los tratamientos en estudios donde pueden existir sesgos potenciales[41].

El uso de puntuaciones de propensión para ajustar factores de confusión medidos ha aumentado en los estudios de cohorte. Sin embargo, su uso en los estudios de caso-control y caso-cohorte ha tenido poca atención. La aplicación de puntuaciones de propensión es menos sencilla en estudios de caso-control y caso-cohorte que en estudios de cohorte. Mänsson reveló que cuando se usan varios enfoques potenciales, hay una modificación artificial del efecto de los odds ratios por nivel de puntuación de

propensión. La magnitud de este fenómeno disminuye si el tamaño de la muestra aumenta. Además, varios enfoques potenciales producen puntuaciones de propensión estimadas que no convergen al valor real si el tamaño de la muestra aumenta, y por lo tanto, puede fallar para ajustar completamente los factores de confusión medidos. Sin embargo, la magnitud de la confusión residual fue modesta en las simulaciones. Los investigadores que estén considerando usar las puntuaciones de propensión en estudios de caso-control y caso-cohorte deben considerar la posibilidad de modificación artificial de los efectos y su capacidad reducida de controlar posibles factores de confusión[44].

Manca analizó tres formas en las cuales la puntuación de propensión puede ser utilizada para controlar la confusión en la estimación de la costo-efectividad promedio. La puntuación de propensión para un individuo es la probabilidad de ser asignado al tratamiento o al control, según el valor de un grupo de covariables observadas. Esto ha sido mostrado bajo el supuesto que no hay factores no observados que podrían originar sesgos de selección (supuesto ignorable de asignación de tratamiento); condicionado en la puntuación de propensión, asegura que la asignación al control o al tratamiento para un individuo dado es independiente del resultado del tratamiento. En otras palabras, condicionado en la puntuación de propensión, permite la estimación no sesgada de efectos promedios de tratamiento [38].

4.5.3. Procedimiento para la aplicación de la Puntuación de Propensión:

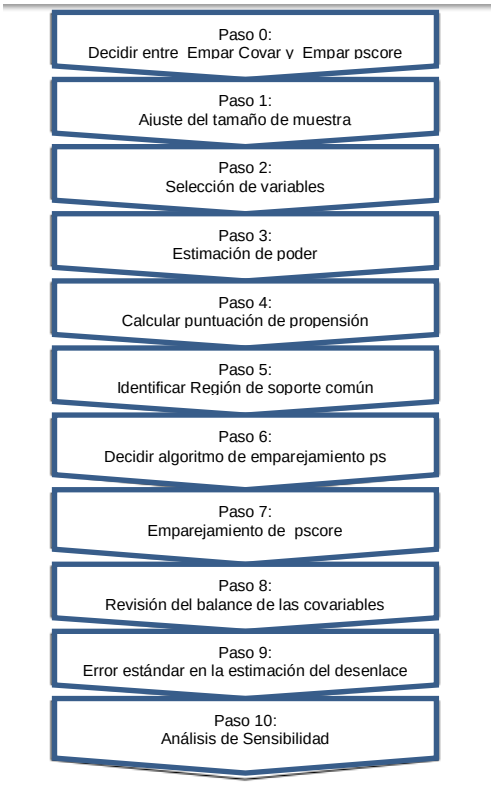
Yang[45] y colaboradores hicieron una revisión de la aplicación de la puntuación de propensión en epidemiología y en desenlaces en salud. En esta revisión describen el procedimiento por etapas para la aplicación, las cuales presentamos a continuación y se resumen en la gráfica 1.

4.5.3.1 Elección entre emparejamiento de covariables y puntuación de propensión:

El objetivo de muchos análisis de bases de datos grandes es obtener inferencias causales acerca de los efectos de acciones, tratamientos o intervenciones. Ejemplos incluyen los efectos de varias opciones disponibles para el médico para tratar un paciente particular, las eficacias relativas de varios proveedores de servicios de salud, y las

consecuencias de implementar una política nacional nueva de salud. Una complicación de usar bases de datos grandes para alcanzar tales objetivos es que sus datos son casi siempre observacionales más que experimentales. Esto es, los datos en la mayoría de las bases grandes no se basan en los resultados de ensayos clínicos aleatorizados conducidos cuidadosamente, sino que representan datos recolectados a través de la observación de sistemas como ellos funcionan en la práctica normal, sin intervenciones implementadas por normas de asignación aleatorizada. Tales datos son relativamente baratos de obtener, sin embargo, frecuentemente representan el espectro de la práctica médica, más que los escenarios de experimentos aleatorizados.

Gráfica 1: Procedimientos para la implementación de la Puntuación de Propensión



Adaptado de: Yang, G., *A review of propensity score application in healthcare outcome and epidemiology* in *PharmaSUG 2007*

Consecuentemente, es razonable ensayar estimar los efectos de tratamientos obtenidos de bases de datos grandes, aun si únicamente ayudan a diseñar un nuevo experimento aleatorizado o dar luz sobre la generalizabilidad de los resultados de experimentos

aleatorizados existentes. Sin embargo los métodos estándar de análisis (como la regresión linear o logística) usando los software estadísticos disponibles pueden ser engañosos para esos objetivos porque ellos no dan advertencias acerca de sus propiedades. Los métodos de puntuación de propensión son herramientas más confiables para abordar tales objetivos porque los supuestos – independencia condicional y soporte común - necesarios para hacer sus respuestas apropiadas son más evaluables y transparentes para el investigador[46].

Cepeda comparó la regresión logística con la puntuación de propensión en términos de sesgos, precisión, probabilidad de cobertura empírica, poder empírico, y robustez cuando el número de eventos es bajo en relación al número de variables de confusión. En la regresión logística, los sesgos disminuyeron en la medida que el número de eventos por variable de confusión aumentó. En la puntuación de propensión, los sesgos disminuyeron en la medida que la fuerza de la asociación de la exposición con el resultado aumentó. La puntuación de propensión produjo estimados que fueron menos sesgados, más robustos y más precisos que los estimados con regresión logística cuando hubo siete o menos eventos por variable de confusión. La probabilidad de cobertura empírica de la regresión logística aumentó en la medida que el número de eventos por variable de confusión aumentó y en la puntuación de propensión después de 8 o más eventos por variable de confusión. En total, la puntuación de propensión muestra más poder empírico que la regresión logística. Las puntuaciones de propensión son una buena alternativa para controlar imbalances cuando hay siete o menos eventos por variable de confusión, sin embargo, el poder empírico podría variar de 35% a 60%. La regresión logística es la técnica de elección cuando hay por lo menos ocho eventos por variable de confusión[47].
El análisis de la puntuación de propensión intenta controlar las covariables de confusión en los estudios no experimentales ajustando la probabilidad que un paciente dado sea expuesto. Tales análisis han sido propuestos para enfrentar la confusión por indicación, pero hay poca evidencia empírica que ellos alcancen mejor control que el modelamiento multivariado convencional de resultados. Stürmer encontró en únicamente 9 de 69 estudios que tenían un efecto estimado que era diferente en más del 20% de los obtenidos con un modelamiento convencional de resultados en todos los análisis de puntuaciones de propensión presentados.[48]

El uso de puntuaciones de propensión para identificar y controlar las covariables de confusión en estudios observacionales que relacionan medicamentos con los resultados se ha aumentado sustancialmente en años recientes. Sin embargo, no queda claro si, y si es así cuando, el uso de las puntuaciones de propensión proporciona estimados de efectos de medicamentos menos sesgados que los obtenidos de modelos multivariados convencionales. En la gran mayoría de los estudios publicados que han usado ambos enfoques, los efectos estimados de la puntuación de propensión y del método de regresión han sido similares. Estudios de simulación también sugieren un desempeño comparable de los dos enfoques en muchos escenarios. El uso de puntuaciones de propensión no corrige los sesgos de las covariables de confusión no medidas, pero puede ayudar a entender los determinantes de uso de medicamentos y llevar a mejorar los estimados de los efectos de los medicamentos en algunos escenarios[49].

Dubberke determinó los costos atribuibles a una enfermedad utilizando regresión linear multivariable y análisis de emparejamiento con puntuaciones de propensión y encontró que ambos métodos son conservadores y pueden subestimar los costos atribuibles reales[50].

4.5.3.2 Ajuste del tamaño de muestra

Es cierto que la puntuación de propensión funciona mejor en muestras grandes. En un estudio con muestra pequeña, un desequilibrio importante de algunas covariables puede ser inevitable a pesar del uso apropiado del algoritmo de emparejamiento y la estimación de la puntuación de propensión. Zhao[51] descubrió que el emparejamiento con puntuación de propensión no es superior al emparejamiento de covariables para muestras pequeñas (n<500), pero funciona mejor en muestras grandes (n>1000)[45].

Un concepto común es que en un estudio se deben preservar en el grupo de expuestos tantos individuos como sea posible; en caso contrario el sesgo de selección ocurre debido a un emparejamiento menos exacto con los individuos control. En la práctica, dado un número suficiente estadísticamente de individuos expuestos, la supresión de algunos valores atípicos en el grupo expuesto es aceptable con poco o ningún efecto sobre la validez del emparejamiento. Así también, para aumentar la posibilidad de encontrar una mejor pareja, los investigadores tienden a mantener tan grande como sea

posible el grupo control. Inevitablemente, los datos malos o eventos raros podrían dar una subestimación de la puntuación de propensión. El criterio establecido para la determinación del tamaño de muestra ha sido recomendado en la literatura, el porcentaje total de individuos expuestos no debe ser menor del 10% [52].

4.5.3.3 Selección de variables

A veces, los investigadores están acostumbrados a incluir numerosos factores en la evaluación del modelo. No siempre es apropiado hacer esto. Incluir más covariables disminuye significativamente el tamaño de muestra debido a las condiciones más restringidas en la estimación de la puntuación de propensión. Como resultado los modelos se vuelven más sensibles. Los investigadores deberían preguntarse qué covariables pueden contener la información importante para el interés de la investigación, este es el primer punto para iniciar la selección de las covariables. Los factores de riesgo relacionados estrechamente con el desenlace y la elección de la exposición deben incluirse. Adicionar esas covariables al modelo usualmente disminuye los sesgos más que aumentar la varianza durante el emparejamiento. Estos factores forman el conjunto de información mínimo relevante. Se ha demostrado que la interacción inadecuada entre factores puede introducir sesgos en la estimación de la puntuación de propensión[45].

4.5.3.4 Estimación de poder

En los casos en donde la estimación de la regresión es adecuada, el conocimiento para tener en cuenta es que, el modelo de regresión pobremente ajustado no crea una estimación exacta del índice de propensión y un buen equilibrio entre los grupos expuesto y de control. Esto lleva a una estimación sesgada del desenlace[45].

Un poder predictivo C-valor(el área bajo la curva operador del receptor [ROC por sus siglas en inglés]) mayor de 0.8 o una prueba Hosmer-Lemeshow insignificante son indicativos de buena clasificación en un modo de puntuación de propensión[53].

4.5.3.5 Región de soporte común

Otro requisito de un uso adecuado de la estimación de la puntuación de propensión es una región de soporte común adecuada entre los grupos de exposición y control. Eso

significa que la mayoría de los individuos caen en un ámbito de probabilidad predictiva común a ambos grupos, de exposición y de control. Los individuos que quedan fuera de la región común tienen poca oportunidad de encontrar un control correspondiente durante el proceso de emparejamiento. Cualquier combinación de características en el grupo de exposición puede también ser observada en el grupo control, es decir la existencia de parejas potenciales en control. El método más sencillo para determinar la región común es el análisis visual de la distribución de la densidad de la puntuación de propensión en ambos grupos. La superposición contiene la mayoría de los individuos de la región de soporte común. El concepto básico de Regla de Mínimos-Máxima es comparar y eliminar todos los individuos cuyas puntuaciones de propensión son menores que el máximo de los mínimos y mayores que el mínimo de los máximos en los dos grupos. Todos los individuos dentro de esta región son seleccionados para la estimación. Esta prueba a dos colas se aplica a los individuos expuestos y de control. Por ejemplo, un rango de puntuación de propensión en el grupo expuesto es [0.3, 0.9] y en el grupo control es [0.2, 0.7]. La región de soporte común entre esos dos grupos es definido como [0.3, el máximo de los mínimos y 0.7 el mínimo de los máximos]. La región común parece funcionar con más poder en la implementación del emparejamiento kernel comparado con el método del vecino más cercano, desde un punto de vista estadístico. En el emparejamiento kernel, todos los individuos control son usados para imputar desenlaces ausentes, mientras que únicamente el más cercano es elegido en el emparejamiento del vecino más cercano. La región de soporte común también puede especificarse por desviaciones estándar de la media, los individuos que caen más allá de tres veces de la desviación estándar, por ejemplo, del puntaje promedio en las colas izquierda y derecha son eliminados antes de la estimación de la puntuación de propensión[54]. Con la región de soporte común surgen ciertos problemas cuando los individuos descartados están muy cerca de los extremos, o cuando hay superposición muy estrecha dentro de la región amplia del soporte común, o cuando la densidad en las colas de la distribución es muy delgada. Una es una muestra con una cola extremadamente sesgada a la derecha para el grupo de expuestos y una cola extremadamente sesgada a la izquierda para el grupo control. El proceso de recorte elimina las estimaciones en la región vacía de los individuos expuestos, lo cual aparentemente da resultados más confiables. Lechner[55] señaló, que ignorando los problemas de la región de soporte común o haciendo una evaluación solamente sobre los individuos dentro de la región de soporte común pueden resultar en estimaciones erróneas. El procedimiento de recorte se propone para

compensar los problemas de la región de soporte común después del análisis visual generalmente recomendado. Después de seleccionar de la región común y de recorte, únicamente los individuos restantes son incluidos en la estimación del desenlace. Si los individuos descartados ocupan únicamente una parte pequeña de la población completa, no hay problema y la eficiencia de la estimación también aumenta. De lo contrario, los resultados obtenidos son cuestionables para los investigadores quienes pueden preguntarse si la muestra es realmente representativa de la población general. Lo más probable, es que los individuos perdidos sean devueltos para ayudar a la investigación y a la interpretación[45].

4.5.3.6 Métodos de emparejamiento de la puntuación de propensión

Existen numerosos factores para considerar cuando se utiliza el emparejamiento con puntuación de propensión; el proceso de emparejamiento se vuelve más complicado por el número de métodos de emparejamiento disponibles. A pesar de su uso frecuente en estudios observacionales, no existe actualmente una matriz de decisión basada en reglas, coherente en el campo clínico o epidemiológico. La posibilidad de aplicar mal esas técnicas es alta y contribuye a la controversia en cuanto al valor de la metodología en sí[45].

A continuación se exponen los diferentes métodos de emparejamiento.

- Emparejamiento uno a uno

Indica el emparejamiento de individuos expuestos y de control en parejas, también llamado emparejamiento por pares. El emparejamiento del vecino más cercano es usado para seleccionar la comparación de sujetos cuya puntuación de propensión está más cerca al sujeto expuesto en cuestión. El emparejamiento es hecho para un individuo expuesto dado sin considerar los individuos control que permanecen. La distancia más cercana entre pares no es necesariamente la más cercana entre ese individuo y todos los individuos control en general. Al contrario, los algoritmos de emparejamiento óptimo, optimizan los objetivos globales más que los locales[45].

- Emparejamiento uno a muchos

Una extensión del emparejamiento por pares es el emparejamiento uno a muchos, el propósito del cual es evitar inflar la varianza o error estándar aumentando el tamaño de muestra en el grupo control. El emparejamiento uno a muchos es el segundo estilo más popular en los métodos de emparejamiento de la puntuación de propensión. A diferencia del emparejamiento uno a uno, el emparejamiento uno a muchos une un individuo en el grupo de expuestos a uno o varios individuos en el grupo control, el número es predefinido por los usuarios, el emparejamiento uno a uno es un caso especial del método uno a muchos[45].

El emparejamiento Kernel y local linear fue desarrollado a partir de los métodos de regresión no paramétricos para la suavización de las curvas. Todos los individuos expuestos son comparados con un promedio ponderado de todos los individuos control. Debido a que todos los individuos control contribuyen a los pesos, la varianza disminuye. En este método es necesario determinar la amplitud de la banda.

Otra variante del emparejamiento kernel es el emparejamiento estratificado. El emparejamiento estratificado también empareja casos expuestos con un promedio ponderado de casos control excepto que la distancia de emparejamiento es definida por el usuario más que por la amplitud de la banda del kernel o núcleo. El emparejamiento se realiza dentro de los estratos respectivamente.

El emparejamiento por radio utiliza solamente los individuos disponibles dentro de un radio predefinido. Un emparejamiento inexacto es inevitable. La ventaja es que cuando no hay buenas parejas disponibles, individuos extras o menores se pueden utilizar. Un radio razonable no es fácil de determinar.

El emparejamiento distancia de Mahalanobis emplea el mismo algoritmo básico antes de que la puntuación de propensión fuera desarrollada y coloca la puntuación de propensión en la medición de distancia.

- Emparejamiento con diferentes razones

La debilidad del emparejamiento de razón fija discutido anteriormente es que usando más controles lleva a mayor sesgo. En general si los individuos control son k veces los

individuos expuestos; el ajuste usando un emparejamiento 1-a-k cantidades no ajusta en absoluto. Cualesquiera que sean sus ventajas para la varianza, la tendencia para utilizar controles fijos de k mayores acompaña una penalización fuerte en términos de sesgos. En la práctica, la razón de individuos expuestos emparejados a individuos control puede variar; el emparejamiento libre con cualquier buen control es obviamente la mejor elección si estos están disponibles[56].

- Emparejamiento muchos a muchos

Este emparejamiento es también llamado el emparejamiento completo. Para algunos conjuntos emparejados, un individuo expuesto puede ser emparejado a dos o más controles. Del mismo modo, otros conjuntos emparejados pueden contener individuos expuestos emparejados con un control. Esto distingue el emparejamiento muchos a muchos del emparejamiento uno a uno, uno a muchos y emparejamiento con un numero diferente de individuos control. El emparejamiento completo a veces coincide con parejas producidas por esos métodos más simples. Entre todos los métodos de emparejamiento para dos grupos, el emparejamiento completo por si solo ha mostrado ser óptimo para producir la mayor similitud dentro de los grupos emparejados[56].

4.5.3.7 Revisión del balance de las covariables

Para la selección de las variables es importante comparar la influencia de las características basales entre los grupos expuesto y control antes del emparejamiento. Es también importante comparar los grupos expuesto y de control después del emparejamiento para comprobar el equilibrio; la literatura sugiere varios métodos para comprobar la existencia de equilibrio para ver si las observaciones con la misma puntuación de propensión, tienen la misma distribución de las covariables observables independiente de la asignación de la exposición. En otras palabras, hay necesidad de probar si el método de puntuación de propensión es una alternativa confiable a los ensayos clínicos aleatorizados en términos de los sesgos introducidos por usar datos no experimentales. Un resultado complejo, sin embargo, es que diferentes métodos pueden generar respuestas inconsistentes[45].

Después de ajustar por la puntuación de propensión, ajustes adicionales no deben dar información nueva. Términos de interacción, de orden superior o diferentes covariables

pueden ser necesarios para mejorar el análisis porque el puntaje estimado actual puede no ser un resumen equilibrado adecuado de todas las covariables. Por supuesto, también es posible que ningún ajuste equilibre la muestra emparejada, donde se concluye que el emparejamiento de la puntuación de propensión no ha resuelto el problema de la selección sesgada.

Baser[52] propuso un conjunto de guías para este propósito.

- Investigar el estadístico-t para todas las variables numéricas y la prueba Chi-cuadrado para todas las variables categóricas entre los grupos expuesto y de control. Valores *p* no significativos indica un emparejamiento adecuado;

- Calcular la diferencia promedio como un porcentaje de la desviación estándar promedio,

$$\frac{100(X_T - X_C)}{\frac{1}{2}(S_{XT} - S_{XC})}$$

donde X es el conjunto de covariables, T y C son los grupos de expuestos y de control, S es la desviación estándar de esas covariables. Valores medios bajos sugieren equilibrio. El valor más bajo sugiere el beneficio potencial del emparejamiento;

- Calcular un porcentaje de reducción del sesgo,

$$\frac{(X_{AT} - X_{AC}) - (\mathrm{X}_{IT} - X_{IC})}{\mathrm{X}_{IT} - X_{IC}} * 100$$

donde X es la media de las covariables, A es después del emparejamiento, I es antes del emparejamiento, T y C son los grupos expuestos y de control. Como se espera, se desea la mayor reducción;

- Utilizar la prueba Kalmogorov-Smirnov para calcular los estimados de densidad para las covariables explicatorias y las puntuaciones de propensión entre los grupos expuestos y de control. Las diferencias no significativas entre los dos grupos expresan un emparejamiento efectivo;

El método de Diferencias Estandarizadas fue respaldado por Love[53] para medir el balance de las covariables antes y después del emparejamiento para las covariables continuas y categóricas. Si el valor absoluto de la diferencia es mayor del 10% o el valor *p* menos que el nivel crítico, representa un índice de un desequilibrio importante.

$$d = \frac{100(\bar{x}_{Treatment} - \bar{x}_{Control})}{\sqrt{\frac{S^2_{Treatment} + S^2_{Control}}{2}}}$$

para variables continuas,

$$d = \frac{100(P_{Treatment} - P_{Control})}{\sqrt{\frac{P_{Treatment}(1 - P_{Treatment}) + P_{Controt}(1 - P_{Control})}{2}}}$$

para variables categóricas.

Si esas pruebas convencionales no son satisfactorias, la versión de permutación de esas pruebas como una técnica estadística intensiva computarizada de bootstraping puede ser una mejor aplicación respecto a un tamaño de muestra pobre.

Por otra parte, cuando no se estima la puntuación de propensión correctamente el poder de simulación permite determinar la utilidad de las pruebas de balanceo, como pruebas con buen tamaño pero bajo poder pueden ser de un uso limitado.

Un hecho interesante es que si el emparejamiento funciona bien, los individuos en los grupos expuesto y de control deben tener muy poco en común. Esto da un indicio que el mejor puntaje predictor no es necesariamente una puntuación de propensión eficiente si no sirve como un resumen adecuado de esas covariables. Además, es importante tener

en cuenta que un balance exitoso para la muestra completa no implica un balance para únicamente la muestra emparejada. La puntuación de propensión es sólo una medida relativa la cual varía dependiendo de la composición del grupo control, y nunca es una etiqueta de identificación permanente para una observación individual.

4.5.3.8 Errores estándar en la estimación del desenlace

La estimación del desenlace en muestras emparejadas merece más revisión en que la varianza de los desenlaces estimados no es únicamente del muestreo normal, sino que también es contribuida por la estimación de la puntuación de propensión, el proceso de emparejamiento en sí, y el orden de los individuos durante el emparejamiento sin reemplazo. Además, la clasificación de las covariables seleccionadas para emparejar también es capaz de adicionar varianza en la evaluación del desenlace[45].

Lechner propuso un método bootstrapping para estimar los errores estándar en los casos donde los estimados son sesgados o no están disponibles. N veces de bootstrapping genera N desenlaces promedio estimados repitiendo todos los pasos del emparejamiento de la puntuación de propensión. La distribución resultante del desenlace se aproxima a los errores estándar del muestreo. Una desventaja de esta técnica es que necesita mucho tiempo y no es factible en todos los casos. El método alterno es la aproximación a la varianza, el cual es más adecuado para el emparejamiento con reemplazo. La ecuación proporcionada para el emparejamiento con el vecino más cercano es la siguiente, donde ω representa cuantas veces se utiliza el individuo control.

$$Var(\hat{T}) = \frac{1}{N_{matched}} Var(Y(1) \mid D=1) + \frac{\sum_{j \in I_0} (\omega_j)^2)}{(N_{matched})^2} . Var(Y(0) \mid D=0),$$

4.5.3.9 Análisis de sensibilidad

Es inevitable que un buen emparejamiento genere un emparejamiento incompleto y que un emparejamiento “máximo” no pueda evitar un emparejamiento inexacto; por lo tanto es necesario determinar un equilibrio entre un emparejamiento incompleto y un emparejamiento inexacto. En la práctica, los sesgos severos debido a un emparejamiento incompleto son de menor importancia que un emparejamiento inexacto.

La selección de un algoritmo de emparejamiento adecuado es un procedimiento importante que afecta la estimación del desenlace. La comparación cuidadosa entre por lo menos dos algoritmos de emparejamiento es importante conforme a pruebas antes y después del emparejamiento[45].

Los individuos dentro de la puntuación de propensión más bajo (casos expuestos) y puntuación de propensión más alto (casos control) no tienen parejas en absoluto porque ellos están fuera de la región de soporte común. El número de esos individuos aparentemente tienen cierta influencia sobre los resultados del emparejamiento. Además, el desenlace es sensible a las covariables condicionales apropiadas en los modelos de estimación de la puntuación de propensión. Más covariables dan una región de superposición más estrecha entre los grupos expuestos y de control. Para probar la sensibilidad del desenlace a la región de soporte común, Lechner[55] describió un enfoque que incorpora información de los individuos no incluidos en la región de soporte común, para calcular el límite no paramétrico del parámetro de interés. Reafirma el concepto errado de ignorar los problemas del soporte común o estimar el desenlace únicamente de los individuos dentro de la región y recomienda un “análisis de límites”.

Factores no observados pueden traer sesgos ocultos adicionales en la estimación del desenlace si ellos están asociados con la exposición y el desenlace o desequilibrada seriamente por factores observables o no correlacionados con la puntuación de propensión. Este sesgo oculto puede crear un efecto de tratamiento falso, enmascarar un efecto de tratamiento verdadero o causar un error en la estimación del efecto del tratamiento en el tamaño, la dirección o ambos. Básicamente, todos los estudios observacionales son afectados potencialmente por los sesgos ocultos, y los análisis de sensibilidad son necesarios en estos estudios. Rosenbaum[57] abordo la solución de este problema estimando el efecto de estos factores en el proceso de selección y análisis de emparejamiento usando un enfoque de límites,

$$\frac{1}{e^r} \leq \frac{P(x_i)(1-P(x_j))}{P(x_j)(1-P(x_i))} \leq e^r$$

donde e^r es una medida del grado de desviación de un estudio que está libre de sesgos ocultos.

Cuando la diferencia en la media y la varianza del logit de la puntuación de propensión y la media y la varianza de los residuales de elementos no observados son muy pequeñas entre los grupos expuesto y control, el estudio puede continuar.

4.5.3.10 Limitaciones y riesgos de la aplicación de la puntuación de propensión

- Limitaciones
  - ❖ Al ajustar solamente por las covariables observables en la estimación de la puntuación de propensión, hay que tener en cuenta que otras covariables ocultas no observables pueden alterar la estimación y la interpretación. Es importante investigar factores ocultos relacionados a la exposición y al desenlace tantos como sea posible.
  - ❖ La puntuación de propensión funciona mejor en muestras grandes. En estudios observacionales pequeños, pueden ser inevitables desequilibrios sustanciales de algunas covariables[58], al igual que la aleatorización inviable en un ensayo clínico aleatorizado.
  - ❖ En teoría, se debe incluir cualquier característica relacionada a la exposición y al desenlace[45] para tener en cuenta la varianza tanto como sea posible. Aunque la inclusión de las covariables extrañas no influye en el sesgo de las estimaciones del emparejamiento, estas introducen más varianza[59]. Trabajos recientes sugieren, que en los estudios moderados o grandes el sesgo de mantener una covariable débilmente predictiva supera la eficiencia ganada de retirarla. Quizá lo más importante es que incluir más covariables hace que la definición de la región de soporte sea mucho más difícil.
  - ❖ Una superposición extremadamente estrecha y con distribuciones desiguales entre los dos grupos puede introducir errores considerables. Por ejemplo, los peores casos en un grupo control y los mejores casos en el grupo expuesto resultarán en una regresión a la media, lo cual hace al grupo control verse mejor y al grupo expuesto verse peor.

- ❖ Se debe emparejar sobre covariables medidas precisamente y estables tanto como sea posible para evitar los valores extremos que también harían regresión hacia la media[45].

- Riesgos

La aplicación de la puntuación de propensión puede debilitar el argumento a favor de los diseños experimentales, que es difícil de controlar e interpretar. Además, se puede sobreestimar la certeza de los resultados por un exceso de la simulación de un diseño experimental.

En resumen, la disminución en el sesgo de selección mediante el uso de las puntuaciones de propensión es una técnica muy útil para aplicar en la evaluación práctica de los desenlaces, particularmente en los estudios observacionales en las áreas farmacéuticas y desenlaces de salud. Su aplicación amplia en muchos principios demuestra que es una arma poderosa para ajustar el sesgo de selección y que es confiable, fácil de usar y muy efectiva [45].

5. Objetivos

5.1 Objetivo general

Determinar si existen diferencias de los costos directos totales de hospitalización entre los pacientes con y sin eventos reportables en el Hospital Universitario de Neiva Colombia, usando la puntuación de propensión para disminuir los sesgos de selección.

5.2 Objetivos específicos

1. Determinar los factores de riesgo de los dos grupos comparados, para presentar ER.

2. Determinar la puntuación de propensión de los individuos de tener o no un ER en el grupo de pacientes con ERs y en el grupo sin ERs según las diferentes covariables medidas.

3. Determinar la reducción del sesgo con el emparejamiento de los grupos de expuestos y no expuestos de acuerdo a la puntuación de propensión.

4. Estimar después del emparejamiento la diferencia del costo total directo promedio institucional por ERs en los dos grupos comparados, expresado en unidades monetarias.

5. Determinar los costos y su diferencia, de los insumos y servicios utilizados en cada hospitalización en ambos grupos como información adicional.

6. Metodología

6.1. Diseño

Estudio auxiliar del estudio inicial – una cohorte prospectiva - sobre incidencia de ERs de Gaitán y colaboradores[1] del cual se obtuvieron datos secundarios de factores de riesgo como la edad, el sexo, estrato socioeconómico, puntaje de comorbilidad de Charlson[60], tipo de seguridad social, y servicio de hospitalización las cuales se consideran como covariables independientes y la presencia de evento reportable o no como desenlace intermedio o variable dependiente. Este desenlace intermedio se convierte en variable explicativa de su consecuencia en costos como desenlace final. Los datos primarios para realizar el costeo, como elementos consumidos, días de estancia hospitalaria, cantidades consumidas, valores unitarios, valores totales y costos institucionales se obtuvieron de las facturas de la atención de cada paciente y de la base de datos de las oficinas de Facturación y Contabilidad del Hospital Universitario de Neiva. Es una cohorte retrospectiva de pacientes hospitalizados durante el año 2006, agrupados en los que tuvieron ERs evaluados por cómite de especialistas y los que no tuvieron ERs, como desenlace intermedio, con el fin de estimar la diferencia de costos directos institucionales promedio – desenlace final - por ERs expresada en unidades monetarias, en pesos colombianos (COP) de 2006. En la figura 2 se presenta un esquema general del diseño de la investigación.

Desde la perspectiva de las evaluaciones económicas, es una evaluación parcial y el estudio se considera un análisis de costos de los eventos reportables. Se tuvo en cuenta la perspectiva del hospital (prestador del servicio de salud) usando los precios de compra y no se hizo tasa de descuento porque toda la muestra se recolectó durante el año de 2006.

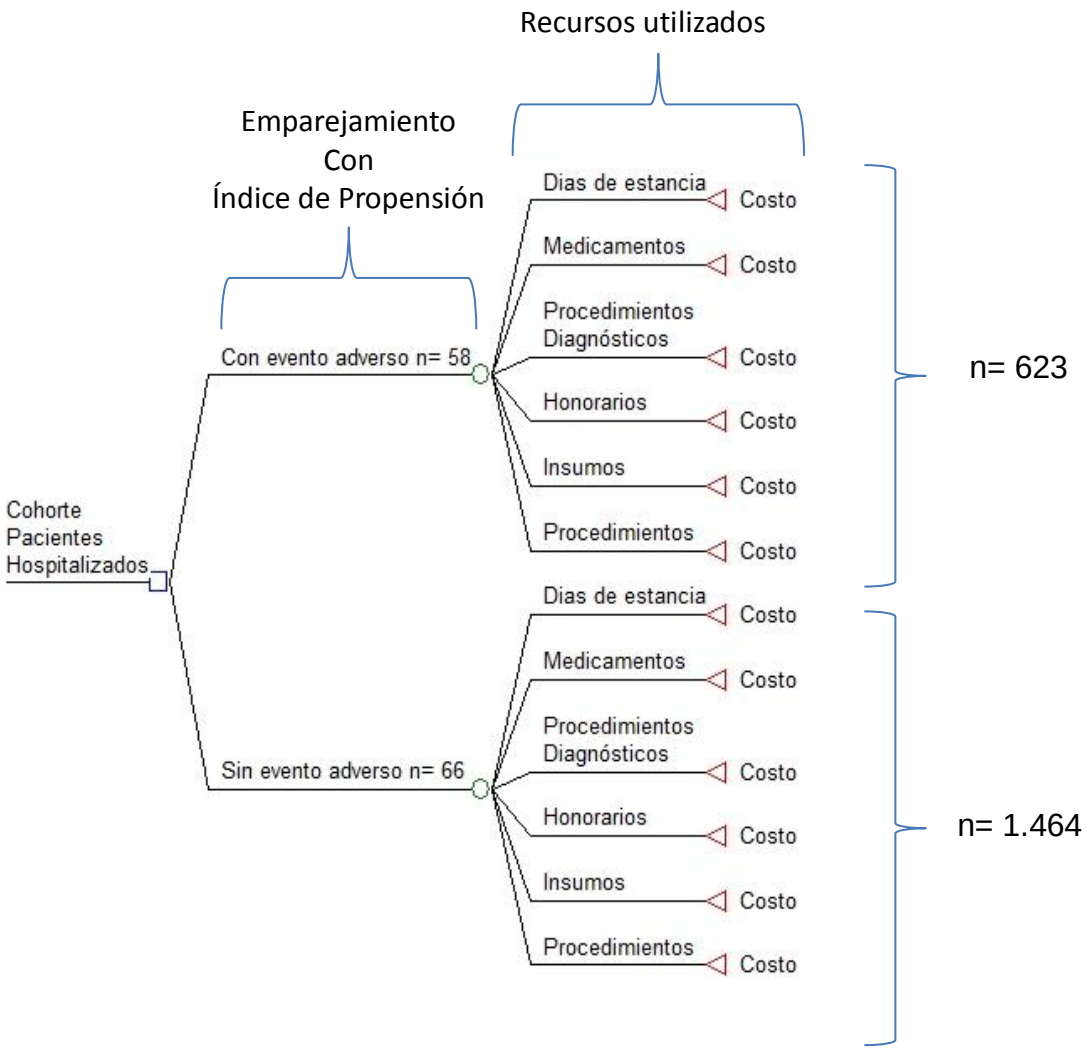
6.2. Población a estudio y lugar:

De las tres instituciones generales incluidas en el estudio de Gaitán y colaboradores[1] se incluyeron los pacientes hospitalizados durante el año de 2006 en los servicios de medicina interna, quirúrgicos o cirugía general y gineco-obstetricia del Hospital Universitario de Neiva, hospital público y general de alto nivel de complejidad, que cumplieron los criterios de inclusión – haber estado hospitalizado mínimo 12 horas – y de exclusión – alteraciones psiquiátricas.

6.3. Tamaño muestral:

Con el programa Epi-Info® para una cohorte asumiendo una frecuencia de ERs para los expuestos en instituciones de alta complejidad del 4%, con un nivel de significancia del 5% y un poder del 80% con una relación 1:.1 de expuestos vs no expuestos respectivamente, se estimó una muestra de 56 individuos con eventos reportables y 56 sin eventos reportables, para un total de 112 individuos. Se hizo un incremento de la muestra a 70 individuos en cada brazo de la cohorte. Se excluyeron 12 individuos en el grupo con eventos reportables y 4 en el grupo sin eventos por ausencia o inconsistencias en las facturas. Finalmente se incluyeron en el análisis 124 individuos, 66 sin eventos reportables y 58 con eventos reportables, 56% (58/103) de todos los que presentaron ER en el Hospital Universitario de Neiva durante el año 2006.

Gráfica 2. Diseño de la cohorte retrospectiva



6.4. Identificación de los recursos utilizados:

De las facturas de los pacientes se determinaron los recursos y servicios utilizados (días de estancia, medicamentos, procedimientos diagnósticos, honorarios, insumos y procedimientos) sus cantidades, valores unitarios y totales respectivos en el grupo sin eventos reportables y en el grupo con eventos reportables y facturados al pagador. La obtención de estos datos la realizó el autor y no hubo un cegamiento en el proceso. Obsérvese en la gráfica 2 que el número de individuos con evento adverso (n=58) y sin evento adverso (66), cuando se consideran en cada grupo los recursos utilizados en cada paciente, las cantidades de observaciones se aumentan a n=623 en el caso de los

que presentaron ER y n=1.464 en los que no presentaron ER. La puntuación de propensión de los individuos se trasladó a estas últimas observaciones y fueron las que finalmente se emparejaron.

6.5. Técnica de costeo de los recursos utilizados:

El estudio se realizó en pacientes que estuvieron hospitalizados y por eso se utilizó una de las formas de estimar los costos de los servicios hospitalarios como es el método de costo por día que nos mejora la precisión de los costos de los pacientes hospitalizados. Para ello se utilizó el enfoque de "Hotel y auxiliares"[34, 35]. Este enfoque tiene en cuenta el costo de la cama/día – días de estancia, que incluye los costos directos no médicos (administración, servicios públicos, alimentación, etc.) - sumado a las diferentes actividades realizadas a los pacientes – costos directos médicos - durante la hospitalización. Teniendo en cuenta los precios de compra se calcularon los costos directos institucionales de acuerdo al costo total de la hospitalización que incluye los insumos y servicios utilizados (días de estancia, medicamentos, procedimientos diagnósticos, honorarios, insumos y procedimientos) en cada paciente sin evento reportable y con evento reportable atribuido a este evento. El enfoque de costeo "Hotel y auxiliares" es semejante al que realiza el Hospital Universitario de Neiva. En el análisis económico se tuvo en cuenta la perspectiva del hospital; por tratarse de una cohorte retrospectiva, recolectada dentro de un mismo año, no se realizó descuento. En el análisis de costos se usaron los costos institucionales de compra que aunque son más difíciles de obtener son los que concuerdan con la perspectiva del hospital, en este caso un hospital público lo cual facilita la consecución de la información y no lo facturado o las tarifas ISS o SOAT que concuerdan mejor con la perspectiva del pagador. Los costos institucionales se expresan en pesos colombianos del año 2006.

6.6 Procedimiento

6.6.1 Verificación preliminar de los costos antes del emparejamiento.

Se hizo una prueba preliminar previamente al cálculo de la puntuación de propensión y su emparejamiento, para determinar en un grupo reducido de individuos la diferencia en

los costos directos institucionales entre el grupo con ER y el grupo sin ER; para ello se determinó el diagnóstico más frecuente y los eventos reportables ocurridos.

6.6.2 Análisis

Antes de determinar las puntuaciones de propensión y realizar el emparejamiento se realizaron pruebas estadísticas *t* de Student y de comparación de proporciones para la determinación de las diferencias significativas en los factores de riesgo, entre los pacientes que tuvieron ER y los que no lo tuvieron. Se encontró que los dos grupos no eran comparables en varias covariables como se muestra más adelante en los resultados (tabla 3) y podrían explicar la posible diferencia en costo entre los dos grupos además de la ocurrencia de los ERs.

6.6.2.1. Emparejamiento con Puntuación de Propensión:

Se decidió utilizar el emparejamiento con la puntuación de propensión para balancear los dos grupos y porque se consideró que la distribución de la variable dependiente continua – costo total promedio institucional – no era normal en parte debido al tamaño de muestra considerado pequeño (n<500) y no se debía usar métodos como la *t* de Student o regresión lineal que puede dar predicciones fuera del intervalo de probabilidad (0,1)[61] y necesitan el supuesto de normalidad. Además el emparejamiento con puntuación de propensión para las cohortes balancea los grupos con un efecto casi como el de la aleatorización en los ensayos clínicos. La implementación de la puntuación de propensión se realizó con Stata 13 (College Station, TX, ®) utilizando el comando *psmatch2*. Este comando por defecto utiliza una regresión probit. Mediante esta regresión se determinaron las covariables o factores de riesgos significativos y relacionados estrechamente con el desenlace y la elección de la exposición. Se estimó la puntuación de propensión para cada individuo incluyendo en el modelo,

$$p(X) = Pr\ (ER=1|\ X_1,\ X_2,\ X_3,\ X_4,\ X_5,\ X_6) = E(ER|X)$$

las covariables independientes o de confusión, edad(X_1), el sexo(X_2), estrato socioeconómico(X_3), puntaje de comorbilidad de Charlson[60](X_4), tipo de seguridad social(X_5), y especialidad médica(X_6) que atendió al paciente; la variable dependiente fue el desenlace intermedio o exposición (evento reportable o no) y el desenlace final fue el

costo total promedio institucional . Se evaluó la capacidad predictiva del modelo mediante análisis de curvas ROC. Se verificó la independencia condicional y la región de soporte común con sus problemas para la puntuación de propensión en el grupo con ER y sin ER. Con el fin de controlar las diferencias existentes en los factores de riesgo, entre el grupo con ERs con el grupo sin ERs, se realizó emparejamiento con la puntuación de propensión mediante el algoritmo *kernel*. Se hizo una revisión del balance de las covariables independientes antes y después del emparejamiento, se calculó el porcentaje de la reducción del sesgo y se estimó la diferencia del costo directo institucional promedio (ATT=Efecto medio del tratamiento en los tratados) con su error estándar de los grupos con ER y sin ER. Para el análisis de sensibilidad se utilizó el enfoque de límites de Rosenbaum para variables continuas.

6.6.3. Análisis estadístico

Se utilizó el software Stata SE, versión 13.0 para el análisis de la información. Sin emparejar, se utilizó la prueba de la suma de rangos de Wilcoxon para muestras independientes para la determinación de la significancia de las diferencias en los costos de los recursos discriminados por categorías.

6.6.4. Consideraciones éticas

El protocolo de investigación fue aprobado por el Comité de Ética e Investigaciones del Hospital Universitario de Neiva por tratarse de un estudio auxiliar sin riesgo, una cohorte retrospectiva con datos obtenidos de la base de datos del estudio de incidencia de eventos reportables aprobado previamente y sin nuevas exposiciones ni intervenciones sobre los pacientes.

7. Resultados

7.1 Población original en Neiva:

En el Hospital Universitario de Neiva 2.054 pacientes cumplieron los criterios de inclusión y de exclusión, de ellos 103 (5%) pacientes presentaron eventos reportables y 1.951 no (95%). De acuerdo al cálculo del tamaño de muestra para la cohorte retrospectiva, mediante muestreo aleatorio simple se seleccionaron 70 pacientes que presentaron ER y 70 pacientes que no presentaron ER. De esta muestra finalmente se analizaron 124 pacientes, 58 con ER (56,31% de los 103 que presentaron ER) y 66 sin ER (3,38% de los 1.951 que no presentaron ER).

Tabla 1. Análisis preliminar de costos por eventos reportables

Evento	Nombre	Código Dx Principal	Dx Principal	Evento	Costo Institucional	Costo Promedio
Si	GPV	O829	Parto por cesárea, sin otra especificación	Readmisión no planeada	5.013.210	8.316.289
Si	MDM	O829	Parto por cesárea, sin otra especificación	Readmisión no planeada	18.900.000	
Si	LMT	O829	Parto por cesárea, sin otra especificación	Infección nosocomial o infección adquirida en el hospital	1.035.658,4	
No	BS	O829	Parto por cesárea, sin otra especificación	-	1.346.450	1’640.347
No	LGE	O829	Parto por cesárea, sin otra especificación	-	959.491	
No	SO	O829	Parto por cesárea, sin otra especificación	-	909.000	
No	MYQ	O829	Parto por cesárea, sin otra especificación	-	3.346.449,6	

Se excluyeron los pacientes en ambos grupos por facturas ausentes, incompletas o con inconsistencias. Se hizo un análisis preliminar de costos con el diagnóstico principal más frecuente que fue parto por cesárea (Tabla 1). Se encontraron con este diagnóstico 4 pacientes que no presentaron ER, 2 que tuvieron una readmisión no planeada y 1 con infección nosocomial o infección adquirida en el hospital como ER; el costo promedio institucional de las hospitalizaciones de los pacientes que tuvieron ER fue $8.316.289 y el de los pacientes sin ER fue $1’.640.347, representando un aumento de 5 veces el costo, con una diferencia de $6.675.942 entre los dos grupos, a simple vista una diferencia

importante. Estos hallazgos preliminares de costos promedios obtenidos por un método simple, con el mismo diagnóstico, se esperan confirmar en la muestra completa calculada para el estudio, controlando los sesgos de selección mediante el emparejamiento con la puntuación de propensión.

Tabla 2 Factores de riesgo asociados a eventos reportables

		Evento Reportable		
Factores de Riesgo		No (n=66)	Si (n=58)	*p*
Relacionados al Paciente				
Edad (años)		37,45 (20,23)	48,48 (20,72)	**0,0033**
Sexo (%)	M - 62.9%	52.5	47.4	0.5218
	H - 37.10%	54.3	45.6	0.4042
Estrato (%)	0 - 29.03%	52.77	47.2	0.6374
	1 - 45.16%	51.7	48.2	0.7055
	2 - 25.1%	56.2	43.7	0.3170
Puntaje de Comorbilidades		0,36 (1,03)	0,87 (1,36)	**0.0184**
Tipo de seguridad social (%)	s - 62.10%	49.3	50.6	0.8720
	c - 5.65%	28.5	71.4	0.1080
	v - 16.13%	50	50	1.0000
	p - 13.71%	76.4	23.5	0.0200
	sd - 2.4%	-	-	
Relacionados al Hospital				
Días de estancia		7,56 (12,40)	17,72 (18,19)	**0,0004**
Servicio de Hospitalización (%)	MI - 19.35%	33.3	66.6	**0.0029**
	Cir - 31.5%	38.4	61.53	**0.0415**
	GO - 49.19%	70.4	29.5	**0.0000**

En la tabla 2 se muestran los factores de riesgo para ER relacionados al paciente y al hospital que caracterizan inicialmente a los dos grupos del estudio antes de estimar las puntuaciones de propensión y realizar el emparejamiento. Se encontraron diferencias estadísticamente significativas entre los dos grupos en la edad (11 años mayores los que presentaron ER, *p=0,0033*), puntaje de comorbilidades (*p=0,0184*), días de estancia (10 días adicionales en el grupo con ER, *p=0,0004*) y en los tres servicios donde estuvieron

hospitalizados los pacientes (medicina interna=MI – *p*=0,0029, cirugía general=Cir – *p*=0,0415 y gineco-obstetricia=GO – *p*=0,0000).

7.2. Selección de variables

Se incluyeron las siguientes covariables medidas por ser algunas de ellas factor de confusión: mujer, edad, estrato socioeconómico, afiliación a seguridad social, puntaje de comorbilidades y especialidad médica que atendió al paciente. Todas las covariables incluidas fueron significativas (*p=0,000)* para la estimación de la puntuación de propensión excepto el puntaje de comorbilidades (*p=0,098)*, sin embargo el modelo es significativo en su conjunto (chi^2=0,0000) (Tabla 3). Con estas covariables se estimó la puntuación de propensión para cada individuo. La covariable días de estancia no se incluyó porque no se pudo discriminar la estancia antes y después de ocurrido el ER.

Tabla 3. Covariables incluidas para la estimación de la puntuación de propensión

Probit regression	Number of obs = 2087				
	LR chi2(7) = 787.69				
	Prob > chi2 = 0.0000				
Log likelihood = -878.38885	Pseudo R2 = 0.3096				
Covariables	Coef.	Std. Err.	z	**P>z**	[Interval Conf 95%]
Mujer	1,333291	0,0990975	13,45	**0,000**	1,139063 - 1,527518
Edad	0,0313056	0,0022391	13,98	**0,000**	0,026917 - 0,0356943
Estrato socioeconómico	-0,3503033	0,047641	-7,35	**0,000**	-0,4436779 - -0,2569287
Afiliación a Seguridad Social	0,6300817	0,0549502	11,47	**0,000**	0,5223812 - 0,7377821
Puntaje de Comorbilidades	0,0474049	0,0286828	1,65	**0,098**	-0,0088123 - 0,1036221
Especialidad Médica	0,9835611	0,0551113	17,85	**0,000**	0,8755449 - 1,091577
_cons	-6,451773	0,3344221	-19,29	**0,000**	-7,107229 - -5,796318

7.3. Estimación de poder

Se evaluó la capacidad predictiva o discriminatoria del modelo mediante análisis de curva ROC para clasificar a los individuos de acuerdo a su mayor o menor probabilidad o puntuación de propensión de tener o no un ER, que se muestra en la gráfica 3. El área bajo la curva fue de 0.8708962 con intervalo de confianza de 95% (0.8556731 - 0.8861192) que nos indica una excelente capacidad discriminatoria del modelo.

7.4. Región de soporte común

Encontramos que las puntuaciones de propensión mínimas y máximas en el grupo sin ER fueron 0,0000369 y 0,7751094, y en el grupo con ER fueron 0,112306 y 0,8044462, lo que nos deja una región de soporte común entre las puntuaciones de propensión de 0,112306 – 0,7751094. (Tabla 4)

Gráfica 3. Curva ROC del modelo predictivo de pscore

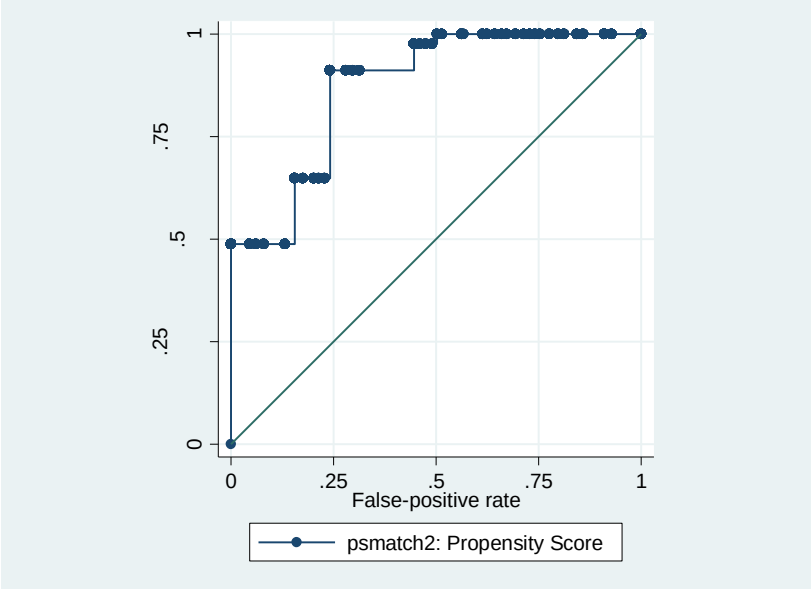


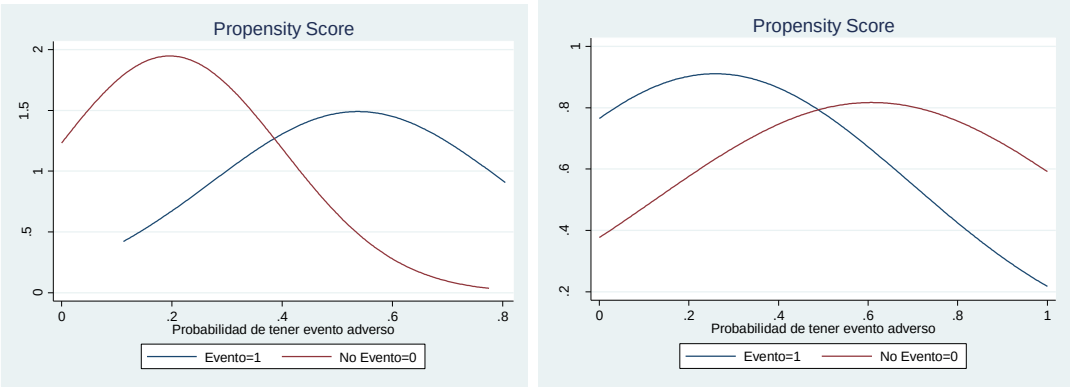
Tabla 4. Límites de Región de Soporte Común

	Mínimos	Máximos
Sin ER	0,0000369	**0,7751094**
Con ER	**0,112306**	0,8044462
Región de Soporte común	**0,112306**	**0,7751094**

En la gráfica 4A antes del emparejamiento, mediante análisis visual simple observamos que la distribución de las puntuaciones de propensión para ambos grupos son diferentes, la del grupo con ER esta sesgada a la derecha (asimetría negativa) y el grupo sin ER esta sesgada a la izquierda (asimetría positiva), sin embargo tiene un espectro similar y una región de superposición con densidades diferentes en los extremos del espectro lo cual era de esperarse para este tipo de exposición. También se puede observar que en el grupo con ER hay un número de puntuaciones de propensión superiores a la máxima puntuación de propensión del grupo sin ER (>0,7751094), así como en el grupo sin ER hay también un número de puntuaciones de propensión menores a la puntuación de propensión mínima en el grupo con ER (<0,112306). En la gráfica 4B después del emparejamiento que se realizó por el algoritmo *kernel*, la distribución de las puntuaciones de propensión muestran una mejor similitud pero no son completamente coincidentes, considerándolas entonces no balanceadas en su totalidad.

La tabla 5 muestra que el 100,00% (2.087/2.087) de las puntuaciones de propensión relacionadas a las observaciones de costos derivadas de los recursos utilizados en cada uno de los paciente incluidos en la muestra, hacían parte del soporte y se incluyeron en el emparejamiento, 1464 del grupo sin ER y 623 del grupo con ER.

Gráfica 4. Distribución de la Puntuación de Propensión Región de Soporte Común



A: **Antes del Emparejamiento** **B: Después del Emparejamiento (No Balanceado)**

Tabla 5. Puntuaciones de Propensión en soporte común

psmatch2: Asignación de exposición	psmatch2: Soporte común	
	On support	Total
Sin ER	1.464	1.464
Con ER	623	623
Total	2.087	2.087

En la gráfica 5 se observa la distribución de las puntuaciones de propensión en los dos grupos de emparejamiento.

7.5 Revisión de balance

Calculando el porcentaje de la reducción del sesgo se hizo la revisión del balance de las covariables antes y después del emparejamiento como se muestra en la tabla 6 y gráfica 6. Se observa que con el emparejamiento hubo una disminución de la proporción del sesgo en las covariables: estrato -931.0%, puntaje de comorbilidad del 46.0%, edad 79.5%, mujer 97.7%, y en especialidad 91.8%; y un aumento de la proporción del sesgo en afiliación al SGSSS del 51.9%. Después del emparejamiento las covariables que presentaron una proporción de sesgo menor del 5%[62] fueron estrato, edad, mujer, y

afiliación a SGSSS. Las covariables puntaje de comorbilidad y especialidad médica aunque tuvieron una reducción importante del sesgo después del emparejamiento, quedaron en 33.8% y 7.8% respectivamente. Esto indica que en general se logró el objetivo de la reducción del sesgo de selección.

Gráfica 5.Puntuación de propensión por estado de exposición zonas de emparejamiento

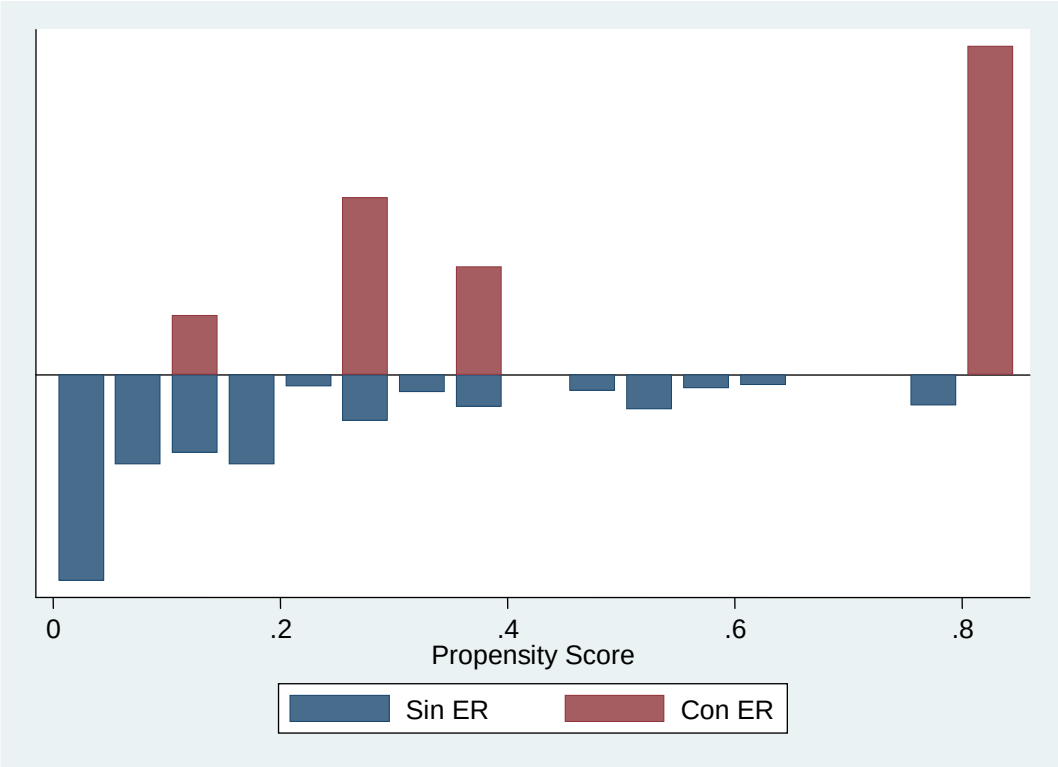
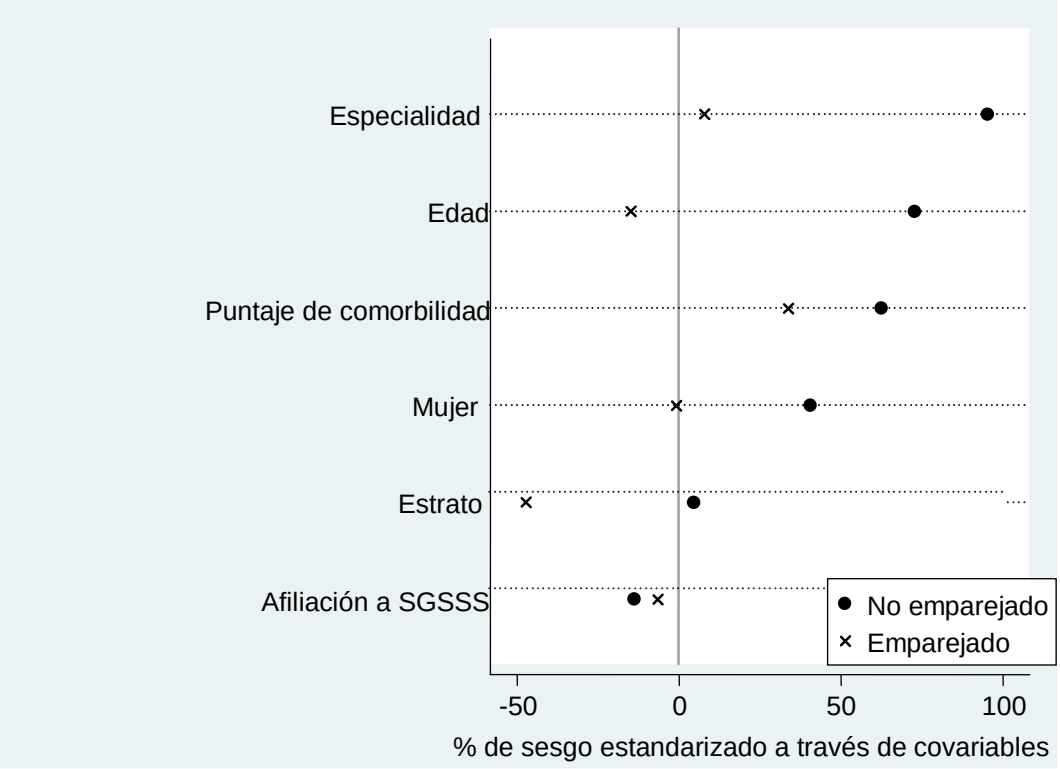


Tabla 6. Proporción de sesgos y su reducción antes y de después del emparejamiento de las covariables

	No emparejado|	Promedio		%reducción|	**t-test**	
Covariable	Emparejado|	Con ER	Sin ER	%sesgo	sesgo |	t	p>|t|
Estrato	No Emparejado|	0.88	0.85	4.6	|	0.98	0.328
	Emparejado|	0.88	1.27	-47.4	-931.0 |	-8.04	0.000
Puntaje Comorb	No Emparejado|	1.38	0.40	62.5	|	14.99	0.000
	Emparejado|	1.38	0.85	33.8	46.0 |	6.11	0.000
Edad	No Emparejado|	60.05	46.68	72.7	|	14.38	0.000
	Emparejado|	60.05	62.79	-14.9	79.5 |	-3.00	0.003
Mujer	No Emparejado|	0.73	0.54	40.5	|	8.26	0.000
	Emparejado|	0.73	0.74	-0.9	97.7 |	-0.18	0.861
Especialidad	No Emparejado|	2.52	1.76	95.3	|	20.31	0.000
	Emparejado|	2.52	2.46	7.8	91.8 |	1.38	0.167
Afiliación	No Emparejado|	2.32	2.45	-13.9	|	-3.04	0.002
	Emparejado|	2.32	2.38	-6.7	51.9 |	-1.26	0.209

Grafica 6. Proporción de sesgos y su reducción antes y después del emparejamiento de las covariables



7.6. Estimación del efecto de la exposición y el error estándar

Mediante emparejamiento de las puntuaciones de propensión utilizando el algoritmo *kernel* se calculó el efecto medio del tratamiento en los tratados (ATT por sus siglas en inglés), para este estudio el costo promedio total institucional en cada paciente; el error estándar se calculó por bootstrapping. El costo promedio total institucional en individuos con ER fue de $30’292.973,6 y en los individuos sin ER fue de $8´138.550,48. Esto representa un incremento de 3.72 veces en el costo promedio. La diferencia del costo promedio total institucional entre los pacientes con ER y sin ER fue de $22’154.423,1. Los resultados anteriores se muestran en la tabla 7.

Tabla 7.Diferencia del costo total institucional con algoritmo *kernel*

Variable	Muestra	Con ER	Sin ER	Diferencia	S.E.	T-stat
Costo total institucional	No emparejado	30’292.973,6	7.867.420,78	22.425.552,8	536.634,07	41,79
	ATT*	**30’292.973,6**	**8´138.550,48**	**22’154.423,1**	**735.645,63**	30,12
ATT* Efecto medio del tratamiento en los tratados						

El error estándar (S.E. por sus siglas en inglés) para el costo institucional promedio (ATT* Efecto medio del tratamiento en los tratados) posterior al emparejamiento fue de $735.645,63; esta cifra representa solamente el 3.32% de la diferencia del costo promedio total institucional entre los individuos con y sin ER. (Tabla 7)

7.7. Análisis de sensibilidad

En la tabla 8 se dan los resultados del análisis de sensibilidad utilizando el método de límites de Rosenbaum, y nos indican que la diferencia del costo promedio total institucional entre los individuos con y sin ER es sensible a la posibilidad de sesgos ocultos que harían que un individuo tuviera 2 a 3 veces mayor probabilidad que su pareja de tener un ER y por supuesto un costo más alto.

Tabla 8. Análisis de Sensibilidad por método de límites de Rosenbaum

Rosenbaum bounds for diff (N = 2408 matched pairs)		
Gamma	**sig+**	**sig-**
1	**0**	**0**
2	**0**	**.00037**
3	**0**	**1**
4	0	1
5	0	1
6	0	1
7	0	1
8	0	1
* gamma - log odds of differential assignment due to unobserved factors sig+ - upper bound significance level sig- - lower bound significance level		

7.8. Costos institucionales según los recursos utilizados

Como información adicional, los recursos discriminados por categorías utilizados en la atención de los pacientes se muestran en la tabla 9. Sin emparejar se estimaron las medianas y sus diferencias en los costos institucionales para las categorías días de estancia, medicamentos, procedimientos diagnósticos, honorarios, insumos y procedimientos en los pacientes que presentaron ER y los que no presentaron ER. Los costos por días de estancia fueron en pacientes sin ER $1’599.499 (Rango IC 1.257.304 - 5.157.464) vs. en pacientes con ER $9’403.270 (Rango IC 5.561.265 - 33.600.000); por medicamentos fueron en pacientes sin ER $5’157.464 (Rango IC 1.599.499 - 9.963.598) vs. en pacientes con ER $33’600.000 (Rango IC 32.300.000 - 47.800.000); por procedimientos diagnósticos fueron en pacientes sin ER $2’487.460 (Rango IC 1.324.164 - 13.900.000) vs. en pacientes con ER $ 32’300.000 (Rango IC 32.300.000 - 33600000); por honorarios fueron en pacientes sin ER $ 1’841.792 (Rango IC 1.335.307 - 5.856.571) vs. en pacientes con ER $33.600.000 (Rango IC 32.300.000 - 33.600.000); por insumos fueron en pacientes sin ER $5’856.571 (Rango IC 1.599.499 - 13.900.000) vs. en pacientes con ER $33’600.000 (Rango IC 9.403.270 - 47.800.000); y por procedimientos fueron en pacientes sin ER $2’487.460 (Rango IC 1.346.450 - 13.900.000) vs. en pacientes con ER $33.600.000 (Rango IC 32.300.000 - 33.600.000). Hubo diferencias significativas (p=0.0000) en los costos institucionales en cada uno de las categorías de recursos entre el grupo que tuvo ER y el grupo que no tuvo ER.

Tabla 9 Costos institucionales discriminados por recursos utilizados

Costo Institucional por Recurso Utilizado								
Evento Adverso								
Recurso	**No ($) Mediana**	**Rango intercuartil**	**%**	**Si ($) Mediana**	**Rango intercuartil**	**%**	**Diferencia ($)**	***p***
Días Estancia	1.599.499,00	1.257.304 - 5.157.464	8,23	9.403.270,00	5.561.265 - 33.600.000	5,34	7.803.771,00	<0.0203
Medicamentos	5.157.464,00	1.599.499 - 9.963.598	26,54	33.600.000,00	32.300.000 - 47.800.000	19,08	28.442.536,00	<0.0000
Procedimientos Diagnósticos	2.487.460,00	1.324.164 - 13.900.000	12,80	32.300.000,00	32.300.000 - 33600000	18,34	29.812.540,00	<0.0000
Honorarios	1.841.792,00	1.335.307 - 5.856.571	9,48	33.600.000,00	32.300.000 - 33.600.000	19,08	31.758.208,00	<0.0000
Insumos	5.856.571,00	1.599.499 - 13.900.000	30,14	33.600.000,00	9.403.270 - 47.800.000	19,08	27.743.429,00	<0.0000
Procedimientos	2.487.460,00	1.346.450 - 13.900.000	12,80	33.600.000,00	32.300.000 - 33.600.000	19,08	31.112.540,00	<0.0000

p calculada con la prueba de la suma de rangos de Wilcoxon para muestras independientes

8. Discusión

Con el objetivo de disminuir los sesgos de selección mediante el emparejamiento, cada vez se utiliza con mayor frecuencia la puntuación de propensión en estudios de costos en salud, en este caso de individuos con y sin ER. En nuestra cohorte retrospectiva, se utilizó para realizar un análisis de costo de los ERs que se presentaron durante el año de 2006 en el Hospital Universitario de Neiva. Además, el estudio se hizo desde la perspectiva del prestador de servicios de salud, con el método de costos por día con el enfoque de "Hotel y auxiliares", que tiene en cuenta el costo de la cama/día – días de estancia - sumado a las diferentes actividades realizadas a los pacientes durante la hospitalización. Se pudo comprobar que la metodología de la puntuación de propensión sirve para seleccionar las variables que deben ser incluidas en el modelo para estimar las puntuaciones de propensión, también que el emparejamiento con la puntuación de propensión hace una función de casi una aleatorización, puede clasificar a los individuos de acuerdo a su mayor o menor probabilidad o puntuación de propensión de tener o no la exposición, disminuye el sesgo de selección y puede calcular un efecto promedio incluido su error estándar y el análisis de sensibilidad. Estas funciones la hacen útil para el análisis de estudios observacionales.

8.1. Resultados claves

En nuestro estudio encontramos que de los factores de riesgo medidos, la edad, el puntaje de comorbilidad, los días de estancia y el servicio de hospitalización fueron los que mostraron diferencia significativa entre los dos grupos, mostrando que los individuos con ER tienen mayor edad (11 años), mayor puntaje de comorbilidad, más días de estancia (10 días), y son hospitalizados en los servicios de medicina interna, cirugía general y gineco-obstetricia en orden descendente. Los dos grupos no se encontraron

comparables antes del emparejamiento por las diferencias en estos factores de los cuales la edad, los días de estancia y el puntaje de comorbilidad se saben son confusores al estar relacionados con los eventos reportables y los costos derivados de ellos.

Las covariables incluidas en el modelo para estimar la puntuación de propensión y observadas antes de la ocurrencia de los eventos reportables se consideraron independientes del grupo asignado, cumpliendo así con el supuesto de independencia condicional. El poder estimado del modelo indicó una excelente capacidad discriminatoria para los individuos de tener o no un ER.

Para el cumplimiento del supuesto del soporte común se encontró una región bastante amplia entre las puntuaciones de propensión 0,112306 y 0,7751094. Mediante análisis visual simple se observó que la distribución de las puntuaciones de propensión para ambos grupos fueron diferentes, la del grupo con ER estaba sesgada a la derecha y la del grupo sin ER estaba sesgada a la izquierda, pero con un espectro similar y una región de superposición amplia con densidades diferentes en los extremos del espectro lo cual era de esperarse para este tipo de exposición. También se observó que en el grupo con ER hubo un número de puntuaciones de propensión superiores a la máxima puntuación de propensión del grupo sin ER (>0,7751094), así como en el grupo sin ER también hubo un número de puntuaciones de propensión menores a la puntuación de propensión mínima en el grupo con ER (<0,112306). Después del emparejamiento la distribución de las puntuaciones de propensión en ambos grupos mostraron una mejor similitud pero sin ser completamente coincidentes, considerándolas entonces no balanceadas.

El 100,00% (2087/2087) de las puntuaciones de propensión correspondientes a las observaciones de costos derivadas de los recursos utilizados en cada uno de los paciente incluidos en la muestra de los dos grupos – sin ER (1.464 observaciones) y con ER (623 observaciones) - se incluyeron en el emparejamiento. La inclusión en el emparejamiento de todas las puntuaciones de propensión garantiza una mayor precisión de los resultados[61], en este estudio la estimación del ATT y por lo tanto la generazibilidad del resultado.

La revisión del balance de las covariables antes y después del emparejamiento se hizo calculando el porcentaje de la reducción del sesgo, observando que con el emparejamiento hubo una disminución importante de la proporción del sesgo en la mayoría de las covariables, excepto en el caso de la afiliación al SGSSS que de todas maneras permaneció con una proporción de sesgos menor al 5%. Este hecho nos permite decir que se logró la reducción del sesgo de selección que es el objetivo más importante que se busca con la utilización de las puntuaciones de propensión.

En el análisis de los costos mediante el emparejamiento de las puntuaciones de propensión utilizando el algoritmo *kernel*, calculando el ATT (por sus siglas en inglés) o efecto medio del tratamiento en los tratados, para este estudio el costo promedio total institucional, se encontró un aumento de los costos promedios institucionales por paciente con ER de $22.154.423,1. Esto representa un incremento de 3.72 veces en el costo promedio. Recordemos que esta diferencia de costos está determinada a partir de la totalidad de las observaciones de costos y esto mejora la posibilidad de generalizar el efecto estimado por su precisión mayor[61]. Mediante bootstrapping se obtuvo un error estándar para el costo institucional promedio de $735.645,63 posterior al emparejamiento; cifra que representa el 3.32% de la diferencia del costo promedio total institucional entre los individuos con y sin ER, considerándose una estimación muy precisa aún en la posibilidad de sesgos. El análisis de sensibilidad realizado utilizando el método de límites de Rosenbaum, nos indica que la diferencia del costo promedio total institucional entre los individuos con y sin ER es sensible a la posibilidad de sesgos ocultos que harían que un individuo tuviera 2 a 3 veces mayor probabilidad que su pareja de tener un ER y por supuesto un costo más alto.

Al estimar sin emparejamiento las medianas de los costos institucionales para los recursos utilizados en las categorías, estancia, medicamentos, procedimientos diagnósticos, honorarios, insumos y procedimientos en los pacientes que presentaron ER y los que no presentaron ER, se encontraron diferencias significativas (p=0.0000) en cada una de las categorías de los recursos utilizados, siendo mayor en los individuos que tuvieron ER.

Utilizando la diferencia estimada por ATT tenemos un total de $1.284’956.539,8 para la muestra de 58 pacientes con ER incluidos en el estudio y de $2.281’905.579,3 para los

103 pacientes que presentaron ER durante el año 2006 en el Hospital Universitario de Neiva, lo cual representa un 3.12% de los $73.013'180.880 facturados por el Hospital Universitario de Neiva durante el año de 2006 por los servicios de salud prestados según comunicación personal de la Oficina de Facturación. Con la evitabilidad de ER del 61% encontrada por Gaitán y colaboradores[1], implementando las intervenciones adecuadas para evitar esta proporción de ER, se hubiera podido ahorrar un total de $1.391'962.403,37, cantidad que se hubieran podido gastar en la prestación de otros servicios de salud con mayor prioridad o invertir en la implementación de un programa de gestión del riesgo.

8.2 Limitaciones

Para este estudio se calculó para una cohorte un tamaño de muestra con una relación 1:.1 de expuestos vs no expuestos respectivamente, con un poder del 80%. Finalmente se analizaron 124 pacientes, 58 con ER (56,31% de los 103 que presentaron ER) y 66 sin ER (3,38% de los que no presentaron ER) y se cumplió con el tamaño de muestra calculado, sin embargo no se utilizó toda la muestra posible de la cohorte del Hospital Universitario de Neiva. Zhao[51] descubrió que el emparejamiento con puntuación de propensión no es superior al emparejamiento de covariables para muestras pequeñas (n<500), pero funciona mejor en muestras grandes (n>1000). Nosotros intentamos que el n para emparejar fuera el número de pacientes que tuvieran o no ER para calcular la diferencia de costos directos de hospitalización, estimando el costo promedio total por paciente pero al estimar las puntuaciones de propensión con n=58 en caso de ER y n=66 sin ER no logramos una región de soporte común adecuada y tuvimos que utilizar los n correspondientes al número de observaciones de recursos utilizados en los dos grupos, n=623 en el grupo con ER y n=1464 en el grupo sin ER como se muestra en la gráfica 2.
En un estudio con muestra pequeña, un desequilibrio importante de algunas covariables puede ser inevitable a pesar del uso apropiado del algoritmo de emparejamiento y la estimación de la puntuación de propensión. En nuestro estudio en la prueba de balance de las covariables no hubo un balance completo de algunas de ellas. Después del emparejamiento las covariables que presentaron una proporción de sesgo menor del 5%[62] fueron estrato, edad, mujer, y afiliación a SGSSS (Sistema General de Seguridad Social en Salud) pero no fue así en el caso de puntaje de comorbilidad y especialidad

médica o servicio de hospitalización. En un estudio se deben preservar en el grupo de expuestos tantos individuos como sea posible, en nuestro caso solo se incluyo el 56% de los expuestos posibles y esto podría disminuir la posibilidad de generalizar el efecto estimado por su precisión menor[61]. También, para aumentar la posibilidad de encontrar una mejor pareja, se tiende a mantener tan grande como sea posible el grupo control, aquí solamente incluimos el 3% de los controles posibles. Sin embargo cumplimos con el criterio establecido para la determinación del tamaño de muestra recomendado en la literatura, el cual establece que el porcentaje total de individuos expuestos no debe ser menor del 10% [52].

Uno de los problemas que pueden surgir con la región de soporte común es cuando la densidad en las colas de la distribución es muy delgada. Una es una muestra con una cola extremadamente sesgada a la derecha para el grupo de expuestos y una cola extremadamente sesgada a la izquierda para el grupo control, en nuestro estudio se presentó esta situación antes del emparejamiento y se corrigió parcialmente debido a que después del emparejamiento la distribución de las densidades de las puntuaciones de propensión no quedaron adecuadamente balanceadas. Lechner[55] señalo que ignorar los problemas de la región de soporte común puede resultar en estimaciones erróneas.
En nuestro estudio no fueron medidos otros factores de riesgo para ER como son raza, escolaridad, severidad de la enfermedad, área de procedencia, órgano o sistema comprometido, signos vitales, automedicación, las deficiencias institucionales, el número y tipo de los medicamentos, la continuidad del control, el tipo de atención y el origen de la remisión y podrían explicar la sensibilidad de la estimación de la diferencia en los costos directos promedio institucionales a la posibilidad de sesgos ocultos derivados de estos factores no medidos. El análisis de sensibilidad realizado, mostró que la diferencia de los costos promedios institucionales entre los individuos con y sin ER es sensible a la posibilidad de sesgos ocultos que harían que un individuo tuviera 2 a 3 veces mayor probabilidad que su pareja de tener un ER y su representación en costos. La variable días de estancia es un factor de confusión y no se incluyó porque no se pudo discriminar la estancia antes y después de ocurrido el ER.

8.3. Interpretación

Este es tal vez el único estudio de análisis de costos de los ER con estas características realizado en Colombia. El estudio de Tribiño[13] no es comparable con nuestro estudio, se limitó a las reacciones adversas a medicamentos, en pacientes hospitalizados en el servicio de medicina interna; utilizó la perspectiva del pagador, el microcosteo, las tarifas SOAT 2004, del ISS 2001 y se realizó en forma prospectiva; encontró que el costo generado por la atención de las reacciones adversas fluctuó entre $93'633.422 y $122'155.406. Pinzón[14] realizó un estudio descriptivo, también limitado a eventos adversos a medicamentos, utilizando la perspectiva del pagador y las tarifas SOAT 2007, mediante análisis de regresión lineal se estableció el modelo de costos generados por eventos prevenibles. Encontró que los costos totales asociados con la atención de estos eventos estuvo entre $ 33´620.346 (US$ 16.687) y $ 37´754.856 (US$ 18.739). Este estudio es algo similar al de Tribiño[13] y tampoco es comparable con el nuestro. Cardona[63] realizo un estudio descriptivo, retrospectivo, limitado a un servicio de urgencias y encontró que los costos de no calidad corresponden al 59,62% de los costos totales de calidad, representados en $8’849.254 y los costos de calidad preventivos al 40,37%, representados en $5’991.018. Ovalle[64] presentó un resumen en la XI Reunión Anual de la Red Cochrane Iberoamericana 2012 en Lima, Perú, donde mostró una cohorte prospectiva de pacientes hospitalizados en los servicios de cirugía y medicina interna; la media de los costos por paciente fue de $ 1.148.618,79, y un costo total de los eventos adversos para el sistema de salud colombiano de $55.713.481.199.

En la literatura existen varios estudios publicados[65-67] de costos de eventos adversos pero ninguno con las mismas características del nuestro. Hoonhout[65], evaluó los costos médicos directos totales asociados a ER y ER prevenibles en 21 hospitales holandeses en forma retrospectiva, midiendo la estancia y procedimientos médicos adicionales después de ocurrir un ER y utilizando precios holandeses de costos estandarizados. Los costos médicos directos anuales en 2004 en hospitales holandeses, se estimaron en un total de 355 millones de euros para todos los ERs y 161 millones de euros para los ER evitables. Las estimaciones de los costos médicos directos totales prevenibles de ER indicaron que formaban una parte importante (1%) de los gastos del presupuesto nacional de salud y eran importantes para la gestión hospitalaria. El factor de costo de los gastos médicos directos fue el exceso de duración de la estancia (incluyendo reingresos) en un hospital. Este estudio es similar al

nuestro en que es retrospectivo, midió la estancia y procedimientos adicionales, y también encontró un aumento de la estancia, difiere en que en nuestro estudio se hizo en un solo hospital de tercer nivel y las diferentes categorías de recursos tuvieron proporciones similares. Los costos no son comparables por estar expresados en monedas y años diferentes. Fuller[66] usando las bases de datos de hospitales de California y Maryland estimó los costos incrementales de 64 categorías de complicaciones potencialmente prevenibles sucedidas durante la hospitalización. El costo incremental fue consistente en ambas bases de datos y explica un aumento del costo total de las hospitalizaciones de 9.39% en los datos de California y de 9.63% en los datos de Maryland debido a complicaciones potencialmente prevenibles. Los autores consideran que esta estimación puede llevar a diseñar sistemas alternativos de pagos que den más incentivos para disminuir significativamente las complicaciones hospitalarias. El estudio de Fuller se diferencia del nuestro en que es realizado en varios hospitales de dos estados americanos, se enfoca en complicaciones prevenibles y no utiliza la puntuación de propensión. Kaushal[67], determinó los costos de ER en pacientes en una unidad de cuidado intensivo médico y en la unidad de cuidado cardíaco, para ello realizó un análisis de caso-control emparejado en forma prospectiva durante un año en un hospital universitario de tercer nivel. Los costos se determinaron a partir de los sistemas de facturación del hospital. Para 56 pacientes de la unidad de cuidados intensivos médicos, el costo de un evento adverso fue de 3.961 dólares (p=0,010) y el aumento de la duración de la estancia fue de 0,77 días (p=0,048). Esto extrapolado a los costos anuales de $ 853.000 para los eventos adversos en la unidad de cuidados intensivos médicos. Del mismo modo, para 52 pacientes de cuidados intensivos cardíacos, el costo de un evento adverso fue de $ 3,857 (p=0.023), correspondiente a $ 630.000 en costos anuales. En promedio, los pacientes con eventos en la unidad de cuidados intensivos cardíacos tuvieron un aumento de 1,08 días de duración de la estancia (p=0,003). Como vemos los pacientes que requieren cuidados intensivos también están en riesgo de ERs y los costos asociados con este tipo de eventos son importantes y pueden justificar una mayor inversión en las estrategias de prevención. En este estudio no definieron la perspectiva del pagador, utilizaron un algoritmo diferente a la puntuación de propensión para realizar el emparejamiento, estuvo limitado a dos unidades de atención especializadas y fue realizado en forma prospectiva; características que lo hacen diferente al nuestro. Brown[12] en Nueva Zelanda estimó en forma retrospectiva en 13 hospitales públicos, el costo del tratamiento de los efectos adversos utilizando los precios cobrados a los pacientes extranjeros para estimar el costo de los recursos sanitarios usados. Encontró que los eventos adversos costaron un promedio de

$NZ 10.264 por paciente. Para Nueva Zelanda, los eventos adversos costarían al sistema médico NZ $870 millones, de los cuales NZ$590 millones fueron hacia el tratamiento de los eventos adversos evitables, lo que sugiere que recursos importantes podrían salvarse mediante la eliminación de los eventos adversos evitables. El estudio de Brown se parece al nuestro en que fue retrospectivo, pero se diferencia en que revisó los casos de 13 hospitales, no define la perspectiva del pagador y el costo se calculó a partir de las facturas cobradas a pacientes extranjeros en una moneda diferente al peso colombiano, por lo tanto tampoco se considera comparable al nuestro.

Varios estudios[68] han utilizado la puntuación de propensión para hacer emparejamiento pero ninguno han utilizado el ATT para estimar la diferencia de costos entre los grupos con ER y sin ER. Chu[68] en pacientes con cáncer colorrectal localmente avanzado y metastásico, tratados con capecitabina o 5-fluorouracil/leucovorin (5-FU/LV) como monoterapia o terapia de combinación con oxaliplatino, cuantificó los costos asociados con la adquisición de la quimioterapia, su administración y el manejo de las complicaciones relacionadas con la quimioterapia y sus efectos sobre el gasto total de los pacientes; los datos fueron obtenidos de una base de datos nacional de servicios médicos. El sesgo de selección fue abordado por emparejamiento mediante la puntuación de propensión, en la evaluación del riesgo de las complicaciones utilizan modelos de Cox, y hacen una estimación de los gastos utilizando modelos lineales generales. Encontró que el costo mensual por paciente durante el tratamiento con capecitabina o capecitabina / oxaliplatino es significativamente más bajo que durante el tratamiento con 5-FU/LV o 5-FU/LV/oxaliplatin debido a costos menores para la administración de la quimioterapia y para el tratamiento de las complicaciones. Este estudio a pesar que utilizó la puntuación de propensión para realizar el emparejamiento como en nuestro caso, es muy diferente a nuestro estudio, ya que incluye costos adicionales a los originados por las complicaciones del tratamiento, los datos no fueron obtenidos de un hospital en particular, la estimación de los gastos se realizó utilizando modelos lineales generales y estuvo limitado a una enfermedad en particular.

Nuestro estudio logró el objetivo de mostrar que si existen diferencias de los costos directos totales de hospitalización entre los pacientes con y sin ERs. El costo fue mayor en 3.72 veces en los pacientes que presentaron ERs. Esta diferencia fue consistente teniendo en cuenta que se mostró de tres formas diferentes y en el mismo sentido. Aunque las tres

comparaciones (comparación preliminar en diagnóstico más frecuente con y sin ER, estimación de ATT con puntuación de propensión y el costo institucional de las diferentes categorías de recursos utilizados sin emparejamiento) que se hicieron para estimar la diferencia de los costos directos promedios institucionales mostraron un aumento significativo de estos costos directos en el grupo de individuos con ER, se considera que esta diferencia estimada por ATT podría estar afectada por las varias limitaciones de nuestro estudio y podría afectar la generazibilidad de la estimación, no obstante la precisión de ella según el error estándar tan pequeño encontrado. Asumimos que la diferencia significativa encontrada por paciente cuando se estimaron los costos institucionales en cada categoría de los recursos utilizados, es explicada en buena parte por los sesgos de selección, teniendo en cuenta que los dos grupos no eran comparables en todas las covariables antes del emparejamiento con la puntuación de propensión.

8.4 Generalizabilidad

Por la características de nuestro estudio, en especial la perspectiva del prestador del servicio de salud, el método de costeo utilizando los precios de compra y las varias limitaciones de nuestro estudio, no se considera posible que los resultados de costos aquí presentados se puedan generalizar a otras instituciones hospitalarias en el país semejantes al Hospital Universitario de Neiva y mucho menos a otros hospitales de tercer nivel por fuera de Colombia. Lo que sí se puede generalizar y se recomienda es el uso de la puntuación de propensión en estudios observacionales como el nuestro para controlar los sesgos de selección en especial en estudios de costos donde la variable de desenlace es continua y no siempre tiene una distribución normal. Recordemos que la puntuación de propensión es útil en el análisis de bases de datos grandes obtenidas usualmente de la práctica habitual donde es difícil realizar ensayos clínicos para realizar inferencias causales. Además por medio de la puntuación de propensión es posible realizar el emparejamiento de un mayor número de covariables observadas que el permitido por el emparejamiento convencional el cual se aplica principalmente en estudios de casos y controles. Corregir las limitaciones de nuestro estudio como es el tamaño de muestra, debe mejorar el emparejamiento, los problemas de soporte común y el balance de las covariables con el fin de obtener mejores estimaciones de la diferencia de costos al estimar el ATT.

8.5 Implicaciones para la investigación y práctica

Se sugiere realizar este tipo de estudios con esta metodología en cohortes prospectivas, con el mayor tamaño de muestra posible y durante un mismo año fiscal con el fin de controlar mejor los posibles sesgos que se puedan presentar. Esto garantizaría unas estimaciones mejores de los costos derivados de la ocurrencia de ER, que nos permita realizar evaluaciones económicas completas validas de las intervenciones implementadas con el fin de disminuir la frecuencia de los diferentes desenlaces producto de los ER y así mejorar la seguridad en la atención de los pacientes y aumentar la calidad en la prestación de los servicios de salud en Colombia. Además aplicando la metodología de la puntuación de propensión se podrían realizar análisis de costos en otras enfermedades y en otros servicios dentro de la Institución aportando un conocimiento que le permita a las directivas gestionar en una forma mejor los recursos.

Es necesario mejorar el conocimiento de la implementación de la puntuación de propensión en el diseño de los estudios observacionales, de tal manera que los resultados obtenidos de estos estudios tengan mayor validez interna y externa. De esta forma podremos mostrar más avances en la investigación científica de los múltiples problemas en salud que actualmente existen en nuestro país y en los cuales no siempre se pueden realizar ensayos clínicos.

En el Hospital Universitario de Neiva durante el año 2006 el aumento del costo total de 103 pacientes que presentaron ERs podría llegar a $2.281’905.579,3, siendo evitables el 61% ($1.391’962.403,37) de acuerdo a lo encontrado por Gaitán y colaboradores[1]. Esta información justifica en el Hospital Universitario de Neiva la implementación de intervenciones que disminuyan la ocurrencia de los ERs, disminuyan las reclamaciones judiciales derivadas de la ocurrencia de los ERs y cuyas pretensiones para indemnizar los daños patrimoniales y extrapatrimoniales pueden llegar a ser muy onerosas; aumenten la seguridad en la atención de los usuarios, reduzcan los costos de la atención para el Sistema de Salud y contribuyan a lograr la acreditación de la Institución.

9. Conclusiones

En nuestro estudio encontramos que de los factores de riesgo medidos, la edad, el puntaje de comorbilidad, los días de estancia y el servicio de hospitalización fueron los que mostraron diferencia significativa entre los dos grupos, mostrando que los individuos con ER tienen mayor edad (11 años) y puntaje de comorbilidad, más días de estancia (10 días), y son hospitalizados en los servicios de medicina interna, cirugía general y gineco-obstetricia.

Se estimaron los costos institucionales para las diferentes categorías de los recursos utilizados encontrando diferencias significativas (p=0.0000) en la mayoría de ellas, siendo mayor en los individuos que tuvieron ER.
Para la estimación de la puntuación de propensión se utilizó un modelo probit y el poder del modelo se estimó mediante un análisis de curva ROC que indicó una excelente capacidad discriminatoria para los individuos de tener o no un ER.

Se cumplió con el supuesto del soporte común. El emparejamiento se realizó utilizando el algoritmo kernel. Después del emparejamiento la distribución de las puntuaciones de propensión en ambos grupos mostraron una mejor similitud pero sin ser completamente coincidentes, considerándolas no balanceadas. Con el emparejamiento hubo una reducción en la proporción de los sesgos. Las covariables que presentaron una proporción de sesgo menor del 5%[62] fueron, estrato, mujer, edad, y afiliación a SGSSS (Sistema General de Seguridad Social en Salud). Se puede decir que la proporción de sesgos de las covariables medidas es mejor después del emparejamiento.

En el análisis de los costos mediante la puntuación de propensión calculando el ATT, para este estudio el costo promedio total institucional, se encontró un aumento de los

costos promedios institucionales en los pacientes con ER de $22.154.423,1 que significa un incremento de 3.72 veces en el costo. Mediante bootstrapping se obtuvo un error estándar para el costo institucional promedio de $735.645,63 posterior al emparejamiento, considerándose una estimación muy precisa aún en la posibilidad de sesgos. El análisis de sensibilidad realizado, dio como resultado odds ratios (OR) entre 2 y 3 e indica que la diferencia de los costos promedios institucionales entre los individuos con y sin ER es sensible a la posibilidad de sesgos ocultos que harían que un individuo tuviera 2 a 3 veces mayor probabilidad que su pareja de tener un ER y por supuesto un costo más alto.

La metodología de puntuación de propensión sirve para seleccionar las variables que deben ser incluidas en un modelo para estimar las puntuaciones de propensión, puede tener un efecto de casi una aleatorización, puede clasificar a los individuos de acuerdo a su mayor o menor probabilidad o puntuación de propensión de tener o no la exposición, disminuye el sesgo de selección y puede calcular un efecto promedio incluido su error estándar y el análisis de sensibilidad. Estas funciones la hacen útil para el análisis de estudios observacionales incluidos los estudios de cohortes y análisis de costo.

El estudio logró el objetivo de mostrar que si existen diferencias de los costos directos totales de hospitalización entre los pacientes con y sin ERs. El costo fue mayor 3.72 veces en los pacientes que presentaron ERs. Este resultado fue consistente teniendo en cuenta que se mostró de tres formas diferentes y en el mismo sentido. También justifica en el Hospital Universitario de Neiva la implementación de intervenciones que disminuyan la ocurrencia de los ERs.

Bibliografía

1. Gaitan-Duarte, H., et al., *[Incidence and preventability of adverse events in patients hospitalised in three Colombian hospitals during 2006].* Rev Salud Publica (Bogota), 2008. **10**(2): p. 215-26.
2. Lifshitz, A. *El Significado actual de "Primum Non Nocere"*. Seminario El Ejercicio Actual de la Medicina 2002 Consultado el 07 de abril de 2010]; Available from: http://www.medicinaysalud.unam.mx/seam2k1/2002/ponencia_jul_2k2.html.
3. Hipocrática, E. *Juramento hipocrático*. (s. V aC) Consultado el 07 de abril de 2010]; Available from: http://www.unav.es/cdb/juramento1.html.
4. PAHO. *Calidad de la atención: seguridad del paciente*. 55ª Asamblea Mundial de la Salud 2002 Consultado el 07 de abril de 2010]; Available from: http://www.paho.org/Spanish/AD/THS/EV/blood-4ta-resolucion.pdf.
5. OMS. *Programa de financiación en investigación en seguridad del paciente*. Alianza Mundial para la Seguridad del Paciente 2008 Consultado el 07 de abril de 2010]; Available from: http://www.who.int/patientsafety/research/grants/Smallgrants_summary_A4Spanish.pdf.
6. Jha, A., *Summary of the evidence on patient safety : implications for research*. 2008, World Alliance for Patient Safety. World Health Organization.
7. MinHealth, *Reportable Events Guidelines*. 2001, Ministry of Health. New Zealand.
8. MinProtección, *Lineamientos para la implementación de la Política de Seguridad del Paciente*. 2008, Ministerio de la Protección Social. República de Colombia.
9. Johnson, W.G., et al., *The economic consequences of medical injuries. Implications for a no-fault insurance plan.* JAMA, 1992. **267**(18): p. 2487-92.
10. Wilson, R.M., et al., *The Quality in Australian Health Care Study.* Med J Aust, 1995. **163**(9): p. 458-71.

11. Thomas, E.J., et al., *Costs of medical injuries in Utah and Colorado.* Inquiry, 1999. **36**(3): p. 255-64.
12. Brown, P., et al., *Cost of medical injury in New Zealand: a retrospective cohort study.* J Health Serv Res Policy, 2002. **7 Suppl 1**: p. S29-34.
13. Tribino, G., et al., *[Direct costs and clinical aspects of adverse drug reactions in patients admitted to a level 3 hospital internal medicine ward].* Biomedica, 2006. **26**(1): p. 31-41.
14. Pinzon, J.F., et al., *[Direct costs and hospital morbimortality impact from preventable adverse drug events].* Biomedica, 2011. **31**(3): p. 307-15.
15. Mechanic, D., *Improving the quality of health care in the United States of America: the need for a multi-level approach.* J Health Serv Res Policy, 2002. **7 Suppl 1**: p. S35-9.
16. Weingart, S.N., et al., *Epidemiology of medical error.* BMJ, 2000. **320**(7237): p. 774-7.
17. Andrews, L.B., et al., *An alternative strategy for studying adverse events in medical care.* Lancet, 1997. **349**(9048): p. 309-13.
18. Collins, T.C., et al., *Risk factors for prolonged length of stay after major elective surgery.* Ann Surg, 1999. **230**(2): p. 251-9.
19. Kane-Gill, S., et al., *Adverse-drug-event rates for high-cost and high-use drugs in the intensive care unit.* Am J Health Syst Pharm, 2006. **63**(19): p. 1876-81.
20. Rothschild, J.M., D.W. Bates, and L.L. Leape, *Preventable medical injuries in older patients.* Arch Intern Med, 2000. **160**(18): p. 2717-28.
21. Bartlett, G., et al., *Impact of patient communication problems on the risk of preventable adverse events in acute care settings.* CMAJ, 2008. **178**(12): p. 1555-62.
22. Wong, R.Y. and W.C. Miller, *Adverse outcomes following hospitalization in acutely ill older patients.* BMC Geriatr, 2008. **8**: p. 10.
23. Vincent, C., G. Neale, and M. Woloshynowych, *Adverse events in British hospitals: preliminary retrospective record review.* BMJ, 2001. **322**(7285): p. 517-9.
24. Wolff, A.M. and J. Bourke, *Detecting and reducing adverse events in an Australian rural base hospital emergency department using medical record screening and review.* Emerg Med J, 2002. **19**(1): p. 35-40.

25. Cho, S.H., et al., *The effects of nurse staffing on adverse events, morbidity, mortality, and medical costs.* Nurs Res, 2003. **52**(2): p. 71-9.
26. Zhan, C. and M.R. Miller, *Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization.* JAMA, 2003. **290**(14): p. 1868-74.
27. Mihaylova, B., et al., *Review of statistical methods for analysing healthcare resources and costs.* Health Econ, 2011. **20**(8): p. 897-916.
28. Rigby, K.D. and J.C. Litt, *Errors in health care management: what do they cost?* Qual Health Care, 2000. **9**(4): p. 216-21.
29. Ehsani, J.P., T. Jackson, and S.J. Duckett, *The incidence and cost of adverse events in Victorian hospitals 2003-04.* Med J Aust, 2006. **184**(11): p. 551-5.
30. Studdert, D.M., et al., *Claims, errors, and compensation payments in medical malpractice litigation.* N Engl J Med, 2006. **354**(19): p. 2024-33.
31. Marchetti, A., et al., *Costs and benefits of an early-alert surveillance system for hospital inpatients.* Curr Med Res Opin, 2007. **23**(1): p. 9-16.
32. Michelle M. Mello, D.M.S., Eric J. Thomas, and a.T.A.B. Catherine S. Yoon, *Who Pays for Medical Errors? An Analysis of Adverse Event Costs, the Medical Liability System, and Incentives for Patient Safety Improvement.* Journal of Empirical Legal Studies, 2007. **4**(4): p. 835–860.
33. Pappas, S.H., *The cost of nurse-sensitive adverse events.* J Nurs Adm, 2008. **38**(5): p. 230-6.
34. Jackson, T., *Cost estimates for hospital inpatient care in Australia: evaluation of alternative sources.* Aust N Z J Public Health, 2000. **24**(3): p. 234-41.
35. Jacobs, P. and J. Bachynsky, *Costing methods in the Canadian literature on the economic evaluation of health care. A survey and assessment.* Int J Technol Assess Health Care, 1996. **12**(4): p. 721-34.
36. Joffe, M.M. and P.R. Rosenbaum, *Invited commentary: propensity scores.* Am J Epidemiol, 1999. **150**(4): p. 327-33.
37. Braitman, L.E. and P.R. Rosenbaum, *Rare outcomes, common treatments: analytic strategies using propensity scores.* Ann Intern Med, 2002. **137**(8): p. 693-5.
38. Manca, A. and P.C. Austin, *Using propensity score methods to analyse individual patient-level cost-effectiveness data from observational studies*. 2008, Centre for Health Economics, The University of York,.

39. Marti, H., et al., *[Analysis with the propensity score of the association between likelihood of treatment and event of interest in observational studies. An example with myocardial reperfusion].* Rev Esp Cardiol, 2005. **58**(2): p. 126-36.
40. Vélez, M., *Uso de la puntuación de propensión (propensity score) en estudios no experimentales.* Iatreia 2013. **26**(1): p. 95-101.
41. D'Agostino, R.B., Jr., *Propensity scores in cardiovascular research.* Circulation, 2007. **115**(17): p. 2340-3.
42. Hall, J.A., K.H. Summers, and R.L. Obenchain, *Cost and utilization comparisons among propensity score-matched insulin lispro and regular insulin users.* J Manag Care Pharm, 2003. **9**(3): p. 263-8.
43. Baser, O., *Too much ado about propensity score models? Comparing methods of propensity score matching.* Value Health, 2006. **9**(6): p. 377-85.
44. Mansson, R., et al., *On the estimation and use of propensity scores in case-control and case-cohort studies.* Am J Epidemiol, 2007. **166**(3): p. 332-9.
45. Yang, G., *A review of propensity score application in healthcare outcome and epidemiology* in *PharmaSUG 2007*. http://www.lexjansen.com/pharmasug/2007/pr/pr02.pdf.
46. Rubin, D.B., *Estimating causal effects from large data sets using propensity scores.* Ann Intern Med, 1997. **127**(8 Pt 2): p. 757-63.
47. Cepeda, M.S., et al., *Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders.* Am J Epidemiol, 2003. **158**(3): p. 280-7.
48. Sturmer, T., et al., *A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods.* J Clin Epidemiol, 2006. **59**(5): p. 437-47.
49. Glynn, R.J., S. Schneeweiss, and T. Sturmer, *Indications for propensity scores and review of their use in pharmacoepidemiology.* Basic Clin Pharmacol Toxicol, 2006. **98**(3): p. 253-9.
50. Dubberke, E.R., et al., *Short- and long-term attributable costs of Clostridium difficile-associated disease in nonsurgical inpatients.* Clin Infect Dis, 2008. **46**(4): p. 497-504.

51. Zhao, Z., *Using matching to estimate treatment effects: Data requirements, Matching metrics, and Monte Carlo Evidence.* Review of Economics and Statistics., 2004. **86**: p. 91-107.
52. Baser, O., *Too Much Ado about Propensity Score Models? Comparing Methods of Propensity Score Matching.* V A L U E I N H E A L T H, 2006. **9**(6).
53. Love, T.E., *Using Propensity Score Methods Effectively*, in *Fall Workshop Cleveland Chapter of the American Statistical Association*. 2004.
54. Guo, S., *Running propensity score matching with STATA/PSMACH2*, in *Workshop conducted at the school of social work, UIUC*. 2005.
55. Lechner, M., *A note on the common support problem in applied evaluation studies. Discussion Paper, SIAW*. 2000.
56. Hansen, B.B., *Full matching in an observational study of coaching for the SAT.* Journal of the American Statistical Association, 2004. **99**: p. 609-618.
57. Rosenbaum, P.R., *Observational Studies*. 2002, New York: Springer.
58. Shadish, W.R., Cook, T.D. and Campbell, D.T. , *Experimental and quasiexperimental designs for generalized causal inference.* 2002, Boston: Houghton Mifflin.
59. Bryson A, R.D., and Susan P, *The use of propensity score matching in the evaluation of active labour market policies*, D.o.W.a. Pensions., Editor. 2002: London.
60. Charlson, M.E., et al., *A new method of classifying prognostic comorbidity in longitudinal studies: development and validation.* J Chronic Dis, 1987. **40**(5): p. 373-83.
61. Bernal Salazar, R., *El método de emparejamiento*, in *Guía práctica para la evaluación de impacto*, E. Uniandes, Editor. 2011: Bogotá.
62. Caliendo, M. and S. Kopeinig, *Some practical guidance for the implementation of propensity score matching*, in *Discussion paper no 1588*. 2005, IZA,: Bonn, Germany.
63. Cardona, S., *Comparacion entre los costos de no calidad y los costos de calidad de los eventos adversos en la atencion en el servicio de urgencias de la ese hospital marco fidel suarez de bello-antioquia.* , in *FACULTAD DE MEDICINA*. 2009, UNIVERSIDAD CES, http://bdigital.ces.edu.co:8080/repositorio/bitstream/10946/829/1/Comparacion_costos_calidad(1).pdf.

64. Ovalle, D., *Vigilancia, evitabilidad y costos asociados a los eventos adversos en la Clínica Carlos Lleras de Bogotá, Colombia 2009-2010.* Rev Med Hered. , 2013. **24**: p. 89.

65. Hoonhout, L.H., et al., *Direct medical costs of adverse events in Dutch hospitals.* BMC Health Serv Res, 2009. **9**: p. 27.

66. Fuller, R.L., et al., *Estimating the costs of potentially preventable hospital acquired complications.* Health Care Financ Rev, 2009. **30**(4): p. 17-32.

67. Kaushal, R., et al., *Costs of adverse events in intensive care units.* Crit Care Med, 2007. **35**(11): p. 2479-83.

68. Chu, E., et al., *Costs associated with capecitabine or 5-fluorouracil monotherapy after surgical resection in patients with colorectal cancer.* Oncology, 2009. **77**(3-4): p. 244-53.