Původní sdělení | Original research article

Proč používat propensity skóre v observačních studiích? Případová studie založená na datech české klinické databáze AHEAD 2006–2009

(Why to use propensity score in observational studies?
Case study based on data from the Czech clinical database AHEAD 2006–09)

Simona Littnerová^a, Jiří Jarkovský^b, Jiří Pařenica^{b,c,d}, Tomáš Pavlík^a,
Jindřich Špinar^{b,c,d}, Ladislav Dušek^a

^a Institut biostatistiky a analýz Lékařské a Přírodovědecké fakulty Masarykovy univerzity, Brno, Česká republika

^b Interní kardiologická klinika, Fakultní nemocnice Brno, Brno, Česká republika

^c Lékařská fakulta Masarykovy univerzity, Brno, Česká republika

^d Interní kardiologická klinika, Fakultní nemocnice u sv. Anny v Brně, Brno, Česká republika

INFORMACE O ČLÁNKU

Historie článku:

Došel do redakce: 24. 1. 2013

Přijat: 1. 4. 2013

Dostupný online: 11. 4. 2013

Klíčová slova:

AHEAD

Akutní srdeční selhání

Mortalita

Propensity skóre

SOUHRN

Randomizované klinické studie jsou považovány za zlatý standard pro medicínu založenou na důkazech, avšak tyto studie nemusejí být vždy proveditelné nebo etické a vědci musejí spoléhat na observační studie nebo výzkumné databáze. Nicméně získání spolehlivých výsledků z těchto studií vyžaduje odstranění potenciálního vlivu dalších faktorů. Naštěstí existuje několik statistických metod schopných identifikovat a snížit vliv dalších faktorů. Jednou z nich je metoda propensity skóre, která je v současné době často používána pro odhad relevantních klinických účinků, jež jsou adjustovány pro dané zkreslení. Tato práce si klade za cíl poskytnout stručný a praktický návod na použití propensity skóre prostřednictvím snadno srozumitelné případové studie. Případová studie je zaměřena na genderové rozdíly v úmrtnosti u pacientů s akutním srdečním selháním v české výzkumné databázi AHEAD (Acute Heart Failure Database).

© 2013, ČKS. Published by Elsevier Urban and Partner Sp. z o.o. All rights reserved.

ABSTRACT

Randomized clinical trials represent the gold standard of the evidence based medicine research; nevertheless they may not always be feasible or ethical and the researchers have to rely on observational studies or research databases. However, obtaining reliable results from these studies requires the elimination of potential influence of confounding factors. Fortunately, several statistical methods capable of identifying and reducing the impact of confounding factors exist. One of them is the propensity score which has been frequently used in recent times to estimate relevant clinical effects adjusted for given confounders. This work aims to provide a concise and practical guide to propensity scores by means of an easily understandable case study. The case study is focused on gender differences in mortality rates of patients with acute heart failure in the Czech research database AHEAD (Acute Heart Failure Database).

Keywords:

Acute heart failure

AHEAD

Mortality

Propensity score

Adresa: RNDr. Simona Littnerová, Institut biostatistiky a analýz Lékařské a Přírodovědecké fakulty Masarykovy univerzity, Kamenice 126/3, 625 00 Brno, e-mail: littnerova@iba.muni.cz

DOI: 10.1016/j.crvasa.2013.04.001

Úvod

Medicína založená na důkazech (evidence based medicine – EBM) by měla být hlavním metodickým přístupem a základním kamenem současného klinického výzkumu a praxe; hraje klíčovou roli při posuzování přínosů a rizik terapie, stejně jako diagnostické testy [1]. Za účelem získání nezaujatých statistických výsledků je nezbytné odlišit faktory odpovědné za kauzální příčiny od proměnných, které ukazují pouze náhodnou korelaci se sledovaným cílovým parametrem [2,3].

Randomizované klinické studie jsou obecně považovány za zlatý standard pro EBM. Počáteční randomizace v těchto studiích zvyšuje pravděpodobnost, že účinek léčby na hodnocený parametr je dán léčbou samotnou, protože ostatní vlivy (jako jsou struktura věku a pohlaví, komorbidita...) jsou náhodně rozděleny mezi srovnávanými skupinami pacientů. Bohužel tyto studie nemusejí být vždy možné či etické nebo použitelné pro reálné životní situace v klinické praxi [2].

V těchto situacích nám mohou pomoci observační studie, které poskytnou dostatečné důkazy; jejich výhodou ve srovnání s randomizovanými klinickými studiemi je schopnost zachytit skutečný stav zdravotní péče, praxe a výsledky, které lze extrapolovat na úroveň populace. Protože předpoklad náhodného rozdělení ovlivňujících faktorů, díky kterému je léčebný účinek nezávislý na dalších faktorech, není vždy sám o sobě splněn, vyžaduje získání dostatečně objektivních výsledků z těchto studií sofistikované statistické metody [3,4].

V dnešní době existuje několik statistických metod pro efektivní práci s několika vzájemně korelujícími proměnnými. Jednou z těchto metod je propensity skóre, metoda zavedená v roce 1983, kdy Rosenbaum a Rubin definovali

propensity skóre na jako „podmíněnou pravděpodobnost zařazení do určité léčby na daném vektoru pozorovaných proměnných“ [5]. Jinými slovy, výsledek vyjadřuje pacientovu pravděpodobnost, že náleží do stejné srovnatelné populace na základě jejich popisných charakteristik, tj. zavádějících faktorů. Hlavním cílem této metody je získat nezaujatý odhad účinku léčby adjustovaný na vliv daných zavádějících faktorů v nerandomizovaných a observačních studiích [3,4,6]. Díky svým vlastnostem se propensity skóre stalo široce používanou metodou v široké škále epidemiologických studií; počet publikací, ve kterých bylo propensity skóre použito, vzrostl z 294 publikací v letech 1998–2002 na 1 111 v roce 2003–2007, a v roce 2008–2012 dokonce na 3 539.

Ačkoliv několik publikací detailně popsalo metodiku propensity skóre [4,6], jiné poskytují příklad analýzy simulovaných dat [7] nebo systematický přehled literatury [8,9], jen málo publikací se zabývá příklady uplatnění propensity skóre na reálná data. Cílem této práce je vysvětlit principy a výhody propensity skóre klinickým výzkumným pracovníkům. Kromě teoretických základů výpočtu propensity skóre práce obsahuje analýzu vlivu genderových rozdílů na mortalitu pacientů s akutním srdečním selháním, která zároveň poslouží jako vzdělávací případové studie (na základě akutní databáze srdečního selhání – AHEAD 2006–2009 [10,11]).

Propensity skóre

Výhodou propensity skóre ve srovnání s vícerozměrnou adjustací je odstranění vlivu zavádějících faktorů z analýzy léčebného účinku [6]. Rosenbaum a Rubin [5] definovali propensity skóre jako podmíněnou pravděpodobnost

Chtěli bychom získat objektivní výsledky ze srovnání dvou skupin pacientů (léčení vs. kontrolní), které se liší v genderové struktuře. Toto je typický příklad aplikace propensity skóre.

Výpočet propensity skóre pomocí kontingenčních tabulek sestává z následujících kroků:

Máme 60 mužů, kteří jsou léčeni ($Z = 1$). Pak pravděpodobnost, že pacient je léčený muž $P(Z = 1 \text{ a } \text{Muž})$, je 60/200 a pravděpodobnost, že pacient je muž $P(\text{Muž})$, je 100/200.

Podmíněná pravděpodobnost (propensity skóre) pro léčeného muže je dána následujícím vztahem:

$$P(Z = 1 | \text{Muž}) = \frac{P(Z = 1 \text{ a } \text{Muž})}{P(\text{Muž})} = \frac{\frac{60}{200}}{\frac{100}{200}} = 0,6$$

Propensity skóre pro neléčeného muže:

$$P(Z = 0 | \text{Muž}) = \frac{P(Z = 0 \text{ a } \text{Muž})}{P(\text{Muž})} = \frac{\frac{40}{200}}{\frac{100}{200}} = 0,4$$

	Skupina léčených ($Z = 1$)	Skupina neléčených ($Z = 0$)	Celkem
Muž	60	40	100
Žena	20	80	100
Celkem	80	120	200

Propensity skóre pro léčenou ženu je dáno stejným výpočtem:

$$P(Z = 1 | \text{Žena}) = \frac{P(Z = 1 \text{ a } \text{Žena})}{P(\text{Žena})} = \frac{\frac{20}{200}}{\frac{100}{200}} = 0,2$$

Propensity skóre pro neléčenou ženu:

$$P(Z = 0 | \text{Žena}) = \frac{P(Z = 0 \text{ a } \text{Žena})}{P(\text{Žena})} = \frac{\frac{80}{200}}{\frac{100}{200}} = 0,8$$

Výpočet propensity skóre pomocí kontingenčních tabulek má většinou edukační význam, jeho praktické použití je však omezené. Propensity skóre pro efektivní adjustaci na zavádějící faktory by mělo zahrnovat více než jednu proměnnou a/nebo spojitou proměnnou. Obr. 2A a 2B zobrazují složitější výpočet propensity skóre, který je použitelný v praxi a využívá přístup logistické regrese.

Obr. 1 – Výpočet propensity skóre pomocí kontingenčních tabulek

Tabulka 1 – Pozorované proměnné před vyvažováním a po něm

N (%) / medián (5.; 95. percentil)	Před vyvažováním			Po vyvažování		
	Ženy N = 1 757	Muži N = 2 396	<i>p</i> ¹	Ženy N = 1 128	Muži N = 1 128	<i>p</i> ¹
Věk ⁺	78 (55; 90)	69 (46; 86)	< 0,001	75 (52; 87)	75 (54; 87)	0,917
BMI	27,6 (20,3; 38,3)	27,7 (21,6; 37,8)	0,056	27,7 (20,5; 38,3)	27,7 (22,1; 37,5)	0,621
Systolický TK	140 (85; 210)	130 (80; 200)	< 0,001	140 (90; 210)	140 (90; 200)	0,050
Diastolický TK	80 (50; 110)	80 (50; 110)	0,551	80 (50; 113)	80 (55; 110)	0,180
Srdeční rytmus	89 (50; 145)	90 (55; 140)	0,154	90 (55; 145)	90 (55; 140)	0,209
Ejekční frakce ⁺	42 (20; 65)	35 (15; 62)	< 0,001	40 (20; 65)	40 (18; 65)	0,367
Clearance kreatininu ⁺	46,8 (18,6; 83,1)	54,2 (22,0; 91,8)	< 0,001	49,7 (21,9; 86,8)	50,5 (21,1; 84,8)	0,824
ASS						
<i>De novo</i>	1 049 (60,9 %)	1 320 (56,4 %)	0,004	683 (60,5 %)	558 (49,5 %)	< 0,001
AKS	585 (33,3 %)	888 (37,1 %)	0,013	400 (35,5 %)	396 (35,1 %)	0,860
ASS syndromy						
ADSS	905 (51,5 %)	1 335 (55,7 %)	0,008	620 (55,0 %)	668 (59,2 %)	0,051
Plicní edém	330 (18,8 %)	418 (17,4 %)	0,265	220 (19,5 %)	200 (17,7 %)	0,279
Kardiogenní šok	234 (13,3 %)	366 (15,3 %)	0,077	132 (11,7 %)	128 (11,3 %)	0,792
NYHA						
I	367 (20,9 %)	585 (24,4 %)	< 0,001	263 (23,3 %)	259 (23,0 %)	0,213
II	569 (32,4 %)	629 (26,3 %)		375 (33,2 %)	316 (28,0 %)	
III	541 (30,8 %)	757 (31,6 %)		364 (32,3 %)	422 (37,4 %)	
IV	184 (10,5 %)	271 (11,3 %)		126 (11,2 %)	131 (11,6 %)	
Anamnéza						
Hypertenze	1 306 (78,2 %)	1 569 (69,3 %)	< 0,001	859 (76,4 %)	849 (75,3 %)	0,546
Diabetes mellitus ⁺	774 (46,2 %)	902 (39,8 %)	< 0,001	507 (44,9 %)	516 (45,7 %)	0,703
Infarkt myokardu ⁺	448 (26,7 %)	815 (36,0 %)	< 0,001	344 (30,5 %)	340 (30,1 %)	0,855
PCI nebo CABG	7 (0,4 %)	14 (0,6 %)	0,399	5 (0,4 %)	7 (0,6 %)	0,562
CMP nebo TIA	302 (18,0 %)	351 (15,5 %)	0,038	186 (16,5 %)	189 (16,8 %)	0,865
Fibrilace nebo flutter	489 (27,8 %)	593 (24,7 %)	0,025	299 (26,5 %)	337 (29,9 %)	0,075

ADSS – akutní dekompenzované srdeční selhání; AKS – akutní koronární syndrom; ASS – akutní srdeční selhání; CABG – aortokoronární bypass; CMP – cévní mozková příhoda; IM – infarkt myokardu; PCI – perkutánní koronární intervence; TIA – transitorní ischemická ataka; TK – tlak krve.

¹ Statistická významnost rozdílů mezi muži a ženami, spojité proměnné byly testovány pomocí Mannova-Whitneyho U-testu, kategoriální pomocí ML χ^2 testu.

⁺ Parametry zahrnuty do modelu propensity skóre.

pro pacienta, že bude léčen (exponován) na základě daného vektoru pozorovaných proměnných. Předpokládáme, že ke každému pacientovi jsou přiřazeny pozorované proměnné $\mathbf{X} = (X_1, X_2, \dots, X_n)$ a ukazatel léčebné skupiny Z ($Z = 1$, pokud je pacient léčen, a $Z = 0$, pokud je kontrola). Propensity skóre, zde označené jako PS, je definováno pro každého pacienta jako pravděpodobnost, že on/ona podstoupil/a léčbu ($Z = 1$), vzhledem k jeho/jejím pozorovaným proměnným $PS = P(Z = 1 | \mathbf{X})$. Pravděpodobnost, že bude pacient léčen, může nabývat u každého pacienta hodnot od 0 do 1, tj. $0 < PS < 1$. Teorie propensity skóre závisí na dvou hlavních předpokladech: 1. sledované proměnné neovlivňují lékařovo rozhodnutí, jak bude s pacientem nakládáno (např. pohlaví/věk pacientů nemá vliv na podávání léčby), 2. nejsou zde žádné nezměřené zkreslující jevy (všechny proměnné potenciálně související s přiřazením k léčbě jsou známy).

V observačních studiích proměnné obvykle nejsou vyváženy mezi oběma skupinami. Rosenbaum a Rubin

[5] ukázali, že pozorované proměnné jsou vyváženy na každou hodnotu propensity skóre, to znamená, že pacienti v experimentální a kontrolní skupině se stejnou hodnotou propensity skóre mají stejné rozdělení pozorovaných proměnných.

Propensity skóre může být chápáno jako zástupce vztahu mezi případy a proměnnými, které ovlivňují expozici, takže jej lze pro zjednodušení analýzy použít místo dalších analýz. Na základě toho propensity skóre jako zástupce proměnných agreguje více zavádějících faktorů do jednoho rozměru [4,5].

Nejjednodušším způsobem, jak odhadnout propensity skóre, je jednoduchá analýza kontingenčních tabulek, kde pro všechny vstupní proměnné získáme poměr pacientů s léčbou a bez ní (obr. 1); nicméně tento přístup není vhodný pro spojité proměnné. Chceme-li tento problém vyřešit, existuje spousta dalších parametrických a neparametrických metod pro odhad propensity skóre. Například D'Agostino [4] používá diskriminační

Chtěli bychom získat objektivní výsledky ze srovnání dvou skupin pacientů (léčení vs kontrolní), které se liší v genderové struktuře a věku. Toto je typický příklad aplikace propensity skóre. Výpočet sestává z následujících kroků:

Obr. 2 – (A) Výpočet propensity skóre pomocí logistické regrese

analýzu pro odhad propensity skóre, zatímco McCaffrey a spol. publikovali článek, kde pro výpočet propensity skóre byly použity zobecněné „boosted modely“ [12]; Setoguchi a spol. publikovali simulační studii, kde bylo propensity skóre odhadnuto pomocí neuronových sítí a klasifikačních stromů. Nicméně tyto metody mají významné nevýhody, jako je jejich značná složitost a problémy s následnou interpretací, zároveň je jejich použití spojeno se značnými komplikacemi z důvodů složitých algoritmů a jejich realizací [12]. Z těchto důvodů se nejčastějším nástrojem pro odhad propensity skóre stala logistická regrese [4,9,13] a ta byla také použita v naší práci.

Logistická regrese je statistická metoda, která odhaduje pravděpodobnost výskytu události (závislá proměnná) na základě známých faktorů (nezávislé proměnné), u kterých se očekává, že mají vliv na výskyt události (pacient je léčen v případě výpočtu propensity skóre). Závislá proměnná proto může nabývat dvou hodnot v závislosti na tom, zda k události (léčbě) došlo nebo ne. Výsledkem logistické regrese je pravděpodobnost vzniku události (v rozmezí od 0 do 1) na základě

hodnoty nezávislých proměnných. Dále je pro většinu vědců logistická regrese známá a poměrně snadno pochopitelná metoda a je aplikována ve všech základních statistických softwarech (SPSS, Stata, Statistica, SAS, R...) [13].

Výsledky logistické regrese pak mohou být použity k výpočtu propensity skóre podle následujícího vzorce:

$$PS = P(Z = 1 | X) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}$$

kde x_1, x_2 až x_n jsou nezávislé proměnné (v příkladu na obr. 2A jde o pohlaví a věk), β_0, β_1 až β_n jsou odpovídající regresní koeficienty, koeficient β_0 představuje vliv absolutní složky (hodnota pravděpodobnosti, když všechny nezávislé proměnné jsou rovny nule). Parametry β_0, β_1 až β_n jsou odhadovány pomocí metody maximální věrohodnosti na základě pozorovaných dat [14], jak je znázorněno na obr. 2A. Po výpočtu se propensity skóre používá k odvození pseudorandomizovaného datového souboru, který nám umožňuje získat nezaujatý odhad efektu expozice (léčby).

Propensity skóre získané z modelu logistické regrese je použito k párování podobných pacientů ze skupiny léčených a ze skupiny neléčených tak, aby bylo dosaženo vyváženého datového souboru. Výpočet se skládá z následujících kroků:

Obr. 2 – (B) Párování pomocí propensity skóre

Aplikace propensity skóre

Hlavní a nejčastější využití propensity skóre před kvantifikací účinku léčby je vytvoření vyváženého datového souboru, který je také popsán v našem textu. Existují další teoretické přístupy (např. stratifikace a nastavení), ale ty se používají méně často [4,5,9,13].

Vyvažování je technika, jež vytváří vyvážený datový soubor, který obsahuje páry z kontrolních a léčených pacientů na základě podobné hodnoty jejich propensity skóre. Avšak jeho hlavní nevýhodou je skutečnost, že někteří pacienti mohou být vyřazeni z datového souboru. Vyvažování spočívá ve výběru léčených pacientů a nalezení kontrolního pacienta k nim se stejnou nebo téměř stejnou hodnotou propensity skóre (obr. 2B). Aby nedošlo k párování u pacientů s velmi odlišnými hodnotami propensity skóre, používá se měřítko, které udává rozsah hodnot propensity skóre, v němž může dojít k vyrovnání hodnot propensity skóre. Tento postup může vést k situaci, kdy získáme nepárové pacienty, kteří jsou vyloučeni z následných analýz.

Účinnost modelu propensity skóre se hodnotí podle absolutního standardizovaného rozdílu v jednotlivých proměnných mezi skupinami pacientů po vyrovnání. Absolutní standardizovaný rozdíl 0 % pro proměnnou znamená nulové reziduální zkreslení pro tuto proměnnou a absolutní standardizovaný rozdíl < 10 % znamená nevýznamné reziduální zkreslení [15].

Vyvažování na základě propensity skóre se často používá v lékařské literatuře [8,9,16]. K popularitě propensity skóre přispívá několik vlastností; ve srovnání s jinými přístupy může vyvažování při odhadování přesnějšího léčebného účinku odstranit větší část zkreslení [17]; vyvažování pomocí propensity skóre vytváří vyvážený datový soubor, který umožňuje jednoduché a přímé porovnání závislých proměnných mezi léčenými a kontrolními pacienty. Z těchto důvodů jsme jako praktický příklad nejběžnějšího použití propensity skóre zvolili případovou studii vyvažování na jeho základě. Na druhé straně bychom neměli zapomínat, že vyvážený soubor dat, získaný pomocí vyvažování, neobsahuje všechny pacienty, a proto jsou některé informace z původního datového souboru ztraceny.

Obr. 3 – Rozložení věku dle pohlaví před vyvažováním a po něm na základě propensity skóre; struktura věku je po vyvažování stejná pro muže i ženy.

Obr. 4 – Absolutní standardizovaný rozdíl (%) pozorovaných proměnných mezi muži a ženami před vyvažováním a po něm na základě propensity skóre (založeno na věku, ejekční frakci, diabetes mellitus, infarktu myokardu, anamnéze a clearance kreatininu); proměnné s absolutním standardizovaným rozdílem < 10 % po vyvažování jsou považovány za úspěšně vyvážené.

Případová studie aplikace propensity skóre ve sledování úmrtnosti u pacientů s akutním srdečním selháním (projekt AHEAD 2006–2009)

Následující text vysvětluje, jak funguje propensity skóre. V tomto příkladě jsme zvolili genderové rozdíly a jejich vliv na mortalitu pacientů. Je známo, že v populaci se srdečním selháním jsou muži mladší než ženy a ženy mají nižší úmrtnost než muži [18–21]. Aplikace propensity skóre nám umožňuje získat vyvážený datový soubor a tímto získat přesnější odhad rozdílu v úmrtnosti pacientů mezi pohlavími (cílový parametr). V této studii pohlaví představuje indikátor léčby zavedený v teoretické části práce ($Z = 1$, pokud je pacient muž, a $Z = 0$, pokud je žena).

Údaje analyzované v tomto příkladu pocházejí z výzkumné databáze o akutním srdečním selháním AHEAD Main. Databáze AHEAD Main obsahuje konsektivní pacienty ze sedmi center s 24hodinovou katetizační laboratoří a s centralizovanou péčí o pacienty s akutním koronárním syndromem (AKS) z oblasti, která má více než tři miliony obyvatel [10,11]. Celý datový soubor obsahuje 4 153 pacientů, z nichž je 1 757 (42,3 %) žen a 2 396 (57,7 %) mužů. Naše výsledky potvrdily, že muži s akutním srdečním selháním jsou mladší než ženy (obr. 3, tabulka 1), což je v souladu s literaturou [18–21]. Kromě toho se muži od žen liší v několika dalších parametrech (tabulka 1).

Výše popsaná metodika propensity skóre byla použita na tato primární data. Před výpočtem propensity skóre pomocí logistické regrese (stejně jako v příkladu na obr. 2A) jsme museli zvolit proměnné, které se liší mezi pohlavími, a mohou tak představovat zavádějící faktory pro analyzovaný cílový parametr. První krok výběru proměnných, který sestával z identifikace důležitých vstupních proměnných, jež ovlivňují úmrtnost a diskriminují pohlaví, byl založen na kombinaci výsledků z analýzy dat, odborných znalostí a publikovaných výsledků. Všechny nezávislé proměnné uvedené v tabulce 1 byly použity v dopředné selekci pro výběr proměnných, které jsou nezbytné pro sestavení modelu. Konečný propensity skóre

Obr. 5 – Mortalita po propuštění před vyvažováním (originální datový soubor) pro muže versus ženy v klinické databázi AHEAD. Nebyl zde pozorován statisticky významný rozdíl v mortalitě mezi pohlavími ($p = 0,239$).

Obr. 6 – Mortalita po propuštění po vyvažování (vyvážený datový soubor) pro muže versus ženy v klinické databázi AHEAD. Byl zde pozorován statisticky významný rozdíl v mortalitě mezi pohlavími ($p = 0,002$).

model se skládal ze čtyř proměnných: věk, ejekční frakce (EF), diabetes mellitus (DM) a infarkt myokardu (IM) v anamnéze a clearance kreatininu při přijetí (CrCl), konečná regresní rovnice má tvar

$$PS = P(Z = 1 \mid \text{věk, EF, DM, IM, CrCl}) = \frac{\exp(3,94 - 0,04\text{věk} - 0,03\text{EF} - 0,18\text{DM} - 0,47\text{IM} - 0,01\text{CrCl})}{1 + \exp(3,94 - 0,04\text{věk} - 0,03\text{EF} - 0,18\text{DM} - 0,47\text{IM} - 0,01\text{CrCl})}$$

Propensity skóre bylo pak použito na vyvažování na základě měřítka $0,01 \cdot \text{standardní odchylka propensity skóre}$, aby vyvážilo strukturu proměnných, které se pro ženy a pro muže liší. Konečný vyvážený datový soubor tvořilo 2 256 pacientů (54,3 % původního datového souboru): 1 128 (50 %) mužů a 1 128 (50 %) žen. Pokles počtu pacientů ve vyváženém datovém souboru vedl k posunu věkové struktury, celkem pacienti ve vyváženém datovém souboru byly starší než v původním datovém souboru, a proto výsledky získané z analýz vyváženého datového souboru by měly být interpretovány pouze pro tyto starší pacienty (obr. 3).

Schopnost propensity skóre efektivně vyvážit skupiny podle vstupních proměnných je potvrzena kombinací výsledků z tabulky 1 a z obr. 4. Tabulka 1 ukazuje, že téměř všechny nezávislé proměnné, které vstoupily do modelu propensity skóre, již nejsou v porovnávání vzorku významné. Absolutní standardizované rozdíly v proměnných mezi oběma pohlavími po párování jsou uvedeny na obr. 4. Z důvodu omezené složitosti modelu propensity skóre nám zbyly proměnné, které se významně lišily mezi pohlavími (tabulka 1, obr. 4). Modely s více proměnnými jsou lépe vyvážené, ale mají za následek menší vyvážený datový soubor, a proto jsou méně reprezentativní pro původní data.

De novo akutní srdeční selhání jako zůstávající významná proměnná mělo jen omezený vliv na výsledek

(jeho vztah k mortalitě byl statisticky nevýznamný). Získali jsme datový soubor skutečně vyvážený na věk, DM, CrCl, EF a předchozí infarkt myokard v anamnéze. Tyto faktory se lišily mezi pohlavími a jsou známy jako nejdůležitější ve vztahu k mortalitě [10,11].

Hlavním cílem naší analýzy bylo srovnání mortality mezi muži a ženami. V původním datovém souboru rozdíl v mortalitě mezi muži a ženami nebyl statisticky významný (obr. 5): pětiletá mortalita byla 58,0 % u mužů a 60,4 % u žen (log-rank test $p = 0,239$). Ve vyváženém datovém souboru však pětiletá mortalita byla 63,1 % u mužů a 56,8 % u žen: tento rozdíl byl statisticky významný ($p = 0,002$) (obr. 6). Ke stejnému výsledku dospěli v norském registru srdečního selhání [20], kde ženy měly významně nižší mortalitu.

Závěr

Propensity skóre se dnes široce používá při statistickém zpracování dat, zejména v analýze observačních klinických studií. Je pravděpodobné, že vzhledem ke stoupajícímu počtu observačních výzkumných databází a dostupnosti jejich dat se jeho použití ještě zvýší [13]. Hlavní výhodou propensity skóre je, že vede k přesnějšímu odhadu léčebné odpovědi. Proto se v současné době propensity skóre doporučuje jako standardní nástroj pro výzkumníky, kteří se snaží odhadnout dopady léčby ve studiích, v nichž mohou existovat případná zkreslení. Další slibnou aplikací propensity skóre je oblast hodnocení zdravotnických technologií při porovnávání různých zdravotnických zařízení nebo poskytovatelů zdravotní péče. Tyto jednotky obvykle pracují na zcela jiném vzorku pacientů a propensity skóre představuje efektivní přístup k získání vzájemně srovnatelných skupin pacientů z reálných klinických dat [22].

V naší případové studii použití vyvažování pomocí propensity skóre odhalilo významné rozdíly v mortalitě mezi oběma pohlavími. To je způsobeno tím, že získaný vyvážený datový soubor podle propensity skóre byl vyvážen na zavádějící faktory, jinými slovy byla odstraněna většina zkreslení ovlivňující mortalitu. Tímto jsme získali přesnější odhady mortality u obou pohlaví.

Poděkování

Děkujeme výzkumným pracovníkům klinické databáze AHEAD za jejich příspěvek ke studii. Participující centra a řešitelé: Česká republika: Fakultní nemocnice Brno, Brno: Roman Miklík, Marián Felšöci, Kateřina Horáková; Fakultní nemocnice u sv. Anny, Brno: Jiří Vitovec, Lenka Špinarová, Tereza Mikušová, Klauďia Židová; Kardiocentrum, Fakultní nemocnice Královské Vinohrady, Praha: Petr Widimský, Filip Roháč, Richard Fojt; Všeobecná fakultní nemocnice, Praha: Aleš Linhart, Jan Bělohávek; Nemocnice Na Homolce, Praha: Filip Málek, Petr Ošťádal, Andreas Krüger; Fakultní nemocnice Olomouc, Olomouc: Marián Fedorco, Jan Václavík, Čestmír Čihalík, David Vindiš; Institut klinické a experimentální medicíny, Praha: Jiří Kettner; Nemocnice T. Bati, Zlín: Zdeněk Coufal, Petr Hrdý, Stanislava Peňasová, Miroslav Bambuch; Nemocnice Frýdek-Místek: Kamil Zeman.

Tato práce byla podpořena grantovým projektem GACR No. 102/09/H083 Masarykovy univerzity [MUNI/IA/0828/2011], Evropským regionálním rozvojovým fondem – Projekt Mezinárodního centra klinického výzkumu, Fakultní nemocnice u sv. Anny, Brno, Česká republika (No. CZ.1.05/1.1.00/02.0123) a projektem Ministerstva zdravotnictví České republiky pro koncepční rozvoj výzkumné organizace 65269705 (Fakultní nemocnice Brno, Brno, Česká republika).

Literatura

- [1] A.S. Elstein, On the origins and development of evidence-based medicine and medical decision making, *Inflammation Research* 53 (2) (2004) S184–S189.
- [2] W.G. Cochran, D.B. Rubin, Controlling bias in observational studies: a review, *Sankhyā: The Indian Journal of Statistics, Series A* 35 (1973) 417–446.
- [3] G. Biondi-Zoccai, E. Romagnoli, P. Agostoni, et al., Are propensity scores really superior to standard multivariable analysis?, *Contemporary Clinical Trials* 32 (5) (2011) 731–740.
- [4] R.B. D'Agostino Jr., Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group, *Statistics in Medicine* 17 (1998) 2265–2281.
- [5] P.R. Rosenbaum, D.B. Rubin, The central role of the propensity score in observational studies for causal effects, *Biometrika* 70 (1) (1983) 41–55.
- [6] R.B. D'Agostino Jr., Propensity scores in cardiovascular research, *Circulation* 115 (2007) 2340–2343.
- [7] S. Setoguchi, S. Schneeweiss, M.A. Brookhart, et al., Evaluating uses of data mining techniques in propensity score estimation: a simulation study, *Pharmacoepidemiology and Drug Safety* 17 (6) (2008) 546–555.
- [8] E. Gayat, R. Pirracchio, M. Resche-Rigon, et al., Propensity scores in intensive care and anaesthesiology literature: a systematic review, *Intensive Care Medicine* 36 (12) (2010) 1993–2003.
- [9] S. Weitzen, K.L. Lapane, A.Y. Toledano, et al., Principles for modeling propensity scores in medical research: a systematic literature review, *Pharmacoepidemiology and Drug Safety* 13 (12) (2004) 841–853.
- [10] J. Spinar, J. Parenica, J. Vitovec, et al., Baseline characteristics and hospital mortality in the Acute Heart Failure Database (AHEAD) Main registry, *Critical Care* 15 (6) (2011) R291.
- [11] J. Parenica, J. Spinar, J. Vitovec, et al., Long-term survival following acute heart failure: The Acute Heart Failure Database Main registry (AHEAD Main), *European Journal of Internal Medicine* 2012 <http://dx.doi.org/10.1016/j.ejim.2012.11.005>
- [12] D.F. McCaffrey, G. Ridgeway, A.R. Morral, Propensity score estimation with boosted regression for evaluating causal effects in observational studies, *Psychological Methods* 9 (4) (2004) 403–425.
- [13] D. Westreich, J. Lessler, M.J. Funk, Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression, *Journal of Clinical Epidemiology* 63 (8) (2010) 826–833.
- [14] D.W. Hosmer, S. Lemeshow, *Applied Logistic Regression*, John Wiley & Sons, New York, 2004.
- [15] P.C. Austin, Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples, *Statistics in Medicine* 28 (25) (2009) 3083–3107.
- [16] P.C. Austin, The performance of different propensity-score methods for estimating relative risks, *Journal of Clinical Epidemiology* 61 (6) (2008) 537–545.
- [17] P.C. Austin, Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations, *Biometrical Journal* 51 (1) (2009) 171–184.
- [18] K.K. Ho, J.L. Pinsky, W.B. Kannel, D. Levy, The epidemiology of heart failure: the Framingham Study, *Journal of the American College of Cardiology* 22 (Suppl-13A) (1993).
- [19] M. Galvao, J. Kalman, T. Demarco, et al., Gender differences in in-hospital management and outcomes in patients with decompensated heart failure: analysis from the Acute Decompensated Heart Failure National Registry (ADHERE), *Journal of Cardiac Failure* 12 (2) (2006) 100–107.
- [20] B. Brandsaeter, D. Atar, S. Agewall, Gender differences among Norwegian patients with heart failure, *International Journal of Cardiology* 146 (3) (2011) 354–358.
- [21] J. Spinar, L. Spinarova, Gender differences in acute heart failure, *Future Cardiology* 5 (2) (2009) 109–111.
- [22] P.A. Kurlansky, M. Argenziano, R. Dunton, et al., Quality, not volume, determines outcome of coronary artery bypass surgery in a university-based community hospital network, *Journal of Thoracic and Cardiovascular Surgery* 143 (2) (2012) 287–293.

Z anglického originálu přeložila autorka.