

مقایسه روش شناسی تحلیل مورد- شاهدی لانه گزیده و هم گروهی بر روی داده‌های مربوط به بیماری سل شهرستان: یک تجربه

کوروش هلاکویی نائینی^۱، افشین استوار^۲، احمد دانش^۲، سیما ساجدی نژاد^۲، لیلا قالیچی^۲، قباد مرادی^۲، محمدعلی منصورنیا^۲، سید سعید هاشمی نظری^۲

^۱ استاد اپیدمیولوژی، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران، دانشکده بهداشت، گروه اپیدمیولوژی و آمار، ایران

^۲ دانشجوی دوره دستیاری اپیدمیولوژی، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران، دانشکده بهداشت، گروه اپیدمیولوژی و آمار، تهران، ایران

نویسنده رابط: محمدعلی منصورنیا، نشانی: تهران، خیابان پورسینا، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران، دانشکده بهداشت، گروه اپیدمیولوژی و آمارزستی،

تلفن: ۰۹۱۲۱۵۰۶۴۳۱، شماره: ۸۸۹۸۹۱۲۷ پست الکترونیک: mansournia_ma@yahoo.com

تاریخ دریافت: ۱۳۸۷/۱۱؛ پذیرش: ۱۳۸۹/۳/۱

مقدمه و اهداف: مطالعه مورد شاهدی لانه گزیده به عنوان جایگزینی کارا برای مطالعه کوهورت دارای محبوبیت است. هدف از این مطالعه مشخص کردن نحوه نمونه گیری مورد - شاهدی لانه گزیده از مجموعه داده‌های یک کوهورت در یک نمونه از داده‌های نظام سلامت کشور و مقایسه برآوردهای شاخص رابطه و واریانس آن‌ها در آنالیزهای مناسب برای هر یک از داده‌ها است.

روش کار: مطالعه حاضر بر روی نمونه‌ای از داده‌های دفتر سل شهرستان‌های یک دانشگاه در یک فاصله زمانی ۱۳۸۱ تا ۱۳۸۴ (۲۷۶ نفر) انجام شد. در تجزیه و تحلیل با رویکرد همگروهی از روش Cox Regression استفاده شد. برای انجام تجزیه و تحلیل با رویکرد مورد شاهدی لانه گزیده، در زمان بروز هر مورد مرگ، سه نمونه تصادفی از بین کسانی که دچار بی‌آمد نشده‌اند انتخاب شد و داده‌ها با روش Conditional logistic Regression تجزیه و تحلیل شد.

نتایج: نتایج تجزیه و تحلیل هر دو رویکرد نشان داد که جنسیت، محل سکونت و سن بر روی رخداد مرگ تأثیری نداشته و متغیر گروه درمانی تنها متغیری است که بر روی رخداد مرگ تأثیر معنی‌داری دارد. برای متغیرهای دو حالتی و با اثر ناچیز مانند جنسیت و محل سکونت، کارایی نسبی مطالعه مورد- شاهد لانه گزیده در حدود ۷۵ درصد است.

نتیجه‌گیری: نتایج تجربی مطالعه حاضر نشان می‌دهد که مطالعه مورد- شاهد لانه گزیده علاوه بر راحت و مقرون به صرفه بودن، دارای اعتباری قابل قیاس با آنالیز کوهورت اصلی در برآورد rate ratio و واریانس آن است.

واژگان کلیدی: آنالیز همگروهی، مورد شاهد لانه گزیده، سل

مقدمه

انجام شده به بررسی تفاوت نتایج بدست آمده از دو نوع رویکرد در تجزیه و تحلیل داده‌های موجود در سیستم بهداشتی پرداخته شده است. این دو رویکرد همگروهی (cohort) و مورد شاهدی لانه گزیده (nested case-control) بر روی داده‌های سل بخشی از کشور ارائه شده و نتایج آن‌ها با هم مقایسه شده است در ادامه، خلاصه‌ای از وضعیت کشور از نظر بیماری سل ارائه گردیده است تا اهمیت انتخاب این داده‌ها به منظور بررسی بیشتر روشن شود.

اگرچه در مجموع شیوع بیماری سل در ایران نسبت به بعضی از مناطق جهان بالا نیست (۲۲ در صد هزار نفر جمعیت)، ولی بعضی از مناطق ایران از شیوع نسبتاً بالایی برخوردار هستند. در این خصوص می‌توان به هفت استان شرقی و جنوبی کشور اشاره کرد

در سیستم متولی سلامت کشور ما داده‌های بسیار متنوعی تولید می‌شود. آمار برنامه‌های جاری سلامت از قبیل سل، مالاریا، تنظیم خانواده، مراقبت از مادران باردار و بسیاری دیگر از فعالیت‌های جاری در قالب بانک‌های اطلاعاتی در سطح مراکز بهداشت شهرستان و دانشگاه‌ها جمع‌آوری می‌شود. در بسیاری از این موارد فقط اطلاعات خاصی از این داده‌ها استخراج گشته و بخش دیگری بدون استفاده باقی می‌مانند و در نتیجه سیستم سلامت از امکان بهره‌گیری از نتایج چنین مطالعاتی محروم می‌ماند. در پاره‌ای از موارد هم که بررسی بر روی این داده‌ها انجام می‌گیرد نتایج حاصل به دلیل هدفمند نبودن تجزیه تحلیل داده‌ها دارای سطوح کیفی مطلوبی نیست. در مطالعه حاضر که بر روی نمونه‌ای از این داده‌ها

که بروز بیماری در این مناطق از ۱۴/۷ در صد هزار در خراسان شمالی تا ۵۵/۸ در صدهزار در سیستان و بلوچستان متغیر است. بیماران شناخته شده در این مناطق ۵۰٪ بیماران سلی را در سطح کشور تشکیل می‌دهند. این در حالی است که این مناطق صرفاً ۲۳٪ جمعیت کشور را در خود جای داده‌اند.

شناخت ماهیت بیماری و عوامل مرتبط با آن می‌تواند در ارائه راه کارهای مناسب به منظور مبارزه با این بیماری از اهمیت خاصی برخوردار باشند. اگرچه تأثیر برخی عوامل از جمله سوء تغذیه و وضعیت اقتصادی-اجتماعی نامناسب در رابطه با ابتلا به بیماری مطرح و به اثبات رسیده است، اما بررسی کامل‌تری در سایر زمینه‌ها نظیر عوامل مرتبط با شکست درمان و یا مرگ بیماران به دلیل این بیماری کمتر مورد توجه قرار گرفته است. برای انجام بهتر این فعالیت، ارزیابی مناسب داده‌های سل در سطوح مختلف کشور اهمیت بنیادی دارد. برای مثال بیماران ثبت شده در دفتر سل شهرستان را می‌توان به صورت یک همگروه (کوهورت) در نظر گرفت و زیرگروه‌های مختلف آن را از نقطه نظر بروز پی آمدهای مورد نظر با هم مقایسه کرد. همچنین می‌توان با بهره‌گیری از روش مورد شاهدهی لانه‌گزیده موارد بروز پی آمد مورد نظر را مورد بررسی قرار داد. در این مقاله قصد بر این است که با بهره‌گیری از اطلاعات دفتر سل شهرستان‌های یکی از دانشگاه‌های کشور و تجزیه و تحلیل آن با استفاده از دو رویکرد همگروهی و مورد شاهدهی لانه‌گزیده نتایج حاصل مقایسه گردد. به طور اختصاصی ضمن مشخص کردن نحوه نمونه‌گیری مورد-شاهدهی لانه‌گزیده از مجموعه داده‌های یک کوهورت، برآوردهای شاخص رابطه و واریانس آن‌ها در آنالیزهای مناسب برای هریک از داده‌ها مقایسه شود.

روش کار

اطلاعات مورد استفاده در این مقاله داده‌های سل مربوط به سال‌های ۱۳۸۱ تا ۱۳۸۵ یکی از دانشگاه‌های کشور است. پی آمد مورد توجه در این تحلیل، مرگ در حین درمان است. عوامل خطر مورد بررسی شامل جنسیت و محل سکونت (روستایی در مقابل شهری) و گروه درمانی (خط اول در مقابل خط دوم) است. تجزیه و تحلیل با رویکرد همگروهی با توجه به زیرگروه‌های تشکیل شده بر اساس متغیرهای فوق‌الذکر انجام شد. داده‌ها با استفاده از روش Cox Regression تجزیه و تحلیل گردید.

برای انجام تجزیه و تحلیل با رویکرد مورد شاهدهی لانه‌گزیده، بانک اطلاعاتی با توجه به الگوی مورد نیاز برای نمونه‌گیری

یافته‌ها

الف- با رویکرد همگروهی

در این تحلیل، زمان مورد نظر تا رخداد مورد بررسی (مرگ) به عنوان پیامد مطالعه کوهورت در نظر گرفته شد. در طول مدت یک سال بررسی بر روی ۲۷۶ بیمار ۱۰ مورد مرگ روی داد. لازم به توضیح است که تعداد ۳۹ نفر از بیماران در طول مدت درمان از مطالعه خارج شدند. در هنگام ورود به مطالعه میانگین سنی این افراد ۵۴/۳۸ سال با انحراف معیار ۲۰/۸۸ بود. ۱۱۲ نفر (۴۰/۶٪) افراد مرد و ۱۶۴ نفر (۵۹/۱٪) زن بودند. ۱۵۶ نفر (۵۶/۵٪) ساکن شهر و ۱۲۰ نفر (۴۳/۵٪) ساکن روستا بودند. در این مطالعه ۲۵۰ نفر (۹۰/۶٪) در گروه یک درمانی و ۲۶ نفر (۹/۴٪) در گروه دو درمانی قرار داشتند. میانگین سنی افراد در گروه فوت شده ۵۴/۴۱ با انحراف معیار ۲۰/۷۵ و در گروه فوت نشده ۵۳/۴۰ با انحراف معیار ۲۵/۴۴ بود. توزیع فراوانی سایر متغیرهای مورد بررسی (جنسیت، محل سکونت، گروه درمانی) در دو گروه فوت شده و فوت نشده در جدول شماره ۱ آمده است.

در تحلیل تک متغیره با استفاده از مدل Cox regression از میان متغیرهای سن، جنس، محل سکونت و گروه درمانی تنها متغیر گروه درمانی با رخداد مرگ ارتباط معنی‌دار از لحاظ بالینی و آماری داشت: $HR (hazard\ ratio) = 4/28$ (۱۶/۵۹ - ۱/۱۰ CI: ۹۵٪) وارد کردن متغیرهای سن، جنس و محل سکونت میزان HR گروه درمانی را چندان تغییری نداد. نتیجه تجزیه و تحلیل تک متغیره مطالعه کوهورت در جدول شماره ۲ آمده است.

ب- با رویکرد مورد-شاهدهی لانه‌گزینی شده

برای هر مورد مرگ سه نفر کنترل به طور تصادفی انتخاب شد و در مجموع ۴۰ نفر در تحلیل داده‌های مورد-شاهدهی لانه‌گزیده وارد شدند. ۱۹ نفر (۴۷/۵٪) افراد مورد مطالعه مذکر و ۲۱ نفر (۵۲/۵٪) بودند. ۲۱ نفر معادل ۵۲/۵٪ ساکن شهر و ۱۹ نفر

۹۵٪ (۲۶/۹۳ - ۰/۷۵). وارد کردن متغیر سن در مدل میزان OR متغیر گروه درمانی را کمی تغییر داد: OR برابر با ۳/۸۵ و فاصله اطمینان ۹۵٪ (۲۴/۲۸ - ۰/۶۱). جنس و محل سکونت میزان OR گروه درمانی را چندان تغییری نداد. نتیجه آنالیزهای تک متغیره مطالعه مورد-شاهدی در جدول ۴ آمده است.

بحث

در تئوری، هر مطالعه مورد-شاهد باید در ارتباط با یک کوهورت مشخص به نام study base صورت گیرد، اگر چه در عمل تعریف کوهورت مرجع برای مطالعات مورد-شاهد کلاسیک مشکل است (۱). مطالعه مورد-شاهد لانه گزیده از جمله مطالعات مورد-شاهد است که ضمن صرفه جویی قابل توجه در هزینه، زمان و پیروی از اصل study base می‌تواند حتی در شرایطی که بیماری نادر نباشد، برآوردی مستقیم از rate ratio در اختیار دهد. همچنین این نوع مطالعه بر خلاف مطالعات مورد-شاهد کلاسیک دچار تورش‌های یادآوری (recall bias) و تقدم و تأخر زمانی (temporal bias) نیست (۲). هدف از مطالعه حاضر مقایسه نتایج تجزیه و تحلیل یک مطالعه مورد-شاهد لانه گزیده با نتایج تجزیه و تحلیل کوهورت اصلی است.

همانطور که در جدول شماره ۵ مشاهده می‌شود، صرفنظر از خطای تصادفی برآورد اثر متغیرهای بررسی حاصل از مدل conditional logistic regression بر روی داده‌های مورد-شاهدی و مدل Cox regression بر روی داده‌های کوهورت کما بیش یکسان هستند. شباهت نتایج آنالیز داده‌های کوهورت و مورد-شاهد به این دلیل است که conditional likelihood مورد استفاده در مدل conditional logistic regression دقیقاً همانند partial likelihood مورد استفاده در مدل Cox regression است، با این تفاوت که در هر زمان وقوع failure مخرج کسر احتمال شرطی به جای این که تمامی افراد در معرض خطر را در برگیرد، تنها شامل افراد انتخاب شده به عنوان گروه کنترل در زمان مذکور است (ضمیمه ۱) (۳). به عبارت دیگر در صورتی که کلیه افراد در معرض خطر را به عنوان گروه کنترل برای هر فرد مورد انتخاب کنیم، تجزیه و تحلیل مطالعه مورد-شاهد و کوهورت کاملاً یکسان خواهد بود. به همین دلیل و با توجه به تقدم زمانی پیدایش تحلیل Cox نسبت به conditional logistic، بسیاری از نرم افزارهای معتبر آماری دستور جدیدی برای تحلیل مطالعه مورد-شاهد لانه گزیده (و یا هر مطالعه دیگر همسان شده) تعریف نکرده‌اند (ضمیمه ۲). باید توجه داشت که برآوردهای حاصل از

جدول شماره ۱- توزیع فراوانی متغیرهای جنس، محل سکونت و گروه درمانی به تفکیک رخدادهای مرگ در داده‌های کوهورت

متغیر	فوت شده		فوت نشده	
	فراوانی	درصد	فراوانی	درصد
جنسیت	مرد	۱۰۷	۴۰/۲	۵
	زن	۱۵۹	۵۹/۸	۵
محل سکونت	شهر	۱۵۰	۵۶/۴	۶
	روستا	۱۱۶	۴۳/۶	۴
گروه درمانی	گروه اول	۲۴۳	۹۱/۴	۷
	گروه دوم	۲۳	۸/۶	۳

جدول شماره ۲- نتایج آنالیزهای انجام شده بر روی داده‌ها با رویکرد کوهورت

متغیر	خطر نسبی	P value	فاصله اطمینان ۹۵٪ برای خطر نسبی
جنسیت	۰/۶۳	۰/۴۶	۰/۱۸ - ۲/۱۸
محل سکونت	۰/۸۴	۰/۷۹	۰/۲۴ - ۲/۹۹
سن (۱۰ سال)	۱/۰۱	۰/۹۵	۰/۷۴ - ۱/۳۸
گروه درمانی	۴/۲۸	۰/۰۳	۱/۱۰ - ۱۶/۵۹

جدول شماره ۳- توزیع فراوانی متغیرهای جنس، محل سکونت و گروه درمانی به تفکیک رخدادهای مرگ در داده‌های مورد-شاهدی

متغیر	گروه مورد		گروه شاهد	
	فراوانی	درصد	فراوانی	درصد
جنسیت	مرد	۱۴	۴۶/۷	۵
	زن	۱۶	۵۳/۳	۵
محل سکونت	شهر	۱۵	۵۰/۰	۶
	روستا	۱۵	۵۰/۰	۴
گروه درمانی	گروه اول	۲۸	۹۳/۳	۷
	گروه دوم	۲	۶/۷	۳

(۴۷/۵٪) ساکن روستا بودند. میانگین سن افراد مورد بررسی ۴۶ سال با انحراف معیار ۲۲/۱۴ سال بود. میانگین سنی گروه کنترل ۴۴/۵۷ با انحراف معیار ۲۰/۹۳ و در گروه مورد ۵۳/۴۵ با انحراف معیار ۲۵/۴۴ بود. توزیع فراوانی سایر متغیرهای مورد بررسی (جنسیت، محل سکونت، گروه درمانی) در دو گروه مورد و شاهد در جدول شماره ۳ آمده است.

در تحلیل تک متغیره با استفاده از مدل Conditional logistic regression از میان متغیرهای سن، جنس، محل سکونت و گروه درمانی تنها متغیر گروه درمانی با رخدادهای مرگ دارای ارتباط معنی دار از لحاظ بالینی و نزدیک به معنی داری از نظر آماری بود (OR برابر ۴/۵۰ با فاصله اطمینان

جدول شماره ۴- نتایج آنالیزهای انجام شده بر روی داده‌ها با رویکرد مورد-شاهدی لانه‌گزینی شده

متغیر	نسبت شانس	P value	فاصله اطمینان ۹۵٪
جنسیت	۰/۸۷	۰/۸۵	۰/۲۰ - ۳/۷۲
محل سکونت	۰/۶۴	۰/۵۶	۰/۱۴ - ۲/۸۰
سن (۱۰ سال)	۱/۲۱	۰/۲۵	۰/۸۶ - ۱/۷۱
گروه درمانی	۴/۵۰	۰/۰۹	۰/۷۵ - ۲۶/۹۳

جدول شماره ۵- خلاصه مقایسه نتایج با استفاده از دو رویکرد

متغیر	خطر نسبی در رویکرد همگروهی	نسبت شانس در رویکرد مورد-شاهدی لانه‌گزیده
جنسیت	۰/۶۳* (۰/۱۸ - ۲/۱۸)	۰/۸۷ (۰/۲۰ - ۳/۷۲) P value: ۰/۸۵
محل سکونت	۰/۸۴ (۰/۲۴ - ۲/۹۹)	۰/۶۴ (۰/۱۴ - ۲/۸۹) P value: ۰/۵۶
سن (۱۰ سال)	۱/۰۱ (۰/۷۴ - ۱/۳۸)	۱/۲۱ (۰/۸۶ - ۱/۷۱) P value: ۰/۲۵
گروه درمانی	۴/۲۸ (۱/۱۰ - ۱۶/۵۹)	۴/۵۰ (۲۶/۹۳ - ۰/۷۵) P value: ۰/۰۹

* فاصله اطمینان ۹۵٪

تجزیه و تحلیل داده‌های مورد-شاهد مانند برآوردهای اصلی حاصل از آنالیز داده‌های کوهورت دارای خواص مناسب مجانبی (asymptotic) مانند سازگاری (consistency) و توزیع نرمال (normality) هستند (۴).

در مطالعه مورد-شاهد لانه‌گزیده، در زمان رخداد هر مورد، تعدادی از افراد در معرض خطر به عنوان گروه کنترل انتخاب می‌شوند. مقیاس زمانی انتخاب شده می‌تواند مدت حضور فرد در مطالعه، سن و یا تقویم زمانی باشد. معمولاً مقیاس زمانی که دارای قوی‌ترین ارتباط با خطر پیامد باشد مورد استفاده قرار می‌گیرد، چرا که تفاوت مقیاس زمانی موجود بین افراد با همسان کردن (matching) کنترل می‌گردد (۵). در مطالعه حاضر (تحلیل مورد-شاهد و کوهورت) مدت حضور افراد در کوهورت (مدت درمان) به عنوان مقیاس زمانی در نظر گرفته شد. همچنین افراد گروه شاهد به طور تصادفی ساده و بدون جایگذاری از میان افراد در معرض خطر در هر زمان انتخاب شدند. این روش به عنوان روش ناریب و بهتر در انتخاب گروه شاهد مطالعات مورد-شاهد لانه‌گزیده شناخته شده است (۶). انتخاب گروه شاهد در زمان‌های مختلف باید به طور مستقل انجام پذیرد و به طور خاص

تحت تأثیر شاهد بودن در زمان‌های قبل و یا مورد شدن در زمان‌های بعد قرار نگیرد. این قاعده نیز انعکاسی از رویکرد تحلیلی در مدل‌های proportional hazards بر روی داده‌های زمان بقا است که تقریباً همه افراد گروه مورد در risk set های قبلی به عنوان گروه شاهد مورد استفاده قرار گرفته اند و همچنین تقریباً همگی افراد شاهد بیش از یکبار مورد استفاده قرار می‌گیرند (۷). در مطالعه حاضر ۴ نفر دو بار به عنوان گروه شاهد انتخاب شدند.

در مطالعات مورد-شاهد گفته شده است که گرفتن بیش از ۴ کنترل به ازای هر مورد کارایی (efficiency) مطالعه را به میزان قابل توجه افزایش نمی‌دهد. کارایی نسبی یک مطالعه مورد-شاهد لانه‌گزیده با m کنترل در مقایسه با مطالعه کوهورت در صورت وجود یک exposure دو حالتی و تحت فرض خنثی بی اثر بودن $m/m+1$ است (۸،۹). عوامل مؤثر دیگر در کارایی نسبی عبارتند از میزان خطر نسبی، شیوع exposure در گروه کنترل و تعداد طبقات (strata) مورد استفاده. به طور کلی در شرایطی که شیوع exposure در گروه کنترل پایین و یا تعداد طبقات کم باشد، ممکن است گرفتن کنترل بیش از ۴ نفر ارزش داشته باشد (۱۰). در مطالعه حاضر کارایی نسبی برای متغیرهای دو حالتی جنس و محل سکونت ۷۲٪ بود که نزدیک به مقدار قابل انتظار ۷۵٪ است. همچنین کارایی نسبی برای متغیر سن و گروه درمان به ترتیب ۸۲٪ و ۵۷٪ بود که با توجه به پیوسته بودن متغیر سن و بزرگی اثر گروه درمان (میزان خطر نسبی معادل ۴/۳) دور از انتظار نیست.

در تجزیه و تحلیل داده‌های کوهورت و مورد-شاهد لانه‌گزیده زمان در معرض خطر به عنوان طبقه در نظر گرفته شده و تعدیل می‌شود. پرسشی که مطرح می‌شود اگر زمان در معرض خطر تعدیل نشود و به جای تحلیل Cox regression و conditional logistic regression از logistic regression استفاده شود، آیا برآوردهای حاصل معتبر هستند یا خیر؟ تحلیل regression logistic روی داده‌های کوهورت در صورتی که بیماری نادر باشد و شیوع exposure در طول زمان ثابت باشد (به ویژه در مورد fixed exposure ها differential loss to follow up نداشته باشیم) و یا تابع hazard در طول زمان ثابت باشد، برآورد odds ratio حاصل تقریبی از rate ratio خواهند بود، ولی برآورد واریانس‌ها در هر صورت بیشتر از برآوردهای حاصل در تحلیل Cox regression خواهد بود (۱۱،۱۲). تحلیل logistic regression روی داده‌های مورد-شاهد لانه‌گزیده به معنای در نظر نگرفتن همسان بودن طراحی مطالعه در تحلیل

می‌باشد، توصیه می‌شود.

ضمیمه ۱- مقایسه partial likelihood در داده‌های کوهورت با conditional likelihood در داده‌های مورد- شاهد لانه گزیده

partial log likelihood برای داده‌های کوهورت از لگاریتم حاصلضرب احتمالات شرطی وقوع رخداد failure در فرد مورد در هر زمان، به شرط وقوع رخداد failure در افراد در معرض خطر در آن زمان بدست می‌آید:

$$\sum_f \log \left(\frac{\exp(\beta TX_f)}{\sum_{i \in r(f)} \exp(\beta TX_i)} \right)$$

به طوری که f و $r(f)$ به ترتیب نشان دهنده رخداد (failure) و افراد در معرض خطر در هر زمان رخداد می‌باشند. همچنین X و β به ترتیب بردار (vector) متغیرهای مستقل و پارامترهای مجهول را نشان می‌دهند. log likelihood فوق در حقیقت یک log profile likelihood است، چرا که پارامترهای baseline hazard توسط تابعی از β که likelihood را به ازای β ثابت ماکزیمم می‌کنند جایگزین شده است.

conditional log likelihood برای داده‌های مورد- شاهد لانه گزیده دارای k کنترل به ازای هر یک مورد و c طبقه به شرط توزیع exposure در هر طبقه عبارتست از:

$$\sum_c \log \left(\frac{\exp(\beta TX_i)}{\sum_{i=1}^{k+1} \exp(\beta TX_i)} \right)$$

به طوری که اندیس ۱ نشان دهنده فرد مورد و اندیس‌های ۲ تا $k+1$ نشان دهنده k فرد شاهد هستند. همچنین X و β مانند قبل به ترتیب بردار (vector) متغیرهای مستقل و پارامترهای مجهول را نشان می‌دهند. از مقایسه ۲ log likelihood فوق می‌توان دریافت که profile log likelihood برای داده کوهورت کاملاً مشابه conditional log likelihood برای داده مورد- شاهد لانه گزیده است، اگر در هر زمان رخداد همه افراد در معرض خطر به عنوان گروه کنترل برای مورد انتخاب شوند.

ضمیمه ۲- تجزیه و تحلیل مطالعه مورد- شاهد لانه گزیده با نرم افزارهای SAS و SPSS

شباهت profile log likelihood برای داده‌های کوهورت و conditional log likelihood برای داده‌های مورد- شاهد لانه گزیده به حدی است که بسیاری از نرم افزارهای آماری معتبر مانند SAS و SPSS دستور جدیدی برای انجام تجزیه و تحلیل مطالعه مورد- شاهد لانه گزیده (و یا هر مطالعه مورد- شاهد

داده‌ها است. در صورتی که شیوع exposure در طول زمان ثابت باشد، برآوردگر odds ratio به طور سازگار (consistent)، ratio rate را برآورد می‌کند. از طرف دیگر واریانس برآوردهای تحلیل logistic regression نسبت به واریانس برآوردهای تحلیل conditional logistic regression کمتر می‌باشد. در مورد داده‌های مطالعه حاضر، تحلیل logistic regression بر روی داده‌های کوهورت و مورد- شاهد لانه گزیده تغییر قابل توجهی در نتایج نمی‌دهد که با توجه به نادر بودن پیامد و ثابت بودن نسبی exposure ها در طی زمان قابل توجیه است (۱۳).

از محدودیت‌های این مقاله می‌توان به تعداد بسیار کم رخداد مرگ در داده‌ها (۱۰ نفر) اشاره نمود که می‌تواند با افزایش خطای تصادفی روی نتایج هر دو مطالعه تاثیر منفی داشته باشد و برآوردهای حاصل را غیر قابل اعتماد سازد. البته از آنجا که همه موارد مرگ وارد هر دو مطالعه می‌گردند، اختلاف نتایج دو مطالعه در تحلیل‌های تک متغیره تنها تحت اثر تعداد افراد کنترل (نسبت کنترل به مورد) قرار می‌گیرد. البته همانطور که در قسمت نتایج مطالعه ذکر شده است، برای مدل‌سازی آماری، مدل‌های رگرسیونی چند گانه Cox و conditional logistic جهت تعدیل اثر مخدوش کنندگی در نظر گرفته شده‌اند. ولی باتوجه به اینکه عملاً تنها متغیر "گروه درمانی" با پیامد مورد نظر در ارتباط می‌باشد و نیز نادر بودن پیامد بیماری (که تعدیل نامناسب در چنین شرایطی می‌تواند به sparsity داده‌ها و غیر معتبر شدن برآوردها بیانجامد) مقایسه برآوردهای شاخص رابطه و واریانس آن‌ها برای هر متغیر به طور جداگانه انجام شد. مطالعات آینده بر روی کوهورت‌های دارای تعداد پیامد کافی و متغیرهای مخدوش کننده اندازه‌گیری شده (برای متغیر یا متغیرهای exposure مورد نظر) می‌توانند برآوردهای شاخص رابطه تعدیل شده و واریانس آن‌ها را میان مدل رگرسیونی چند گانه Cox برای داده کوهورت اصلی و مدل رگرسیونی چند گانه conditional logistic برای داده مورد- شاهدی لانه گزیده مقایسه نمایند.

نتیجه گیری

نتایج تجربی مطالعه حاضر نشان می‌دهد که مطالعه مورد- شاهد لانه گزیده علاوه بر راحت و مقرون به صرفه بودن، دارای اعتباری قابل قیاس با آنالیز کوهورت اصلی در برآورد rate ratio و واریانس آن می‌باشد. انجام این مطالعه به ویژه در شرایطی که بیماری نادر است و پیامد برای همه افراد قابل دستیابی بوده، ولی جمع‌آوری متغیرهای دیگر برای همه اعضای کوهورت گران

به طوری که در هر طبقه افراد شاهد زمان بیشتری از زمان فرد مورد داشته باشند (به طور مثال می‌توان در هر طبقه به فرد مورد و افراد شاهد به ترتیب زمان ۱ و ۲ را تخصیص داد) و زمان‌های افراد شاهد به عنوان زمان censoring در نظر گرفته شوند. در چنین شرایطی آنالیز stratified Cox regression با متغیرهای مذکور نتایجی دقیقاً معادل conditional logistic regression خواهد داشت.

همسان شده) ارائه نکرده‌اند. در حقیقت آنالیز regression conditional logistic را می‌توان با تغییراتی توسط دستورهای مرتبط با آنالیز Cox regression (به طور مثال PROC PHREG در SAS و COXREG در SPSS) انجام داد. به طور خلاصه و در صورت وجود یک مورد همسان شده با یک یا بیشتر شاهد (مانند مطالعه حاضر) باید گروه‌های همسان شده را به عنوان طبقه در نظر گرفت و متغیر جعلی زمان بقا برای افراد مورد مطالعه ساخت.

منابع

- 1- Wacholder S, McLaughlin JK, Silverman DT, et al. Selection of controls in case control studies: I: principles. *Am J Epidemiol* 1992; 135: 1019–28.
- 2- Langholz B. Case-Control Study, Nested In: Armitage P, Colton T(eds). *Encyclopedia of Biostatistics*. Chichester, UK: John Wiley & Sons, 2005: 646-65.
- 3- Prentice RL, Breslow NE: Retrospective studies and failure time models. *Biometrika* 1978, 65: 153-8.
- 4- Goldstein L, Langholz B. Asymptotic theory for nested case-control sampling in the Cox regression model. *Ann Stat* 1992; 20: 1903–28.
- 5- Clayton D, Hills M. *Statistical Models in Epidemiology*. Oxford: OUP, 1993.
- 6- Robins JM, Gail MH, Lubin JH. More on “biased selection of controls for case-control analyses of cohort studies.” *Biometrics* 1986; 42: 293–9.
- 7- Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case control studies: III: design options. *Am J Epidemiol* 1992; 135: 1042–50.
- 8- Ury H. Efficiency of case-control studies with multiple controls per case: continuous or dichotomous data. *Biometrics* 1975; 31: 643–9.
- 9- Breslow NE, Day NE. *Statistical Methods in Cancer Research: Vol II. The Design and Analysis of Cohort Studies*. Lyon: IARC, 1987.
- 10- Pang D. A relative power table for nested matched case-control studies. *Occup Environ Med* 1999; 56: 67–9.
- 11- Callas PW, Pastides H, Hosmer DW. Empirical comparisons of proportional hazards, poisson, and logistic regression modeling of occupational cohort data. *Am J Ind Med* 1998; 33: 33-47.
- 12- Jewell NP. *Statistics for Epidemiology*. Boca Raton, FL: Chapman & Hall/CRC, 2004.
- 13- Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *Am J Epidemiol* 1982; 116: 547–53.

Iranian Journal of Epidemiology 2010; 6(2): 1-6.

Original Article

Comparison of Nested Case-control and Cohort Analysis Methodologies using a District TB Registry Data

Holakouie Naieni K¹, Ostovar A², Danesh A², Sadjedinejad S², Ghalichee L², Moradi Gh², Mansournia M A², Hashemi Nazari SS²

1- Professor of Epidemiology, Department, Department of Epidemiology and Biostatistics, School of public health, Tehran university of Medical Sciences, Tehran, Iran

2- PHD candidate of epidemiology, Department of Epidemiology and Biostatistics, School of public health, Tehran university of Medical Sciences, Tehran, Iran

Corresponding author: Mansournia MA., mansournia_ma@yahoo.com

Background and objective: The nested case-control study has become popular as an efficient alternative to the full-cohort design. This study compares the results of a nested case-control analysis approach with the full cohort analysis.

Methods: A cohort of 276 subjects (new cases from a TB registry) was used for this study. Cox Regression model was used for the full cohort analysis. In order to do the nested case-control analysis, for each death, three random controls were selected from those who did not suffer from the outcome at the time of the outcome took place. Case control data was analyzed by the conditional logistic regression model.

Results: Results from both cohort and nested case-control analyses show that treatment group is the only variable that affects on the outcome. Gender, place of residence, and age has no effect on the outcome. For binary exposure variables with trivial effects (e.g. Gender and place of residence), the relative efficiency of nested case-control study design is approximately 75%.

Conclusion: Results of this study show that nested case-control study is not only an easy and cost-effective method for data analysis but also is as robust as cohort analysis in rate ratio and its variance estimation.

Keywords: Cohort, Nested case control, Tuberculosis