

## روش شناسی مطالعات مورد-هم گروهی در مقایسه با مورد-شاهدی لانه گزیده؛ یک مطالعه مروری

کوروش هلاکوئی نائینی<sup>۱</sup>، محمدعلی منصورنیا<sup>۱</sup>، محمدحسین پناهی<sup>۲</sup>، عادل الدوما<sup>۲</sup>، شهرزاد نعمت الهی<sup>۲</sup>

<sup>۱</sup> گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

<sup>۲</sup> دانشجوی دکتری تخصصی اپیدمیولوژی، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

نویسنده رابط: شهرزاد نعمت الهی، نشانی: تهران، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، طبقه پنجم، گروه اپیدمیولوژی، تلفن: ۰۹۲۲۷۰۳۸۴۰۵

پست الکترونیک: shahrzadnema@yahoo.com

تاریخ دریافت: ۹۴/۹/۷؛ پذیرش: ۹۵/۲/۴

### چکیده

مطالعات مورد-شاهدی لانه گزیده و مورد-همگروهی یکی از شایع ترین راهکارها برای کاهش هزینه در مطالعات همگروهی آینده نگر به شمار می آیند. به دلیل اهمیت و نیاز به این نوع مطالعات در حوزه های پزشکی و بهداشت طی سال های اخیر، در این مطالعه سعی شده است مرور جامعی از اطلاعات به روز در خصوص این دو نوع مطالعه با محوریت تخمین اندازه اثر، نمونه گیری، روش های تجزیه و تحلیل و ملاحظات آماری و متدولوژیک ارائه شود. بطور خلاصه مزایای طرح مورد-همگروهی ناشی از نمونه گیری تصادفی شاهدها می باشد. از سوی دیگر مزایای طرح لانه گزیده علاوه بر مقرون به صرفه بودن آن از منظر اقتصادی، ناشی از همسان سازی روی واحد زمان نیز است. در نهایت، قضاوت در خصوص طرح برتر به عواملی نظیر اهداف مطالعه، چگونگی انتخاب اعضای همگروه، چگونگی تعیین موردها و شاهدها از درون همگروه و چگونگی جمع آوری اطلاعات متغیرها بستگی دارد. در یک موقعیت خاص عواملی که سبب اولویت بندی نکات مذکور و کارآیی آماری می شوند بستگی به خصوصیات و ماهیت همان مطالعه خواهند داشت.

واژگان کلیدی: مورد-همگروهی، مورد-شاهدی لانه گزیده، طرح مطالعه اپیدمیولوژیک، اندازه اثر

### مقدمه

دینامیک<sup>۳</sup> است (۱). براین اساس موردها و شاهدها می توانند از یک جمعیت ثابت (مانند متولدین یک سال تقویمی)، یک جمعیت دینامیک (جمعیتی که تحت تأثیر میرایی و تولد و مهاجرت قرار می گیرد) و یا از یک جمعیت پایا<sup>۴</sup> (جمعیتی که ترکیب نسبتاً ثابت توزیع مواجهه طی زمان دارد) انتخاب شوند (۲). از طرف دیگر، جمعیت های دینامیک می توانند طی زمان های کوتاه یا برای مواجهات خاصی (مانند عوامل ژنتیکی) پایا باشند. درون جمعیت های ثابت<sup>۳</sup> راهکار به شرح زیر برای نمونه گیری شاهدها وجود دارد:

الف) کورنفیلد در سال ۱۹۵۱ نشان داد که شاهدها را می توان از اشخاصی که در انتهای دوره پیگیری سالم مانده اند انتخاب کرد.

مطالعات مورد شاهدی<sup>۱</sup> از مهم ترین انواع مطالعات مشاهده ای هستند. در این نوع مطالعات گروه "مورد" یعنی افراد مبتلابه بیماری یا وضعیتی خاص با گروه "شاهد" یعنی افرادی که مبتلابه آن بیماری نیستند از نظر سطوح مواجهه با عوامل خطر مشکوک در گذشته مورد بررسی قرار می گیرند. مطلب کلیدی در مطالعات مورد - شاهدی راهکار مورد استفاده برای تعیین موردها و شاهدها است. ملاحظه اول تعیین ماهیت موردها از نظر بروز یا شیوع است. اگر موردها بروز هستند ملاحظه دوم تعیین منبع انتخاب موردها و شاهدها (از یک هم گروه ثابت<sup>۲</sup> یا یک جمعیت

<sup>۳</sup> Dynamic-Open

<sup>۴</sup> Stable population

<sup>۱</sup> Case-control study

<sup>۲</sup> Fixed-Closed

با وجود مزایای بسیار طرح مورد-شاهدی مانند دوره کوتاه مطالعه و کارایی؛ نادر بودن واقعه موردنظر یا گران‌قیمت بودن ارزیابی مواجهه سودمندی این طرح را کاهش می‌دهند (۱). نویسندگان مختلف پیشنهاد می‌کنند که تخمین نسبت شانس<sup>۸</sup> به‌عنوان مهم‌ترین اندازه اثر به‌دست‌آمده در مطالعات مورد-شاهدی هر زمان که بروز جمعی واقعه در طول مطالعه بیش از ۱۰٪ (محدوده ۲۰٪-۵٪) باشد دارای تورش خواهد بود (۱۱، ۱۰). بنابراین بستگی به ماهیت موردها، نوع جمعیت منبع (Source population)، استراتژی نمونه‌گیری، و فرضیات زیربنایی؛ نسبت شانس به‌دست‌آمده از مطالعه مورد-شاهدی می‌تواند نسبت شانس باقی‌مانده یا به‌عنوان نسبت خطر<sup>۹</sup>، یا نسبت میزان<sup>۱۰</sup> تفسیر شود.

فرض کنید که تعداد کل جمعیت در معرض خطر در زمان  $t$  برابر با  $N(t)$  بوده که تناسب  $P1(t)$  نفر از آن‌ها مواجهه دارند. بر این اساس می‌توان نسبت میزان را به‌صورت  $IDR = ID1(t)/ID0$  نوشت (۱۱)؛ و نسبت خطر را به‌صورت  $RR (CIR) = CI1(t) / CI0(t)$  نوشت که  $CIi(t)$  و  $IDi(t)$  به ترتیب بروز جمعی و تراکمی، و  $i$  نشان‌دهنده وضعیت مواجهه (دو حالتی) است (۱۱). وقتی نمونه‌گیری‌ها در یک هم‌گروه ثابت در آغاز دوره پیگیری (طرح مورد-هم‌گروهی) انجام شود با فرض اینکه گمشدگی<sup>۱۲</sup> مرتبط با مواجهه و یا خطر پیامد نباشد نسبت شانس تخمین نسبت خطر است. همچنین اگر  $P1(t)$  در طول دوره خطر تغییر نکند (جمعیت پایا باشد) می‌توان نسبت شانس را تخمینی از نسبت میزان به‌حساب آورد (۱۲)؛ شایان‌ذکر است که هیچ‌کدام از شرایط مذکور نیازمند اثبات فرضیه نادر بودن واقعه نیستند (۱).

## ۲-۱- روش نمونه‌گیری

یک مطالعه هم‌گروه با  $N$  فرد را در نظر بگیرید. فرد  $i$  در زمان  $t_i$  وارد مطالعه می‌شود. این فرد در زمان  $T_i$  دچار واقعه می‌شود درحالی‌که زمان گمشدگی  $W_i$  دارد. همچنین فرض می‌شود که زمان واقعه و زمان گمشدگی از هم مستقل هستند.  $X_i(t)$  نیز

این طرح نمونه‌گیری کلاسیک مورد - شاهدی است که بنام طرح انحصاری<sup>۱</sup> و یا نمونه‌گیری تجمعی بروز نامیده می‌شود (۳).

ب) طرحی که اولین بار توسط کوپر در سال ۱۹۷۵ و سپس توسط پرنیس در سال ۱۹۸۶ (۴) مبنی بر انتخاب شاهدی در آغاز دوره پیگیری از جمعیت در معرض خطر ارائه شد که بنام طرح گنجانده<sup>۲</sup>، یا مطالعه مورد - هم‌گروهی<sup>۳</sup> نامیده می‌شود (۵). ج) انتخاب شاهدی به‌نحوی که هر زمان یک مورد بیماری تشخیص داده می‌شود (یا برای موردهایی که تاکنون مشخص شده‌اند) یک شاهد از جمعیت در معرض خطر در همان زمان انتخاب می‌شود. به‌عبارت‌دیگر شاهدی از مجموعه شخص - زمان‌های در معرض خطر انتخاب می‌شوند و لذا با موردها روی متغیر زمان همسان شده‌اند. بر این اساس این طرح نمونه‌گیری تراکمی<sup>۴</sup> (مجموعه در معرض خطر<sup>۵</sup>) و مطالعه متناظر آن مطالعه مورد - شاهدی لانه‌گزیده<sup>۶</sup> نامیده می‌شود (۶).

مطالعات مورد-شاهدی لانه‌گزیده و مورد-هم‌گروهی یکی از شایع‌ترین راهکارها برای کاهش هزینه در مطالعات هم‌گروهی آینده‌نگر به شمار می‌آیند. البته بی‌شک ارائه برآوردهای با خطای استاندارد کوچک‌تر، و اجازه استفاده از مدل‌های آنالیز آماری با توان بیشتر از دیگر مزایای این طرح‌ها نسبت به مطالعه مورد-شاهدی کلاسیک می‌باشند (۷).

از سوی دیگر مطالعات مروری حکایتی<sup>۷</sup> ابزاری قوی برای اهداف آموزشی و باهدف ارائه یک مرور جامع در خصوص موضوع موردنظر به‌حساب می‌آیند (۸). به دلیل اهمیت و نیاز به این نوع مطالعات در حوزه‌های پزشکی و بهداشت طی سال‌های اخیر (۹)، بر آن شدیم که مرور جامعی از اطلاعات به‌روز در خصوص مطالعات "مورد-شاهدی لانه‌گزیده و "مورد-هم‌گروهی" با تأکید بر تخمین‌های اندازه اثر، نمونه‌گیری، روش‌های تجزیه و تحلیل، و ملاحظات آماری و روش‌شناختی داشته باشیم.

## مطالعه مورد - هم‌گروهی

### ۱-۱ - تخمین اندازه اثر

<sup>۱</sup> Exclusive design

<sup>۲</sup> Inclusive design

<sup>۳</sup> Case-cohort study

<sup>۴</sup> Density sampling

<sup>۵</sup> Risk-set sampling

<sup>۶</sup> Nested case-control study

<sup>۷</sup> Narrative Review Article

<sup>۸</sup> Odds Ratio

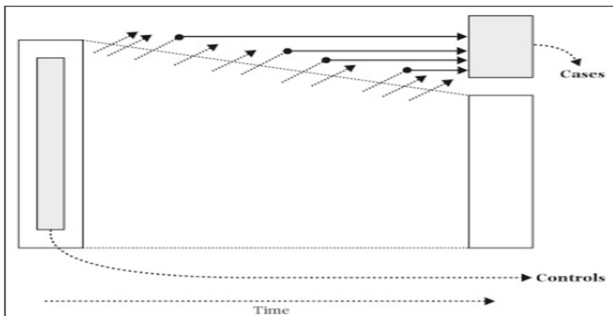
<sup>۹</sup> Risk Ratio

<sup>۱۰</sup> Rate Ratio

<sup>۱۱</sup> Incidence-Density Ratio

<sup>۱۲</sup> Censoring

جانشین<sup>۵</sup> (که پیش‌بینی کننده مواجهه اصلی هستند) برای کل اعضای هم‌گروه موجود است. بطور مثال گروه خونی ممکن است برای کل هم‌گروه معلوم باشد در حالی که تایپینگ DNA برای تعیین آل‌های یک ژن خاص در هر فرد مورد نیاز باشد. اگر فراوانی ژن وابسته به گروه خونی باشد، می‌توان از گروه خونی بعنوان یک جانشین در هنگام نمونه‌گیری شاهدها برای اجرای تایپینگ DNA استفاده کرد. با لایه‌بندی کردن نمونه تحت هم‌گروه برحسب متغیرهای جانشین، طرح مطالعه مورد-هم‌گروهی قوی‌تر می‌شود. چنین طرحی به‌عنوان طرح لایه‌بندی شده<sup>۶</sup> نیز شناخته می‌شود. درون هر لایه، نمونه‌گیری از تحت هم‌گروه با نمونه‌گیری تصادفی ساده انجام می‌شود. سپس در مرحله دوم، لایه‌ها و کسرهای نمونه‌گیری بعد از مشاهده اینکه چه افرادی بیمار می‌شوند دوباره تعریف شده و برآوردها تخمین زده می‌شوند (۱۵).



شکل شماره ۱- شمای تئوریک نمونه‌گیری در طرح مورد-هم‌گروهی (۱۶)

### ۳-۱- روش وزن‌دهی و تجزیه و تحلیل

پیامد در مطالعه مورد-هم‌گروهی به‌صورت "زمان تا واقعه" تعریف می‌شود. بر این اساس روش‌های آنالیز برای مطالعه مورد-هم‌گروهی توسط پرنیتیس (۴)، و بعدها با همکاری سیلف (۱۷) در سال ۱۹۸۸ تحت عنوان راهکار "درست‌نمایی نما"<sup>۷</sup> ارائه شد که تعدیلی از راهکار درست‌نمایی جزئی<sup>۸</sup> کاکس با وزن‌دهی متفاوت به موردها و تحت هم‌گروه بود (۱۸). بنابراین در مدل رگرسیون کاکس وزن گرفته، هر واقعه‌ای که در زمان  $t$  برای فرد  $i$  اتفاق می‌افتد به‌صورت تابع زیر وارد مدل می‌شود:

بُرداری از متغیرهای کمکی وابسته به زمان  $t$  است. فرض کنید که تابع مخاطره  $\lambda_i(t)$  از زمان واقعه، از مدل زیر پیروی می‌کند:

$$\lambda_i(t) = \lambda_0(t) \exp(X_i(t)\beta)$$

بنابراین  $\lambda_0(t)$  تابع اولیه نامشخص و  $\beta$  بُرداری از پارامترهای موردنظر می‌باشند. درنهایت، استنتاج‌ها بر اساس حداکثر کردن تابع درست‌نمایی جزئی کاکس به شرح زیر به دست می‌آیند که در آن:

فرمول

$$L_c(\beta) = \prod_{i=1}^N \left[ \frac{e^{X_i(Y_i)\beta}}{\sum_{j \in R_i} e^{X_j(Y_i)\beta}} \right]^{\delta_i}$$

$R_i = \{j: Y_j \geq Y_i > a_j\}$  مجموعه در معرض خطر<sup>۱</sup> است ((یعنی مجموعه افرادی که در زمان دچار شدن فرد  $i$  ام به واقعه در معرض خطر بوده‌اند). مطالعه مورد-هم‌گروهی، همه موردها و یک نمونه به تصادف انتخاب انتخاب‌شده از شاهدها (تحت هم‌گروه ابتدای<sup>۲</sup> کار را در نظر می‌گیرد یعنی شاهدها تنها یک‌بار به‌طور تصادفی انتخاب می‌شوند. فرض کنید  $C_i$  نشان‌دهنده شاهدهایی باشد که در زمان  $R_i$  انتخاب شده‌اند درحالی که  $C$  نشان‌دهنده جمعیت شاهدها باشد؛ تحت هم‌گروه برای طرح مورد-هم‌گروهی به معنی  $C=C_i$  است (۷). طرح مورد-هم‌گروهی (شکل شماره ۱) برای آنالیز داده‌های هم‌گروهی که در آنها واقعه نادر باشد، مفید است؛ زیرا اطلاعات متغیرها برای همه وقایع و فقط یک نمونه تصادفی (با احتمال نمونه‌گیری  $\alpha$ ) از مشاهدات سانسور شده (تحت هم‌گروه) استفاده می‌شود. همچنین اگر جمع آوری جزئیات پیگیری، هزینه و زمانبر باشند این طرح کارآ است. با انتخاب یک نمونه تصادفی تحت هم‌گروه، پیگیری بیماری هنوز برای کل هم‌گروه مورد نیاز است درحالی که پیگیری برای دیگر وقایع سانسور شده محدود به اعضای تحت هم‌گروه می‌شود (۱۳). از سوی دیگر استفاده از نمونه تحت هم‌گروه بجای کل اعضای هم‌گروه، به معنی نادیده گرفتن ارتباطات حاشیه‌ای<sup>۳</sup> است. برسלו نشان داد در شرایط مواجهه نادر، بیش نمونه‌گیری<sup>۴</sup> موردها می‌تواند منجر به تخمین نادرست نسبت خطر شود (۱۴). از سویی دیگر در اغلب موارد، اطلاعات متغیرهایی مانند متغیرهای

<sup>۵</sup> Surrogate variables

<sup>۶</sup> Stratified case-cohort/Nested case-control design

<sup>۷</sup> Pseudo Likelihood

<sup>۸</sup> Partial-Likelihood

<sup>۱</sup> Risk-set

<sup>۲</sup> Sub-cohort

<sup>۳</sup> Marginal

<sup>۴</sup> Over-sampling

$$\frac{Y_i(t_j)e^{Z_i(t_j)\beta}}{Y_i(t_j)w_i(t_j)e^{Z_i(t_j)\beta} + \sum_{K \neq i} Y_k(t_j)W_k(t_j)e^{Z_k(t_j)\beta}}$$

جدول شماره ۱- وزن‌های اختصاص داده‌شده به موردها و شاهد‌ها در طرح

مورد-هم‌گروهی

روش وزن دهی			نوع پیامد و زمان بندی
بارلو	پرنتمیس و سلف	پرنتمیس	
۰	۰	۰	موردهای خارج از تحت هم‌گروه پیش از واقعه
۱	۰	۱	موردهای خارج از تحت هم‌گروه در هنگام واقعه
۱/α	۱	۱	موردهای درون تحت هم‌گروه پیش از واقعه
۱	۱	۱	موردهای درون تحت هم‌گروه در هنگام واقعه
۱/α	۱	۱	شاهد‌ها در تحت هم‌گروه

مخرج کسر شامل دو قسمت است که قسمت اول آن مورد با وزن  $w_i$ ، و قسمت دوم مجموع اعضای تحت هم‌گروه در معرض خطر در زمان رخداد واقعه (شامل موردها قبل از رخداد واقعه و شاهد‌ها) با وزن  $W_k$  است (۱۳).

سه روش وزندهی متفاوت بر اساس این تابع پیشنهاد شده‌اند (۲۰، ۱۹، ۱۷، ۴) (جدول شماره ۱). در روش اول که توسط پرنتمیس توصیف شد، تمامی اعضای تحت هم‌گروه صرف‌نظر از این‌که درآینده به بیماری مبتلا شوند یا خیر به‌طور برابر وزن می‌گیرند. بنابراین موردهایی که خارج از تحت هم‌گروه به وجود می‌آیند قبل از واقعه وزنی نمی‌گیرند و در زمان واقعه وزنی مشابه با اعضای تحت هم‌گروه می‌گیرند (۴). مدتی بعد پرنتمیس و سلف روش دوم را پیشنهاد کردند که در آن، همه اعضای تحت هم‌گروه وزن برابر می‌گیرند درحالی‌که موردهای خارج از تحت هم‌گروه وزنی نمی‌گیرند (۱۷). این بدین معنی است که موردهای خارج از تحت هم‌گروه در هیچ نقطه‌ای از زمان نقشی در مخرج تابع درست‌نمایی ندارند ولی در صورت کسر منظور می‌شوند. روش بارلو بنام روش وزن گرفته<sup>۱</sup> به موردهای خارج از تحت هم‌گروه و موردهای درون تحت هم‌گروه پیش از واقعه به‌اندازه معکوس تناسب نمونه‌گیری (۱/α) که α: تناسب نمونه‌گیری موردها از هم‌گروه اصلی است) وزن می‌دهد. همه موردها چه خارج و چه داخل تحت هم‌گروه در زمان واقعه وزنی معادل "یک" می‌گیرند اما موردهای خارج از تحت هم‌گروه قبل از واقعه وزنی نمی‌گیرند (۲۱).

نشان داده‌شده که سه روش آنالیز داده‌های مورد-هم‌گروهی در جدول شماره ۱ منجر به برآورد و خطای معیارهای مشابه می‌شوند. می‌توان نتیجه‌گیری کرد که تنها در شرایطی که اندازه تحت هم‌گروه یک درصد یا کمتر باشد برآوردها و خطاهای معیار روش پرنتمیس نسبت به دو روش دیگر برآوردهای دقیق‌تر (نسبت به کل هم‌گروه) را ایجاد می‌کند. همچنین بیشترین شباهت بین خطای معیار این روش و خطای معیار کل هم‌گروه وجود دارد. از طرف دیگر چنانچه اندازه تحت هم‌گروه خیلی کوچک باشد (کمتر از ۱۵٪ از کل هم‌گروه ۱۰۰۰ نفری) یا وقتی که هم‌گروه تقریباً کوچکی مورد استفاده قرار گیرد (کمتر از ۱۲۵۰ مشاهده) اختلافات بین سه روش وزندهی چشمگیر خواهند بود. انتخاب بین این سه روش فقط وقتی اهمیت پیدا می‌کند که اندازه هم‌گروه مورد استفاده، کارآیی طرح مورد - هم‌گروهی را محدود کند (۱۳). اگرچه اخیراً روش‌های وزن دهی معکوس احتمال<sup>۲</sup> معرفی شده‌اند که از داده‌های بیشتری مخصوصاً موارد خارج از تحت هم‌گروه برای تخمین توزیع متغیرها در تابع درست‌نمایی نما استفاده می‌کنند و بنابراین توانمندتر از روش‌های استاندارد ذکر شده هستند. (۷).

مدل‌های مخاطره نسبی کاکس (۱۹۷۲) پایه اکثر روش‌های استفاده‌شده برای محاسبه خطر نسبی در طرح مورد-هم‌گروهی است (۷). خصوصیت اصلی طرح مورد-هم‌گروهی به حساب آوردن تمامی موارد واقعه در هم‌گروه اصلی و صرف‌نظر از وضعیت حضور آن‌ها در تحت هم‌گروه است. این ویژگی، یک تورش نمونه‌گیری با توجه به وضعیت واقعه به حساب می‌آید که می‌تواند تخمین برآورد

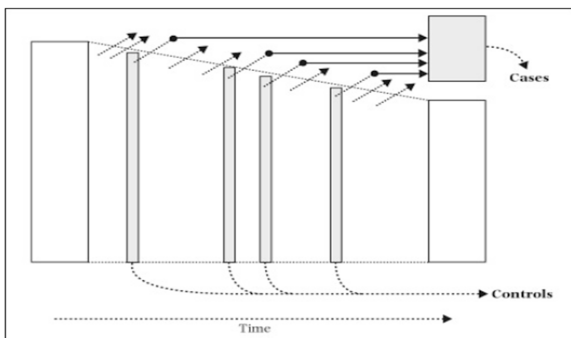
<sup>۲</sup>Invers Probability Weighting(IPW)

<sup>۱</sup> Weighted method

استراتژی همسان‌سازی چه به‌صورت هم‌زمان چه به‌صورت استفاده از یک مبدأ زمانی؛ باعث می‌شود نسبت شانس تقریبی از نسبت میزان شود (۱۰، ۱۲).

## ۲-۲- روش نمونه‌گیری

شکل شماره ۲ شمای تئوریک طرح لانه‌گزیده را نشان می‌دهد. در مطالعات مورد-شاهدی لانه‌گزیده، برای هر فرد مورد،  $m$  شاهد بدون جایگذاری در هر  $Y_i$  اخذ می‌شوند درحالی‌که  $\delta_i=1$  است (از افرادی که در زمان دچار شدن فرد  $i$  به واقعه، در معرض خطر بوده‌اند). توجه داشته باشید که شاهد‌ها هم از افراد واقعه‌دار و هم از افراد بدون واقعه هستند. نکته مهم در این طرح همسان‌سازی شاهد‌ها با موردها بر روی زمان است. اگرچه این زمان می‌تواند تعاریف متفاوتی مانند زمان از آغاز حضور در مطالعه، سن، یا تاریخ ورود به هم‌گروه را شامل شود (۷). در صورت وجود یک مواجهه دوحالتی و تحت فرض بی‌اثر بودن مواجهه (no effect)، کارآیی یک مطالعه لانه‌گزیده با  $m$  شاهد در مقایسه با مطالعه هم‌گروه برابر با  $m/m+1$  است. عواملی مانند مقدار خطر نسبی، شیوع مواجهه در گروه‌های شاهد و تعداد طبقات (strata) مورد استفاده روی کارآیی مطالعه لانه‌گزیده اثر بسزایی دارند (۲۷، ۲۸).



شکل شماره ۲- شمای تئوریک طرح مورد-شاهدی لانه‌گزیده (۱۶)

## ۲-۳- طرح Counter-matched

عدم توجه به نادر بودن مواجهه یا چولگی توزیع آن در انتخاب شاهد‌ها می‌تواند منجر به از دست رفتن کارآیی آماری شود. در این شرایط روش‌های انتخاب شاهد‌ها بر پایه مواجهه می‌توانند مفید باشند. مطالعات counter-matched اولین بار توسط لانگولز در سال ۱۹۹۳ ارائه شده است (۲۹). هدف این طرح، مطالعه اثر مخدوشگرها یا تعدیل‌کننده‌های اثر (بررسی تعامل<sup>۴</sup>) روی مواجهه

درست‌نمایی نما را دچار خطا کند. به همین دلیل تخمین بدون خطا از واریانس موضوع مهمی در طرح مورد-هم‌گروهی به حساب می‌آید (۲۲).

بر این اساس پرنتمیس و سیلف (۱۷) شرایط سازگاری<sup>۱</sup> و توزیع نرمال<sup>۲</sup> را برای تخمین‌های درست‌نمایی نما ارائه کردند. لین و پینگ (۲۳)، بارلو (۲۱)، و وکولدر (۲۴) تخمین‌های واریانس برای آنالیز رگرسیون کاکس تحت طرح مورد-هم‌گروهی را ارائه دادند و مشکلات محاسباتی تخمین واریانس توسط ثرنیو و لی (۲۵) برطرف شد. اکثر آنالیزها امروزه از راهکار "استوار"<sup>۳</sup> که توسط لی و وین پیشنهاد شده، استفاده می‌کنند. این راهکار شامل رگرسیون کاکس با وزن دهی معکوس احتمال انتخاب شدن به موردها و شاهد‌ها است.

## ۴-۱- توان آماری طرح مورد-هم‌گروهی

توان طرح مورد-هم‌گروهی وابسته به اندازه همپوشانی در مجموعه‌های افراد در معرض خطر در هر زوج از زمان‌های وقایع است. بنابراین توان در طرح مورد-هم‌گروهی به راحتی قابل محاسبه نیست مگر اینکه هم‌گروه دوباره جمع شود یعنی اینکه اعضای هم‌گروه اصلی شناسایی شده، تاریخ ورود و خروجشان از هم‌گروه مشخص شود و تاریخ تشخیص بیماری برای تمامی موردها تعیین شود. با این حال، محاسبات ساده‌ای برای حجم نمونه و توان تحت فرض نادر بودن واقعه ارائه شده است (۲۶). همچنین طرح مورد-هم‌گروهی به دلیل افزودن کارآیی در وقایع نادر، باید در هم‌گروه‌هایی استفاده شود که بزرگ باشند بنابراین در هم‌گروه‌های کوچک‌تر از ۱۲۵۰ مشاهده، آنالیز با استفاده از تحت هم‌گروه منجر به تخمین‌های دارای خطا خواهد شد (۱۳).

## مطالعه مورد-شاهدی لانه‌گزیده

### ۲-۱- تخمین اندازه اثر

در طرح لانه‌گزیده، شاهد‌ها همسان شده با موردها برحسب زمان انتخاب می‌شوند بنابراین به شرط آنکه همسان‌سازی در آنالیز نیز لحاظ شود نسبت شانس تقریبی از نسبت میزان خواهد بود (۱۲). از آنجایی‌که اغلب هم‌گروه‌ها خصوصیات یک جمعیت دینامیک را دارند، و صرف‌نظر از پایایی جمعیت، استفاده از

<sup>۱</sup> Consistency

<sup>۲</sup> Asymptotic Normality

<sup>۳</sup> Robust estimator

<sup>۴</sup> Interaction

به‌عبارت‌دیگر به دلیل عدم استفاده از یک مجموعه در معرض خطر برای وقایع متوالی (تحت هم‌گروه) در طرح لانه‌گزیده، محاسبه توان در این طرح ارتباطی به‌اندازه هم‌گروه اصلی و مقدار همپوشانی بین مجموعه‌های در معرض خطر ندارد. استفاده از این طرح در مواردی که ورودی‌های گسسته مانند گمشدگی راست تصادفی<sup>۴</sup> وجود دارد مفید است زیرا با بیش نمونه‌گیری شاهدهایی که پیگیری طولانی‌تری دارند مخصوصاً زمانی که تناسب بروز خیلی کوچک نباشد طرح لانه‌گزیده منجر به افزایش توان می‌شود (۷).

### ۳- مطالعه مورد-هم‌گروهی در مقابل لانه‌گزیده:

به‌طورکلی مزایای طرح مورد-هم‌گروهی را می‌توان ناشی از نمونه‌گیری تصادفی شاهدها از هم‌گروه و مزایای طرح لانه‌گزیده را به دلیل همسان‌سازی روی زمان (سن، زمان حضور در مطالعه، زمان ورود به مطالعه، یا ترکیبی از آن‌ها) دانست (۲۴). مقایسه کارایی آماری این دو طرح نشان می‌دهد طرح لانه‌گزیده در مطالعاتی که ورودی‌های تأخیری<sup>۵</sup> یا گمشدگی زیادی دارند ارجحیت دارد. بعلاوه اگر گمشدگی تصادفی باشد استفاده از روش وزن دهی معکوس احتمال (IPW<sup>۶</sup>) باعث افزایش کارایی طرح لانه‌گزیده خواهد شد (۳۴). روش وزن دهی معکوس احتمال به دلیل استفاده از موردهای خارج از تحت هم‌گروه در مجموعه‌های در معرض خطر در زمان‌های قبل‌تر، روش ارجحی در هر دو طرح به حساب می‌آید اگرچه کارایی آن در طرح لانه‌گزیده بیشتر است (۷). از دیگر عوامل تأثیرگذار بر اختلافات بین دو طرح، تناسب‌های بروز هستند. زمانی که تناسب بروز وقایع ناچیز باشد (کمتر از ۰.۱٪) اختلاف بین روش‌ها قابل‌چشم‌پوشی است. کارایی طرح لانه‌گزیده در شرایطی که تناسب رخداد وقایع کوچک‌تر باشد (کمتر از ۰.۱۰٪) افزایش‌یافته و انتخاب روش‌های آنالیز اختصاصی درون هر نوع طرح اهمیت کمتری پیدا می‌کند. اگرچه وقتی پیشگو کننده‌ها دوتایی باشند طرح مورد-هم‌گروهی استاندارد اغلب قوی‌تر از روش لجستیک شرطی برای طرح لانه‌گزیده است (۷، ۱۳).

طرح مورد-هم‌گروهی را می‌توان یک نوع همسان نشده از طرح

است. در این طرح لایه‌های مواجهه بر پایه تعداد شاهدهایی که برای هر مورد باید انتخاب شوند تعیین می‌شوند. در هر مجموعه در معرض خطر، در هر لایه‌ای از مواجهه که در آن موردی وجود ندارد یک شاهد انتخاب می‌شود (۳۰). این طرح را می‌توان گسترشی از طرح لانه‌گزیده دانست زیرا شاهدها از مجموعه در معرض خطر هر مورد انتخاب می‌شوند.

### ۴-۲- روش وزن دهی و آنالیز

اگر در زمان هر واقعه تمامی افراد در معرض خطر به‌عنوان گروه شاهد برای مورد انتخاب شوند تابع درستنمایی شرطی در طرح لانه‌گزیده مشابه تابع درستنمایی نیم‌رخ<sup>۱</sup> برای طرح هم‌گروه است. به‌عبارت‌دیگر شباهت نتایج آنالیز داده‌های هم‌گروه و لانه‌گزیده به این دلیل است که عملکرد تابع درستنمایی شرطی در مدل رگرسیون لجستیک شرطی در طرح لانه‌گزیده دقیقاً همانند تابع درستنمایی جزئی در مدل رگرسیون کاکس است. تنها با این تفاوت که مخرج کسر احتمال شرطی در هر زمان واقعه، بجای شامل شدن تمامی افراد در معرض خطر فقط شامل افراد گروه شاهد می‌شود (۲۸). کاتگ نشان داد برای طرح‌هایی که بیش از یک پیامد را می‌سنجند مدل حاشیه‌ای مخاطرات تناسبی جمعی<sup>۲</sup> بهتر از مدل رگرسیون کاکس عمل می‌کند (۳۱). محاسبه واریانس استوار در هر سه روش وزن‌دهی ضروری است زیرا به دلیل متد نمونه‌گیری، نقش افراد مطالعه برای حداکثر کردن تابع درستنمایی نما مستقل از هم نیست (۲۰). در طرح لانه‌گزیده انتخاب تصادفی و بدون جایگذاری افراد شاهد ضمانتی برای انتخاب ناریب آن‌ها به حساب می‌آید. قانون استقلال انتخاب شاهدها در زمان‌های واقعه از یکدیگر، نشان‌دهنده رویکرد مخاطرات تناسبی در مدل‌های بقا است (۳۲). همچنین استفاده از روش حداقل مربعات وزن گرفته<sup>۳</sup> برای مطالعه لانه‌گزیده counter-matched نیز نشان داده که افزایش کارایی در این نوع طرح با بکار بردن نمونه‌گیری دو مرحله‌ای با لحاظ کردن تناسب‌های هم‌گروه به نمونه قابل‌دستیابی است (۳۳).

### ۵-۲- توان آماری طرح

توان طرح لانه‌گزیده مستقل از اندازه هم‌گروه است.

<sup>۴</sup> Random right censoring

<sup>۵</sup> Late entry

<sup>۶</sup> Inverse Probability Weighting

<sup>۱</sup> Profile Likelihood Function

<sup>۲</sup> Marginal Additive hazards model

<sup>۳</sup> Weighted Least Square

کسب اطلاعات دقیق و صحیح از مواجهه به دلیل گذشت زمان از آغاز مواجهه، به نفع طرح مورد-هم گروهی است. به طور مثال اگر پرونده‌های جدید بیمارستانی کامپیوتری شده‌اند و پرونده‌های قدیمی‌تر نیاز به خلاصه‌سازی کاغذی دارند، طرح مورد-هم گروهی به دلیل عدم نیاز به پرونده‌های قدیمی‌تر بهتر است (۲۴).

#### ۴- نرم‌افزارهای کامپیوتری

نرم‌افزارهای کامپیوتری برای روش‌های آنالیز طرح مورد-هم گروهی و لانه گزیده در بسته‌های آماری STATA و R (۳۶)، SAS PHREG (۳۸، ۳۷)، SPSS موجود می‌باشند. کدهای آنالیز برای این نرم‌افزارها نیز توسط نویسندگان مختلف ارائه شده‌اند (۳۹، ۱۵). همچنین بسته کامپیوتری EPITOM توسط بویس و همکاران در انستیتو ملی سرطان برای داده‌های مورد-هم گروهی ارائه شده است (۲۴). The R survey package تنوعی از آنالیزها شامل تخمین ضرایب رگرسیونی در مدل‌های خطی تعمیم‌یافته را ارائه می‌نماید. همچنین The Nested Cohort Package نیز نرم‌افزار محدود به رگرسیون کاکس و تطبیق وزن‌ها با روش‌های استاندارد است (۴۰).

#### نتیجه‌گیری

به طور خلاصه مزایای طرح مورد-هم گروهی ناشی از نمونه‌گیری تصادفی شاهد‌ها است که امکان تخمین برآوردهای جمعیتی، بررسی پیامدهای متعدد و آنالیزهای متنوع را فراهم می‌کند. طرح مورد-هم گروهی در زمان کوتاه‌تری نمونه‌گیری شده و پیگیری‌های بعدی نمونه در آن عملی است. از سوی دیگر مزایای طرح لانه گزیده ناشی از همسان‌سازی روی واحد زمان است که امکان برآورد تخمین‌های اثر مانند نسبت میزان را فراهم می‌آورد. همچنین این طرح توجیه اقتصادی قوی‌تری داشته و در مواردی که هم‌گروه پویا باشد یا از دست رفتن زیادی حین پیگیری رخ بدهد طرح ارجح به حساب می‌آید. در نهایت، قضاوت در خصوص طرح برتر به عواملی نظیر اهداف مطالعه، چگونگی انتخاب اعضای هم‌گروه، چگونگی تعیین موردها و شاهد‌ها از درون هم‌گروه، و چگونگی جمع‌آوری اطلاعات متغیرها بستگی دارد. در یک موقعیت خاص عواملی که سبب اولویت‌بندی نکات مذکور و کارایی آماری می‌شوند بستگی به خصوصیات و ماهیت همان مطالعه خواهند داشت.

لانه گزیده در نظر گرفت. در طرح مورد-هم گروهی یک تحت هم‌گروه که یک نمونه تصادفی از هم‌گروه اصلی است منبع تمامی شاهد‌ها است درحالی‌که در طرح لانه گزیده شاهد‌هایی انتخاب می‌شوند که همسان شده با موردها روی زمان باشند. تعداد شاهد‌هایی که بدین ترتیب انتخاب می‌شوند می‌توانند تا ۴-۵ نفر به ازای هر مورد نیز افزایش یابند (۳۵).

از مزایای انتخاب یک تحت هم‌گروه در مطالعه مورد-هم گروهی می‌توان به توانایی تخمین فراوانی‌های جمعیتی متغیرها (مانند ژنوتیپ‌ها)، انتخاب شاهد‌ها برای پیامدهای متعدد در زمان واقعه، و انجام آنالیزها با مقیاس‌های زمانی متعدد (مانند زمان حضور در مطالعه یا سن) اشاره کرد؛ درحالی‌که در طرح لانه گزیده همسان‌سازی روی یک متغیر خاص مانند سن یا زمان از ورود به مطالعه بوده و لذا همه آنالیزها باید از آن متغیر به‌عنوان مقیاس زمانی اولیه استفاده کنند. بنابراین طرح لانه گزیده، دست محقق را برای مراحل آنالیز محدود می‌کند درحالی‌که در طرح مورد-هم گروهی همه آنالیزهای در دسترس برای کل هم‌گروه قابل اجرا هستند. نکته دیگر این است که در طرح مورد-هم گروهی یک تحت هم‌گروه می‌تواند برای مطالعه چندین پیامد مختلف مورد بررسی قرار گیرد درحالی‌که برای طرح‌هایی نظیر طرح لانه گزیده همسان‌سازی جدید موردها و شاهد‌ها برای هر پیامد مجزا مورد نیاز است (۲۴). جدا از جزییات آنالیز، طرح مورد-هم گروهی ذاتاً ساده‌تر از طرح لانه گزیده است؛ درحالی‌که ملاحظات مقتصدانه در مطالعات لانه گزیده یکی از مزایای مهم این طرح بشمار می‌آید. از طرف دیگر در طرح لانه گزیده انتخاب شاهد‌ها به دلیل انتظار برای تعیین و تأیید موردها ممکن است زمان‌بر باشد؛ اما در طرح مورد-هم گروهی انتخاب شاهد‌ها مستقل از خصوصیات موردها است و در نتیجه می‌تواند بلافاصله انجام شود (۲۴، ۴).

امکان پیگیری بعدی در همان هم‌گروه یکی از گزینه‌های مهم در انتخاب طرح مطالعه است. در مطالعه مورد-هم گروهی یک تحت هم‌گروه می‌تواند برای دوره‌های بعدی پیگیری نیز مورد استفاده قرار بگیرد. موردهای جدید پیداشده در طول پیگیری بعدی نیاز به تعیین شاهد‌های جدید ندارند و فقط اطلاعات مواجهه‌شان از زمان پایان پیگیری اولیه در اعضای تحت هم‌گروهی که قبلاً تعیین شده بودند و پس از پایان دوره پیگیری در معرض خطر بودند جمع‌آوری می‌شود. در مطالعه لانه گزیده نیز به‌روز کردن اطلاعات مواجهه برای شاهد‌هایی که از پیش تعیین شده بودند ضروری نیست؛ اما برای هر مورد جدید تعیین سابقه کامل مواجهه برای شاهد‌های جدید نیاز است. دشوار بودن

1. Knol MJ, Vandenbroucke JP, Scott P, Egger M. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *American journal of epidemiology*. 2008; 168: 1073-81.
2. Last JMe. *A Dictionary of Epidemiology*. 4th ed. New York, NY: Oxford University Press; 2000. 2000.
3. Cornfield J. A method of estimating comparative rates from clinical data. *J Natl Cancer Inst*. 1951; 11: 1269-75.
4. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986; 73: 1-11.
5. Kupper LL, McMichad AJ, Spinal R. A hybrid epidemiologic design useful in estimating relative risk. *J Am Stat Assoc*. 1975; 70: 524-28.
6. Rodrigues L, Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *Int J Epidemiol*. 1990; 19: 205-13.
7. Kim RS. A new comparison of nested case-control and case-cohort designs and methods. *Eur J Epidemiol*. 2015; 30: 197-207.
8. Green BN, Johnson CD, Adams A. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *J Sports Chiropr Rehabil*. 2001; 15: 5-19.
9. Hutchinson BG. Critical Appraisal of Review Articles. *Can Fam Physician*. 1993;39:1097-1102.
10. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *Am J Epidemiol*. 1982; 116: 547-53.
11. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Wolters Kluwer, Lipincott Williams and Wilkins, Philadelphia, USA 2008.
12. Greenland S, Thomas DC, Morgenstern H. The rare-disease assumption revisited. A critique of "estimators of relative risk for case-control studies.". *Am J Epidemiol*. 1986; 124: 869-83.
13. Onland-Moret NC, Van-der-A DL, Van-der-Schouw YT, Buschersa W, Elias SG, Van-Gils CH, et al. Analysis of case-cohort data: A comparison of different methods. *Journal of Clinical Epidemiology*. 2007; 60: 350e5.
14. Breslow NE, Cain KC. Logistic regression for two-stage case-control data. *Biometrika*. 1988; 75: 11-20.
15. Samuelsen SO, Ånestad H, Skrandal A. Stratified Case-Cohort Analysis of General Cohort Sampling Designs. *Scandinavian Journal of Statistics*. 2007;34:103-19.
16. Szklo M, Nieto FJ. *Epidemiology: beyond the basics*: Jones & Bartlett Publishers; 2014.
17. Self SG, Prentice RL. Asymptotic Distribution Theory and Efficiency Results for Case-Cohort Studies,". *The Annals of Statistics*, 1988; 16: 64-81.
18. COX DR. Partial likelihood. *Biometrika*. 1975; 62: 269-76.
19. van-den- Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. *J Clin Epidemiol*. 1990; 43: 285-95.
20. Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case-cohort designs. *Journal of Clinical Epidemiology*. 1999; 12: 1165-72.
21. Barlow W. Robust variance estimation for the case-cohort design. *Biometrics*. 1994; 50: 1064-72.
22. Cologne J, Preston DL, Imai K, Misumi M, Yoshida K, Hayashi T, et al. Conventional case-cohort design and analysis for studies of interaction. *International journal of epidemiology*. 2012; 41: 1174-86.
23. Lin D, Ying Z. Cox regression with incomplete covariate measurements. *J Amer Statist Assoc*. 1993; 88: 1341-9.
24. Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology*. 1991; 2: 155-8.
25. Therneau T, Li H. Computing the Cox model for case cohort designs. *Lifetime Data Anal*. 1999; 5: 99-112.
26. Cai J, Zeng D. Sample size/power calculation for case-cohort studies. *Biometrics*. 2004; 60: 1015-24.
27. Ury H. Efficiency of case-control studies with multiple controls per case: continuous or dichotomous data. *Biometrics*. 1975; 31: 643-9.
28. Holakoei -Naeini K, Ostovar A, Danesh A, SajediNezhad S, Ghalichi L, Moradi G, et al. The comparison between nested case-control and cohort analysis on Tuberculosis data: an experiment. *Iranian journal of Epidemiology*. 2010; 6: 1-6.
29. Langholz B, Borgan O. Counter-Matching: A Stratified Nested Case-Control Sampling Method. *stat res*. 1993; 3.
30. Cologne JB, Sharp GB, Neriishi K, Verkasalo PK, Land CE, Nakachi K. Improving the efficiency of nested case-control studies of interaction by selecting controls using counter matching on exposure. *International Epidemiological Association*. 2004; 33: 3.(
31. Kang S, Cai J, Chambless L. Marginal additive hazards model for case-cohort studies with multiple disease outcomes: an application to the Atherosclerosis Risk in Communities (ARIC) study *Biostatistics*. 2011; 14: 28-41.
32. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies: III. Design options. *American journal of epidemiology* 1992; 135: 1042-50.
33. White JE. A two stage design for the study of the relationship between a rare exposure and a rare disease. *Am J Epidemiol*. 1982; 115: 119-28.
34. Samuelsen S. A pseudo-likelihood approach to analysis of nested case-control studies. *Biometrika*. 1997; 84: 379-94.
35. Ernster VL. Nested case-control studies. *Preventive medicine*. 1994; 23: 587-90.
36. Team RDC: R: A language and environment for statistical computing 2006 [http://www.R-project.org]. Vienna, Austria: R Foundation for Statistical Computing.
37. Inc SI: SAS/STAT User's Guide, Version 9.2. 2004.
38. Langholz B, Jiao J. Computational methods for case-cohort studies. *Computational Statistics & Data Analysis*. 2007; 51: 3737 - 48.
39. Kulathinal S, Karvanen J, Saarela O, Kuulasmaa K. Case-cohort design in practice - experiences from the MORGAM Project. *Epidemiologic Perspectives & Innovations*. 2007; 4.
40. Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. Using the Whole Cohort in the Analysis of Case-Cohort Data. *American Journal of Epidemiology*. 2009; 169.



# Nested-Case Control Versus Case-Cohort Studies: A Narrative Review on Methodological Considerations

Holakouie Naieni K<sup>1</sup>, Mansournia MA<sup>1</sup>, Panahi MH<sup>2</sup>, Elduma A<sup>2</sup>, Nematollahi SH<sup>2</sup>

1- Epidemiology and Biostatistics Department, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

2- PhD Student in Epidemiology, Epidemiology and Biostatistics Department, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

**Corresponding author:** Nematollahi SH, shahrzadnema@yahoo.com

Case-Cohort and nested case-control designs are the most well-known and cost-effective derivatives of prospective studies. Due to the popularity of these designs in recent years, the aim of this study is to provide a comprehensive and up-to-date description of them with respect to methodological issues such as sampling scheme, effect measure estimation, and analysis strategies. Overall, random sampling of controls in case-cohort and matching for time in nested design are considered their hallmarks. Eventually, decision on a better design depends on study objectives, sub-cohort selection, case and control determination, and method of data collection. It should be kept in mind that any particular study requires unique elaboration and delicate examination to choose from the designs.

**Keywords:** Case-cohort, Nested case-control, Epidemiologic study design, Effect measure