

# برآورد میزان بروز سرطان معده در استان اردبیل با استفاده از روش صید- بازصید براساس اطلاعات ثبت سرطان مبتنی بر جمعیت

محمود خدادوست<sup>۱</sup>، پروین یآوری<sup>۲</sup>، سید سعید هاشمی نظری<sup>۳</sup>، مسعود بابایی<sup>۴</sup>، علیرضا ابدی<sup>۵</sup>، فاطمه سروی<sup>۶</sup>

<sup>۱</sup> کارشناس ارشد اپیدمیولوژی، گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی شهید بهشتی، ایران

<sup>۲</sup> استاد اپیدمیولوژی، گروه بهداشت و پزشکی اجتماعی، دانشکده پزشکی، دانشگاه علوم پزشکی شهید بهشتی، ایران

<sup>۳</sup> عضو انجمن علمی اپیدمیولوژیست‌های ایران

<sup>۴</sup> مرکز تحقیقات پیشگیری از مصدومیت‌ها و ارتقای ایمنی، دانشگاه علوم پزشکی شهید بهشتی، ایران

<sup>۵</sup> استادیار اپیدمیولوژی، گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی شهید بهشتی، ایران

<sup>۶</sup> پزشک عمومی، معاونت سلامت، دانشگاه علوم پزشکی اردبیل، ایران

<sup>۷</sup> مرکز تحقیقات سرطان آلمان (DKFZ)

<sup>۸</sup> دانشیار آمار زیستی، گروه بهداشت و پزشکی اجتماعی، دانشکده پزشکی، دانشگاه علوم پزشکی شهید بهشتی، ایران

<sup>۹</sup> دانشجوی دکتری آمار زیستی، گروه اپیدمیولوژی و آمار، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، ایران

نویسنده رابط: پروین یآوری، نشانی: گروه بهداشت و پزشکی اجتماعی، دانشکده پزشکی، دانشگاه علوم پزشکی شهید بهشتی، اوین، تهران، تلفن: ۲۳۸۷۲۵۶۷

پست الکترونیک: p.yavari-grc@sbmu.ac.ir

تاریخ دریافت: ۹۳/۲/۲۷؛ پذیرش: ۹۳/۴/۷

**مقدمه و اهداف:** آگاهی از میزان بروز سرطان برای برنامه‌های پیش‌گیری و کنترل سرطان ضروری است. روش صید- بازصید برای کاهش سوگرایی و افزایش دقت برآورد بروز سرطان توصیه شده است. هدف از این مطالعه، برآورد میزان بروز سرطان معده با روش صید- بازصید با استفاده از داده‌های ثبت سرطان مبتنی بر جمعیت در استان اردبیل بوده است.

**روش کار:** تمامی موارد جدید سرطان معده گزارش شده از سه منبع پاتولوژی، مستندات بیمارستانی و گواهی مرگ به مرکز ثبت سرطان مبتنی بر جمعیت استان اردبیل در سال‌های ۱۳۸۵ و ۱۳۸۷ به مطالعه وارد گردید. موارد مشترک بین منابع براساس تشابه نام، نام خانوادگی و نام پدر شناسایی شدند و در نهایت برآورد بروز توسط روش لگ خطی و با استفاده از نرم‌افزار آماری Stata 12 محاسبه شد.

**نتایج:** در مجموع ۸۵۷ مورد جدید سرطان معده از سه منبع گزارش شدند. میزان بروز گزارش شده پس از حذف موارد تکراری برای سال‌های ۱۳۸۵ و ۱۳۸۷ به ترتیب برابر با ۳۵/۳ و ۳۲/۵ درصد هزار نفر بود و میزان بروز برآورد شده برای این سال‌ها به وسیله روش لگ خطی به ترتیب ۹۶/۲ و ۹۰/۴ در صدهزار نفر برآورد شد.

**نتیجه‌گیری:** نتایج این مطالعه نشان داد که هیچ‌یک از منابع پاتولوژی، مستندات بیمارستانی و ثبت مرگ به تنهایی و یا در مجموع پوشش کاملی از موارد بروز سرطان معده را نداشته‌اند و با استفاده از روش صید- بازصید می‌تواند میزان بروز دقیق‌تری را برآورد نمود.

**واژگان کلیدی:** صید- بازصید، برآورد بروز، لگ خطی، سرطان معده، ثبت سرطان

## مقدمه

که مبنای سیاست‌ها و تصمیم‌گیری‌های آنان را تشکیل می‌دهد؛ اعتماد کنند. بنابراین، مقدار کامل‌شماری ثبت<sup>۱</sup> به عنوان یکی از معیارهای سنجش کیفیت ثبت سرطان مورد استفاده قرار می‌گیرد (۲،۳). کامل‌شماری به عنوان نسبت موارد بروزی که ثبت شده‌اند، تعریف می‌شود (۲). کامل بودن ثبت سرطان به عنوان یکی از

آگاهی از اطلاعات میزان دقیق بروز سرطان برای برنامه‌ریزی، پایش و ارزشیابی برنامه‌های ملی و منطقه‌ای کنترل سرطان ضروری است (۱). هدف از ثبت سرطان مبتنی بر جمعیت، برآورد بار ناشی از سرطان در جمعیت تحت پوشش به منظور مشاهده روند و تفاوت‌های منطقه‌ای بیماری و همچنین تهیه یک بانک اطلاعات به منظور پژوهش‌های اپیدمیولوژیک است (۲). سیاست‌گذاران بهداشتی نیاز دارند بدانند که چگونه به اطلاعاتی

<sup>۱</sup>Completeness of registration

نفر برای زنان به خود اختصاص داده است. در ایران، بیش‌تر مناطق شمالی و شمال‌غربی مناطق با خطر بالای سرطان معده هستند؛ در حالی‌که استان‌های مرکزی و غربی خطر متوسط، و مناطق جنوبی کم‌خطر محسوب می‌شوند (۱۱).

پژوهش حاضر با هدف برآورد میزان بروز سرطان معده در استان اردبیل با مقایسه سه منبع ثبت داده‌ها (شامل اطلاعات پاتولوژی، گواهی مرگ و مستندات بیمارستانی) و استفاده از روش صید- بازصید انجام شده است و از نخستین مطالعه‌های در زمینه سرطان با این روش در کشور محسوب می‌شود. با توجه به این‌که بر اساس گزارش وزارت بهداشت در سال ۱۳۸۷، سرطان معده شایع‌ترین سرطان در مردان و دومین سرطان شایع در زنان در استان اردبیل محسوب می‌شود، این سرطان برای انجام این مطالعه انتخاب گردید.

## روش کار

این تمامی موارد جدید سرطان معده در سال‌های ۱۳۸۵ و ۱۳۸۷ که توسط سه منبع اطلاعات پاتولوژی، مستندات بیمارستانی و گواهی‌های مرگ در استان اردبیل ثبت شده بودند؛ وارد مطالعه شد و از آن‌جایی‌که اطلاعات ثبت شده در سال ۱۳۸۶ ناقص بود و از کیفیت مناسب برخوردار نبود، از مطالعه خارج شد. ابتدا موارد تکراری در هر منبع با استفاده از نرم‌افزار اکسل ۲۰۰۷ شناسایی و حذف شدند. سپس موارد مشترک بین سه منبع با استفاده از اطلاعات بیماران مانند نام، نام خانوادگی، نام پدر و کد ICD مربوط به نوع سرطان آن‌ها و در صورت نیاز تاریخ تولد، محل تولد و سکونت آن‌ها در نرم‌افزار اکسل شناسایی شدند. موارد مشترک بین سه منبع با مقایسه دو به دو آن‌ها و پس از ردیف کردن اسامی به ترتیب‌های مختلف (به ترتیب نام خانوادگی، نام پدر و شهر محل سکونت) با دستور Sort و مقایسه هر مورد با افراد ثبت شده در منبع دیگر شناسایی شدند. میزان بروز سرطان معده توسط روش‌گیر و بازگیر و مدل لگاریتم خطی برآورد زده شد. در این روش، دو فرض اصلی باید برقرار باشد. اول این‌که منابع اطلاعات باید مستقل باشند و دوم این‌که همه افرادی که در هر منبع اطلاعاتی وجود دارند؛ باید شانس برابری برای حضور در مطالعه داشته باشند (۹ و ۷)، اما در بیش‌تر مطالعه‌های جوامع انسانی و حوزه علوم پزشکی معمولاً این فرض‌ها به طور کامل برقرار نیستند، و منابع مختلف مستقل از هم نیستند. تجزیه

اجزای مهم کنترل کیفیت در این منبع ثبت محسوب می‌شود (۴). در متون علمی، چندین روش مختلف به منظور ارزیابی مقدار کامل‌شماری بیان شده است که به دو گروه کلی تقسیم می‌شوند: روش‌های کمی و روش‌های کیفی. روش‌های کیفی مورد استفاده شامل نسبت مرگ‌ومیر به بروز و نسبت موارد تأیید میکروسکوپی، و از بین روش‌های کمی آن‌هایی که مورد استفاده قرار گرفته‌اند؛ شامل روش صید- بازصید<sup>۱</sup>، گواهی‌های مرگ و نسبت مرگ‌ومیر به بروز، و روش جریان<sup>۲</sup> است (۵). از آن‌جایی‌که بیش‌تر سامانه‌های ثبت سرطان از بیش از یک منبع اطلاعات برای تشخیص موارد بیماری استفاده می‌کنند، روش‌های صید- بازصید ممکن است برای برآورد تعداد موارد بروز در جمعیت مورد استفاده قرار گیرند. از این رو، به منظور بررسی مقدار کامل بودن ثبت موارد بیماری مورد استفاده قرار می‌گیرد (۶). روش صید- بازصید به طور گسترده در سرشماری حیات وحش مورد استفاده قرار گرفته است (۷). کاربرد مهم دیگر این روش در پژوهش‌های اپیدمیولوژی برآورد میزان شیوع یک بیماری خاص و هم‌چنین برآورد مقدار کامل‌شماری ثبت بیماری است، اما روش صید- بازصید می‌تواند به ویژه در شرایطی دو و یا بیش از دو فهرست ناکامل از بیماران وجود داشته باشد؛ مورد استفاده قرار گیرد (۸). در هنگام استفاده از روش صید- بازصید ساده دو فرض باید مورد توجه قرار گیرند: اول این‌که منابع مستقل از یکدیگر باشند، دوم این‌که تمامی افراد درون یک منبع مشابه شانس برابری برای حضور داشته باشند (۹، ۷). استفاده از روش صید- بازصید برای کاهش هزینه‌های ثبت بیماری بسیار کارآیی دارد و هم‌چنین در کاهش سوگرایی در برآورد بروز و در موارد مقایسه زیرگروه‌های جمعیت کاربرد دارد (۴). ثبت سرطان مبتنی بر اطلاعات پاتولوژی از سال ۱۳۷۸ در استان اردبیل راه‌اندازی شد. برای نخستین بار در ایران ثبت سرطان مبتنی بر جمعیت در سال ۱۳۸۲ در اردبیل راه‌اندازی شد. فعالیت‌های مرکز ثبت سرطان اردبیل شامل جمع‌آوری اطلاعات میزان بروز سرطان از منابع پاتولوژی، بیمارستانی و گواهی‌های مرگ است. هدف اصلی مرکز ثبت سرطان اردبیل اندازه‌گیری میزان بروز و مرگ‌ومیر سرطان در ساکنان استان اردبیل است (۱۰). استان اردبیل که در شمال‌غربی کشور واقع شده است، بیش‌ترین میزان سرطان معده را در ایران با  $ASR=49/1$  در صدهزار نفر برای مردان و  $ASR=25/4$  در صدهزار

<sup>۱</sup> Capture recapture

<sup>۲</sup> Flow method

<sup>۳</sup> Age Specific Rate

$$G^2 = -2 \sum \text{Obs}_j \ln[\text{Obs}_j / \text{Exp}_j]$$

که در آن  $\text{Obs}_j$  تعداد افراد مشاهده شده در هر خانه  $j$ ، و  $\text{Exp}_j$  تعداد موارد مورد انتظار افراد در هر خانه  $j$  تحت مدل  $i$  است که هر چقدر مقدار آماره  $G^2$  کم تر باشد، مدل برازش شده مناسب تر خواهد بود.

$$\text{AIC} = G^2 - 2 [\text{df}]$$

اصطلاح اول "G2" معیاری است برای این که چگونه مدل متناسب با داده ها برازش شده است و اصطلاح دوم "df" به منظور پارامترهای اضافه و هم چنین پیچیدگی مدل است.

$$\text{BIC} = G^2 - [\ln N_{\text{obs}}] [\text{df}]$$

که در آن  $N_{\text{obs}}$  برابر است با تعداد کل افراد مشاهده شده. BIC نسبت به AIC به طور سنگین تری تحت الشعاع پیچیدگی مدل ها قرار می گیرد (۱۲).

شایع ترین معیار برای انتخاب مدل مناسب AIC است؛ اگر چه برخی از پژوهشگران پیشنهاد کرده اند در صورتی که حدود اطمینان مدل اشباع کم باشد، بهتر است از برآورد مدل اشباع استفاده شود (۱۵-۱۳).

در مطالعه حاضر به منظور انتخاب مدلی که بهترین برازش را برای داده ها دارد، از آماره های AIC، BIC و  $G^2$  استفاده شد و در نهایت براساس آماره AIC که بیشترین کاربرد را توسط پژوهشگران دارد، مدلی که شامل یک اثر متقابل بین منابع پاتولوژی و مستندات بیمارستانی و یک اثر مستقل منبع ثبت مرگ بود (مدل شماره ۲) انتخاب شد.

در این مطالعه برای کنترل کردن سوگرایی ناشی از عدم استقلال منابع از روش لگاریتم خطی با استفاده از سه منبع اثر متقابل بین منابع می توان وابستگی بین منابع مختلف را کنترل نمود. از آن جایی که ممکن است افراد بیمار به علت متفاوت بودن شدت و مرحله بیماری، شانس برابری برای حضور در منبع ثبت مرگ در مقایسه با منابع ثبت پاتولوژی و مستندات بیمارستانی را نداشته باشند، در این مطالعه سرطان معده که میزان کشندگی بالایی دارد؛ انتخاب گردید تا شانس حضور افراد در منبع ثبت

و تحلیل به صورت سه منبعی و با استفاده از مدل لگ خطی<sup>۱</sup> با نرم افزار Stata 12 (StataCorp, Texas, USA) انجام شد. با سه منبع ثبت ۸ ترکیب احتمالی از این منابع وجود داشت، که در آن احتمال حضور یا عدم حضور موارد در این منابع وجود دارد. به منظور تجزیه و تحلیل ابتدا یک جدول توافقی با ۸ خانه ( $2^3 = 8$ ) که ۷ خانه آن مشخص بودند؛ تشکیل شد (نمودار شماره ۱). برای برآورد فراوانی خانه هشتم در جدول توافقی (افرادی که در هیچ کدام از ۳ منبع ثبت نشده اند)، ۸ مدل لگ خطی مختلف روی این هفت خانه جدول برازش شدند و با استفاده از هر یک از این مدل ها، فراوانی منتظره خانه هشتم برآورد گردید. هنگامی که از سه منبع یا بیش تر استفاده می شود، در صورت برقرار نبودن فرض استقلال منابع، می توان با وارد کردن اثر متقابل بین منابع مختلف در مدل لگ خطی، سوگرایی ناشی از برقرار نبودن این پیش فرض را کنترل کرد. برحسب نوع و تعداد اثرهای متقابل، هشت مدل مختلف لگ خطی وجود داشت:

یک مدل مستقل، سه مدل دارای یک اثر متقابل؛ سه مدل دارای دو اثر متقابل و یک مدل دارای سه اثر متقابل (مدل اشباع). برای ۳ منبع ثبت، A با i سطح، B با j سطح، C با k سطح، لگاریتم طبیعی (Ln یا log e) فراوانی مورد انتظار  $F_{ijk}$  برای خانه  $ijk$ ، به صورت زیر نمایان شد:

$$\ln F_{ijk} = \theta + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

که در آن  $\theta$  پارامتر مشترک، A، B، C پارامترهای اثر اصلی، AB، AC و BC پارامترهای برهم کنش های دو طرفه و ABC پارامتر بر هم کنش سه طرفه (بالاترین اثر مورد نظر) می باشد. برای ارزیابی چگونگی برازش مدل های لگ خطی با داده ها (برازش مدل)<sup>۲</sup> و انتخاب بهترین مدل، از لگاریتم نسبت درست نمایی- که به عنوان آماره  $G^2$  نیز شناخته می شود، معیار اطلاعات آکائیک (AIC)<sup>۳</sup> و معیار اطلاعات بیزین (BIC)<sup>۴</sup> استفاده شد که نحوه محاسبه آن ها به صورت زیر است:

<sup>۱</sup> Loglinear  
<sup>۲</sup> Model fitting  
<sup>۳</sup> Akaike's Information Criterion  
<sup>۴</sup> Bayesian Information Criterion

مبنای آماره BIC نیز همین مدل دارای بیشترین میزان برازندگی مدل بود. از میان معیارهای برازندگی AIC بیش از BIC توسط پژوهشگران استفاده و توصیه می‌شود. بنابراین با استفاده از AIC که برآوردهای محافظه‌کارانه‌تری ارائه می‌دهد، تعداد کل موارد بیماری که توسط هیچ‌یک از منابع ثبت نشده‌اند ۱۵۰۰ نفر تخمین زده شد. در نتیجه برآورد تعداد کل موارد سرطان معده در سال‌های ۱۳۸۵ و ۱۳۸۷ برابر  $2357/35$  (حدود اطمینان ۹۵ درصد:  $2019/12 - 2794/04$ ) نفر شد. بر همین اساس، مقدار کامل بودن ثبت برای هر سه منبع پس از حذف موارد تکراری (۸۵۷ مورد) برابر است با  $36/35$  درصد و برای هر یک از منابع پاتولوژی با ۴۳۹ مورد، مستندات بیمارستانی با ۳۶۱ مورد، و گواهی‌های مرگ با ۳۷۸ مورد به ترتیب  $18/62$ ،  $15/31$  و  $16/04$  درصد برآورد شد. میزان بروز سرطان معده در استان اردبیل براساس تعداد موارد برآورد شده بیماری در سال‌های ۱۳۸۵ و ۱۳۸۷ و هم‌چنین بر اساس جمعیت کل ساکن استان اردبیل در این سال‌ها، یک‌بار به صورت کلی و بار دیگر در زیرگروه‌های سنی و جنسی مختلف به صورت مجزا محاسبه گردید (جدول شماره ۳). برآورد میزان بروز سرطان معده در سال ۱۳۸۵ برابر با  $96/2$  در صد هزار نفر (حدود اطمینان ۹۵ درصد:  $78/4 - 121/9$ ) و در سال ۱۳۸۷ برابر با  $90/4$  در صد هزار نفر (حدود اطمینان ۹۵ درصد:  $72/5 - 116/4$ ) محاسبه گردید. بیشترین میزان بروز برآورد شده در بین گروه‌های سنی مربوط به افراد ۶۵ سال و بالاتر و برابر با  $1238/3$  در صد هزار نفر (حدود اطمینان ۹۵ درصد:  $942/6 - 1700/9$ ) برای سال ۱۳۸۵ و  $1042/4$  در صد هزار نفر (حدود اطمینان ۹۵ درصد:  $812/7 - 1397/7$ ) برای سال ۱۳۸۷ بود.

بیش‌تر باشد و بتوان منبع ثبت را نیز مورد استفاده قرار داد. هم‌چنین در این مطالعه بروز بیماری بر اساس لایه‌های مختلف متغیرهای تعدیل‌کننده مانند سن و گروه‌های سنی به طور جداگانه محاسبه شد.

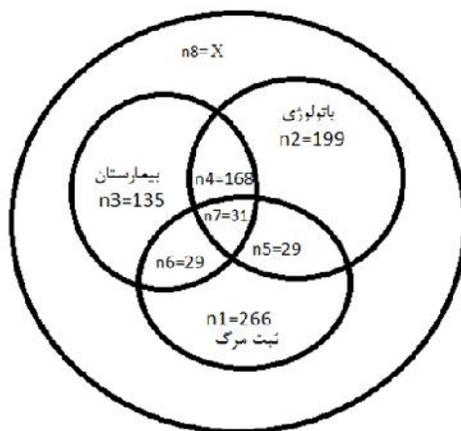
در تمامی مراحل این مطالعه اطلاعات افراد شامل نام و نام‌خانوادگی، شماره تلفن، آدرس و سایر اطلاعات آن‌ها به صورت محرمانه باقی ماند. به منظور انتخاب مدل و برآورد میزان بروز از نرم‌افزار Stata 12 استفاده شد.

## یافته‌ها

پس از بررسی و حذف موارد تکراری بین سه منبع ثبت سرطان، در مجموع ۸۵۷ مورد جدید سرطان معده در سال‌های ۱۳۸۵ و ۱۳۸۷ به مرکز ثبت سرطان استان اردبیل گزارش شده است که ۴۳۹ مورد توسط منبع پاتولوژی، ۳۶۱ مورد توسط مستندات بیمارستانی و ۳۷۸ مورد توسط گواهی‌های مرگ گزارش شده‌اند. از این تعداد ۵۸۷ ( $68/49$  درصد) مورد آن‌ها مرد و ۲۷۰ ( $31/51$  درصد) مورد زن بوده‌اند، و نسبت مرد به زن برابر با  $2/17$  بود. میانگین کلی سن افراد  $66/95 \pm 12/16$  سال ( $67/37 \pm 11/68$  سال برای مردان و  $66/04 \pm 13/13$  سال برای زنان) بود.

نمودار شماره ۱ تعداد موارد مشترک بین سه منبع پاتولوژی، مستندات بیمارستانی و ثبت مرگ را نشان می‌دهد که کل موارد گزارش شده (با یک‌بار احتساب موارد مشترک) ۸۵۷ نفر می‌باشد. جدول شماره ۱ مقایسه برخی ویژگی‌های دموگرافیک بیماران که توسط سه منبع ثبت سرطان گزارش شده‌اند (پس از حذف موارد تکراری) را بین سال‌های ۱۳۸۵ و ۱۳۸۷ نشان می‌دهد. ملاحظه می‌شود که توزیع سنی و جنسی بیماران در سال ۱۳۸۵ و ۱۳۸۷ تفاوت آماری معنی‌داری با هم ندارند و به طور کلی بیشترین موارد بیماری در جنس مرد با ۵۸۷ ( $68/49$  درصد) مورد و گروه سنی بالای ۶۵ سال با ۵۶۱ ( $65/46$  درصد) مورد بوده است.

نتایج تجزیه و تحلیل سه منبعی توسط مدل لگ خطی در جدول شماره ۲ آمده است. از بین ۸ مدل برازش شده بر اساس آماره AIC، مدل دوم شامل یک اثر متقابل بین منابع پاتولوژی و مستندات بیمارستانی و یک اثر مستقل منبع ثبت مرگ که بیشترین برازندگی را داشت انتخاب گردید. در صورتی که بر



نمودار شماره ۱- نمودار ون تعداد موارد مشترک سرطان معده در استان اردبیل که در سه منبع پاتولوژی، مستندات بیمارستانی و گواهی مرگ ثبت شده‌اند.

جدول شماره ۱- مقایسه اطلاعات دموگرافیک بیماران گزارش شده از سه منبع ثبت سرطان در سال‌های ۱۳۸۵ و ۱۳۸۷

متغیر	سال ۱۳۸۵ (n=۴۴۴)		سال ۱۳۸۷ (n=۴۱۳)		p.value	تعداد کل (n=۸۵۷)	
	تعداد	درصد	تعداد	درصد		تعداد	درصد
جنس	مرد	۳۰۱	۶۷/۷۹	۲۸۶	۰/۶۴	۵۸۷	۶۸/۴۹
	زن	۱۴۳	۳۲/۲۱	۱۲۷	۳۰/۷۵	۲۷۰	۳۱/۵۹
سن	۵۰≤	۴۸	۱۰/۸۱	۳۹	۰/۶۸	۸۷	۱۰/۱۵
	۶۴-۵۱	۱۱۱	۲۵/۰۰	۹۸		۲۰۹	۲۴/۳۹
	۶۵≥	۲۸۵	۶۴/۱۹	۲۷۶		۵۶۱	۶۵/۴۶

جدول شماره ۲- مشخصات مدل‌های لگ خطی برازنده شده به داده‌های سه منبع ثبت سرطان معده در سال‌های ۱۳۸۵ و ۱۳۸۷

*AIC	*BIC	*G2	**df	فاصله اطمینان ۹۵ درصد برای N	***N	***X	مدل لگ خطی
۲۴۳/۲۷	۲۴۳/۰۶	۱۹۱/۳۱	۴	(۱۴۷۳/۹۴- ۱۲۸۸/۲۴)	۱۳۷۲/۸۰	۵۱۵/۸۰	P/C/D
۵۵/۸۶	۵۵/۵۹	۱/۹۰	۵	(۲۷۹۴/۰۴- ۲۰۱۹/۱۲)	۲۳۵۷/۳۵	۱۵۰۰/۳۵	PC/D
۱۷۵/۶۶	۱۷۵/۳۹	۱۲۱/۷۰	۵	(۱۲۲۵/۲۸- ۱۰۸۹/۳۰)	۵۰. ۱۱۴۹	۲۹۲/۵۰	PD/C
۲۱۳/۹۶	۲۱۳/۶۹	۱/۶۰	۵	(۱۳۲۱/۰۹- ۱۱۶۰/۵۰)	۱۲۳۲/۳۰	۳۷۵/۳۰	CD/P
۵۶/۶۷	۵۶/۳۵	۰/۷۱	۶	(۲۷۳۹/۲۸- ۱۶۷۱/۶۰)	۲۰۹۵/۲۷	۱۲۳۸/۲۷	PC/PD
۵۶/۲۴	۵۵/۹۲	۰/۲۸	۶	(۳۶۰۱/۰۷- ۲۰۷۱/۱۶)	۲۶۸۲/۳۱	۱۸۲۵/۳۱	PC/CD
۱۰۵/۸۲	۱۰۵/۴۹	۴۹/۸۶	۶	(۱۰۶۵/۵۹- ۹۷۹/۵۹)	۱۰۱۶/۹۱	۱۵۹/۹۱	PD/CD
۵۷/۹۶	۵۷/۵۸	۰/۰۰	۷	(۳۹۷۸/۰۷- ۱۶۴۴/۶۷)	۲۴۲۴/۹۲	۱۵۶۷/۹۲	PC/PD/CD

\* Akaike's Information Criterion /Bayesian Information Criterion/ Goodness of fit

\*\* درجه آزادی

\*\*\* برآورد تعداد موارد سرطان معده که در هیچ‌یک از منابع ثبت نشده‌اند.

\*\*\*\* برآورد تعداد کل موارد سرطان معده در استان اردبیل در سال‌های ۱۳۸۵ و ۱۳۸۷

جدول شماره ۳- برآورد میزان بروز سرطان معده با استفاده از روش لگ خطی بر مبنای جمعیت استان اردبیل در سال ۱۳۸۵ و ۱۳۸۷

سال	زیرگروه‌ها	بروز گزارشی شده* (در ۱۰۰۰۰۰ نفر) (I1)	برآورد شده بروز (I2)	فاصله اطمینان ۹۵ درصد برای I2
۱۳۸۵	جنسیت	مرد	۱۱۶/۲	(۹۲/۷- ۱۵۱/۶)
		زن	۸۱/۵	(۵۲/۴- ۱۳۹/۵)
	گروه‌های سنی	≤ ۵۰	۱۳/۶۲	(۷/۲- ۳۴/۲)
		۵۱-۶۴	۲۵۷/۲	(۱۸۲/۱- ۳۷۳/۳)
		≥ ۶۵	۱۲۳۸/۳	(۹۴۲/۶- ۱۷۰۰/۹)
	کل		۹۶/۲	(۷۸/۱- ۱۲۱/۹)
۱۳۸۷	جنسیت	مرد	۱۰۷/۴	(۸۵/۰- ۱۴۱/۸)
		زن	۸۰/۹	(۴۹/۱- ۱۴۸/۴)
	گروه‌های سنی	≤ ۵۰	۱۲/۲	(۵/۸- ۳۶/۵)
		۵۱-۶۴	۲۷۵/۲	(۱۷۳/۱- ۵۱۳/۶)
		≥ ۶۵	۱۰۴۲/۴	(۸۱۲/۷- ۱۳۹۷/۷)
	کل		۹۰/۴	(۷۲/۵- ۱۱۶/۴)

\* بروز گزارش شده توسط سه منبع پاتولوژی، مستندات بیمارستانی و گواهی‌های مرگ پس از حذف موارد تکراری/بروز گزارش شده برحسب گروه‌های سنی و جنسی و همچنین بروز کل بر حسب تعداد افراد در معرض خطر در استان اردبیل طی سال‌های ۱۳۸۵ و ۱۳۸۷ محاسبه شده است.

## بحث

ارجاعی از استان‌های مجاور باشد. این نسبت نشان دهنده بروز بسیار متفاوت در زیر گروه‌های جنسی در اردبیل است؛ در حالی که در فنلاند و اتحادیه اروپا این مقدار به ترتیب برابر با ۱/۳۴ و ۱/۶۶ است و تفاوت کمتری در بروز جنسی در اروپا دیده می‌شود (۲۲).

میزان بروز گزارش شده بر اساس سه منبع پاتولوژی، مستندات بیمارستانی و ثبت مرگ و پس از حذف موارد تکراری برای سال‌های ۱۳۸۵ و ۱۳۸۷ به ترتیب برابر با ۳۵/۳۲ و ۳۲/۵۰ در صد هزار نفر بود. با استفاده از مدل لگاریتم خطی میزان بروز برای این سال‌ها به ترتیب ۹۶/۲۲ و ۹۰/۴۰ برآورد شد. همچنین بروز گزارش شده (بر مبنای سه منبع ثبت و پس از حذف موارد تکراری) برای مردان و زنان به ترتیب برابر با ۴۶/۶ و ۲۳/۳ در صد هزار نفر برای سال ۱۳۸۷ بود که برآورد این میزان‌ها با استفاده از روش لگ خطی برای مردان و زنان به ترتیب ۱۱۶/۲ و ۸۱/۵ در صد هزار نفر برای سال ۱۳۸۵ و ۱۰۷/۴ و ۸۰/۹ در صد هزار نفر برای سال ۱۳۸۷ بود. طبق گزارش GLOBOCAN در سال ۲۰۰۸ میلادی، میزان برآورد شده بروز سرطان معده برای مردان و زنان ایرانی به ترتیب ۲۱/۹ و ۹ در صد هزار نفر بوده است که خیلی کمتر از میزان محاسبه شده در این مطالعه است (۲۳)، اما باید

در مطالعه حاضر میزان بروز سرطان معده توسط روش گیر و بازگیر و مدل لگاریتم خطی برآورد شد. میانگین سنی برای مردان و زنان به ترتیب  $67/37 \pm 11/68$  و  $66/04 \pm 13/13$  سال بود که تفاوت زیادی از نظر توزیع سنی بروز بیماری بین مردان و زنان مشاهده نشد، اما نسبت به مطالعه‌های مشابه انجام شده در ایران که میانگین سنی افراد ۶۵-۵۸ سال گزارش شده، کمی بالاتر است (۱۹-۱۶)، اما مطالعه مشابه در تهران که یک مرکز ارجاعی از سایر استان‌ها در ایران محسوب می‌شود؛ میانگین سنی برای مردان و زنان به ترتیب ۶۸/۲ و ۶۷/۷ سال گزارش شده است که با مطالعه حاضر هم‌خوانی دارد (۲۰).

نسبت مرد به زن برابر با ۲/۱۷ بود که این مقدار با گزارش کشوری سال‌های ۱۳۸۴ و ۱۳۸۵ (به ترتیب ۲/۵۹ و ۲/۶۱) هم‌خوانی دارد. همچنین با نسبت جنسی گزارش شده در استان اردبیل در سال‌های ۱۳۸۳ تا ۱۳۸۵ هم‌خوانی دارد (۲۱)، که نشان دهنده‌ی عدم تغییر بروز جنسی در طی سال‌های ۱۳۸۵ و ۱۳۸۷ در این استان است. اما در مطالعه مشابه انجام شده در تهران در سال‌های ۸۵-۱۳۸۲ نسبت جنسی مرد به زن ۱/۹۹ گزارش شده است، که این نسبت کمتر می‌تواند ناشی از موارد

دقیق‌تری از میزان بروز که به واقعیت نزدیک‌تر باشد؛ را نشان می‌دهد.

میزان بروز برآورد شده در این مطالعه با روش صید- بازصید و براساس سه منبع ثبت سرطان صورت گرفته است که به نظر می‌رسد به واقعیت نزدیک‌تر باشد، و در مقایسه با میزان بروز گزارش شده از ثبت سرطان مبتنی بر جمعیت استان اردبیل دقیق‌تر است، و وضعیت سرطان معده را در استان اردبیل بهتر نشان می‌دهد. باید توجه داشت که در این مطالعه با توجه به این‌که ممکن است تعدادی از موارد مرگ ثبت شده که در منبع پاتولوژی همان سال حضور ندارند؛ توسط پاتولوژی سال‌های قبل ثبت شده باشند، و همچنین تعدادی از افرادی که در منبع پاتولوژی حضور دارند و در ثبت مرگ همان سال حضور ندارند، ممکن است توسط منبع ثبت مرگ سال‌های بعد ثبت شده باشند، و با توجه به محدودیت مطالعه در دسترسی به داده‌های سال‌های قبل و سال‌های پس از مطالعه، ممکن است میزان بروز برآورد شده دارای مقداری بیش برآورد<sup>۱</sup> از مقدار واقعی-به سبب زیاد شدن تعداد افراد مربوط به خانهای از جدول که در ثبت مرگ هستند، اما در پاتولوژی و مستندات بیمارستانی ثبت نشده‌اند- باشد. بنابراین پیشنهاد می‌شود در مطالعه‌های آینده با مطابقت دادن اطلاعات ثبت مرگ در هر سال با ثبت پاتولوژی و مستندات بیمارستانی سال‌های قبل و بعد همان سال و حذف موارد تکراری، از میزان بیش برآورد بروز بیماری کاسته شود. توصیه می‌شود با کاربرد این روش در ثبت سرطان سایر نقاط کشور توسط پژوهشگران، برآوردهای دقیق‌تری از وضعیت سرطان در سایر استان‌های کشور به دست آید.

### تشکر و قدردانی

این مقاله برگرفته از پایان‌نامه آقای محمود خدادوست برای اخذ درجه کارشناسی ارشد اپیدمیولوژی از دانشکده‌ی بهداشت دانشگاه علوم پزشکی شهید بهشتی می‌باشد. نویسندگان این مقاله مراتب تشکر و قدردانی را از ریاست محترم و کارکنان مرکز ثبت سرطان اردبیل که ما را در انجام این مطالعه یاری فرموده‌اند؛ اعلام می‌دارند.

توجه داشت که این میزان گزارش شده مربوط به کل جمعیت ایران است؛ در صورتی‌که بسیاری از نواحی شمالی و شمال‌غربی کشور از جمله استان اردبیل از استان‌های با خطر بالای سرطان معده، استان‌های مرکزی و غربی در خطر متوسط و استان‌های جنوبی از استان‌های با خطر کم سرطان معده محسوب می‌شوند (۱۱). بابایی و همکاران، میزان بروز سرطان معده را در استان اردبیل در طی سال‌های ۸۵-۱۳۸۳ برای مردان و زنان به تفکیک ۵۱/۸ و ۲۴/۴ در صدهزار نفر گزارش کرده است که با بروز گزارش شده (و نه برآورد شده) بر اساس سه منبع ثبت مطالعه حاضر هم‌خوانی دارد، اما میزان بروز برآورد شده در این مطالعه با روش لگ خطی بسیار بالاتر بود، که این اختلاف به سبب مقداری کم‌شماری ثبت سرطان در استان اردبیل است که توسط هیچ‌یک از منابع ثبت نشده‌اند. در مطالعه انجام شده در تهران در سال‌های ۸۵-۱۳۸۱، میزان بروز برآورد شده سرطان معده در مردان و زنان ۳۳/۹ و ۱۹/۷ گزارش شده است (۲۰)، که این اختلاف با توجه به این‌که استان اردبیل از استان‌های شمال غربی کشور و با خطر بالای سرطان معده و تهران از استان‌های با خطر متوسط محسوب می‌شوند؛ توجیه‌پذیر است.

بروز گزارش شده ( $I_1$ ) و برآورد شده ( $I_2$ ) به تفکیک سال‌های تقویمی (۱۳۸۵ و ۱۳۸۷) تفاوت زیادی را نشان نداد و در سال ۱۳۸۵ کمی بیش‌تر از ۱۳۸۷ بود.

بروز برآورد شده بر اساس زیر گروه‌های سنی نشان دهنده‌ی افزایش میزان بروز بیماری با افزایش سن است، و بیش‌ترین بروز مربوط به گروه سنی ۶۵ سال و بالاتر است که با مطالعه انجام شده در استان اردبیل در ۱۳۸۳ تا ۱۳۸۵ و مطالعه انجام شده در تهران و همچنین مطالعه انجام شده در ژاپن و سایر گزارش‌ها از سراسر جهان هم‌خوانی دارد (۲۰، ۲۱، ۲۴، ۲۵). میزان بروز برآورد شده به تفکیک زیرگروه‌های سنی بیش‌تر از بروز گزارش شده در مطالعه استان اردبیل در ۱۳۸۳ تا ۱۳۸۵ و مطالعه انجام شده در تهران (۲۰، ۱۰) و همچنین مطالعه‌های انجام شده در ژاپن و گزارش‌های جهانی است (۲۴، ۲۵).

### نتیجه‌گیری

نتایج این مطالعه لزوم استفاده از روش صید- بازصید برای اصلاح کم‌شماری میزان بروز سرطان معده و به دست آوردن برآورد

<sup>۱</sup>Over estimate

## منابع

1. Kamo K-i, Kaneko S, Satoh K, Yanagihara H, Mizuno S, Sobue T. A mathematical estimation of true cancer incidence using data from population-based cancer registries. *Japanese Journal of Clinical Oncology*. 2007; 37: 150-5.
2. Schmidtmann I. Estimating Completeness in Cancer Registries—Comparing Capture Recapture Methods in a Simulation Study. *Biometrical Journal*. 2008; 50: 1077-92.
3. Parkin D, Chen V, Ferlay J, Galceran J, Storm H, Whelan S. Comparability and Quality Control in Cancer Registration: IARC Technical Report No19. International Agency for Research on Cancer, Lyon, France. 1994.
4. Mosavi-Jarrahi A, Ahmadi-Jouibari T, Najafi F, Mehrabi Y, Aghaei A. Estimation of Esophageal Cancer Incidence in Tehran by Log-linear Method using Population-based Cancer Registry Data. *Asian Pac J Cancer Prev*. 2013; 14: 5367-70.
5. Castro CSFCd. Estimating Completeness of Cancer Registration—Na application to gastric cancer in the North Region Registry of Portugal 2011.
6. Robles SC, Marrett LD, Aileen Clarke E, Risch HA. An application of capture-recapture methods to the estimation of completeness of cancer registration. *Journal of Clinical Epidemiology*. 1988; 41: 495-501.
7. Suwanrungruang K, Sriplung H, Attasara P, Temiyasathit S, Buasom R, Waisri N, et al. Quality of Case Ascertainment in Cancer Registries: A Proposal for a Virtual Three-source Capture-recapture Technique. *Asian Pacific Journal of Cancer Prevention*. 2011; 12: 173-8.
8. Poorolajal J, Haghdoost AA, Mahmoodi M, Majdzadeh R, Nasser-Moghaddam S, Fotouhi A. Capture-recapture method for assessing publication bias. *Journal of Research in Medical Sciences: the official Journal of Isfahan University of Medical Sciences*. 2010; 15: 107.
9. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. *European Journal of Cancer*. 2009; 45: 756-64.
10. Babaei M, Pourfarzi F, Yazdanbod A, Chiniforush MM, Derakhshan MH, Mousavi SM, et al. Gastric cancer in Ardabil, Iran--a review and update on cancer registry data. *Asian Pac J Cancer Prev*. 2010; 11: 595-9.
11. Radmard AR. Five common cancers in Iran. *Archives of Iranian medicine*. 2010; 13: 143-6.
12. van Hest NAH. Capture-recapture methods in surveillance of tuberculosis and other infectious diseases: Erasmus University Rotterdam; 2007.
13. Motevalian A, Holakoei naeini K, Mahmoodi M, Majdzadeh R, Akbari M. Estimating deaths due to traffic accidents in Kerman using capture\_recapture method. *Journal of Health Faculty and Health Research*. 2007; 5: 61-72.
14. Hook EB, Regal RR. Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation. *American Journal of Epidemiology*. 1997; 145: 1138-44.
15. Hook EB, Regal RR. Accuracy of alternative approaches to capture-recapture estimates of disease frequency: internal validity analysis of data from five sources. *American Journal of Epidemiology*. 2000; 152: 771-9.
16. Rajaiefard A, Moghimi B, Tabatabaie SH, Safaie A, Tabeie SZ. Epidemiological and clinical features of gastric cancer: descriptive study of cancer registry cases of Fars province (2001-2006). *ISMJ*. 2011; 14: 114-21.
17. Mehrabian A, Esna-Ashari F, Zham H, Hadizadeh M, Bohlooli M, Khayamzadeh M, et al. Gastric cancer prevalence, according to survival data in Iran (National Study-2007). *Iranian Journal of Public Health*. 2010; 39.
18. Biglarian A, Hajizadeh E, Kazemnejad A, Zali M. Survival analysis of gastric cancer patients using Cox model: a five year study. *Tehran University Medical Journal*. 2009; 67.
19. Khodabakhshi R, Gouhari M, Biglarian A, Hajizadeh E. Survival analysis of patients with gastric adenocarcinoma INTERNATI. 2009; 17: 1.
20. Aghaei A, Ahmadi-Jouibari T, Baiki O, Mosavi-Jarrahi A. Estimation of the gastric cancer incidence in Tehran by two-source capture-recapture. *Asian Pac J Cancer Prev*. 2013; 14: 673-7.
21. Babaei M, Jaafarzadeh H, Sadjadi A, Samadi F, Yazdanbod A, Fallah M, et al. Cancer incidence and mortality in Ardabil: Report of an ongoing population-based cancer registry in Iran, 2004-2006. *Iranian Journal of Public Health*. 2009; 38, 35-45.
22. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh J, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *European Journal of Cancer*. 2013; 49: 1374-403.
23. GLOBOCAN. 2008; Available at: <http://globocan.iarc.fr/factsheets/cancers/stomach.asp>.
24. Boyle P, Levin B. World cancer report 2008: IARC Press, International Agency for Research on Cancer; 2008.
25. Matsuda T, Marugame T, Kamo K-i, Katanoda K, Ajiki W, Sobue T. Cancer incidence and incidence rates in Japan in 2002: based on data from 11 population-based cancer registries. *Japanese Journal of Clinical Oncology*. 2008; 38: 641-8

**Original Article**

# Estimation of Gastric Cancer Incidence in Ardabil Province by Capture-Recapture Method Using Population-Based Cancer Registry Data

Khodadost M<sup>1</sup>, Yavari P<sup>2,3</sup>, Hashemi Nazari SS<sup>4,5</sup>, Babaei M<sup>6,7</sup>, Abadi A<sup>8</sup>, Sarvi F<sup>9</sup>

1- Msc in Epidemiology, faculty of health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2- Professor of Epidemiology, Dept. of Health & Community Medicine, School of Medicine, Shahid Beheshti University of Medical Sciences, Evin, Tehran, Iran

3- Iranian Epidemiological Association

4- Safety Promotion and Injury Prevention Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

5- Assistant professor of Epidemiology, Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

6- Health Deputy, Ardabil University of Medical sciences, Ardabil, Iran

7- German Cancer Research Center (DKFZ)

8- Associate Professor, Dept of Health & Community Medicine, School of Medicine, Shahid Beheshti University of Medical Sciences, Evin, Tehran, Iran

9- Msc in Biostatistics, Paramedical Department, Shahid Beheshti university of Medical Sciences, Tehran, Iran

**Corresponding author** Yavari P, parvinyavari@yahoo.com

**Background and Objectives:** Awareness of the cancer incidence is essential for cancer prevention and control programs. Capture-recapture methods have been recommended for reducing bias and increasing the accuracy of cancer incidence estimation. This study aimed to estimate the incidence of gastric cancer by the capture-recapture method based on Ardabil population-based cancer registry data.

**Methods:** All new cases of gastric cancer reported by three sources, i.e. pathology reports, death certificates, and medical records, reported to Ardabil population-based cancer registry between 2006 and 2008 were enrolled in the study. The duplicate cases based on the similarity of the first name, surname, and father's name were identified among sources. The estimated incidence was calculated by the log-linear method using the Stata 12 software.

**Results:** A total 857 new cases of gastric cancer were reported from three sources. After removing duplicates, the reported incidence rate was 35.3 and 32.5 per 100,000 population for the years 2006 and 2008, respectively. The estimated incidence rate calculated by the log-linear method for these years was 96.2 and 90.4 per 100,000 population, respectively.

**Conclusion:** The results showed that none of the sources of pathology reports, death certificates, and medical records, individually or collectively, fully covered the incidence of gastric cancer. We can obtain more accurate estimates of the incidence rate using the capture-recapture method.

**Keywords:** Capture-recapture, Incidence estimation, Log-linear, Gastric cancer, Cancer registration