Accurately estimating breast cancer survival in Spain: cross-matching local cancer registries with the National Death Index

M. Carmen Martos,1,2 Carme Saurina,3,4 Cristina Feja,1,3 Marc Saez,3,4 M. Carmen Burriel,1 Maria Antònia Barceló,3,4 Pilar Gómez,1 Gemma Renart,3,4 Tomás Alcalá,2,5 and Rafael Marcos-Gragera6

Objectives. To assess the impact of using data from the National Death Index (NDI) of Spain to estimate breast cancer survival rates among residents of Girona and Zaragoza diagnosed in 1995–1999.

Methods. This was an observational, longitudinal epidemiologic study, using two population-based cancer registries. Data collected were of female residents of Girona or Zaragoza who had been diagnosed with breast cancer in 1995–1999. Observed and relative 5-year survival rates were estimated, first using the information available from the Girona and Zaragoza cancer registries, and then with the inclusion of NDI data. The 5-year relative survival rate and corresponding 95% Confidence Intervals were estimated using the Hakulinen method. The Kaplan-Meier method and Log Rank test were used to compare survival curves.

Results. No statistically significant difference in survival curves was observed in Girona for the data obtained before and after cross-matching with the NDI. However, there was a significant difference in Zaragoza. A comparison of the relative survival rates of each of the two registries before NDI cross-matching showed differences of 3.9% (5-year) and 16.1% (10-year) between the two, whereas after the cross-match, the difference was only 0.5% (5-year) and 1.2% (10-year).

Conclusions. In Spain it is imperative that there be systematic use of NDI data to supplement cancer registries, so that comparisons of relative survival rates between registries can be improved.

Key words Breast neoplasms, surveillance; medical record linkage; vital statistics; cancer survival rate; Spain.
Zaragoza and in many towns in the province of Girona (2).

Survival is one way to measure the overall organizational efficiency of health services that manage cancer patients. Population models of survival based on cancer registries provide a useful tool for evaluating health systems and the effectiveness of interventions (3). In Spain, survival rates for breast cancer have risen in recent years, mortality rates have remained stable or decreased, and overall incidence has increased moderately (4). However, increases in survival rates are not easy to interpret because they could be due, at least in part, to improvements in treatment, more effective treatment due to earlier diagnoses, or over-diagnosis due to screening programs, or simply to lead-time bias due to earlier diagnoses (5–8). Furthermore, because of certain problems, especially in terms of coverage and continued follow-up of patients, the quality of data from cancer registries may compromise the interpretation of the geographical variations of survival rates (7, 9).

There are 11 population-based cancer registries in Spain that were included by the International Agency for Research on Cancer in its publication, *Cancer Incidence in Five Continents*. Zaragoza and Girona were among them (1). The use of standardized criteria for choosing sources and collecting information, along with the systematic elaboration of quality indicators, ensures the comparability of incidence estimates among different registries (10). However, since follow-up procedures are not standardized and actually vary widely due to issues with accessing individual mortality data, comparability of survival rates is greatly compromised.

Currently, information on cancer survival rates is scarce. As such, only data provided by registries participating in the EUROCare4 study are available. According to this data, 5-year relative survival for female breast cancer diagnosed in 1995–1999 in Spain was 80.8% (7).

Since 2006, Spain has made available its National Death Index (NDI). The index may contribute to improving the geographic comparability of estimated survival rates. The NDI is an information system containing personal data on each and every one of the deaths inscribed in the civil register throughout the country. It has the structure of a database and its variables include: name of the deceased, national identity card number (since 1996), date and place of birth, sex, marital status, nationality, home address, date of death, and place of death. According to the user’s manual (1), it contains data from 1987 onwards. Access is restricted to research centers, government health boards, and health centers and other establishments, all of which are state-controlled.

The objective of this article is to assess the impact using NDI data to estimate survival rates among patients who were residents of Girona and Zaragoza and were diagnosed with breast cancer in 1995–1999.

**MATERIALS AND METHODS**

An observational, longitudinal, epidemiologic study was carried out using the population-based the cancer registries of Girona and Zaragoza, two areas in northeastern Spain, to identify female residents diagnosed with breast cancer (International Classification of Diseases—10th Edition, code C50) (11) in 1995–1990. In 2001, the Girona registry covered a population of 553 661, and the registry of Zaragoza, 861 855. The sources of information for each of the registries are their respective regional and local hospitals, as well as hematology and pathology departments. However, the patient health status and, if applicable, date of death, were monitored in different ways by each registry. In Girona, which is one of four provinces in Catalonia, data from the Catalan Mortality Registry and the Girona Cancer Registry are cross-matched annually through record-linkage. In Zaragoza, monitoring was done through death certificates in which the cause of death specified “cancer,” “carcinoma,” or “neoplasia,” as well as by the Health Identity Card Registration and hospital discharges due to death.

For the purposes of this study, all cases of breast cancer from the two registries were cross-matched, through record-linkage, with data from the NDI (12). The cross-matching was first done automatically via the website of the Ministry of Health, followed by a manual search of the database for cases not identified previously. The variables used for the automatic cross-matching were: last name(s); first name(s); and date of birth. Patients not found automatically were sought manually by: first name(s), last name(s), date of birth, and National Identity Number, if available (recorded in the databases since 1996). For the survival rate analysis, cases obtained only from the death certificates were excluded: 26 in Girona (1.9%) and 68 in Zaragoza (3.5%). Patients were followed up until 31 December 2005. The 5-year observed and relative survival rates in Girona and Zaragoza were subsequently estimated using information available previously from the registries; later, the NDI data (date of death) were included as well. To that end, the Kaplan-Meier method and the Log Rank test were employed for comparison (13). The 5-year relative survival rates and 95% Confidence Intervals (95%CI), estimated using the Hakulin method, were obtained using WAERS-ICO software, available from Catalan Institute of Oncology website (14).

**RESULTS**

In 1995–1999, there were 1 381 new cases of invasive breast cancer registered in Girona; of these, 26 (1.9%) were reported only by means of a death certificate. In Zaragoza, 1 951 cases were identified, of which 68 (3.5%) were reported only by death certificate. When cross-matching the data in the NDI, the date of death was modified or completed in 13 cases in Girona and 210 cases in Zaragoza, representing 0.9% and 10.8% of cases, respectively.

No statistically significant difference in survival curves was observed in Girona for the data obtained before and after the cross-matching with the NDI (Figure 1). However, there was a significant difference ($P < 0.001$) in Zaragoza (Figure 2).

When comparing the relative survival rates of each of the two registries prior to the NDI cross-matching, there was a difference $3.9\%$, whereas after the cross-matching the difference was only $0.5\%$. The difference prior to using NDI was much higher at the 10-year relative survival mark, with $16.1\%$ and $1.2\%$ respectively (Table 1).

**DISCUSSION**

The differences found between the 5- and 10-year relative survival estimates for breast cancer in Girona and Zaragoza are explained, at least in part, by whether or not there is access to individualized mortality data. While the population-based
cancer registries in Catalonia (which includes Girona) have access to the Catalonia Mortality Registry and systematically cross-match each registry, this is not possible in Aragon (which includes Zaragoza) and other autonomous areas, such as Andalucía. The main reason for this in Aragon is that legislation regarding health and demographic statistics is not as progressive as it is in Catalonia, making it difficult to access data from information systems outside the central health administration. The slight overestimation of breast cancer survival produced by data from the Girona Cancer Registry was due to patient emigration following cancer diagnosis. Access to the NDI revealed the true status of these patients.

By using NDI data in conjunction with both cancer registries to obtain the vital status of breast cancer patients and, if applicable, the date of death, 5- and 10-year relative survival rates were found to be similar: Girona, 80.5% and 70.2%; and Zaragoza, 81.7% and 69.5%. Five-year relative survival figures are slightly higher than the average for Europe (79.2%) described by the EUROCARE-4 project for the same period (15), although France and all of the Scandinavian countries except Denmark have better 5-year relative breast cancer survival rates than do Girona and Zaragoza (8). Since Scandinavian cancer registries have national coverage and systematic record-linkage with mortality registries and population databases, differences in survival rates were presumed to be reliable. However, in France, just as in Spain, there is no national cancer registry, and survival is estimated from regional registers; furthermore, there are difficulties in accessing individualized mortality data. Therefore, survival rate estimates based on these records must be interpreted with caution.

Data on survival rates collected from the population-based cancer registries allow global assessment of prevention and control activities (16). Moreover, the health authorities need cancer survival statistics in order to plan and execute local cancer control programs and to optimize the allocation of resources. However, differences in the quality of data may jeopardize geographic and temporal comparability. This may also influence the interpretation of the results obtained (17).

Most of the population-based cancer registries in Spain depend on the governments of autonomous regions. Legislative differences and variations in their health systems, priorities, and resources may compromise the comparability of data obtained from their information systems. It is therefore necessary to develop initiatives at the national level that will improve the comparability of health indicators across autonomous communities.

In summary, these results show that systematic use NDI data to supplement cancer registries is imperative to improving comparisons of relative survival rates between registries.

![FIGURE 1. Survival functions before and after cross-matching with the National Death Index, Girona, 1995–1999](image1)

![FIGURE 2. Survival functions before and after cross-matching with the National Death Index, Zaragoza, 1995–1999](image2)

<table>
<thead>
<tr>
<th>Registry</th>
<th>Before NDI cross-matching</th>
<th>After NDI cross-matching</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girona</td>
<td>81.0</td>
<td>80.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Zaragoza</td>
<td>85.6</td>
<td>81.7</td>
<td>3.9</td>
</tr>
<tr>
<td>10-year relative survival rates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girona</td>
<td>71.4</td>
<td>70.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Zaragoza</td>
<td>85.6</td>
<td>69.5</td>
<td>16.1</td>
</tr>
</tbody>
</table>

*95% Confidence Interval.
RESUMEN

Estimación precisa de la supervivencia de cáncer de mama en España: emparejamiento de los registros locales de cáncer con el Índice Nacional de Defunciones

Objetivo. Evaluar el efecto de utilizar los datos del Índice Nacional de Defunciones (IND) de España para estimar las tasas de supervivencia de cáncer de mama en las mujeres residentes en Girona y Zaragoza que recibieron el diagnóstico de cáncer de mama en 1995–1999.

Métodos. Se realizó un estudio epidemiológico observacional y longitudinal basado en el empleo de los registros de cáncer de mujeres residentes en Girona y Zaragoza que habían recibido el diagnóstico de cáncer de mama en 1995–1999. Se estimaron las tasas de supervivencia observada y relativa a 5 años, primero según la información disponible en los registros de cáncer de Girona y Zaragoza y luego con la inclusión de los datos del IND. Se calcularon las tasas de supervivencia relativa a 5 años y sus correspondientes intervalos de confianza de 95% por el método de Hakulinen. Las curvas de supervivencia se compararon por el método de Kaplan-Maier y la prueba de rangos logarítmicos.

Resultados. No se encontraron diferencias estadísticamente significativas entre las curvas de supervivencia de Girona antes y después de emparejar los datos locales con los del IND; sin embargo, hubo diferencias significativas entre las curvas de Zaragoza. Al comparar las tasas de supervivencia relativa de cada uno de los registros antes de emparejarlos con los datos del IND se encontraron diferencias de 3,9% (a 5 años) y 16,1% (a 10 años), mientras que después del emparejamiento, la diferencia entre ellas fue solamente de 0,5% (a 5 años) y 1,2% (a 10 años).

Conclusiones. En España es imperativo el empleo sistemático de los datos del IND para complementar los registros de cáncer de manera de mejorar las comparaciones de las tasas de supervivencia relativa cuando se utilizan diferentes registros.

Palabras clave
Neoplasias de la mama, vigilancia; registro médico coordinado; estadísticas vitales; tasa de supervivencia; España.