Funnel plots and choropleth maps in cancer risk communication: a comparison of tools for disseminating population-based incidence data to stakeholders

Walter Mazzucco,1,2 Rosanna Cusimano,1,3 Maurizio Zarcone,1,4 Sergio Mazzola,1,4 Francesco Vitale1,2,4

ABSTRACT

Background: Population-based cancer registries provide epidemiological cancer information, but the indicators are often too complex to be interpreted by local authorities and communities, due to numeracy and literacy limitations. The aim of this paper is to compare the commonly used visual formats to funnel plots to enable local public health authorities and communities to access valid and understandable cancer incidence data obtained at the municipal level.

Methods: A funnel plot representation of standardised incidence ratio (SIR) was generated for the 82 municipalities of the Palermo Province with the 2003–2011 data from the Palermo Province Cancer Registry (Sicily, Italy). The properties of the funnel plot and choropleth map methodologies were compared within the context of disseminating epidemiological data to stakeholders.

Results: The SIRs of all the municipalities remained within the control limits, except for Palermo city area (SIR=1.12), which was sited outside the upper control limit line of 99.8%. The Palermo Province SIRs funnel plot representation was congruent with the choropleth map generated from the same data, but the former resulted more informative as shown by the comparisons of the weaknesses and strengths of the 2 visual formats.

Conclusions: Funnel plot should be used as a complementary valuable tool to communicate epidemiological data of cancer registries to communities and local authorities, visually conveying an efficient and simple way to interpret cancer incidence data.

BACKGROUND

Cancer is the second major cause of death in the developed countries.1 In the past few decades, the increasing burden of disease has caused major concerns in local communities, requiring local health authorities to develop risk communication plans that address cancer incidence, survival and the potential impact of environmental exposure.3 Apart from the presumed effects of lifestyle changes and environmental factors on cancer trends,3,4–6 the global increase in cancer prevalence could be largely attributable to a combination of improved cancer survival7 and ageing population.8 Local communities possess a variable degree of literacy and numeracy, which, in turn, influence their understanding of such demographical and epidemiological concepts.9–10 Local public health and political authorities regularly engage in finding better ways to satisfy the growing demand for information on the
impact of cancer by the general public.\textsuperscript{11} In particular, citizens often question if they live in an area at high risk for environmental exposure.\textsuperscript{2}

The Centers for Diseases Control and Prevention (CDC) define public health surveillance as the “Ongoing, systematic collection, analysis, interpretation, and dissemination of data regarding a health-related event for use in public health action to reduce morbidity and mortality and to improve health.”\textsuperscript{12} Population-based cancer registries (PBCRs) carry out cancer surveillance by continuously collecting and classifying information on all new cancer cases within a defined population, and providing statistics on its occurrence for the purpose of assessing and controlling the impact of this disease on the community.\textsuperscript{13} The mission of PBCRs includes the translation and dissemination of evidences to enable informed decision-making and to empower the general population or other stakeholders, while preserving a rigorous methodological approach and facilitating a truthful interpretation of the data obtained. PBCR publications use validated and internationally shared measurement systems and employ terminology and visual formats that are easily understood by the scientific community, but often difficult to interpret for other stakeholders, particularly at the local level.\textsuperscript{14, 15}

The most commonly used format for reporting geographic comparisons of cancer epidemiological data is an atlas, which includes thematic maps, such as choropleth maps (CMs), representing cancer incidence rates (standardised rates, standardised ratios, etc) computed for specific areas.\textsuperscript{16, 17}

While data are available on how the context\textsuperscript{18} and the content of such communications influence individual risk perception,\textsuperscript{19} little is known about the effects of risk communications at a group level, particularly in small communities.\textsuperscript{20}

The Italian Association of Cancer Registries (AIRTumb), a national network of 41 local PBCRs, including the Palermo Province Cancer Registry (PPCR), has greatly emphasised improving communication tools.\textsuperscript{21}

The aim of this paper is to propose the use of funnel plots (FPs) for reporting local cancer incidence data, as a complement to the more common visual formats employed by the PPCR to address local public health authorities and communities, in order to facilitate the dissemination and interpretation of measures of cancer statistics at the municipal level.

**METHODS**

The study population consists of the 51,951 new cancer cases, excluding non-melanoma skin cancers, registered between 2003 and 2011 by the PPCR among the 1,244,239 residents of the 82 municipalities of the Palermo Province (PP); 679,850 inhabitants within the Palermo metropolitan area only).\textsuperscript{22} Cancer incidence in the PP municipalities was measured by using standardised incidence ratio (SIR), defined as the ratio between observed cases (Oi) and expected cases (Ei).\textsuperscript{23} The Oi were assumed to follow a homogeneous Poisson distribution with parameter $\theta = E_i$. The Ei were estimated by indirect method,\textsuperscript{24} considering the entire population time under study (the PP) as the reference population, with $\sum O_i = \sum E_i$.\textsuperscript{25} The resident population was reported using the intercensus estimates, provided by the Italian National Statistical Institute (ISTAT), also considering the annual municipal data on migration.\textsuperscript{22} For each SIR, the 95% CI was calculated by using the normal approximation method.\textsuperscript{26}

Graphic FP representation\textsuperscript{26} was used to highlight any municipality with a higher cancer incidence compared with the reference population (entire PP population). The following elements were included to generate the FP (figure 1A): the SIRs of the 82 municipalities, on the y-axis; the target line ($\theta=1$), representing the reference value for the indicator of interest ($O_i=E_i$); the Ei precision parameter, measuring the accuracy of the indicator of interest (Poisson variance parameter, using the hypothesis $\theta=1$), represented on the x-axis; the 95% and 99.8% CIs, calculated with the normal approximation method, defining the control limits.\textsuperscript{26} The two sets of control limit lines define three different areas within the graph (figure 1B): the ‘undercontrol’ area (in green), the ‘warning’ area (in yellow) and the ‘alert’ area (in red).\textsuperscript{27}

As the data distribution was not congruent with the underlying assumption (variance equal to the expected value), in order to check for any potential overdispersion\textsuperscript{28} both additive and multiplicative approaches were adopted. Overdispersion coefficients ($\tau$ for the additive approach and $\phi$ for the multiplicative approach) were calculated. Overdispersion was addressed by considering the winsorised estimates too.\textsuperscript{27} Moreover, Z-score\textsuperscript{29} and the winsorisation method (by testing for different levels of Z-score quantiles\textsuperscript{28}) were applied for the direct selection of extreme values. Furthermore, to define the level of winsorisation, an R-script routine was developed to set a cut-off for the quantile between the acceptance and rejection of the overdispersion test (see online supplementary material).

The map representing the PP municipalities was generated by using the ISTAT Shapefile vector format,\textsuperscript{30} released in the ED50 (European Datum - 1950) UTM Zone 32N reference system, and converted in plane coordinates (decimal degrees), providing georeferenced data in addition to the coordinates of geographic objects and their borders (for polygons), also including the information on the location of each municipality. Although traditional geographical analyses use the centroids as geostatistical units, considering that some centroid could fall outside the municipal bounds, the coordinates of the city hall were used instead.\textsuperscript{31}

The PP cancer incidence variation was also shown in a CM,\textsuperscript{32} representing the SIRs of each municipality. To distinguish potential high-risk and low-risk areas, a central interval of 0.95-1.05 for the colour scale was fixed, irrespective of statistical significance. Values above 1.05 and below 0.95 were divided in tertiles.\textsuperscript{33}
Cluster analysis was performed by using the scan statistics obtained with Openshaw’s Geographical Analysis Machine (GAM), with varying radiuses, in order to detect potentials high-risk clusters and hot spot locations, setting the p value at 0.002. The analysis for hot spot research was performed using circles with a 3 km radius for each point of a grid, covering the study region by steps of 600 m (radius/5). The RStudio IDE (RStudio Team. RStudio: Integrated Development for R. 2015. http://www.rstudio.com/ (accessed 18 Jan 2016)) for the R software, V.3.1.0 (2014-04-10)—‘Spring Dance’ (R Core Team. R: A language and environment for statistical computing 2015. http://www.R-project.org/ (accessed 18 Jan 2016)), was used to perform statistical analysis.

Finally, the weaknesses and strengths of the FP and CM methodological approaches were compared using the available literature as reference.29 33 35 36

Figure 1 (A) Funnel plot of the SIRs in the 82 Palermo Province municipalities (study period 2003–2011); (B) cancer attention areas: ‘undercontrol’ area (in green), ‘warning’ area (in yellow) and ‘alert’ area (in red). Ninety-five per cent CIs (‘blue’ control lines) and 99.8% CIs (‘red’ control lines); $\phi$=overdispersion, calculated with multiplicative approach. SIR, standardised incidence ratio.
RESULTS

Figure 1A represents the FP of 82 municipality-specific SIRs, corrected for overdispersion ($\phi=13.46$) and adjusted using the multiplicative approach. All of the SIRs lay within the control limits, except for the Palermo city 1 (SIR=1.12), which resulted above the upper control limit line of 99.8%. Figure 1B identifies the three different cancer risk areas within the graph.

Overdispersion test results were concordant and the routine did not find out any valid value for winsorisation (see online supplementary material, section B).

Figure 2 displays the CM for cancer incidence in the 82 PP municipalities, generated by using the SIRs. The map highlights three different municipality areas (ISTAT code: 082042, 082053 and 082061; see table 1) with SIRs higher than 1.05.

Table 1 represents the expected cases (both men and women) and SIRs with 95% CIs in the 82 PP municipalities: most of the SIRs are lower than 1 and only six municipalities present SIRs higher than 1. Among them only Palermo had a statistically significant value higher than 1 (SIR=1.12; 95% CIs 1.11 to 1.14) while Isnello, the municipality showing the highest SIR, failed to meet the conventional criteria for statistical significance (SIR=1.22; 95% CI 0.99 to 1.45).

No clusters were identified by the GAM approach, while a hot spot corresponding to Palermo city was highlighted (figure 3).

Table 2 summarises a comparison of the weaknesses and strengths, as per the available literature, between the different visual formats explored within the context of disseminating epidemiological data to stakeholders.

As shown in the table 2, in terms of strengths, FP differed from CM in its ability to disseminate epidemiological data to stakeholders, in particular in the capability to show the scope of the phenomenon under investigation and the precision of estimates, and to highlight the significance of the estimates. On the other hand, CM, unlike FP, was able to define the spatial location of the risk and to locate the presence of any cluster. Both FP and CM were able to identify hot spots.

Figure 2  Choropleth map of the SIRs in the 82 Palermo Province municipalities (study period 2003–2011). Circles represent the locations of city halls. SIR, standardised incidence ratio.
DISCUSSION

FPs are commonly used in process control and, in particular, in the healthcare field to compare institutional performance data, however, this format is used for survival and standardised mortality ratio in public health surveillance. We explored the use of FPs as a supplementary tool to local provide authorities and communities with synthetic access to valid and

Table 1 Expected cases and SIRs (shown in a descending order) with 95% CIs in the 82 PP municipalities (study period: 2003–2011, reference: the entire PP population)

<table>
<thead>
<tr>
<th>ISTAT code</th>
<th>Municipality</th>
<th>Expected SIR 95% CI</th>
<th>ISTAT code</th>
<th>Municipality</th>
<th>Expected SIR 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>082042</td>
<td>Isnello</td>
<td>104.1 0.99 to 1.45</td>
<td>082031</td>
<td>Cinisi</td>
<td>441.0 0.74 to 0.90</td>
</tr>
<tr>
<td>082053</td>
<td>Palermo City</td>
<td>2737.1 1.10 to 1.14</td>
<td>082007</td>
<td>Balestrate</td>
<td>292.4 0.72 to 0.91</td>
</tr>
<tr>
<td>082061</td>
<td>Roccamena</td>
<td>81.4 0.80 to 1.23</td>
<td>082065</td>
<td>San Giuseppe Jato</td>
<td>379.0 0.73 to 0.89</td>
</tr>
<tr>
<td>082070</td>
<td>Termini Imerese</td>
<td>116.4 0.99 to 1.11</td>
<td>082058</td>
<td>Polizzi Generosa</td>
<td>212.6 0.69 to 0.91</td>
</tr>
<tr>
<td>082027</td>
<td>Cefalù</td>
<td>685.3 1.01 to 1.08</td>
<td>082026</td>
<td>Gangi</td>
<td>406.6 0.72 to 0.87</td>
</tr>
<tr>
<td>082044</td>
<td>Lascari</td>
<td>152.2 0.85 to 1.17</td>
<td>082036</td>
<td>San Giuseppe Jato</td>
<td>379.0 0.73 to 0.89</td>
</tr>
<tr>
<td>082014</td>
<td>Caccamo</td>
<td>396.9 0.88 to 1.07</td>
<td>082059</td>
<td>Pollina</td>
<td>148.0 0.68 to 0.94</td>
</tr>
<tr>
<td>082035</td>
<td>Ficarazzi</td>
<td>357.5 0.87 to 1.07</td>
<td>082023</td>
<td>Casteldaccia</td>
<td>212.6 0.72 to 0.87</td>
</tr>
<tr>
<td>082038</td>
<td>Giardiniello</td>
<td>84.1 0.76 to 1.17</td>
<td>082054</td>
<td>Partinico</td>
<td>1291.1 0.73 to 0.82</td>
</tr>
<tr>
<td>082056</td>
<td>Petralia Sottana</td>
<td>168.6 0.81 to 1.09</td>
<td>082004</td>
<td>Altavilla Milicia</td>
<td>242.8 0.68 to 0.88</td>
</tr>
<tr>
<td>082012</td>
<td>Bompietro</td>
<td>104.6 0.97 to 1.13</td>
<td>082005</td>
<td>Altopiano</td>
<td>379.8 0.70 to 0.86</td>
</tr>
<tr>
<td>082049</td>
<td>Monreale</td>
<td>1319.0 0.94 to 1.07</td>
<td>082020</td>
<td>Alimena</td>
<td>301.8 0.70 to 0.87</td>
</tr>
<tr>
<td>082052</td>
<td>Palazzo Adriano</td>
<td>123.4 0.93 to 1.10</td>
<td>082025</td>
<td>Castronovo di Sicilia</td>
<td>175.4 0.67 to 0.70</td>
</tr>
<tr>
<td>082079</td>
<td>Villabate</td>
<td>631.9 0.93 to 1.06</td>
<td>082050</td>
<td>Monreale</td>
<td>258.1 0.68 to 0.87</td>
</tr>
<tr>
<td>082007</td>
<td>Balestrate</td>
<td>81.4 0.80 to 1.23</td>
<td>082064</td>
<td>Palermo City</td>
<td>312.6 0.69 to 0.88</td>
</tr>
<tr>
<td>082067</td>
<td>San Giuseppe Jato</td>
<td>396.9 0.88 to 1.07</td>
<td>082058</td>
<td>Palermo City</td>
<td>212.6 0.69 to 0.91</td>
</tr>
<tr>
<td>082026</td>
<td>Gangi</td>
<td>357.5 0.87 to 1.07</td>
<td>082036</td>
<td>Palermo City</td>
<td>212.6 0.69 to 0.88</td>
</tr>
<tr>
<td>082023</td>
<td>Casteldaccia</td>
<td>84.1 0.76 to 1.17</td>
<td>082054</td>
<td>Partinico</td>
<td>1291.1 0.73 to 0.82</td>
</tr>
<tr>
<td>082054</td>
<td>Pollina</td>
<td>104.1 0.99 to 1.45</td>
<td>082023</td>
<td>Casteldaccia</td>
<td>212.6 0.72 to 0.87</td>
</tr>
<tr>
<td>082026</td>
<td>Palermo City</td>
<td>2737.1 1.10 to 1.14</td>
<td>082007</td>
<td>Balestrate</td>
<td>292.4 0.72 to 0.91</td>
</tr>
<tr>
<td>082065</td>
<td>San Giuseppe Jato</td>
<td>379.0 0.73 to 0.89</td>
<td>082058</td>
<td>Polizzi Generosa</td>
<td>212.6 0.69 to 0.91</td>
</tr>
<tr>
<td>082026</td>
<td>Gangi</td>
<td>357.5 0.87 to 1.07</td>
<td>082036</td>
<td>Palermo City</td>
<td>212.6 0.69 to 0.88</td>
</tr>
<tr>
<td>082038</td>
<td>Giardiniello</td>
<td>84.1 0.76 to 1.17</td>
<td>082054</td>
<td>Partinico</td>
<td>1291.1 0.73 to 0.82</td>
</tr>
<tr>
<td>082056</td>
<td>Petralia Sottana</td>
<td>168.6 0.81 to 1.09</td>
<td>082004</td>
<td>Altavilla Milicia</td>
<td>242.8 0.68 to 0.88</td>
</tr>
<tr>
<td>082012</td>
<td>Bompietro</td>
<td>104.6 0.97 to 1.13</td>
<td>082005</td>
<td>Altopiano</td>
<td>379.8 0.70 to 0.86</td>
</tr>
<tr>
<td>082049</td>
<td>Monreale</td>
<td>1319.0 0.94 to 1.07</td>
<td>082020</td>
<td>Alimena</td>
<td>301.8 0.70 to 0.87</td>
</tr>
<tr>
<td>082052</td>
<td>Palazzo Adriano</td>
<td>123.4 0.93 to 1.10</td>
<td>082025</td>
<td>Castronovo di Sicilia</td>
<td>175.4 0.67 to 0.70</td>
</tr>
<tr>
<td>082079</td>
<td>Villabate</td>
<td>631.9 0.93 to 1.06</td>
<td>082050</td>
<td>Monreale</td>
<td>258.1 0.68 to 0.87</td>
</tr>
</tbody>
</table>

Bold typeface indicates a significant SIR value.
ISTAT, Italian National Statistical Institute; SIR, standardised incidence ratio; PP, Palermo Province.


Open Access

BMJ Open: first published as 10.1136/bmjopen-2016-011502 on 30 March 2017. Downloaded from http://bmjopen.bmj.com on September 23, 2021 by guest. Protected by copyright.
understandable cancer incidence data (SIRs) obtained at the municipal level.

Given that SIR is an effective and well-established measure in the descriptive cancer epidemiology, we used this parameter to compare the use of FPs and the more common formats for reporting cancer epidemiological data.

Whereas scale-risk tables are easy to understand, readers do not usually take notice of the CI, which is a critically important measure of the precision of SIR estimates. By displaying sample statistics together with the corresponding sample size, in relation to the control limits, FPs allow visualising both information and precision levels without the need for processing several numeric values (in this study, we used 82 point estimates and 164 confidence boundaries). Moreover, while it is common knowledge that the numeracy skills of the general public are limited, that this obviously reduces...
the general understanding of public health statistics, studies have also documented that understanding of the CIs is poor even among physicians, as heuristic reasoning often prevails on sample size. Therefore, in order to facilitate comprehension of the epidemiological message, we have chosen the FP as a visual display method to allow the reader to identify the SIR for each municipality within the plot, and the different attention-level areas (represented by different colours) under which each location falls (figure 1B).

Reading a CM may be misleading for stakeholders since the fear of being overexposed to environmental and other risk factors may lead to misinterpretation of the differences in colour scale, which do not properly display the potential inaccuracy in the estimation of cancer indicators (figure 2). On the other hand, the conservative choice of reporting only statistically significant increased cancer risks, as shown for the Palermo city hot spot (figure 3), excludes from the discussion the residents of most municipalities who would certainly be interested in knowing ‘what is going on in their back yard’. The combination of FP and CM, supported by tabulation of the numeric results, allows to identify locations where cancer incidence may deserve further attention, such as the municipality of Isnello, with a high SIR but a 95% CI including the null value. Clear understanding by the relevant stakeholders and their productive engagement may clarify whether such borderline findings simply reflect inadequate sample size, chance or a departure from the expected incidence that deserves further investigation.

Within the context of the chosen sample population and data, it has to be considered the presence of a single area containing a large proportion of the entire study population must be highlighted. This obviously influences each SIR value, but its potential effects are related to the study population used in the calculation of SIRs, and do not influence the FP methodology itself. Moreover, the graphic FP representation, differently from the more commonly used visual formats, allows the reader to observe, simultaneously, the situation of the municipality of interest in relation to the entire study population and to three specific areas (under control, warning and alert) representing the different attention levels. Moreover, it should also be kept in mind that the SIR values have been standardised using the EU population as external reference, allowing adjustment for age. Finally, the presence of a single area with a substantial population (Palermo city) implies an overestimation of expected cases, but the epidemiological message did not change even after the exclusion of the Palermo city area from the analysis (data not shown).

Following the methodological approach proposed, representation of the PP SIRs through FP seemed to be congruent with CM generated using the same data, with the former resulting more informative dealing with some of the dimensions explored, as shown by the comparisons of the weaknesses and strengths between the two visual formats (table 2). In particular, with regard to the strengths of the proposed visual format, FP shows the scope of the phenomenon under investigation and the precision and significance of estimates simultaneously, by simply positioning the indicator of interest in one of the three cancer attention areas; on the contrary, the more commonly used CMs monodimensionally represent the parameters of interest by using a different colour gradation based on the frequency distribution of the values. The highlighted difference could be considered the main reason for making FP more comprehensive to stakeholders than CM. However, the weaknesses of FP also need to be taken into account. FP cannot be considered the ideal visual format to highlight the geographical position of the indicator of interest (SIR) and, consequently, to define any spatial cluster. Finally, both FP and CM had the ability to identify potential hot spots, even though for CM, it is necessary to further validate the hot spot by using suitable statistical tests (eg, the GAM approach). All of the previous considerations have led us to believe that FP could be used as a complement to CM, according to its properties, particularly in terms of validity and in terms of interpretability.

However, the proposed complementary dissemination approach needs to be further validated in the field both by involving local communities and by administering the two different visual formats to a sample of stakeholders according to the Delphi consensus process. In fact, it can be presumed that the efficacy of a presentation format depends both on the type of format, and on the context in which the format is used (scientific vs general public).

CONCLUSIONS

According to the proposed comparison between the two explored methodological approaches, we concluded that FP should be considered as a complement to the current and commonly used graphical and visual formats (CMs, tables, GAM maps) to effectively communicate cancer registry statistics, particularly incidence rate, to communities and local authorities, visually conveying an efficient and simple to interpret cancer epidemiological data.

Future research on cancer risk communication should concentrate on the presentation format and on the framework in which the message is presented. From this perspective, the FP could represent a useful tool for empowering health communications to local communities and other stakeholders (patients’ associations, physicians, pharmacists, local administration, etc).

Author affiliations
1Registro Tumori di Palermo e Provincia, University of Palermo, Palermo, Italy
2Dipartimento Scienze Promozione della Salute e Materno Infantile “G. D’Alessandro”, Sezione di Igiene, University of Palermo, Palermo, Italy
3Azienda Sanitaria Provinciale 6 di Palermo—Dipartimento di Prevenzione, Palermo, Italy
4Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Azienda Ospedaliera Universitaria Policlinico “Paolo Giaccone”, Palermo, Italy

REFERENCES


