BMJ Open

Funnel Plot as useful alternative tool to communicate epidemiological data of population based cancer registries.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011502
Article Type:	Research
Date Submitted by the Author:	03-Apr-2016

Complete List of Authors:	MAZZUCCO, WALTER; University of Palermo, Scienze Promozione Salute e Materno Infantile Cusimano, Rosanna ; Azienda Sanitaria Provinciale 6 di Palermo, Dipartimento Prevenzione Zarcone, Maurizio; Azienda Ospedaliera Universitaria Policlinico “Paolo Giaccone”, Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Mazzola, Sergio; Azienda Ospedaliera Universitaria Policlinico “Paolo Giaccone”, Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Vitale, Francesco; University of Palermo, Department of Sciences for Health Promotion “G. D’Alessandro” – Hygiene section

Primary Subject Heading:	Public health
Secondary Subject Heading:	Public health, Epidemiology, Oncology
Keywords:	Funnel plot, cancer epidemiology, cancer registry;, cancer burden communication, Standardized Incidence Ratio

SCHOLARONE™ Manuscripts

TITLE

The funnel plot as a useful tool for communicating epidemiological data of population based cancer registries**.**

Authors

Mazzucco Walter[1,4], *Cusimano Rosanna*[1,3], *Zarcone Maurizio*[1,2], *Mazzola Sergio*[1,2], *Vitale Francesco*[1,2,4].

1. Registro Tumori di Palermo e Provincia, University of Palermo, Palermo, Italy.

2. Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Azienda Ospedaliera Universitaria Policlinico "Paolo Giaccone", Palermo, Italy

3. Azienda Sanitaria Provinciale 6 di Palermo - Dipartimento di Prevenzione, Palermo, Italy

4. Dipartimento Scienze Promozione della Salute e Materno Infantile "G. D'Alessandro", Sezione di Igiene, University of Palermo, Palermo, Italy

Mazzucco Walter (Corresponding author), Dipartimento Scienze Promozione della Salute e Materno Infantile "G. D'Alessandro", Sezione di Igiene, Università degli Studi di Palermo. Via del Vespro, 133. 90127 Palermo. E-mail: walter.mazzucco@unipa.it.

Cusimano Rosanna, Dipartimento Prevenzione - Azienda Sanitaria Provinciale 6 di Palermo. Via Siracusa, 45. 90100 Palermo. E-mail: rosannacusimano55@gmail.com.

Zarcone Maurizio, Registro Tumori di Palermo e Provincia - Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Azienda Ospedaliera Universitaria Policlinico "Paolo Giaccone". Via del Vespro, 133. 90127 Palermo. E-mail: zarcone7@gmail.com.

Mazzola Sergio, Registro Tumori di Palermo e Provincia - Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Azienda Ospedaliera Universitaria Policlinico "Paolo Giaccone". Via del Vespro, 133. 90127 Palermo. E-mail: mazzolasergio3@gmail.com.

Vitale Francesco, Dipartimento Scienze Promozione della Salute e Materno Infantile "G. D'Alessandro", Sezione di Igiene, Università degli Studi di Palermo. Via del Vespro, 133. 90127 Palermo. E-mail: francesco.vitale@unipa.it.

Keywords

Funnel plot; cancer registry; cancer epidemiology; cancer burden communication; Standardized Incidence Ratio.

Word count: 1834

Number of figures: 3

Number of tables: 2

Number of references: 41

Details of preferred reviewers

Maurizio Macaluso, Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA - Email: Maurizio.Macaluso@cchmc.org

Chiara De Waure, Public Health Institute, Medicine Faculty, Catholic University of Sacred Hearth (UCSC), Rome, Italy - Email: chiara.dewaure@rm.unicatt.it

Maria Rosaria Gualano, Department of Public Health and Pediatrics, University of Turin, Turin, Italy - Email: mariarosaria.gualano@unito.it

ABSTRACT

Background: Population based cancer registries produce epidemiologic cancer information, but indicators are often difficult to be interpreted by local authorities and communities, according to numeracy and literacy limitations. Aim of the paper is to propose an alternative communication format to common visual formats, using funnel plots to address local Public Health authorities and communities to access valid and understandable measures of burden of cancer at the municipal level.

Methods: Funnel plot graphical representation of Standardized Incidence Ratio was applied in the 82 municipalities of the Palermo Province, by using Palermo Province Cancer Registry (Sicily, Italy) 2003-2011 data. Funnel plot and choropleth map methodologies were compared by properties to communicate epidemiological data to stakeholders.

Results: The SIRs of the municipalities laid all within the control lines, except the Palermo city one (SIR = 1.12), sited outside the upper control line of 99.8%. The Palermo Province SIRs representation through Funnel plot was concordant, but more informative than the same shown by the choropleth map, as resumed by the comparisons of weaknesses and strengths between the two visual formats.

Conclusions: Funnel plot should be used as useful tool to communicate epidemiological data of cancer registries to communities and local authorities, visually conveying a valid and and easy way to interpret measure of the cancer burden.

Article summary

Population based cancer registries are in charge of cancer surveillance by continuously recording and monitoring cancer incidence, mortality, survival and prevalence of a resident population in a specific administrative area (in Italy usually corresponding to the district level). Cancer registries are frequently queried by health authorities or citizens to answer specific questions with regard to the perception of a potential local high risk exposure, aiming to facilitate the correct interpretation of the results and to preserve a rigorous methodological approach at the same time. Unfortunately, the different visual and graphical formats currently used to represent and publish data on cancer epidemiology (choropleth maps, tables, Scan statistic - Geographical Analysis Machine) can often result in a low informative or misleading message to non-expert stakeholders.

Strengths and limitations of the study

1) As far as our knowledge, this study has explored for the first time the application of funnel plot methodology to represent cancer standardized incidence ratio at the municipal level, comparing it with the commonly used visual format, the choropleth map.
2) The study supports the use of funnel plot as an alternative tool to choropleth map in order to communicate epidemiological data of cancer registries to local communities and authorities.
3) The proposed communication approach needs to be further validated on the field. To this purpose the Palermo Province Cancer Registry had defined 82 municipal risk maps, one for each municipality belonging to the Palermo province, and along a one year period it'll be involved in on site meetings to experience the presentation to stakeholders (citizens, local authorities, general practitioners, specialized physicians, pharmacists, etc) of cancer incidence data by using funnel plots. Delphi consensus process will be explored as well by involving public health operators.

BACKGROUND

Cancer is the second cause of death in the developed countries.[1] In the last 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 exposures.[2] Although the estimated contribution of changes in lifestyles and environmental factors to cancer trends,[3-6] the global increase in cancer prevalence is largely attributable to a combination of improving cancer survival[7] and population aging.[8] Local communities have a variable degree of literacy and numeracy, which in turn influence their understanding of these demographical and epidemiological concepts.[9, 10] Local public health and political authorities are regularly engaged in finding better ways to satisfy the growing demand for information on cancer causes and impact that is expressed by the general public.[11]

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*".[12] Population-based cancer registries (PBCRs) carry out cancer surveillance by continuously recording and monitoring cancer incidence, mortality, survival and prevalence of a resident population in a specific administrative area. The mission of PBCRs includes the translation and communication of evidence to inform decision making and to empower the general population or others stakeholders, preserving a rigorous methodological approach and facilitating a correct interpretation of results. PBCR publications report validated and internationally shared measurements and use terminology and visual formats that are easily understood by the scientific community but often difficult to interpret for other stakeholders.[13, 14]

The most common format used to report geographic comparisons in cancer epidemiology is the atlas, which includes thematic maps, as choropleth map (CM), representing measures of cancer frequency (standardized rates, standardized ratios, etc.) computed for areas with defined administrative boundaries.[15]

While data are available on how the context[16] and the content level influence individual risk perception,[17] little is known about the effects of risk communications at the group level, particularly in small communities.[18]

The Italian Association of Cancer Registries (AIRTum), a national network of 41 local population-based cancer registries, including the Palermo Province Cancer Registry (PPCR), has paid great attention in communication tools improving.[19]

The aim of this paper is to propose the use of funnel plot (FP) as an alternative communication format, as compared to the common visual ones, employed by the PPCR to address local Public Health authorities and communities to facilitate dissemination and interpretation of measures of cancer burden at the municipal level.

METHODS

The 51,951 incident cancer cases, excluding non-melanoma skin cancers, registered between 2003 and 2011 by PPCR among the 1.244.239 resident inhabitants in the Palermo Province (PP) 82 municipalities (679.850 resident inhabitants only in the Palermo metropolitan area),[20] comprise the study population.

Cancer incidence in the PP municipalities was measured by using Standardized Incidence Ratio (SIR), defined as the ratio between the observed cases (O_i) and the expected cases (E_i).[21] The O_i have been assumed under a homogeneous Poisson distribution with parameter $\lambda = \theta_0 \cdot E_i$. The E_i have been estimated by indirect method[22] with $\Sigma O_i = \Sigma E_i$.[23] The resident population has been reported by using the inter-census estimates, provided by the National Statistical Institute (ISTAT), considering also the annual municipal data on migration.[20] For each SIR the 95% confidence interval (CI) was calculated by using the normal approximation.[24]

FP graphical representation[24] was used to highlight any municipality with an excess in cancer incidence as compared to the reference population. The following elements have been included to generate the FP: the 82 municipalities' SIRs on the y-axis; the target line ($\theta_0 = 1$), representing the reference value for the indicator of interest ($O_i = E_i$); the E_i precision parameter, measuring the accuracy of the indicator of interest (Poisson variance parameter, under the hypothesis $\theta_0 = 1$), represented on the x-axis; the 95% and 99.8% CIs, calculated with the normal approximation, defining the control lines.[24]

As data distribution could not agree with the underlying assumption (variance equal to the expected value), in order to control any possible over-dispersion[25] both additive and multiplicative approaches were adopted. Over-dispersion coefficients (τ for additive approach and ϕ for multiplicative approach) were calculated. Over-dispersion was treated by considering the "winsorized" estimates too.[26] In addition, the direct selection of units with extremes Z-score[27] and the winsorization method (by testing for different levels of Z-score quantiles[25]) were applied. Furthermore, to define the level of winsorization, a R-script routine was developed to set a cut-off for the quantile between the acceptance of the over-dispersion test and its rejection.

The map representing the PP municipalities was generated by using ISTAT Shapefile vector format,[28] released in the reference system ED50 (European Datum – 1950) UTM Zone 32N and converted in plane coordinates (decimal degrees), providing geo-referenced data in addition to the coordinates of geographic objects and their borders (for polygons), also including the information about the spatial location of each municipality. Although traditional geographical analyses use the centroids as geo-statistical units, considering that some centroid could fall outside its municipal bounds, the coordinates of the city hall were adopted instead.[29]

PP cancer incidence variation was also described by a CM,[30] representing the SIR of each municipality. To distinguish potential high and low risk areas, a central interval of 0.95-1.05 for the color scale was fixed, irrespective of statistical significance. Both the values above 1.05 and below 0.95 were divided in tertiles.[31]

Cluster analysis was performed by using the scan statistics Openshaw's Geographical Analysis Machine (GAM), with varying radiuses, in order to detect potentials high risk clusters and hot spots, setting p-value at 0.002.[32] The analysis for hot spots research was performed by using circles of radius 3 Km for each point of a grid covering the study region by steps of 600 meters (radius/5).

To perform statistical analysis the RStudio IDE[33] for the R software, version 3.1.0 (2014-04-10) - "Spring Dance",[34] was used.

Lastly, weaknesses and strengths of FP and CM methodological approaches were compared according to the available literature.[31, 35, 36]

RESULTS

Figure 1 represents the FP of 82 municipality-specific SIRs, corrected for over-dispersion (ϕ=13.46) and adjusted using the multiplicative approach.[25] All of the SIRs lay within the control lines, except the Palermo city one (SIR= 1.12), which is above the upper control line (UCL) of 99.8%. Over-dispersion test results were concordant and the routine did not find out any valid value for winsorization (data not shown).

Figure 2 depicts the CM of cancer incidence in the 82 municipalities of the PP, generated by using 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. Expected cases and SIRs (sorted by descendent order) with 95% CIs in the 82 Palermo Province municipalities (Study period 2003-2011).

ISTAT code	Municipality	Expected	SIR	95% CI	ISTAT code	Municipality	Expected	SIR	95% CI
082042	Isnello	104.1	1.22	0.99-1.45	082031	Cinisi	441.0	0.82	0.74-0.90
082053	***Palermo City***	***27371.4***	***1.12***	***1.11-1.14***	082007	Balestrate	292.4	0.81	0.72-0.91
082061	Roccamena	81.4	1.06	0.83-1.29	082067	Santa Flavia	414.6	0.81	0.73-0.89
082070	Termini Imerese	1166.4	1.05	0.99-1.11	082059	Pollina	148.0	0.81	0.68-0.94
082027	Cefalù	685.3	1.01	0.93-1.08	082030	Ciminna	209.5	0.81	0.70-0.92
082044	Lascari	152.2	1.01	0.85-1.17	082064	San Giuseppe Jato	379.0	0.81	0.73-0.89
082014	Caccamo	396.1	0.98	0.88-1.07	082058	Polizzi Generosa	212.6	0.80	0.69-0.91
082035	Ficarazzi	357.5	0.97	0.87-1.07	082036	Gangi	406.6	0.80	0.72-0.87
082038	Giardinello	84.1	0.96	0.76-1.17	082023	Casteldaccia	416.4	0.79	0.72-0.89
082056	Petralia Sottana	168.6	0.95	0.81-1.09	082004	Altavilla Milicia	242.8	0.78	0.68-0.88
082012	Bompietro	104.6	0.95	0.77-1.13	082005	Altofonte	379.8	0.78	0.70-0.86
082049	Monreale	1319.0	0.94	0.89-0.99	082046	Marineo	310.8	0.78	0.70-0.87
082052	Palazzo Adriano	123.4	0.93	0.77-1.10	082025	Castronovo di Sicilia	175.4	0.78	0.67-0.90
082079	Villabate	631.9	0.93	0.85-1.00	082050	Montelepre	258.1	0.78	0.68-0.87
082008	Baucina	102.0	0.92	0.74-1.10	082034	Corleone	520.3	0.78	0.71-0.84
082006	Bagheria	2068.7	0.92	0.88-0.96	082054	Partinico	1291.1	0.78	0.73-0.82
082076	Valledolmo	212.5	0.92	0.79-1.04	082051	Montemaggiore Belsito	208.5	0.77	0.66-0.87
082017	Campofelice di Roccella	272.6	0.91	0.80-1.02	082078	Vicari	156.2	0.76	0.64-0.88
082074	Trappeto	151.5	0.91	0.77-1.06	082001	Alia	225.4	0.76	0.66-0.86
082020	Capaci	390.4	0.92	0.82-0.99	082010	Bisacquino	272.7	0.75	0.66-0.84
082019	Camporeale	157.5	0.90	0.76-1.04	082063	San Cipirello	222.7	0.75	0.65-0.85
082071	Terrasini	446.0	0.90	0.86-0.98	082002	Alimena	134.9	0.74	0.62-0.87
082048	Misilmeri	969.4	0.90	0.84-0.95	082065	San Mauro Castelverde	120.1	0.74	0.61-0.87
082045	Lercara Friddi	335.8	0.90	0.80-0.99	082082	Blufi	76.5	0.73	0.57-0.90
082028	Cerda	243.1	0.89	0.78-1.01	082003	Aliminusa	73.8	0.73	0.57-0.90
082043	Isola delle Femmine	235.3	0.88	0.77-0.99	082026	Cefalà Diana	50.6	0.73	0.53-0.93
082032	Collesano	220.9	0.88	0.77-0.99	082072	Torretta	143.6	0.73	0.61-0.85
082024	Castellana Sicula	200.3	0.87	0.75-0.99	082039	Giuliana	121.2	0.72	0.59-0.85
082041	Gratteri	65.2	0.87	0.66-1.09	082047	Mezzojuso	146.5	0.71	0.60-0.82
082060	Prizzi	283.8	0.87	0.77-0.98	082055	Petralia Soprana	201.3	0.71	0.61-0.81
082021	Carini	1146.5	0.87	0.82-0.92	082062	Roccapalumba	139.0	0.71	0.59-0.82
082029	Chiusa Sclafani	181.4	0.86	0.73-0.99	082013	Borgetto	261.3	0.70	0.62-0.79
082009	Belmonte Mezzagno	372.3	0.86	0.77-0.94	082037	Geraci Siculo	112.7	0.69	0.56-0.82
082073	Trabia	382.2	0.86	0.77-0.94	082066	Santa Cristina Gela	40.6	0.69	0.48-0.90
082057	Piana degli Albanesi	305.0	0.85	0.76-0.95	082018	Campofiorito	74.2	0.67	0.52-0.83

082081	Scillato	36.6	0.85	0.57-1.12	082075	Ustica	66.6	0.66	0.50-0.82
082015	Caltavuturo	229.0	0.84	0.73-0.95	082033	Contessa Entellina	101.6	0.66	0.53-0.79
082080	Villafrati	166.2	0.84	0.71-0.96	082077	Ventimiglia di Sicilia	116.9	0.62	0.50-0.73
082022	Castelbuono	469.4	0.83	0.76-0.91	082040	Godrano	52.2	0.61	0.45-0.78
082068	Sciara	117.2	0.83	0.68-0.98	082069	Sclafani Bagni	29.4	0.58	0.37-0.79
082011	Bolognetta	162.6	0.82	0.70-0.95	082016	Campofelice di Fitalia	34.6	0.46	0.31-0.62

Table 1, represents the expected cases (males and females combined) and SIRs with 95% CIs in the 82 PP municipalities: a statistically significant value higher than 1 was documented only for the city of Palermo (SIR=1,12; 95% CIs= 1.11-1.14). Moreover, most of CIs lied below the value of 1.

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

Table 2 summarizes the comparison of weaknesses and strengths between the different visual formats explored in the prospective to communicate epidemiological data to stakeholders.

Table 2. Comparison of weaknesses (-) and strengths (+) between Funnel Plot and Choropleth Map in the prospective to communicate epidemiological data to stakeholders.

Properties explored	**Weaknesses and Strengths of Visual format**	
	Funnel Plot	**Choropleth Map**
To define the spatial location of the risk	-	+
To identify hot spots	+	+
To locate clusters	-	+
To show the dimension of phenomenon under investigation	+	-
To show the precision of estimates	+	-
To communicate the significance of the estimates	+	-

In terms of strengths, FP differed from CM by property to communicate epidemiological data to stakeholders and particularly with regard to the capability a) to show the dimension of phenomenon under investigation, b) to show the precision of estimates and c) to highlight the significance of the estimates. On the opposite, CM as compared to FP was able a) to define the spatial location of the risk and b) to locate the presence of any cluster. Both FP and CM had the property to identify any hot spot.

DISCUSSIONS

FP are commonly used in process control and, particularly, in the health care field to compare institutional performance measures[27] but rarely in public health surveillance.[37] We explored FPs to address local authorities and communities and provide them with synthetic access to valid and understandable measures of cancer incidence (SIRs) at the municipal level.

As the SIR is a valid and well established measure in the descriptive epidemiology of cancer,[21] we used this measure to compare its presentation via the FP with more common formats used to represent epidemiological data on cancer.

Whereas scale risk tables are easy to understand,[17] readers do not usually pay attention to the confidence interval, which is a critically important measure of the precision of SIR estimates.[38] By displaying sample statistics together with the corresponding sample size, and their relation to control limits, FPs allow visualizing both information and uncertainty without the need for processing several numeric values (in this study, 82 point estimates and 164 confidence boundaries).[37] Moreover, while it is common knowledge that the numeracy skills of the general public are limited, and reduce general understanding of public health statistics, studies have also documented how understanding of the confidence intervals is poor even among physicians, as heuristic reasoning often prevails on sample size.[39]

Reading a CM may be misleading for stakeholders[40] as the fear of being over-exposed to environmental and other risk factors may lead to over-interpreting differences in color scale, which do not represent properly the uncertainty in the estimation of measures of cancer burden (Figure 2). On the other hand, the conservative choice implied by reporting only statistically significant excess cancer risks, as shown for the Palermo city hot spot (Figure 3), excludes from the discussion the residents of most municipalities who would be interested in knowing “what 's going on in their back yard” even if it does not represent excess cancer risk.

According to the methodological approach proposed, the Palermo Province SIRs representation through FP seemed to be concordant, but more informative than the same shown by the CM, as resumed by the comparisons of weaknesses and strengths between the two visual formats. Particularly, all of the previous considerations led us to believe that FP could be preferable to CM both in terms of validity and in terms of interpretability.

However, the proposed alternative communication approach needs to be further validated on the field by administering the two different visual formats to a sample of stakeholders according to Delphi consensus process [41]. In fact, it can be assumed that the effects of a presentation format depend not only on the type of format, but also on the context in which the format is used (scientific versus general public).[16]

CONCLUSIONS

According to the proposed comparison between the two explored methodological approaches, we concluded that FP should be considered as an alternative to the current and commonly used graphical and visual formats (CMs, tables, Scan statistic GAM) to effectively communicate cancer registry statistics to communities and local authorities, visually conveying a valid and easy way to interpret measure of the cancer burden.

Future research on cancer risk communication should not only concentrate on the presentation format, but also on the framework in which the message is presented. In this perspective, the FP could represent a useful tool for empowering health communication with local communities and other stakeholders (patients' associations, physicians, pharmacists, local administration, etc.).

ACKNOWLEDGEMENTS

Contributorship statement

All individuals listed as authors have contributed substantially to design, performing or reporting of the work and every specific contribution is indicated as follows.

Conception and design of the study: WM, RC, MZ, SM

Statistical analysis: MZ, SM

Interpretation of data: WM, RC, MZ, SM

Manuscript writing and drafting: WM, RC

Revision of the manuscript: FV, WM, RC

Approval of the final version of the manuscript: WM, RC, MZ, SM, FV

Data sharing statement

Extra data (results of over-dispersion tests, R-script to detect the greatest cut-off for the winsorization procedure, and others statistical results) can be available by emailing to walter.mazzucco@unipa.it.

COMPETING INTERESTS

The authors declare they have no conflict of interest.

FUNDING

None.

REFERENCE

1 Summary tables: deaths by cause, age and sex 2000-2012. In: Global health estimates 2014. WHO. http://www.who.int/topics/global_burden_of_disease/en/ (accessed 18 Jan 2016).

2 Barchitta M, Fragapane S, Quattrocchi A, et al. Environmental health risk communication in the case "Terra dei Fuochi": content analysis of online newspaper articles. *Ann Ig* 2015;Jan-Feb;27(1):30-8.

3 Pirastu R, Comba P, Conti S, et al. Studio epidemiologico dei territori e degli insediamenti esposti a rischio da inquinamento. Mortalità incidenza oncologica e ricoveri ospedalieri. *Epidemiol Prev* 2004;38(2)(Suppl 1).

4 K. Straif, A. Cohen, J. Samet. Air pollution and cancer. IARC Scientific publication no.161. https://www.iarc.fr/en/publications/books/sp161/AirPollutionandCancer161.pdf (accessed 18 Jan 2016).

5 Monographs on the Evaluation of Carcinogenic Risk to Humans. International Agency for Research on Cancer (IARC). http://monographs.iarc.fr/ENG/Monographs/PDFs/index.php (accessed 18 Jan 2016).

6 Boffetta P and Nyberg F. Contribution of environmental factors to cancer risk. *Br Med Bull* 2003;68:71-94.

7 Allemani C, Weir HK, Carreira H, et al. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* 2015 Mar 14;385(9972):977-1010.

8 Zahedi A1, Rafiemanesh H, Enayatrad M, Ghoncheh M, Salehiniya H. Incidence, Trends and Epidemiology of Cancers in North West of Iran. Asian Pac J Cancer Prev 2015;16(16):7189-93.

9 Melissa G. Health literacy and Numeracy. Workshop Summary. The National Academy Press. Washington 2014. http://www.nap.edu/catalog/18660/health-literacy-and-numeracy-workshop-summary (accessed 18 Jan 2016).

10 Reyna VF, Nelson WL, Han PK and Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull* 2009;135(6):943-73.

11 Edwards A, Elwyn G and Mulley A. Explaining risks: turning numerical data into meaningful pictures. *BMJ* 2002;324;827-830.

12 CDC Updated Guidelines for evaluating Public Health Surveillance Systems. MMWR 2001/50(RR13);1-35. http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm (accessed 18 Jan 2016).

13 CANCER IN FRIULI VENEZIA GIULIA INCIDENCE, SURVIVAL AND PREVALENCE: 1995-2005. Agenzia Regionale della Sanità Friuli Venezia Giulia. http://www.cro.it/opencms/export/sites/cro/it/azienda_informa/_docs/registro-tumori-1995-2005.pdf (accessed 18 Jan 2016).

14 Rapporto sui tumori a Palermo e provincia. Anni 2003 – 2007. Registro Tumori Palermo e Provincia. Anno 2011. http://www.unipa.it/dipartimenti/dipartimentodiscienzeperlapromozionedellasaluteematernoinfantileg.dalessandro/.content/documenti/Rapporto-sui-tumori-a-Palermo-e-Provincia-anni-2003-2007-1.pdf (accessed 18 Jan 2016).

15 Atlante Sanitario della Sicilia. Suppl. monografico Notiziario Dipartimento per le Attività Sanitarie ed Osservatorio Epidemiologico. Assessorato della Salute, Regione Sicilia. Anno 2013: http://pti.regione.sicilia.it/portal/page/portal/PIR_PORTALE/PIR_LaStrutturaRegionale/PIR_AssessoratoSalute/PIR_AreeTematiche/PIR_Atlanteoncologico (accessed 18 Jan 2016).

16 Visschers VH, Meertens RM, Passchier WW, de Vries NN. Probability information in risk communication: a review of the research literature. *Risk Anal* 2009;29(2):267-87.

17 Lipkus IM, Hollands JG. The visual communication of risk. *Natl Cancer Inst Monogr* 1999;25:149-63.

18 Mclnerny GJ, Chen M, Freeman R, Gavaghan D, Meyer M, Rowland F et al. Information visualization for science and policy: engaging users and avoiding bias. *Trends Ecol Evol* 2014;Vol 29,Iss 7:148-157. doi:10.1016/j.tree.2014.01.003 [published Online First: 21 February 2014] (accessed 18 Jan 2016).

19 Mappa dei Registri Tumori di popolazione. Associazione Italiana Registri Tumori (AIRTUM) 2015. http://www.registri-tumori.it/cms/it/copertura (accessed 18 Jan 2016).

20 Ricostruzione intercensuaria della popolazione 2002-2011. Istituto Nazionale di Statistica. 2015; www.demo.istat.it (accessed 18 Jan 2016).

21 Greenland S, Rothman KJ and Lash TL. Measures of Effect and Measures of Association. In: Rothman KJ, Greenland S and Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins 2008:67-69.

22 Mantel N and Stark CR. Computation of indirect-adjusted rates in the presence of confounding. Biometrics, Vol. 24, No. 4 (Dec., 1968), pp. 997-1005

23 Morris SE, Wakefield JC. Assessment of disease risk in relation to a pre-specified source. In: Elliot P, Wakefield JC, Best NG, Briggs DJ, editors. Spatial epidemiology: methods and applications. New York: Oxford University Press 2000:53–184.

24 Spiegelhalter DJ. Funnel plots for institutional comparisons. *Qual Saf Health Care* 2002;11;390-391.

25 Silcocks P. Hazard ratio funnel plots for Survival. *J Epidemiol Community Health* 2009;63:856-861.

26 Spiegelhalter DJ. Handling Over-Dispersion of Performance Indicators. *Qual Saf Health Care* 2005;14;347-351.

27 Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med* 2005;24;1185-1202.

28 Confini amministrativi, dei sistemi locali del lavoro e delle NUTS2 in versione generalizzata. Istituto Nazionale di Statistica 2010. http://www.istat.it/it/files/2011/04/com2011_g.zip (accessed 18 Jan 2016).

29 Mazumdar, Soumya. "Shape and scale in detecting disease clusters." PhD (Doctor of Philosophy) thesis, University of Iowa, 2008. http://ir.uiowa.edu/etd/208 (accessed 18 Jan 2016).

30 Richards TB, Berkowitz Z, Thomas CC, Foster SL, Gardner A, King JB et al. Choroplet map design for cancer incidence, part 2. *Prev Chronic Dis* 2010;7(1).

31 Cromley G and Ellen K Cromley. Choropleth map legend design for visualizing comunity heath disparities International. Journal of Health Geographics 2009, 8:52.

32 Besag J and Newell J. The Detection of Clusters in Rare Diseases. Wiley and Royal Statistical Society 1991;154,Part 1,143-155.

33 RStudio Team. RStudio: Integrated Development for R. 2015. http://www.rstudio.com/ (accessed 18 Jan 2016).

34 R Core Team. R: A language and environment for statistical computing 2015. http://www.R-project.org/ (accessed 18 Jan 2016).

35 Sui DZ, Holt JB: Visualizing and analysing public-health data using value-by-area cartograms: toward a new synthetic framework. Cartographica 2008, 43:3-20.

36 Kumar N: Frequency histogram legend in the choropleth map: a substitute to traditional legends. Cartogr Geogr Inform Sci 2004, 31:217-236.

37 Dover DC and Schopflocher DP. Using funnel plot in public health surveillance. *Popul Health Metr* 2011,9:58.

38 Hildon Z, Allwood D and Black N. Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review. *Int J Qual Health Care* 2012;24(1):55-64.

39 Rakow T, Wright RJ, Spieghelhalter D and Bull C. The pros and cons of funnel plots as an aid to risk communication and patient decision making. *Br J Psychol* 2015;106:327-348.

40 Ocana-Riola R. Common errors in disease mapping. *Geospat Health* 2010;4(2)139-54.

41 Chia-Chien Hsu and Brian A. The Delphi Technique: Making Sense of Consensus. Practical Assessment, Research & Evaluation 2007;Vol 12, No 10.

Figure 1. Funnel plot of the SIRs in the 82 Palermo Province municipalities (Study period 2003-2011).

Figure legend:

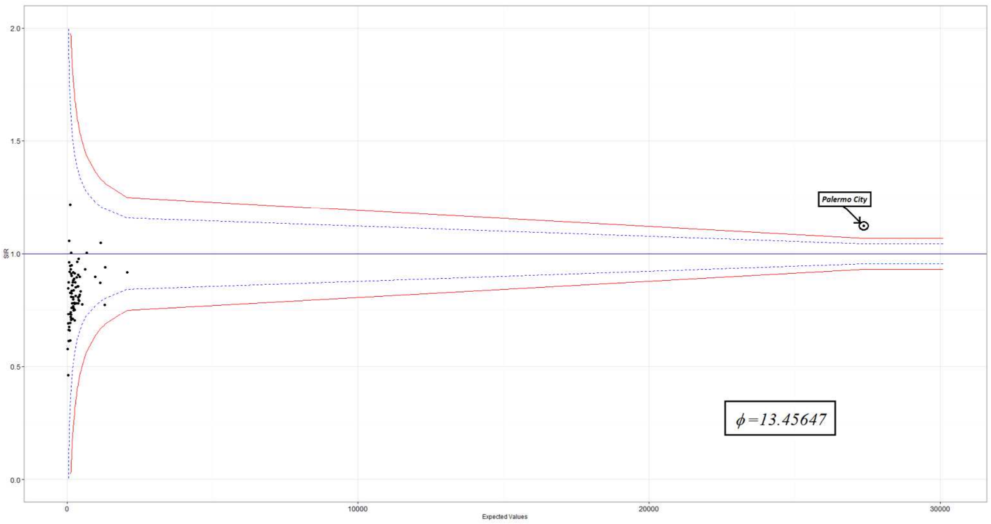
95% CIs ("blue" control lines) and 99.8% CIs ("red" control lines)

ϕ = over-dispersion, calculated with multiplicative approach

Figure 2. Choropleth map of the SIRs in the 82 Palermo Province municipalities^ (Study period 2003-2011).

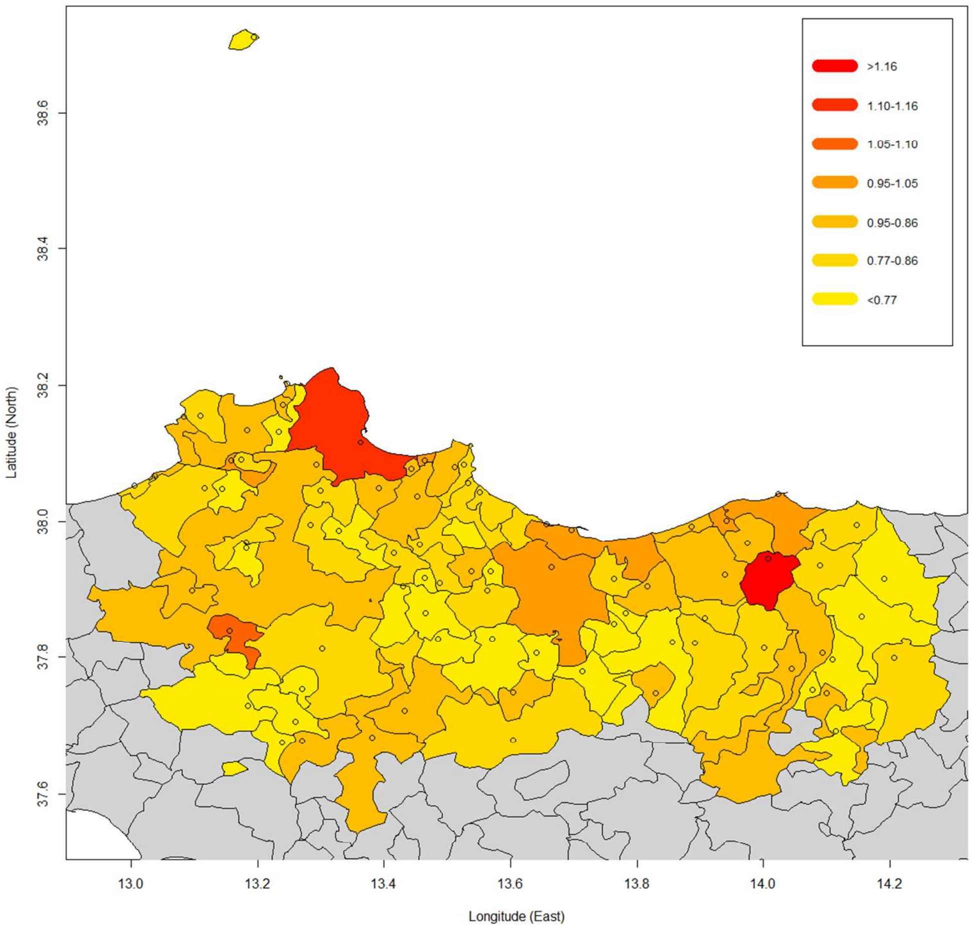
^Circles represent the position of city halls.

Figure 3. Scan statistic GAM of Palermo Province (Study period 2003-2011).



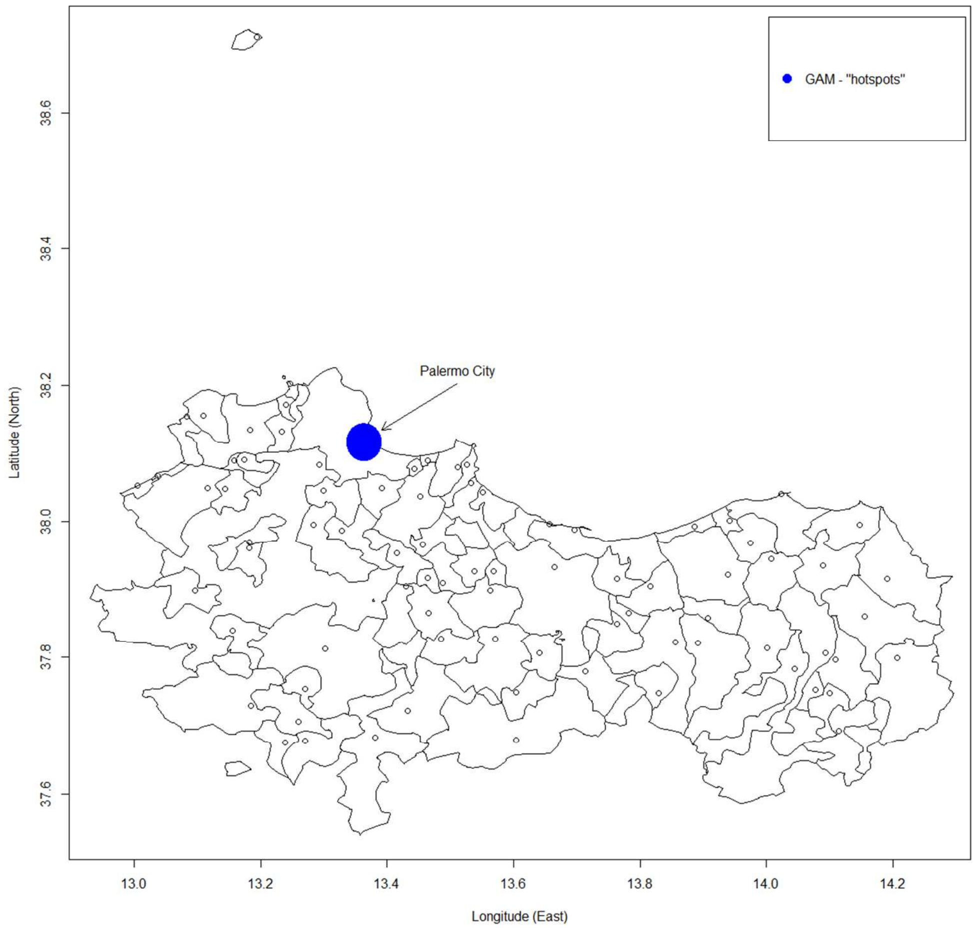
95% CIs (“blue” control lines) and 99.8% CIs (“red” control lines)
φ = over-dispersion, calculated with multiplicative approach

169x89mm (300 x 300 DPI)



^Circles represent the position of city halls

101x101mm (300 x 300 DPI)



101x101mm (300 x 300 DPI)

BMJ Open

Use of funnel plot for disseminating incidence data of population-based cancer registries: the Palermo Cancer Registry experience.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011502.R1
Article Type:	Research
Date Submitted by the Author:	04-Aug-2016
Complete List of Authors:	MAZZUCCO, WALTER; University of Palermo, Scienze Promozione Salute e Materno Infantile Cusimano, Rosanna ; Azienda Sanitaria Provinciale 6 di Palermo, Dipartimento Prevenzione Zarcone, Maurizio; Azienda Ospedaliera Universitaria Policlinico “Paolo Giaccone”, Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Mazzola, Sergio; Azienda Ospedaliera Universitaria Policlinico “Paolo Giaccone”, Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Vitale, Francesco; University of Palermo, Department of Sciences for Health Promotion “G. D’Alessandro” – Hygiene section
Primary Subject Heading:	Public health
Secondary Subject Heading:	Public health, Epidemiology, Oncology, Communication
Keywords:	Funnel plot, cancer epidemiology, cancer registry;, Standardized Incidence Ratio, cancer data dissemination

SCHOLARONE™ Manuscripts

TITLE

Use of funnel plot for disseminating incidence data of population-based cancer registries: the Palermo Cancer Registry experience.

Authors

Mazzucco Walter[1,4]*, Cusimano Rosanna*[1,3]*, Zarcone Maurizio*[1,2]*, Mazzola Sergio*[1,2]*, Vitale Francesco*[1,2,4].

1. Registro Tumori di Palermo e Provincia, University of Palermo, Palermo, Italy.

2. Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Azienda Ospedaliera Universitaria Policlinico "Paolo Giaccone", Palermo, Italy

3. Azienda Sanitaria Provinciale 6 di Palermo - Dipartimento di Prevenzione, Palermo, Italy

4. Dipartimento Scienze Promozione della Salute e Materno Infantile "G. D'Alessandro", Sezione di Igiene, University of Palermo, Palermo, Italy

Mazzucco Walter (Corresponding author), Dipartimento Scienze Promozione della Salute e Materno Infantile "G. D'Alessandro", Sezione di Igiene, Università degli Studi di Palermo. Via del Vespro, 133. 90127 Palermo. E-mail: walter.mazzucco@unipa.it.

Cusimano Rosanna, Dipartimento Prevenzione - Azienda Sanitaria Provinciale 6 di Palermo. Via Siracusa, 45. 90100 Palermo. E-mail: rosanna.cusimano.55@gmail.com.

Zarcone Maurizio, Registro Tumori di Palermo e Provincia - Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Azienda Ospedaliera Universitaria Policlinico "Paolo Giaccone". Via del Vespro, 133. 90127 Palermo. E-mail: zarcone7@gmail.com.

Mazzola Sergio, Registro Tumori di Palermo e Provincia - Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Azienda Ospedaliera Universitaria Policlinico "Paolo Giaccone". Via del Vespro, 133. 90127 Palermo. E-mail: mazzolasergio3@gmail.com.

Vitale Francesco, Dipartimento Scienze Promozione della Salute e Materno Infantile "G. D'Alessandro", Sezione di Igiene, Università degli Studi di Palermo. Via del Vespro, 133. 90127 Palermo. E-mail: francesco.vitale@unipa.it.

Keywords

Funnel plot; cancer registry; cancer epidemiology; cancer data dissemination; Standardized Incidence Ratio.

Word count: 2397

Number of figures: 3

Number of tables: 2

Number of references: 44

Details of preferred reviewers

Maurizio Macaluso, Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA - Email: Maurizio.Macaluso@cchmc.org

Chiara De Waure, Public Health Institute, Medicine Faculty, Catholic University of Sacred Hearth (UCSC), Rome, Italy - Email: chiara.dewaure@rm.unicatt.it

Maria Rosaria Gualano, Department of Public Health and Pediatrics, University of Turin, Turin, Italy - Email: mariarosaria.gualano@unito.it

ABSTRACT

Background: Population based cancer registries provide epidemiologic 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 propose an alternative to common visual formats: using 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 Standardized Incidence Ratio 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 the 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 two visual formats.

Conclusions: Funnel plot should be used as 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.

Article summary

Population-based cancer registries perform cancer surveillance through a continuous recording and monitoring of cancer incidence, survival and prevalence in the resident population of a specific administrative area (in Italy, this usually corresponds to the district level). Cancer registries are frequently queried by health authorities or citizens to answer specific questions, often related to a perception of a potentially high local exposure risk; these registries aim to facilitate correct interpretation of the acquired data and, at the same time, to preserve a rigorous methodological approach. Unfortunately, the different visual and graphical formats (choropleth maps, tables, Geographical Analysis Machine) currently used to represent and publish data on cancer epidemiology can often result in a poorly informative or misleading message to non-expert stakeholders.

Strengths and limitations of the study

1) To the best of our knowledge, this study explores for the first time the application of the funnel plot methodology to represent standardized cancer incidence ratio at the municipal level through a comparison with the commonly used visual format, choropleth map.
2) The results of this study support the use of funnel plot as an alternative to choropleth map for disseminating epidemiological data of cancer registries to local communities and authorities.
3) The proposed communication approach needs to be further validated in the field. To this end, the Palermo Province Cancer Registry has generated 82 municipal risk maps, one for each municipality of the province, and for a period of one year, qualified personnel from the registry will be involved in on-site meetings to share cancer incidence data with stakeholders (citizens, local authorities, general practitioners, specialized physicians, pharmacists, etc) using funnel plots. The Delphi consensus process will be explored as well by involving public health operators.

BACKGROUND

Cancer is the second cause of death in the developed countries.[1] In the last 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.[2] Apart from the presumed effects of lifestyle changes and environmental factors on cancer trends,[3-6] the global increase in cancer prevalence could be largely attributable to a combination of improved cancer survival[7] and aging 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.[11] In particular, citizens often question if they live in an area at high risk for environmental exposure.[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*".[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.[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 measurements 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.[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 (CM), representing cancer incidence rates (standardized rates, standardized ratios, etc.) computed for specific areas.[16, 17] While data are available on how the context[18] and the content of such communications influence individual risk perception,[19] little is known about the effects of risk communications at a group level, particularly in small communities.[20]

The Italian Association of Cancer Registries (AIRTum), a national network of 41 local population-based cancer registries, including the Palermo Province Cancer Registry (PPCR), has greatly emphasized improving communication tools.[21]

The aim of this paper is to propose the use of funnel plots (FPs) for reporting local cancer incidence data, as an alternative 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).[22] Cancer incidence in the PP municipalities was measured by using Standardized Incidence Ratio (SIR), defined as the ratio between observed cases (O_i) and expected cases (E_i).[23] The O_i were assumed to follow a homogeneous Poisson distribution with parameter $\lambda = \theta_0 \cdot E_i$. The E_i were estimated by indirect method,[24] considering the entire population-time under study (the PP) as the reference population, with $\Sigma O_i = \Sigma E_i$.[25] The resident population was reported using the inter-census estimates, provided by the Italian National Statistical Institute (ISTAT), also considering the annual municipal data on migration.[22] For each SIR, the 95% confidence interval (CI) was calculated by using the normal approximation method.[26]

Graphic FP representation[26] was used to highlight any municipality with a higher cancer incidence compared to 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 ($\theta_0 = 1$), representing the reference value for the indicator of interest ($O_i = E_i$); the E_i precision parameter, measuring the accuracy of the indicator of interest (Poisson variance parameter, using the hypothesis $\theta_0 = 1$), represented on the x-axis; the 95% and 99.8% CIs, calculated with the normal approximation method, defining the control limits.[26] The two sets of control limit lines define three different areas within the graph (Figure 1b): the “under-control” area (in green), the “warning” area (in yellow) and the “alert” area (in red).[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[28] both additive and multiplicative approaches were adopted. Overdispersion coefficients (τ for the additive approach and ϕ for the multiplicative approach) were calculated. Overdispersion was addressed by considering the winsorized estimates too.[27] Moreover, Z-score[29] and the winsorization method (by testing for different levels of Z-score quantiles[28]) were applied for the direct selection of extreme values. Furthermore, to define the level of winsorization, an R-script routine was developed to set a cut-off for the quantile between the acceptance and rejection of the overdispersion test (Supplementary material).

The map representing the PP municipalities was generated by using the ISTAT Shapefile vector format,[30] released in the ED50 (European Datum – 1950) reference system. UTM Zone 32N, and converted to plane coordinates (decimal degrees), providing geo-referenced 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 geo-statistical units, considering that some centroid could fall outside the municipal bounds, the coordinates of the city hall were used instead.[31]

The PP cancer incidence variation was also shown in a CM,[32] representing the SIRs of each municipality. To distinguish potential high- and low-risk areas, a central interval of 0.95-1.05 for the color scale was fixed, irrespective of statistical significance. Values above 1.05 and below 0.95 were divided in tertiles.[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.[34] The analysis for hot spot research was performed using circles with a 3 kilometer radius for each point of a grid, covering the study region by steps of 600 meters (radius/5).

The RStudio IDE[35] for the R software, version 3.1.0 (2014-04-10) - "Spring Dance",[36] 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, 37, 38]

RESULTS

Figure 1a represents the FP of 82 municipality-specific SIRs, corrected for overdispersion (ϕ=13.46) and adjusted using the multiplicative approach.[28] All of the SIRs lay within the control limits, except for the Palermo city one (SIR= 1.12), which resulted above the upper control limit line (UCL) 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 winsorization (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 males and females) and SIRs with 95% CIs in the 82 PP municipalities: a statistically significant value higher than 1 was documented only for Palermo city (SIR=1,12; 95% CIs= 1.11-1.14). Moreover, most of the CIs resulted below the value of 1.

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

ISTAT code	Municipality	Expected	SIR	95% CI	ISTAT code	Municipality	Expected	SIR	95% CI
082042	Isnello	104.1	1.22	0.99-1.45	082031	Cinisi	441.0	0.82	0.74-0.90
082053	***Palermo City***	***27371.4***	***1.12***	***1.11-1.14***	082007	Balestrate	292.4	0.81	0.72-0.91
082061	Roccamena	81.4	1.06	0.83-1.29	082067	Santa Flavia	414.6	0.81	0.73-0.89
082070	Termini Imerese	1166.4	1.05	0.99-1.11	082059	Pollina	148.0	0.81	0.68-0.94
082027	Cefalù	685.3	1.01	0.93-1.08	082030	Ciminna	209.5	0.81	0.70-0.92
082044	Lascari	152.2	1.01	0.85-1.17	082064	San Giuseppe Jato	379.0	0.81	0.73-0.89
082014	Caccamo	396.1	0.98	0.88-1.07	082058	Polizzi Generosa	212.6	0.80	0.69-0.91
082035	Ficarazzi	357.5	0.97	0.87-1.07	082036	Gangi	406.6	0.80	0.72-0.87
082038	Giardinello	84.1	0.96	0.76-1.17	082023	Casteldaccia	416.4	0.79	0.72-0.89
082056	Petralia Sottana	168.6	0.95	0.81-1.09	082004	Altavilla Milicia	242.8	0.78	0.68-0.88
082012	Bompietro	104.6	0.95	0.77-1.13	082005	Altofonte	379.8	0.78	0.70-0.86
082049	Monreale	1319.0	0.94	0.89-0.99	082046	Marineo	310.8	0.78	0.70-0.87
082052	Palazzo Adriano	123.4	0.93	0.77-1.10	082025	Castronovo di Sicilia	175.4	0.78	0.67-0.90
082079	Villabate	631.9	0.93	0.85-1.00	082050	Montelepre	258.1	0.78	0.68-0.87
082008	Baucina	102.0	0.92	0.74-1.10	082034	Corleone	520.3	0.78	0.71-0.84
082006	Bagheria	2068.7	0.92	0.88-0.96	082054	Partinico	1291.1	0.78	0.73-0.82
082076	Valledolmo	212.5	0.92	0.79-1.04	082051	Montemaggiore Belsito	208.5	0.77	0.66-0.87
082017	Campofelice di Roccella	272.6	0.91	0.80-1.02	082078	Vicari	156.2	0.76	0.64-0.88
082074	Trappeto	151.5	0.91	0.77-1.06	082001	Alia	225.4	0.76	0.66-0.86
082020	Capaci	390.4	0.92	0.82-0.99	082010	Bisacquino	272.7	0.75	0.66-0.84

082019	Camporeale	157.5	0.90	0.76-1.04	082063	San Cipirello	222.7	0.75	0.65-0.85
082071	Terrasini	446.0	0.90	0.86-0.98	082002	Alimena	134.9	0.74	0.62-0.87
082048	Misilmeri	969.4	0.90	0.84-0.95	082065	San Mauro Castelverde	120.1	0.74	0.61-0.87
082045	Lercara Friddi	335.8	0.90	0.80-0.99	082082	Blufi	76.5	0.73	0.57-0.90
082028	Cerda	243.1	0.89	0.78-1.01	082003	Aliminusa	73.8	0.73	0.57-0.90
082043	Isola delle Femmine	235.3	0.88	0.77-0.99	082026	Cefalà Diana	50.6	0.73	0.53-0.93
082032	Collesano	220.9	0.88	0.77-0.99	082072	Torretta	143.6	0.73	0.61-0.85
082024	Castellana Sicula	200.3	0.87	0.75-0.99	082039	Giuliana	121.2	0.72	0.59-0.85
082041	Gratteri	65.2	0.87	0.66-1.09	082047	Mezzojuso	146.5	0.71	0.60-0.82
082060	Prizzi	283.8	0.87	0.77-0.98	082055	Petralia Soprana	201.3	0.71	0.61-0.81
082021	Carini	1146.5	0.87	0.82-0.92	082062	Roccapalumba	139.0	0.71	0.59-0.82
082029	Chiusa Sclafani	181.4	0.86	0.73-0.99	082013	Borgetto	261.3	0.70	0.62-0.79
082009	Belmonte Mezzagno	372.3	0.86	0.77-0.94	082037	Geraci Siculo	112.7	0.69	0.56-0.82
082073	Trabia	382.2	0.86	0.77-0.94	082066	Santa Cristina Gela	40.6	0.69	0.48-0.90
082057	Piana degli Albanesi	305.0	0.85	0.76-0.95	082018	Campofiorito	74.2	0.67	0.52-0.83
082081	Scillato	36.6	0.85	0.57-1.12	082075	Ustica	66.6	0.66	0.50-0.82
082015	Caltavuturo	229.0	0.84	0.73-0.95	082033	Contessa Entellina	101.6	0.66	0.53-0.79
082080	Villafrati	166.2	0.84	0.71-0.96	082077	Ventimiglia di Sicilia	116.9	0.62	0.50-0.73
082022	Castelbuono	469.4	0.83	0.76-0.91	082040	Godrano	52.2	0.61	0.45-0.78
082068	Sciara	117.2	0.83	0.68-0.98	082069	Sclafani Bagni	29.4	0.58	0.37-0.79
082011	Bolognetta	162.6	0.82	0.70-0.95	082016	Campofelice di Fitalia	34.6	0.46	0.31-0.62

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

Table 2 summarizes a comparison of the weaknesses and strengths, as per the available literature,[29, 33, 37, 38] between the different visual formats explored within the context of disseminating epidemiological data to stakeholders.

Table 2. Comparison of the weaknesses (-) and strengths (+) of the Funnel Plot and Choropleth Map within the context of disseminating epidemiological data to stakeholders.

Properties explored	**Weaknesses and Strengths of Visual format**	
	Funnel Plot	**Choropleth Map**
Definition of the spatial location of the risk	-	+
Identification of hot spots	+	+
Locating clusters	-	+
Displaying the scope of the phenomenon under investigation	+	-
Showing the precision of estimates	+	-
Communicating the significance of estimates	+	-

As show 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.

DISCUSSION

FP are commonly used in process control and, in particular, in the health care field to compare institutional performance data[29]; however, this format is used for survival[39] and standardized mortality ratio[29] in public health surveillance.[40] We explored the use of FPs to local provide authorities and communities with synthetic access to valid and 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,[23] we used this 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,[19] readers do not usually take notice of the confidence interval, which is a critically important measure of the precision of SIR estimates.[41] By displaying sample statistics together with the corresponding sample size, in relation to the control limits, FPs allow visualizing 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).[40] 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 confidence intervals is poor even among physicians, as heuristic reasoning often prevails on sample size.[42] 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 colors) under which each location falls (Figure 1b).

Reading a CM may be misleading for stakeholders[43] since the fear of being overexposed to environmental and other risk factors may lead to misinterpretation of the differences in color 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" even if the local data do not indicate an increased cancer risk.

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. Lastly, 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 Palermo Province SIRs through FP seemed to be congruent with CM generated using the same data, with the former resulting more informative, 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;[29] on the contrary, the more commonly used CMs monodimensionally represent the parameters of interest by using a different color gradation based on the frequency distribution of the values.[33, 37, 38] 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.[29] Lastly, 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 (for example the GAM approach).[34] All of the previous considerations have led us to believe that FP could be preferable to CM both in terms of validity and in terms of interpretability.

However, the proposed alternative 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.[44] In fact, it can be presumed that the efficacy of a presentation format depends not only on the type of format, but also on the context in which the format is used (scientific versus general public).[18]

CONCLUSIONS

According to the proposed comparison between the two explored methodological approaches, we concluded that FP should be considered as an alternative 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 not only concentrate on the presentation format, but also 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.).

ACKNOWLEDGEMENTS

Contributorship statement

All individuals listed as authors have contributed substantially to designing, performing or reporting of the study and every specific contribution is indicated as follows.

Conception and design of the study: WM, RC, MZ, SM

Statistical analysis: MZ, SM

Interpretation of data: WM, RC, MZ, SM

Manuscript writing and drafting: WM, RC

Revision of the manuscript: FV, WM, RC

Approval of the final version of the manuscript: WM, RC, MZ, SM, FV

The document has been reviewed and corrected by a native English speaker with extensive scientific editorial experience to ensure a high level of spelling, grammar and punctuation.

Data sharing statement

Supplementary data (results of over-dispersion tests, R-script to detect the greatest cut-off for the winsorization procedure) have been provided as a supplementary file. Other statistical results are available by emailing walter.mazzucco@unipa.it.

COMPETING INTERESTS

The authors declare they have no conflict of interest.

FUNDING

None.

REFERENCES

1 Summary tables: deaths by cause, age and sex 2000-2012. In: Global health estimates 2014. WHO. http://www.who.int/topics/global_burden_of_disease/en/ (accessed 18 Jan 2016).

2 Barchitta M, Fragapane S, Quattrocchi A, et al. Environmental health risk communication in the case "Terra dei Fuochi": content analysis of online newspaper articles. *Ann Ig* 2015;Jan-Feb;27(1):30-8.

3 Pirastu R, Comba P, Conti S, et al. Studio epidemiologico dei territori e degli insediamenti esposti a rischio da inquinamento. Mortalità incidenza oncologica e ricoveri ospedalieri. *Epidemiol Prev* 2004;38(2)(Suppl 1).

4 K. Straif, A. Cohen, J. Samet. Air pollution and cancer. IARC Scientific publication no.161. https://www.iarc.fr/en/publications/books/sp161/AirPollutionandCancer161.pdf (accessed 18 Jan 2016).

5 Monographs on the Evaluation of Carcinogenic Risk to Humans. International Agency for Research on Cancer (IARC). http://monographs.iarc.fr/ENG/Monographs/PDFs/index.php (accessed 18 Jan 2016).

6 Boffetta P and Nyberg F. Contribution of environmental factors to cancer risk. *Br Med Bull* 2003;68:71-94.

7 Allemani C, Weir HK, Carreira H, et al. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* 2015 Mar 14;385(9972):977-1010.

8 WHO Global Status Report on Non-Communicable Diseases 2010. Geneva 2011.

9 Melissa G. Health literacy and Numeracy. Workshop Summary. The National Academy Press. Washington 2014. http://www.nap.edu/catalog/18660/health-literacy-and-numeracy-workshop-summary (accessed 18 Jan 2016).

10 Reyna VF, Nelson WL, Han PK and Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull* 2009;135(6):943-73.

11 Edwards A, Elwyn G and Mulley A. Explaining risks: turning numerical data into meaningful pictures. *BMJ* 2002;324;827-830.

12 CDC Updated Guidelines for evaluating Public Health Surveillance Systems. MMWR 2001/50(RR13);1-35. http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm (accessed 18 Jan 2016).13 Bray F, Soerjomataram I. The Changing Global Burden of Cancer: Transitions in Human Development and Implications forCancer Prevention and Control. In: Gelband H, Jha P, Sankaranarayanan R, Horton S, editors. Cancer: Disease Control Priorities, Third Edition (Volume 3). Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2015 Nov 01. Chapter 2.

14 Rapporto sui tumori a Palermo e provincia. Anni 2003 – 2007. Registro Tumori Palermo e Provincia. Anno 2011. http://www.unipa.it/dipartimenti/dipartimentodiscienzeperlapromozionedellasaluteematernoinfantileg.dalessandro/.content/documenti/Rapporto-sui-tumori-a-Palermo-e-Provincia-anni-2003-2007-1.pdf (accessed 18 Jan 2016).

15 Cancer in Friuli Venezia Giulia Incidence, Survival and Prevalence: 1995-2005. Agenzia Regionale della Sanità Friuli Venezia Giulia. http://www.cro.it/opencms/export/sites/cro/it/azienda_informa/_docs/registro-tumori-1995-2005.pdf (accessed 18 Jan 2016).

16 Jemal A., Vineis P, Bray F, Torre L, Forman D. The Cancer Atlas, second edition. Non series publications IARC Lyon 2014.

17 Atlante Sanitario della Sicilia. Suppl. monografico Notiziario Dipartimento per le Attività Sanitarie ed Osservatorio Epidemiologico. Assessorato della Salute, Regione Sicilia. Anno 2013:

http://pti.regione.sicilia.it/portal/page/portal/PIR_PORTALE/PIR_LaStrutturaRegionale/PIR_AssessoratoSalute/PIR_AreeTematiche/PIR_Atlanteoncologico (accessed 18 Jan 2016).

18 Visschers VH, Meertens RM, Passchier WW, de Vries NN. Probability information in risk communication: a review of the research literature. *Risk Anal* 2009;29(2):267-87.

19 Lipkus IM, Hollands JG. The visual communication of risk. *Natl Cancer Inst Monogr* 1999;25:149-63.

20 Mclnerny GJ, Chen M, Freeman R, Gavaghan D, Meyer M, Rowland F et al. Information visualization for science and policy: engaging users and avoiding bias. *Trends Ecol Evol* 2014;Vol 29,Iss 7:148-157. doi:10.1016/j.tree.2014.01.003 [published Online First: 21 February 2014] (accessed 18 Jan 2016).

21 Mappa dei Registri Tumori di popolazione. Associazione Italiana Registri Tumori (AIRTUM) 2015. http://www.registri-tumori.it/cms/it/copertura (accessed 18 Jan 2016).

22 Ricostruzione intercensuaria della popolazione 2002-2011. Istituto Nazionale di Statistica. 2015; www.demo.istat.it (accessed 18 Jan 2016).

23 Greenland S, Rothman KJ and Lash TL. Measures of Effect and Measures of Association. In: Rothman KJ, Greenland S and Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins 2008:67-69.

24 Mantel N and Stark CR. Computation of indirect-adjusted rates in the presence of confounding. Biometrics, Vol. 24, No. 4 (Dec., 1968), pp. 997-1005

25 Morris SE, Wakefield JC. Assessment of disease risk in relation to a pre-specified source. In: Elliot P, Wakefield JC, Best NG, Briggs DJ, editors. Spatial epidemiology: methods and applications. New York: Oxford University Press 2000:53–184.

26 Spiegelhalter DJ. Funnel plots for institutional comparisons. *Qual Saf Health Care* 2002;11;390-391.

27 Spiegelhalter DJ. Handling Over-Dispersion of Performance Indicators. *Qual Saf Health Care* 2005;14;347-351.

28 Silcocks P. Hazard ratio funnel plots for Survival. *J Epidemiol Community Health* 2009;63:856-861.

29 Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med* 2005;24;1185-1202.

30 Confini amministrativi, dei sistemi locali del lavoro e delle NUTS2 in versione generalizzata. Istituto Nazionale di Statistica 2010. http://www.istat.it/it/files/2011/04/com2011_g.zip (accessed 18 Jan 2016).

31 Mazumdar, Soumya. "Shape and scale in detecting disease clusters." PhD (Doctor of Philosophy) thesis, University of Iowa, 2008. http://ir.uiowa.edu/etd/208 (accessed 18 Jan 2016).

32 Richards TB, Berkowitz Z, Thomas CC, Foster SL, Gardner A, King JB et al. Choroplet map design for cancer incidence, part 2. *Prev Chronic Dis* 2010;7(1).

33 Cromley G and Ellen K Cromley. Choropleth map legend design for visualizing community health disparities International. Journal of Health Geographics 2009, 8:52.

34 Besag J and Newell J. The Detection of Clusters in Rare Diseases. Wiley and Royal Statistical Society 1991;154,Part 1,143-155.

35 RStudio Team. RStudio: Integrated Development for R. 2015. http://www.rstudio.com/ (accessed 18 Jan 2016).

36 R Core Team. R: A language and environment for statistical computing 2015. http://www.R-project.org/ (accessed 18 Jan 2016).

37 Sui DZ, Holt JB: Visualizing and analysing public-health data using value-by-area cartograms: toward a new synthetic framework. Cartographica 2008, 43:3-20.

38 Kumar N: Frequency histogram legend in the choropleth map: a substitute to traditional legends. Cartogr Geogr Inform Sci 2004, 31:217-236.

39 Quaresma M, Coleman MP, Rachet B. Funnel plots for population-based cancer survival: principles, methods and applications. *Stat Med* 2014, 33(6):1070-80. doi: 10.1002/sim.5953 [published Online First: 4 September 2013].

40 Dover DC and Schopflocher DP. Using funnel plot in public health surveillance. *Popul Health Metr* 2011,9:58.

41 Hildon Z, Allwood D and Black N. Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review. *Int J Qual Health Care* 2012;24(1):55-64.

42 Rakow T, Wright RJ, Spieghelhalter D and Bull C. The pros and cons of funnel plots as an aid to risk communication and patient decision making. *Br J Psychol* 2015;106:327-348.

43 Ocana-Riola R. Common errors in disease mapping. *Geospat Health* 2010;4(2)139-54.

44 Chia-Chien Hsu and Brian A. The Delphi Technique: Making Sense of Consensus. Practical Assessment, Research & Evaluation 2007;Vol 12, No 10.

Figure 1. a) Funnel plot of the SIRs in the 82 Palermo Province municipalities (Study period 2003-2011); b) Cancer attention areas: "under-control" area (in green), "warning" area (in yellow) and "alert" area (in red).

Figure legend:

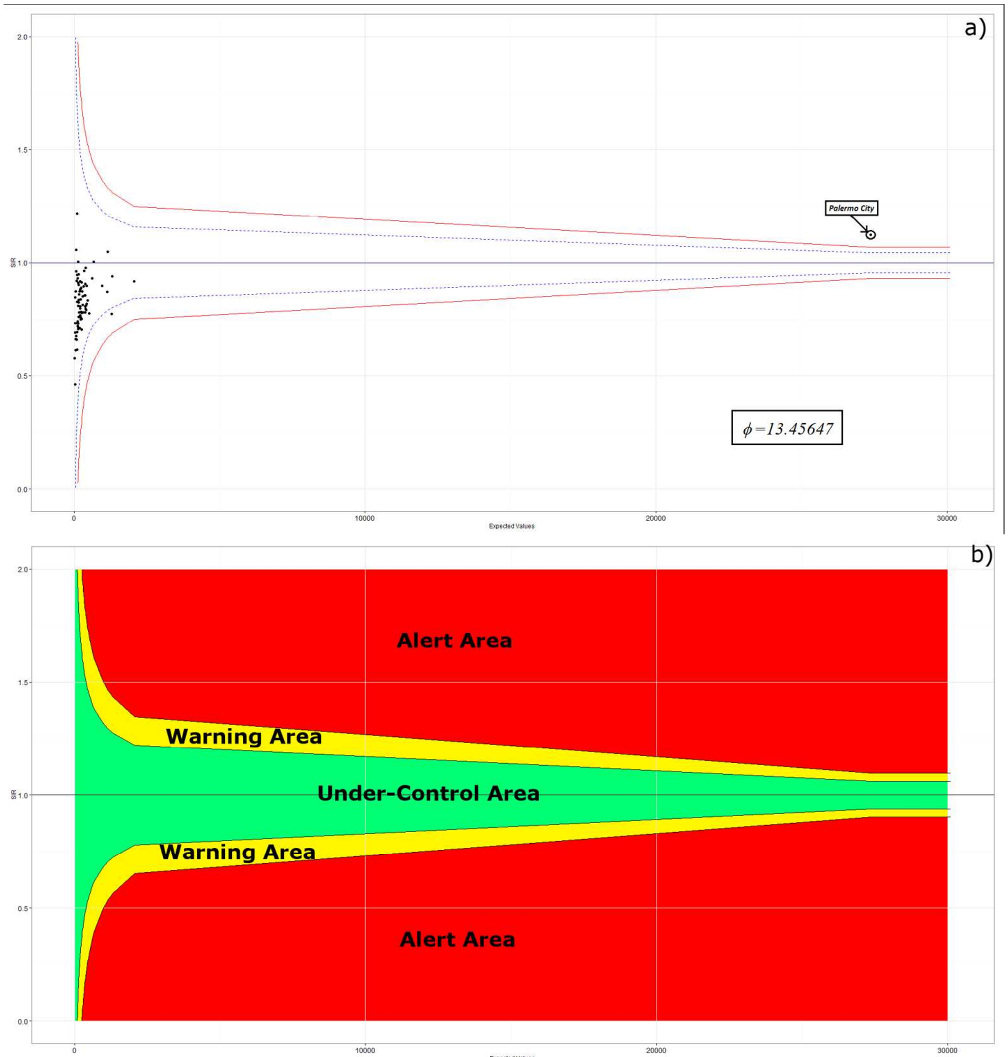
95% CIs ("blue" control lines) and 99.8% CIs ("red" control lines)

ϕ = overdispersion, calculated with multiplicative approach

Figure 2. Choropleth map of the SIRs in the 82 Palermo Province municipalities^ (Study period 2003-2011).

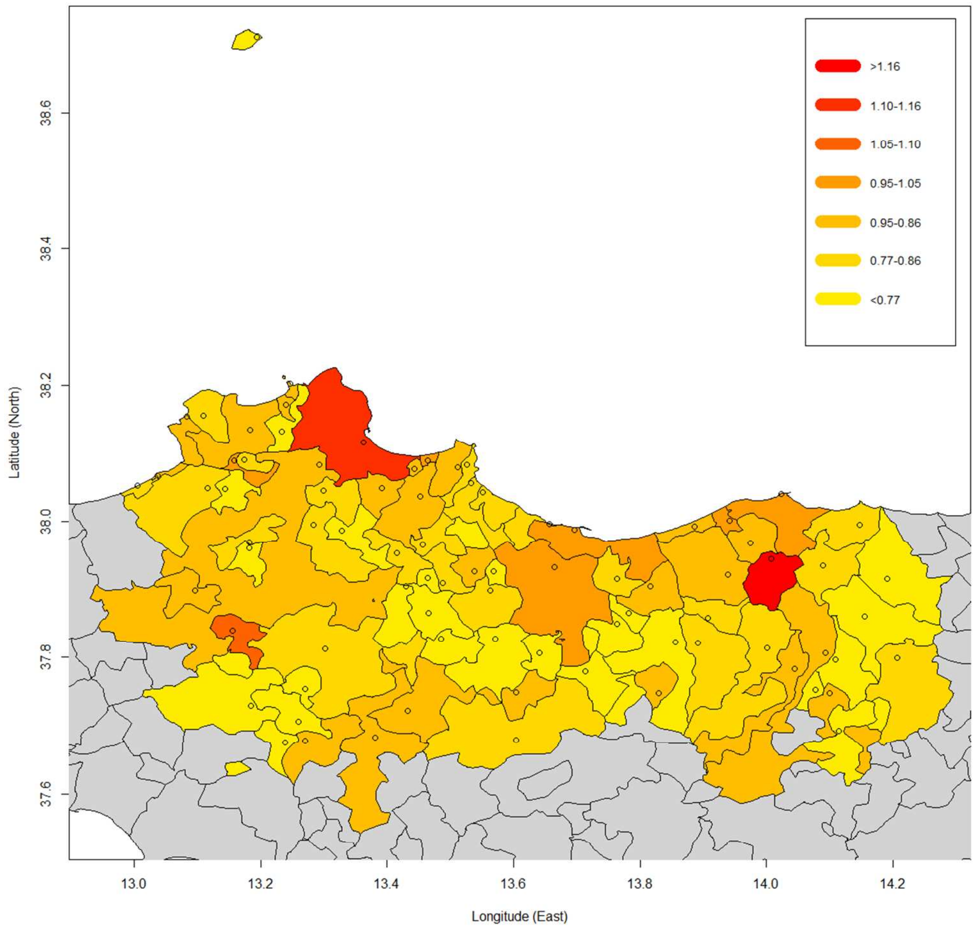
^Circles represent the locations of city halls

Figure 3. Geographical Analysis Machine (GAM) map of the Palermo Province (Study period 2003-2011).



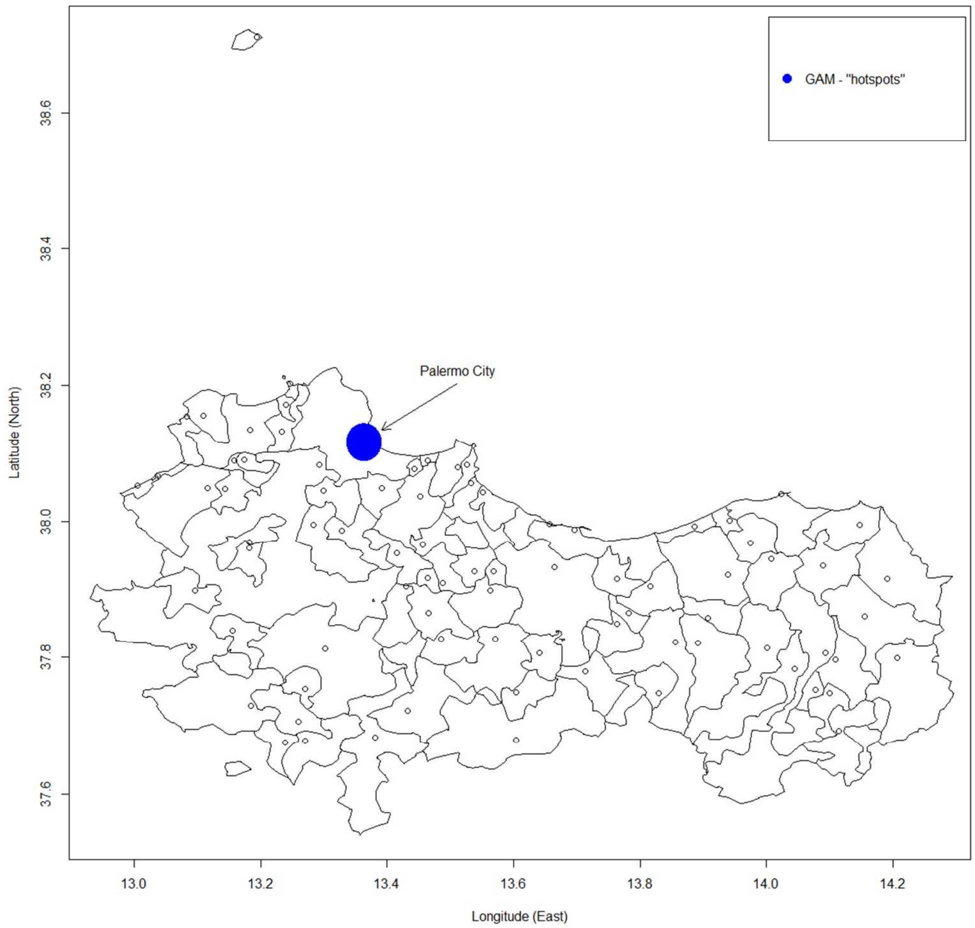
95% CIs ("blue" control lines) and 99.8% CIs ("red" control lines)
φ = overdispersion, calculated with multiplicative approach

169x179mm (300 x 300 DPI)



^Circles represent the locations of city halls

101x101mm (300 x 300 DPI)



101x101mm (300 x 300 DPI)

SUPPLEMENTARY MATERIAL

A) Over-dispersion test (multiplicative approach).[27]

The multiplicative approach introduces an over-dispersion factor ϕ that inflates the null variance. Suppose we have a sample of "I" units (i.e. I = 82 municipalities for the Palermo Province) that we shall assume (for the present) all to be in-control. ϕ may be estimated as follows:

$$\hat{\phi} = \frac{1}{I}\sum_i \frac{(y_i - \theta_0)^2 \rho_i}{g(\theta_0)} = \frac{1}{I}\sum_i z_i^2$$

where z_i is the standardized Pearson residual defined as

$$z_i = \frac{y_i - \theta_0}{\sqrt{\mathbb{V}(Y|\theta_0)}}$$

where y_i is the indicator of interest (SIR of municipality) and θ_0 is the value of target (SIR_{target} = 1). The current control limits can then be inflated by a factor $\sqrt{\hat{\phi}}$ around θ_0. For example, based on the approximate normal control limits, over-dispersed control limits can then be plotted as

$$y_p(\theta_0, \rho) = \theta_0 + z_p\sqrt{\hat{\phi} g(\theta_0)/\rho}$$

where g(θ_0) can be choice equal to θ_0, when y_i is a standardised ratio. There is an element of circularity, in that if out-of-control units are included in this estimation process, they will tend to increase the estimate of θ_0, widen the funnel limits and hence make it more difficult to detect the very cases in which we are interested. Therefore. when estimating ϕ, we may want to 'robustify' the analysis by minimizing the influence of outlying cases that the system is designed to detect. There follows the 'Winsorised' estimation algorithm in which the most extreme cases are shrunk to pre-specified percentiles of the distribution:

1. Rank cases according to their naive Z-scores.

2. Identify Z_q and Z_{1-q}, the 100q per cent most extreme top and bottom naive Z-scores, where q might, for example, be 0.1.

3. Set the lowest 100q per cent of Z-scores to Z_q, and the highest 100q per cent of Z-scores to Z_{1-q}. Denote the resulting set of Z-scores, both those left unchanged and those that have been 'pulled-in', by Z^w.

4. Calculate the estimate ϕ using Z^w, so that

$$\hat{\phi}W = \frac{1}{I}\sum_i Z_i^W(q)^2$$

If there is no true over-dispersion, then $I\phi$ has approximately a χ^2 distribution, with I degrees of freedom, which means that E(ϕ) = 1, V(ϕ) = 2/I. Rather than applying the over-dispersion adjustment to all data by default, it may therefore be better to:

1. not assume under-dispersion: i.e. if $\phi < 1$, assume $\phi = 1$;
2. demand a 'statistically significant' ϕ before including an adjustment for over-dispersion: i.e. assume $\phi = 1$ unless the estimated $\phi > 1 + 2\sqrt{2/I}$.

In our example ϕ (= 13.45647) is more than 10 times the value of $1 + 2\sqrt{2/I}$=1.312348).

B) R-script developed to detect the greatest cut-off for the winsorization procedure.

We have written the following r-script in order to automatically detect the "maximum" q value of winsorization:

```
qWINZORING <- function(Znu){
 c<-0
 n<-length(Znu)
 qseq <- seq(from = 0, to = 1, by = 0.001)
 for(q in qseq) {
  c <- c+1
  Zq <- quantile(Znu, probs = q)
  Z1_q <- quantile(Znu, probs = 1-q)
  Ziu <- Znu
  Ziu[Znu<Zq] <- Zq
  Ziu[Znu>Z1_q] <- Z1_q
  phiW <- mean(Ziu^2)
  if(phiW <= 1+2*sqrt(2/n)) break
 }
 qqnorm(Znu); qqline(Znu)
 return(qseq[c-1])
}
```

Where Znu is the vector of Z-scores. This function return the "maximum" q level for the winsorization.

This R-function applied to our data returns the value 1, i.e. any q < 1 is suitable.

C) Over-dispersion test (additive approach).[27]

The random-effects approach assumes that Y_i has expectation $E(Y_i) = \phi_i$ and variance $V(Y_i) = \sigma_i^2$, and that for 'on-target' trusts ϕ_i is distributed with mean ϕ_0 and standard deviation τ. Hence the null hypothesis is a distribution rather than a point. τ can be estimated using a standard 'method of moments' estimator

$$\hat{\tau}^2 = \frac{I\hat{\phi} - (I-1)}{\sum_i w_i - \sum_k w_i^2 / \sum_i w_i}$$

where $w_i = 1/\sigma_i^2$, and ϕ is the test for heterogeneity: if $\phi < (I-1)/I$, then τ^2 is set to 0 and complete homogeneity is assumed. The funnel plot boundaries are then given by

$$\theta_0 \pm z_p \sqrt{\mathbb{V}(Y|\theta_0, \rho) + \tau^2}$$

In our data, τ^2 results equal to 0.0007151463.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011502.R2
Article Type:	Research
Date Submitted by the Author:	28-Oct-2016
Complete List of Authors:	MAZZUCCO, WALTER; University of Palermo, Scienze Promozione Salute e Materno Infantile Cusimano, Rosanna ; Azienda Sanitaria Provinciale 6 di Palermo, Dipartimento Prevenzione Zarcone, Maurizio; Azienda Ospedaliera Universitaria Policlinico “Paolo Giaccone”, Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Mazzola, Sergio; Azienda Ospedaliera Universitaria Policlinico “Paolo Giaccone”, Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Vitale, Francesco; University of Palermo, Department of Sciences for Health Promotion “G. D’Alessandro” – Hygiene section
Primary Subject Heading:	Public health
Secondary Subject Heading:	Public health, Epidemiology, Oncology, Communication
Keywords:	Funnel plot, cancer epidemiology, cancer registry;, Standardized Incidence Ratio, cancer data dissemination

SCHOLARONE™
Manuscripts

TITLE

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

Authors

Mazzucco Walter[1,4], *Cusimano Rosanna*[1,3], *Zarcone Maurizio*[1,2], *Mazzola Sergio*[1,2], *Vitale Francesco*[1,2,4].

1. Registro Tumori di Palermo e Provincia, University of Palermo, Palermo, Italy.

2. Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Azienda Ospedaliera Universitaria Policlinico "Paolo Giaccone", Palermo, Italy

3. Azienda Sanitaria Provinciale 6 di Palermo - Dipartimento di Prevenzione, Palermo, Italy

4. Dipartimento Scienze Promozione della Salute e Materno Infantile "G. D'Alessandro", Sezione di Igiene, University of Palermo, Palermo, Italy

Mazzucco Walter (Corresponding author), Dipartimento Scienze Promozione della Salute e Materno Infantile "G. D'Alessandro", Sezione di Igiene, Università degli Studi di Palermo. Via del Vespro, 133. 90127 Palermo. E-mail: walter.mazzucco@unipa.it.

Cusimano Rosanna, Dipartimento Prevenzione - Azienda Sanitaria Provinciale 6 di Palermo. Via Siracusa, 45. 90100 Palermo. E-mail: rosanna.cusimano.55@gmail.com.

Zarcone Maurizio, Registro Tumori di Palermo e Provincia - Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Azienda Ospedaliera Universitaria Policlinico "Paolo Giaccone". Via del Vespro, 133. 90127 Palermo. E-mail: zarcone7@gmail.com.

Mazzola Sergio, Registro Tumori di Palermo e Provincia - Unità Operativa Complessa di Epidemiologia Clinica con Registro Tumori, Azienda Ospedaliera Universitaria Policlinico "Paolo Giaccone". Via del Vespro, 133. 90127 Palermo. E-mail: mazzolasergio3@gmail.com.

Vitale Francesco, Dipartimento Scienze Promozione della Salute e Materno Infantile "G. D'Alessandro", Sezione di Igiene, Università degli Studi di Palermo. Via del Vespro, 133. 90127 Palermo. E-mail: francesco.vitale@unipa.it.

Keywords

Funnel plot; cancer registry; cancer epidemiology; cancer data dissemination; Standardized Incidence Ratio.

Word count: 3075

Number of figures: 3

Number of tables: 2

Number of references: 44

Details of preferred reviewers

Maurizio Macaluso, Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA - Email: Maurizio.Macaluso@cchmc.org

Chiara De Waure, Public Health Institute, Medicine Faculty, Catholic University of Sacred Hearth (UCSC), Rome, Italy - Email: chiara.dewaure@rm.unicatt.it

Maria Rosaria Gualano, Department of Public Health and Pediatrics, University of Turin, Turin, Italy - Email: mariarosaria.gualano@unito.it

ABSTRACT

Background: Population based cancer registries provide epidemiologic cancer information, but the indicators are often too complex to be interpreted by local authorities and communities, due to numeracy and literacy limitations. 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 Standardized Incidence Ratio 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 the 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 two 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.

Article summary

Population-based cancer registries perform cancer surveillance through a continuous recording and monitoring of cancer incidence, survival and prevalence in the resident population of a specific administrative area (in Italy, this usually corresponds to the district level). Cancer registries are frequently queried by health authorities or citizens to answer specific questions, often related to a perception of a potentially high local exposure risk; these registries aim to facilitate correct interpretation of the acquired data and, at the same time, to preserve a rigorous methodological approach. Unfortunately, the different visual and graphical formats (choropleth maps, tables, Geographical Analysis Machine) currently used to represent and publish data on cancer epidemiology can often result in a poorly informative or misleading message to non-expert stakeholders.

Strengths and limitations of the study

1) To the best of our knowledge, this study explores for the first time the application of the funnel plot methodology to represent standardized cancer incidence ratio at the municipal level through a comparison with the commonly used visual format, as choropleth map.
2) The results of this study support the use of funnel plot as a complement to choropleth map for disseminating epidemiological data of cancer registries to local communities and authorities.
3) The proposed communication approach needs to be further validated in the field. To this end, the Palermo Province Cancer Registry has generated 82 municipal risk maps, one for each municipality of the province, and for a period of one year, qualified personnel from the registry will be involved in on-site meetings to share cancer incidence data with stakeholders (citizens, local authorities, general practitioners, specialized physicians, pharmacists, etc) using funnel plots. The Delphi consensus process will be explored as well by involving public health operators.

BACKGROUND

Cancer is the second cause of death in the developed countries.[1] In the last 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.[2] Apart from the presumed effects of lifestyle changes and environmental factors on cancer trends,[3-6] the global increase in cancer prevalence could be largely attributable to a combination of improved cancer survival[7] and aging 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.[11] In particular, citizens often question if they live in an area at high risk for environmental exposure.[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*".[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.[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 measurements 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.[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 (CM), representing cancer incidence rates (standardized rates, standardized ratios, etc.) computed for specific areas.[16, 17] While data are available on how the context[18] and the content of such communications influence individual risk perception,[19] little is known about the effects of risk communications at a group level, particularly in small communities.[20]

The Italian Association of Cancer Registries (AIRTum), a national network of 41 local population-based cancer registries, including the Palermo Province Cancer Registry (PPCR), has greatly emphasized improving communication tools.[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).[22] Cancer incidence in the PP municipalities was measured by using Standardized Incidence Ratio (SIR), defined as the ratio between observed cases (O_i) and expected cases (E_i).[23] The O_i were assumed to follow a homogeneous Poisson distribution with parameter $\lambda = \theta_0 \cdot E_i$. The E_i were estimated by indirect method,[24] considering the entire population-time under study (the PP) as the reference population, with $\Sigma O_i = \Sigma E_i$.[25] The resident population was reported using the inter-census estimates, provided by the Italian National Statistical Institute (ISTAT), also considering the annual municipal data on migration.[22] For each SIR, the 95% confidence interval (CI) was calculated by using the normal approximation method.[26]

Graphic FP representation[26] was used to highlight any municipality with a higher cancer incidence compared to 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 ($\theta_0 = 1$), representing the reference value for the indicator of interest ($O_i = E_i$); the E_i precision parameter, measuring the accuracy of the indicator of interest (Poisson variance parameter, using the hypothesis $\theta_0 = 1$), represented on the x-axis; the 95% and 99.8% CIs, calculated with the normal approximation method, defining the control limits.[26] The two sets of control limit lines define three different areas within the graph (Figure 1b): the “under-control” area (in green), the “warning” area (in yellow) and the “alert” area (in red).[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[28] both additive and multiplicative approaches were adopted. Overdispersion coefficients (τ for the additive approach and ϕ for the multiplicative approach) were calculated. Overdispersion was addressed by considering the winsorized estimates too.[27] Moreover, Z-score[29] and the winsorization method (by testing for different levels of Z-score quantiles[28]) were applied for the direct selection of extreme values. Furthermore, to define the level of winsorization, an R-script routine was developed to set a cut-off for the quantile between the acceptance and rejection of the overdispersion test (Supplementary material).

The map representing the PP municipalities was generated by using the ISTAT Shapefile vector format,[30] released in the ED50 (European Datum – 1950) reference system. UTM Zone 32N, and converted to plane coordinates (decimal degrees), providing geo-referenced 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 geo-statistical units, considering that some centroid could fall outside the municipal bounds, the coordinates of the city hall were used instead.[31]

The PP cancer incidence variation was also shown in a CM,[32] representing the SIRs of each municipality. To distinguish potential high- and low-risk areas, a central interval of 0.95-1.05 for the color scale was fixed, irrespective of statistical significance. Values above 1.05 and below 0.95 were divided in tertiles.[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.[34] The analysis for hot spot research was performed using circles with a 3 kilometer radius for each point of a grid, covering the study region by steps of 600 meters (radius/5).

The RStudio IDE[35] for the R software, version 3.1.0 (2014-04-10) - "Spring Dance",[36] 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, 37, 38]

RESULTS

Figure 1a represents the FP of 82 municipality-specific SIRs, corrected for overdispersion (ϕ=13.46) and adjusted using the multiplicative approach.[28] All of the SIRs lay within the control limits, except for the Palermo city one (SIR= 1.12), which resulted above the upper control limit line (UCL) 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 winsorization (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 males and females) 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= 1.11-1.14) while Isnello, the municipality showing the highest SIR, failed to meet the conventional criteria for statistical significance (SIR=1.22; 95% CIs 0.99-1.45).

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

ISTAT code	Municipality	Expected	SIR	95% CI	ISTAT code	Municipality	Expected	SIR	95% CI
082042	Isnello	104.1	1.22	0.99-1.45	082031	Cinisi	441.0	0.82	0.74-0.90
082053	**Palermo City**	**27371.4**	**1.12**	**1.11-1.14**	082007	Balestrate	292.4	0.81	0.72-0.91
082061	Roccamena	81.4	1.06	0.83-1.29	082067	Santa Flavia	414.6	0.81	0.73-0.89
082070	Termini Imerese	1166.4	1.05	0.99-1.11	082059	Pollina	148.0	0.81	0.68-0.94
082027	Cefalù	685.3	1.01	0.93-1.08	082030	Ciminna	209.5	0.81	0.70-0.92
082044	Lascari	152.2	1.01	0.85-1.17	082064	San Giuseppe Jato	379.0	0.81	0.73-0.89
082014	Caccamo	396.1	0.98	0.88-1.07	082058	Polizzi Generosa	212.6	0.80	0.69-0.91
082035	Ficarazzi	357.5	0.97	0.87-1.07	082036	Gangi	406.6	0.80	0.72-0.87
082038	Giardinello	84.1	0.96	0.76-1.17	082023	Casteldaccia	416.4	0.79	0.72-0.89
082056	Petralia Sottana	168.6	0.95	0.81-1.09	082004	Altavilla Milicia	242.8	0.78	0.68-0.88
082012	Bompietro	104.6	0.95	0.77-1.13	082005	Altofonte	379.8	0.78	0.70-0.86
082049	Monreale	1319.0	0.94	0.89-0.99	082046	Marineo	310.8	0.78	0.70-0.87
082052	Palazzo Adriano	123.4	0.93	0.77-1.10	082025	Castronovo di Sicilia	175.4	0.78	0.67-0.90
082079	Villabate	631.9	0.93	0.85-1.00	082050	Montelepre	258.1	0.78	0.68-0.87

082008	Baucina	102.0	0.92	0.74-1.10	082034	Corleone	520.3	0.78	0.71-0.84
082006	Bagheria	2068.7	0.92	0.88-0.96	082054	Partinico	1291.1	0.78	0.73-0.82
082076	Valledolmo	212.5	0.92	0.79-1.04	082051	Montemaggiore Belsito	208.5	0.77	0.66-0.87
082017	Campofelice di Roccella	272.6	0.91	0.80-1.02	082078	Vicari	156.2	0.76	0.64-0.88
082074	Trappeto	151.5	0.91	0.77-1.06	082001	Alia	225.4	0.76	0.66-0.86
082020	Capaci	390.4	0.92	0.82-0.99	082010	Bisacquino	272.7	0.75	0.66-0.84
082019	Camporeale	157.5	0.90	0.76-1.04	082063	San Cipirello	222.7	0.75	0.65-0.85
082071	Terrasini	446.0	0.90	0.86-0.98	082002	Alimena	134.9	0.74	0.62-0.87
082048	Misilmeri	969.4	0.90	0.84-0.95	082065	San Mauro Castelverde	120.1	0.74	0.61-0.87
082045	Lercara Friddi	335.8	0.90	0.80-0.99	082082	Blufi	76.5	0.73	0.57-0.90
082028	Cerda	243.1	0.89	0.78-1.01	082003	Aliminusa	73.8	0.73	0.57-0.90
082043	Isola delle Femmine	235.3	0.88	0.77-0.99	082026	Cefalà Diana	50.6	0.73	0.53-0.93
082032	Collesano	220.9	0.88	0.77-0.99	082072	Torretta	143.6	0.73	0.61-0.85
082024	Castellana Sicula	200.3	0.87	0.75-0.99	082039	Giuliana	121.2	0.72	0.59-0.85
082041	Gratteri	65.2	0.87	0.66-1.09	082047	Mezzojuso	146.5	0.71	0.60-0.82
082060	Prizzi	283.8	0.87	0.77-0.98	082055	Petralia Soprana	201.3	0.71	0.61-0.81
082021	Carini	1146.5	0.87	0.82-0.92	082062	Roccapalumba	139.0	0.71	0.59-0.82
082029	Chiusa Sclafani	181.4	0.86	0.73-0.99	082013	Borgetto	261.3	0.70	0.62-0.79
082009	Belmonte Mezzagno	372.3	0.86	0.77-0.94	082037	Geraci Siculo	112.7	0.69	0.56-0.82
082073	Trabia	382.2	0.86	0.77-0.94	082066	Santa Cristina Gela	40.6	0.69	0.48-0.90
082057	Piana degli Albanesi	305.0	0.85	0.76-0.95	082018	Campofiorito	74.2	0.67	0.52-0.83
082081	Scillato	36.6	0.85	0.57-1.12	082075	Ustica	66.6	0.66	0.50-0.82
082015	Caltavuturo	229.0	0.84	0.73-0.95	082033	Contessa Entellina	101.6	0.66	0.53-0.79
082080	Villafrati	166.2	0.84	0.71-0.96	082077	Ventimiglia di Sicilia	116.9	0.62	0.50-0.73
082022	Castelbuono	469.4	0.83	0.76-0.91	082040	Godrano	52.2	0.61	0.45-0.78
082068	Sciara	117.2	0.83	0.68-0.98	082069	Sclafani Bagni	29.4	0.58	0.37-0.79
082011	Bolognetta	162.6	0.82	0.70-0.95	082016	Campofelice di Fitalia	34.6	0.46	0.31-0.62

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

Table 2 summarizes a comparison of the weaknesses and strengths, as per the available literature,[29, 33, 37, 38] between the different visual formats explored within the context of disseminating epidemiological data to stakeholders.

Table 2. Comparison of the weaknesses (-) and strengths (+) of the Funnel Plot and Choropleth Map within the context of disseminating epidemiological data to stakeholders.

Properties explored	**Weaknesses and Strengths of Visual format**	
	Funnel Plot	**Choropleth Map**
Definition of the spatial location of the risk	-	+
Identification of hot spots	+	+
Locating clusters	-	+
Displaying the scope of the phenomenon under investigation	+	-
Showing the precision of estimates	+	-
Communicating the significance of estimates	+	-

As show 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.

DISCUSSION

FP are commonly used in process control and, in particular, in the health care field to compare institutional performance data[29]; however, this format is used for survival[39] and standardized mortality ratio[29] in public health surveillance.[40] We explored the use of FPs as a supplementary tool to local provide authorities and communities with synthetic access to valid and 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,[23] 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,[19] readers do not usually take notice of the confidence interval, which is a critically important measure of the precision of SIR estimates.[41] By displaying sample statistics together with the corresponding sample size, in relation to the control limits, FPs allow visualizing 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).[40] 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 confidence intervals is poor even among physicians, as heuristic reasoning often prevails on sample size.[42] 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 colors) under which each location falls (Figure 1b).

Reading a CM may be misleading for stakeholders[43] since the fear of being overexposed to environmental and other risk factors may lead to misinterpretation of the differences in color 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 cloropleth map, supported by tabulation of the numeric results, allows to identify locations where cancer incidence may deserves 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. Lastly, 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 Palermo Province 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;[29] on the contrary, the more commonly used CMs monodimensionally represent the parameters of interest by using a different color gradation based on the frequency distribution of the values.[33, 37, 38] 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.[29] Lastly, 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 (for example the GAM approach).[34] 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.[44] In fact, it can be presumed that the efficacy of a presentation format depends not only on the type of format, but also on the context in which the format is used (scientific versus general public).[18]

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 not only concentrate on the presentation format, but also 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.).

ACKNOWLEDGEMENTS

Contributorship statement

All individuals listed as authors have contributed substantially to designing, performing or reporting of the study and every specific contribution is indicated as follows.

Conception and design of the study: WM, RC, MZ, SM

Statistical analysis: MZ, SM

Interpretation of data: WM, RC, MZ, SM

Manuscript writing and drafting: WM, RC

Revision of the manuscript: FV, WM, RC

Approval of the final version of the manuscript: WM, RC, MZ, SM, FV

The document has been reviewed and corrected by a native English speaker with extensive scientific editorial experience to ensure a high level of spelling, grammar and punctuation.

Data sharing statement

Supplementary data (results of over-dispersion tests, R-script to detect the greatest cut-off for the winsorization procedure) have been provided as a supplementary file. Other statistical results are available by emailing walter.mazzucco@unipa.it.

COMPETING INTERESTS

The authors declare they have no conflict of interest.

FUNDING

None.

REFERENCES

1 Summary tables: deaths by cause, age and sex 2000-2012. In: Global health estimates 2014. WHO. http://www.who.int/topics/global_burden_of_disease/en/ (accessed 18 Jan 2016).

2 Barchitta M, Fragapane S, Quattrocchi A, et al. Environmental health risk communication in the case "Terra dei Fuochi": content analysis of online newspaper articles. *Ann Ig* 2015;Jan-Feb;27(1):30-8.

3 Pirastu R, Comba P, Conti S, et al. Studio epidemiologico dei territori e degli insediamenti esposti a rischio da inquinamento. Mortalità incidenza oncologica e ricoveri ospedalieri. *Epidemiol Prev* 2004;38(2)(Suppl 1).

4 K. Straif, A. Cohen, J. Samet. Air pollution and cancer. IARC Scientific publication no.161. https://www.iarc.fr/en/publications/books/sp161/AirPollutionandCancer161.pdf (accessed 18 Jan 2016).

5 Monographs on the Evaluation of Carcinogenic Risk to Humans. International Agency for Research on Cancer (IARC). http://monographs.iarc.fr/ENG/Monographs/PDFs/index.php (accessed 18 Jan 2016).

6 Boffetta P and Nyberg F. Contribution of environmental factors to cancer risk. *Br Med Bull* 2003;68:71-94.

7 Allemani C, Weir HK, Carreira H, et al. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* 2015 Mar 14;385(9972):977-1010.

8 WHO Global Status Report on Non-Communicable Diseases 2010. Geneva 2011.

9 Melissa G. Health literacy and Numeracy. Workshop Summary. The National Academy Press. Washington 2014. http://www.nap.edu/catalog/18660/health-literacy-and-numeracy-workshop-summary (accessed 18 Jan 2016).

10 Reyna VF, Nelson WL, Han PK and Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull* 2009;135(6):943-73.

11 Edwards A, Elwyn G and Mulley A. Explaining risks: turning numerical data into meaningful pictures. *BMJ* 2002;324;827-830.

12 CDC Updated Guidelines for evaluating Public Health Surveillance Systems. MMWR 2001/50(RR13);1-35. http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm (accessed 18 Jan 2016).13 Bray F, Soerjomataram I. The Changing Global Burden of Cancer: Transitions in Human Development and Implications forCancer Prevention and Control. In: Gelband H, Jha P, Sankaranarayanan R, Horton S, editors. Cancer: Disease Control Priorities, Third Edition (Volume 3). Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2015 Nov 01. Chapter 2.

14 Rapporto sui tumori a Palermo e provincia. Anni 2003 – 2007. Registro Tumori Palermo e Provincia. Anno 2011. http://www.unipa.it/dipartimenti/dipartimentodiscienzeperlapromozionedellasaluteematernoinfantileg.dalessandro/.content/documenti/Rapporto-sui-tumori-a-Palermo-e-Provincia-anni-2003-2007-1.pdf (accessed 18 Jan 2016).

15 Cancer in Friuli Venezia Giulia Incidence, Survival and Prevalence: 1995-2005. Agenzia Regionale della Sanità Friuli Venezia Giulia. http://www.cro.it/opencms/export/sites/cro/it/azienda_informa/_docs/registro-tumori-1995-2005.pdf (accessed 18 Jan 2016).

16 Jemal A., Vineis P, Bray F, Torre L, Forman D. The Cancer Atlas, second edition. Non series publications IARC Lyon 2014.

17 Atlante Sanitario della Sicilia. Suppl. monografico Notiziario Dipartimento per le Attività Sanitarie ed Osservatorio Epidemiologico. Assessorato della Salute, Regione Sicilia. Anno 2013:

http://pti.regione.sicilia.it/portal/page/portal/PIR_PORTALE/PIR_LaStrutturaRegionale/PIR_AssessoratoSalute/PIR_AreeTematiche/PIR_Atlanteoncologico (accessed 18 Jan 2016).

18 Visschers VH, Meertens RM, Passchier WW, de Vries NN. Probability information in risk communication: a review of the research literature. *Risk Anal* 2009;29(2):267-87.

19 Lipkus IM, Hollands JG. The visual communication of risk. *Natl Cancer Inst Monogr* 1999;25:149-63.

20 Mclnerny GJ, Chen M, Freeman R, Gavaghan D, Meyer M, Rowland F et al. Information visualization for science and policy: engaging users and avoiding bias. *Trends Ecol Evol* 2014;Vol 29,Iss 7:148-157. doi:10.1016/j.tree.2014.01.003 [published Online First: 21 February 2014] (accessed 18 Jan 2016).

21 Mappa dei Registri Tumori di popolazione. Associazione Italiana Registri Tumori (AIRTUM) 2015. http://www.registri-tumori.it/cms/it/copertura (accessed 18 Jan 2016).

22 Ricostruzione intercensuaria della popolazione 2002-2011. Istituto Nazionale di Statistica. 2015; www.demo.istat.it (accessed 18 Jan 2016).

23 Greenland S, Rothman KJ and Lash TL. Measures of Effect and Measures of Association. In: Rothman KJ, Greenland S and Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins 2008:67-69.

24 Mantel N and Stark CR. Computation of indirect-adjusted rates in the presence of confounding. Biometrics, Vol. 24, No. 4 (Dec., 1968), pp. 997-1005

25 Morris SE, Wakefield JC. Assessment of disease risk in relation to a pre-specified source. In: Elliot P, Wakefield JC, Best NG, Briggs DJ, editors. Spatial epidemiology: methods and applications. New York: Oxford University Press 2000:53–184.

26 Spiegelhalter DJ. Funnel plots for institutional comparisons. *Qual Saf Health Care* 2002;11;390-391.

27 Spiegelhalter DJ. Handling Over-Dispersion of Performance Indicators. *Qual Saf Health Care* 2005;14;347-351.

28 Silcocks P. Hazard ratio funnel plots for Survival. *J Epidemiol Community Health* 2009;63:856-861.

29 Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med* 2005;24;1185-1202.

30 Confini amministrativi, dei sistemi locali del lavoro e delle NUTS2 in versione generalizzata. Istituto Nazionale di Statistica 2010. http://www.istat.it/it/files/2011/04/com2011_g.zip (accessed 18 Jan 2016).

31 Mazumdar, Soumya. "Shape and scale in detecting disease clusters." PhD (Doctor of Philosophy) thesis, University of Iowa, 2008. http://ir.uiowa.edu/etd/208 (accessed 18 Jan 2016).

32 Richards TB, Berkowitz Z, Thomas CC, Foster SL, Gardner A, King JB et al. Choroplet map design for cancer incidence, part 2. *Prev Chronic Dis* 2010;7(1).

33 Cromley G and Ellen K Cromley. Choropleth map legend design for visualizing community health disparities International. Journal of Health Geographics 2009, 8:52.

34 Besag J and Newell J. The Detection of Clusters in Rare Diseases. Wiley and Royal Statistical Society 1991;154,Part 1,143-155.

35 RStudio Team. RStudio: Integrated Development for R. 2015. http://www.rstudio.com/ (accessed 18 Jan 2016).

36 R Core Team. R: A language and environment for statistical computing 2015. http://www.R-project.org/ (accessed 18 Jan 2016).

37 Sui DZ, Holt JB: Visualizing and analysing public-health data using value-by-area cartograms: toward a new synthetic framework. Cartographica 2008, 43:3-20.

38 Kumar N: Frequency histogram legend in the choropleth map: a substitute to traditional legends. Cartogr Geogr Inform Sci 2004, 31:217-236.

39 Quaresma M, Coleman MP, Rachet B. Funnel plots for population-based cancer survival: principles, methods and applications. *Stat Med* 2014, 33(6):1070-80. doi: 10.1002/sim.5953 [published Online First: 4 September 2013].

40 Dover DC and Schopflocher DP. Using funnel plot in public health surveillance. *Popul Health Metr* 2011,9:58.

41 Hildon Z, Allwood D and Black N. Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review. *Int J Qual Health Care* 2012;24(1):55-64.

42 Rakow T, Wright RJ, Spieghelhalter D and Bull C. The pros and cons of funnel plots as an aid to risk communication and patient decision making. *Br J Psychol* 2015;106:327-348.

43 Ocana-Riola R. Common errors in disease mapping. *Geospat Health* 2010;4(2)139-54.

44 Chia-Chien Hsu and Brian A. The Delphi Technique: Making Sense of Consensus. Practical Assessment, Research & Evaluation 2007;Vol 12, No 10.

Figure 1. a) Funnel plot of the SIRs in the 82 Palermo Province municipalities (Study period 2003-2011); b) Cancer attention areas: “under-control” area (in green), “warning” area (in yellow) and “alert” area (in red).

Figure legend:

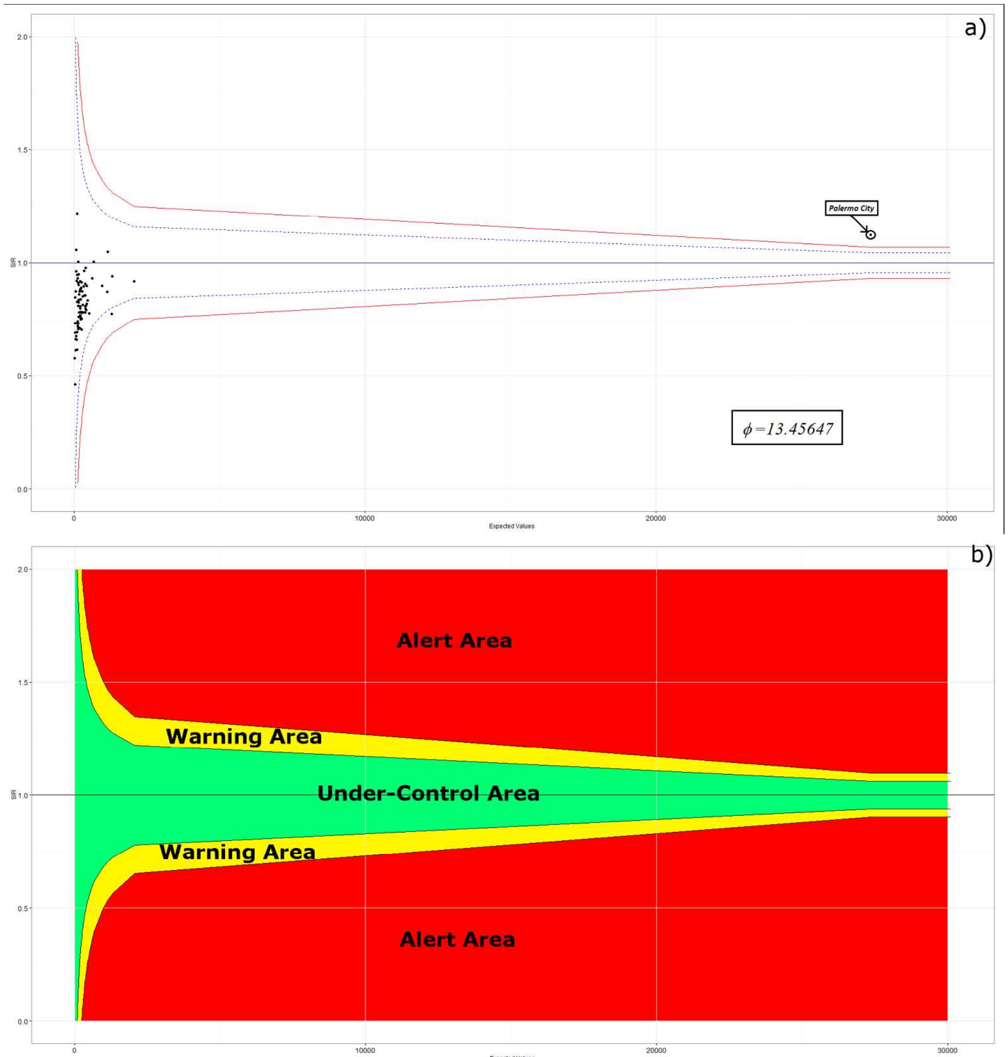
95% CIs (“blue” control lines) and 99.8% CIs (“red” control lines)

ϕ = overdispersion, calculated with multiplicative approach

Figure 2. Choropleth map of the SIRs in the 82 Palermo Province municipalities^ (Study period 2003-2011).

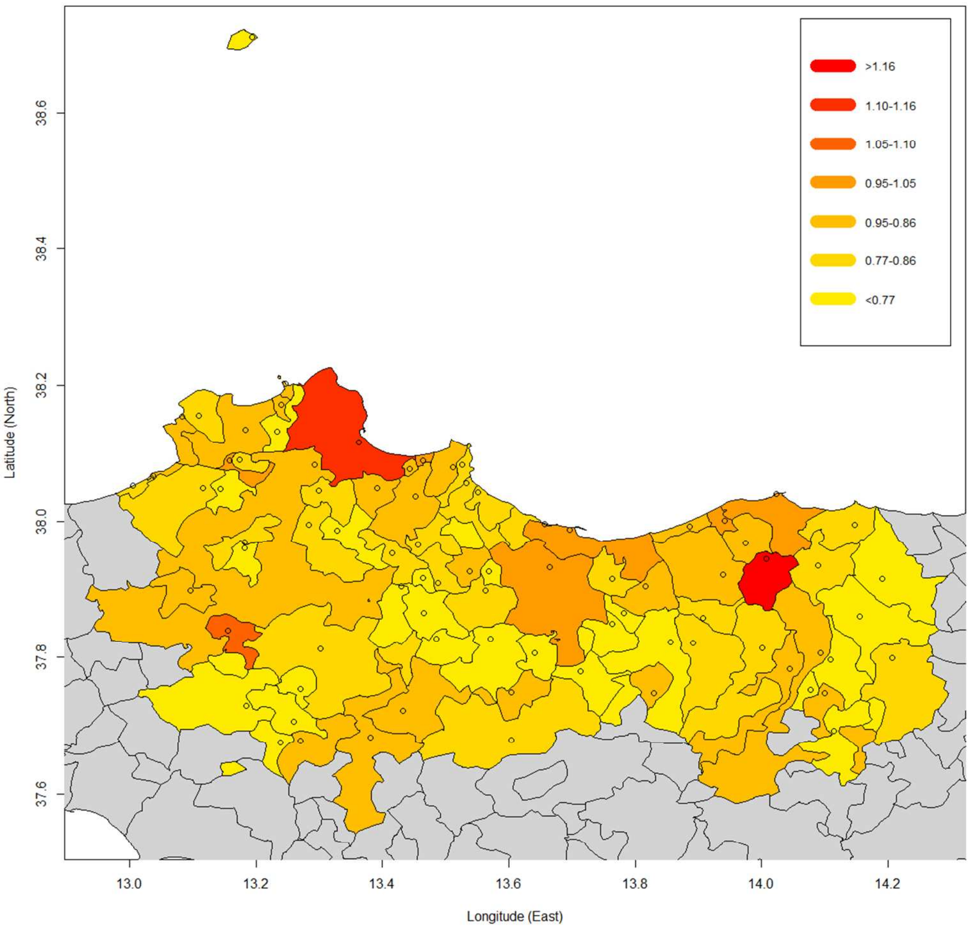
^Circles represent the locations of city halls

Figure 3. Geographical Analysis Machine (GAM) map of the Palermo Province (Study period 2003-2011).



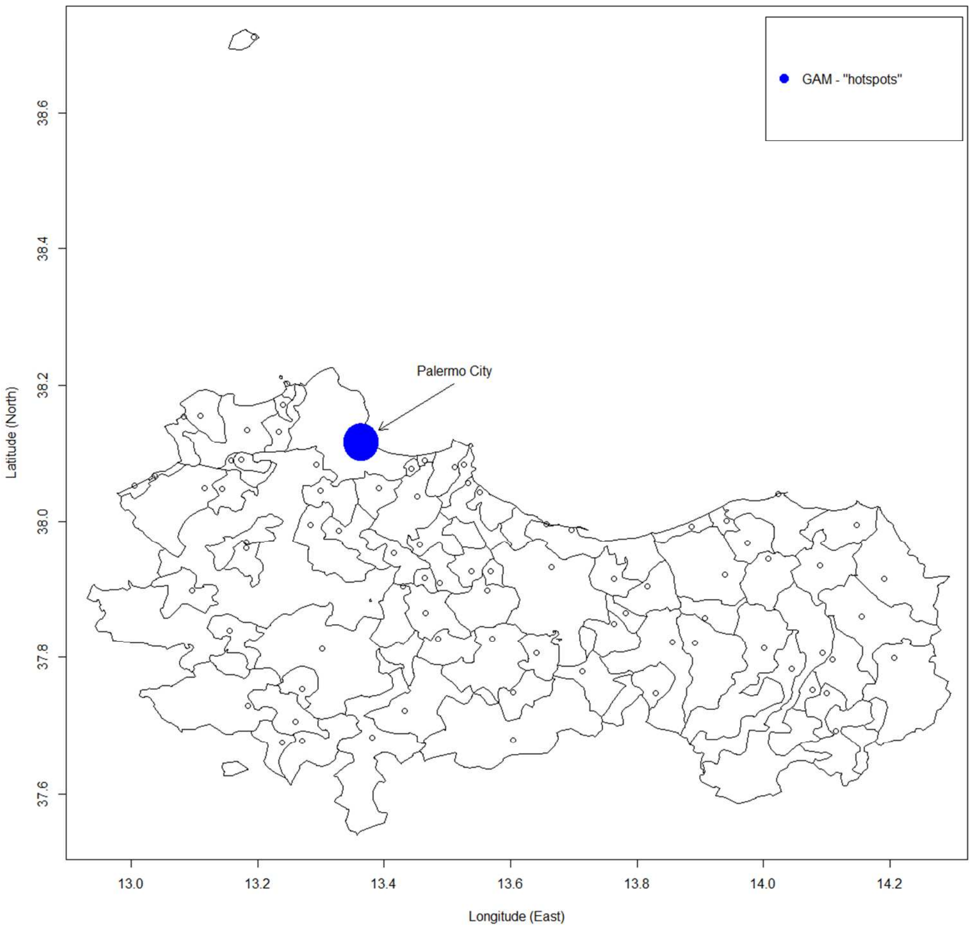
95% CIs ("blue" control lines) and 99.8% CIs ("red" control lines)
φ = overdispersion, calculated with multiplicative approach

169x179mm (300 x 300 DPI)



^Circles represent the locations of city halls

101x101mm (300 x 300 DPI)



101x101mm (300 x 300 DPI)

SUPPLEMENTARY MATERIAL

A) Over-dispersion test (multiplicative approach).[27]

The multiplicative approach introduces an over-dispersion factor ϕ that inflates the null variance. Suppose we have a sample of "I" units (i.e. I = 82 municipalities for the Palermo Province) that we shall assume (for the present) all to be in-control. ϕ may be estimated as follows:

$$\hat{\phi} = \frac{1}{I}\sum_i \frac{(y_i - \theta_0)^2 \rho_i}{g(\theta_0)} = \frac{1}{I}\sum_i z_i^2$$

where z_i is the standardized Pearson residual defined as

$$z_i = \frac{y_i - \theta_0}{\sqrt{\mathbb{V}(Y|\theta_0)}}$$

where y_i is the indicator of interest (SIR of municipality) and θ_0 is the value of target ($SIR_{target} = 1$). The current control limits can then be inflated by a factor $\sqrt{\hat{\phi}}$ around θ_0. For example, based on the approximate normal control limits, over-dispersed control limits can then be plotted as

$$y_p(\theta_0, \rho) = \theta_0 + z_p\sqrt{\hat{\phi} g(\theta_0)/\rho}$$

where g(θ_0) can be choice equal to θ_0, when y_i is a standardised ratio. There is an element of circularity, in that if out-of-control units are included in this estimation process, they will tend to increase the estimate of θ_0, widen the funnel limits and hence make it more difficult to detect the very cases in which we are interested. Therefore. when estimating ϕ, we may want to 'robustify' the analysis by minimizing the influence of outlying cases that the system is designed to detect. There follows the 'Winsorised' estimation algorithm in which the most extreme cases are shrunk to pre-specified percentiles of the distribution:

1. Rank cases according to their naive Z-scores.

2. Identify Z_q and Z_{1-q}, the 100q per cent most extreme top and bottom naive Z-scores, where q might, for example, be 0.1.

3. Set the lowest 100q per cent of Z-scores to Z_q, and the highest 100q per cent of Z-scores to Z_{1-q}. Denote the resulting set of Z-scores, both those left unchanged and those that have been 'pulled-in', by Z^w.

4. Calculate the estimate ϕ using Z^w, so that

$$\hat{\phi}W = \frac{1}{I}\sum_i Z_i^W(q)^2$$

If there is no true over-dispersion, then $I\phi$ has approximately a χ^2 distribution, with I degrees of freedom, which means that E(ϕ) = 1, V(ϕ) = 2/I. Rather than applying the over-dispersion adjustment to all data by default, it may therefore be better to:

1. not assume under-dispersion: i.e. if $\phi < 1$, assume $\phi = 1$;
2. demand a 'statistically significant' ϕ before including an adjustment for over-dispersion: i.e. assume $\phi = 1$ unless the estimated $\phi > 1 + 2\sqrt{2/I}$.

In our example ϕ (= 13.45647) is more than 10 times the value of $1 + 2\sqrt{2/I}$=1.312348).

B) R-script developed to detect the greatest cut-off for the winsorization procedure.

We have written the following r-script in order to automatically detect the "maximum" q value of winsorization:

```
qWINZORING <- function(Znu){
 c<-0
 n<-length(Znu)
 qseq <- seq(from = 0, to = 1, by = 0.001)
 for(q in qseq) {
  c <- c+1
  Zq <- quantile(Znu, probs = q)
  Z1_q <- quantile(Znu, probs = 1-q)
  Ziu <- Znu
  Ziu[Znu<Zq] <- Zq
  Ziu[Znu>Z1_q] <- Z1_q
  phiW <- mean(Ziu^2)
  if(phiW <= 1+2*sqrt(2/n)) break
 }
 qqnorm(Znu); qqline(Znu)
 return(qseq[c-1])
}
```

Where Znu is the vector of Z-scores. This function return the "maximum" q level for the winsorization.

This R-function applied to our data returns the value 1, i.e. any q < 1 is suitable.

C) Over-dispersion test (additive approach).[27]

The random-effects approach assumes that Y_i has expectation $E(Y_i) = \phi_i$ and variance $V(Y_i) = \sigma_i^2$, and that for 'on-target' trusts ϕ_i is distributed with mean ϕ_0 and standard deviation τ. Hence the null hypothesis is a distribution rather than a point. τ can be estimated using a standard 'method of moments' estimator

$$\hat{\tau}^2 = \frac{I\hat{\phi} - (I-1)}{\sum_i w_i - \sum_k w_i^2 / \sum_i w_i}$$

where $w_i = 1/\sigma_i^2$, and ϕ is the test for heterogeneity: if $\phi < (I - 1)/I$, then τ^2 is set to 0 and complete homogeneity is assumed. The funnel plot boundaries are then given by

$$\theta_0 \pm z_p \sqrt{\mathbb{V}(Y|\theta_0, \rho) + \tau^2}$$

In our data, τ^2 results equal to 0.0007151463.