

## RESEARCH COMMUNICATION

## Quality of Case Ascertainment in Cancer Registries: A Proposal for a Virtual Three-source Capture-recapture Technique

Krittika Suwanrungruang<sup>1,2\*</sup>, Hucha Sriplung<sup>1</sup>, Pattarawin Attasara<sup>3</sup>, Somnuk Temiyasathit<sup>4</sup>, Rangsiya Buasom<sup>3</sup>, Narate Waisri<sup>5</sup>, Karnchana Daoprasert<sup>6</sup>, Supot Kamsa-ard<sup>2</sup>, Cheamchit Tasanapitak<sup>7</sup>

### Abstract

**Background:** The ability and behaviour of the capture-recapture method using a virtual three-source model for evaluation of the level of completeness of case ascertainment requires exploration. **Methods:** Cancer cases obtained from 9 population-based cancer registries in Thailand during 2003 to 2007 were applied for capture-recapture using a model based on clinical, pathological and mortality data. These three virtual sources were derived from three actual items common to all cancer registries: the basis of diagnosis, ICD-O morphology code, and last known patient status. Poisson regression models were fit to the data to estimate parameters which were then transformed into demographic values. A linear model was used to determine the predictors and estimated percentage of completeness (EPC) in case ascertainment among the cancer registries. **Results:** The EPC was greater than 97% in 5 and less than 90% in 4 registries. The worst had an EPC of 70%. The percentage death certificate only (%DCO) and the interaction between %DCO and morphological verification (MV) were significantly associated with EPC. Other factors intrinsic to registries also exerted influence on the EPC. **Conclusions:** In addition to other standard indicators to monitor completeness of cancer registries, the present virtual three-source capture-recapture model can be routinely used to estimate the level of completeness of case ascertainment in cancer registries.

**Keywords:** Capture-recapture - estimated percentage completeness (EPC) - virtual three-source model - Thailand

*Asian Pacific J Cancer Prev*, 12, 173-178

### Introduction

A population-based cancer registry routinely collects information on patients diagnosed with cancer from various sources, such as hospitals and pathological laboratories, and provides estimates of the magnitude of cancer problems in a community. However, cancer registries have their own strategic and logistic autonomy to an extent where the standard procedures are appropriated. Cancer statistics might be distorted when a cancer registry violates the standard registration procedures, i.e., data abstraction, coding, and other processes. Monitoring the sources of cancer cases is a useful method for evaluating the completeness of case ascertainment for all registries. Therefore, the capture-recapture method which handles data sources is considered as an indicator for completeness in terms of coverage of case ascertainment (International working group for disease monitoring and forecasting, 1995; Parkin and Bray, 2009).

Capture-recapture is the method widely used in wildlife population censuses (Cormack, 1968). Since cancer registries employ multiple data sources for case-finding, capture-recapture methods can utilize these 'incomplete' lists of patients to assess completeness of case ascertainment. Two assumptions have to be made when using the simple capture-recapture method. Firstly, the sources are independent, and secondly, all individuals within the same source have an equal chance of being included (Parkin and Bray, 2009).

Traditionally, sources of cancer data in capture-recapture modeling are actual sources, such as out-patient departments (OPD), cancer clinics and laboratories. One patient may appear in one or more sources as he or she passes the process of clinical examination, laboratory investigation, treatment, and documentation of death at the final source. Thus, the number of sources able to be included in capture-recapture analyses can vary from one investigation to another due to the sources used by

<sup>1</sup>Epidemiology Unit, Faculty of Medicine, Prince of Songkla University, Songkhla, <sup>2</sup>Cancer Unit, Srinagarind Hospital, Faculty of Medicine, Khon Kaen University, Khon Kaen, <sup>3</sup>National Cancer Institute, Ministry of Public Health, Bangkok, <sup>4</sup>Udon Thani Cancer Registry, Udon Thani Cancer Center, Ministry of Public Health, Udon Thani, <sup>5</sup>Cancer Unit, Faculty of Medicine, Chiang Mai University, Chiang Mai, <sup>6</sup>Lampang Cancer Registry, Lampang Cancer center, Ministry of Public Health, Lampang, <sup>7</sup>Songkla Registry, Faculty of Medicine, Prince of Songkla University, Songkhla, Thailand \*For correspondence: krisuw@kku.ac.th

investigators. Hence, the level of completeness obtained from the analysis can vary when different sources are chosen. However, a number of cancer registries (CRs) do not collect or have difficulty in identifying the actual sources of data. In this study, we assessed a virtual three source model to perform capture-recapture method, where clinical, pathological and mortality data are derived from existing, standard variables in any population-based cancer registry (MacLennan, 1991). The source derivation is explained in detail in the methods section.

The aim of this study was to explore the ability and behaviour of the capture-recapture method using a virtual three-source model in evaluation of the level of completeness of case ascertainment in cases that actual sources are lacking, and furthermore, to ensure comparability of completeness of data across registries and time periods.

## Materials and Methods

### Sources of data

Cancer registration in Thailand was initiated in 1971 by the National Cancer Institute (NCI) as a hospital-based registry. The first population based cancer registry started in Chiang Mai, representing the northern region. In 1988, Khon Kaen provincial cancer registry, representing the north-eastern region, was established. In 1990, Songkhla cancer registry at the Prince of Songkla University was setup as a proxy for the southern part of Thailand. The Bangkok registry, located in the National Cancer Institute, was the representative of the central region. Lampang provincial cancer registry, located in the northern region was set-up in 1995. Cancer registries are regulated to follow the registration procedure outlined by the International Agency for Research on Cancer (IARC) and International Association of Cancer Registries (IACR) (Vatanasapt et al., 1993; Deerasamee et al., 1999; Sriplung et al., 2003; Khuhaprema et al., 2007)

In 2007, there were 9 cancer registries operating in Thailand; 3 of which were managed by Chiang Mai University, Khon Kaen University and Prince of Songkla

University, while Lampang, Udon Thani, Nakhon Phanom, Bangkok, Rayong and Prachuap Khiri Khan were under the administrative structure of the Department of Medical Services, Ministry of Public Health (MOPH).

Information on cancer cases was retrieved from 9 population-based cancer registries in four regions of Thailand from January 1, 2003 to December 31, 2007. Cancer registries were anonymously labeled as A to I, not in the order above. Characteristics of cancer cases in the 9 registries are shown in Table 1.

Since actual data sources are not recorded in cancer registries in Thailand, we compiled three virtual data sources by considering three standard variables in cancer registries; basis of diagnosis, International Classification of Diseases for Oncology (ICD-O) morphology code and status at last seen. A virtual clinical source means a source where patients visit any medical care facility for symptoms or signs suspected of cancer. In our study this is derived from the basis of diagnosis variable for codes between 1 and 7 (clinical only to histology of primary tumor). Virtual morphology source includes any source where patients have a morphological diagnosis of cancer. It is derived from the basis of diagnosis variable for codes between 5 and 8 and/or ICD-O morphology code greater than 8000 (Fritz A et al, 2000). A virtual mortality source includes sources in which the patient's last known status is dead and/or the basis of diagnosis is death certificate only. The distribution of the three virtual sources in the 9 cancer registries is shown in Table 2.

By this definition of sources we assumed a temporal sequence of capture from clinical diagnosis, pathological confirmation, and death respectively. It was assumed that a case could not have a pathology result without a previous clinical diagnosis (rows 2 and 3 in Table 2). Thus, dependencies among the three virtual sources were not avoidable.

### Statistical methods

The Rcapture package (Baillargeon and Rivest, 2009) in R software (R Development Core Team, 2010). was used to fit Poisson regression models using the generalized

**Table 1. Number of Cases According to Characteristics of Data from the 9 Cancer Registries**

Characteristic	Registry								
	A	B	C	D	E	F	G	H	I
Sex									
Male	6,209	6,475	6,451	4,174	4,244	2,631	1,361	2,316	20,420
Female	5,858	7,483	5,724	4,215	3,893	2,644	1,403	2,360	22,188
Basis of diagnosis									
0 Death certificate only	597	1,193	1,409	620	275	2,886	974	920	19,006
1 Clinical only	464	178	1,843	477	157	565	650	1,575	2,322
2 Clinical investigation	4,045	2,289	2,190	1,533	887	437	123	346	1,330
3 Surgery without histology	237	189	221	21	24	0	3	8	30
4 Specific tumor markers	490	25	1,473	27	218	0	3	2	14
5 Cytology	414	651	594	456	239	76	98	156	1,044
6 Histology of metastasis	352	1,133	348	547	129	102	49	81	933
7 Histology of primary tumor	5,468	8,300	4,097	4,708	6,208	1,209	864	1,588	17,929
8 Autopsy with histology	0	0	0	0	0	0	0	0	0
Last known status									
Alive	5,763	7,025	5,425	3,387	3,363	1,983	1,402	3,127	21,136
Dead	6,304	6,933	6,750	5,002	4,774	3,292	1,362	1,549	21,472
Total number of cases	12,067	13,958	12,175	8,389	8,137	5,275	2,764	4,676	42,608

**Table 2. Number of Cases Falling in the 3 Virtual Sources in the 9 Cancer Registries**

Source			Registry								
Clinical	Morphology	Mortality	A	B	C	D	E	F	G	H	I
no	no	yes	597	1,197	1,409	620	275	2,886	974	920	19,006
no	yes	no	0	0	0	0	0	0	0	0	0
no	yes	yes	0	0	0	0	0	0	0	0	0
yes	no	no	1,587	919	2,362	611	387	751	602	1,609	3,384
yes	no	yes	3,701	1,810	3,506	1,447	959	267	203	370	600
yes	yes	no	4,710	5,900	3,063	2,772	2,976	1,232	798	1,517	17,752
yes	yes	yes	2,010	4,136	1,835	2,939	3,540	139	187	260	1,866
no	no	no	-	-	-	-	-	-	-	-	-

linear modeling framework (Baillargeon and Rivest, 2007; 2009; Rivest, 2008). Estimates of the demographic parameters of interest are derived from these log-linear parameter estimates. The model can incorporate up to three sources of variation among capture probabilities: a temporal effect, a heterogeneity effect between units and a behavioral effect, which are the characteristics of the cancer registries data as mentioned above. Hence, the method of capture-recapture analysis used by Rcapture package compromises the two main assumptions of the traditional capture-recapture method; source independency, and equality of chance to be captured. In addition to the results of the analysis, the package also reports the degree of heterogeneity of sources. All available models were evaluated and fitted, including Chao (Chao, 2001), Poisson2 (Baillargeon and Rivest, 2009) and Darroch (Baillargeon and Rivest, 2009) methods which account for the effect of time and heterogeneity in the Monte Carlo comparison of estimators of abundance models. From these three models, the best one was selected by considering deviances, degrees of freedom, and Akaike's Information Criterion (AIC) (Baillargeon and Rivest, 2007). The model with the lowest AIC was considered to be the best model. The total number of cases that might have been present in the cancer registries were estimated from the best fitting model.

The estimated percentage of completeness (EPC) is the proportion of observed to estimated cases. In all

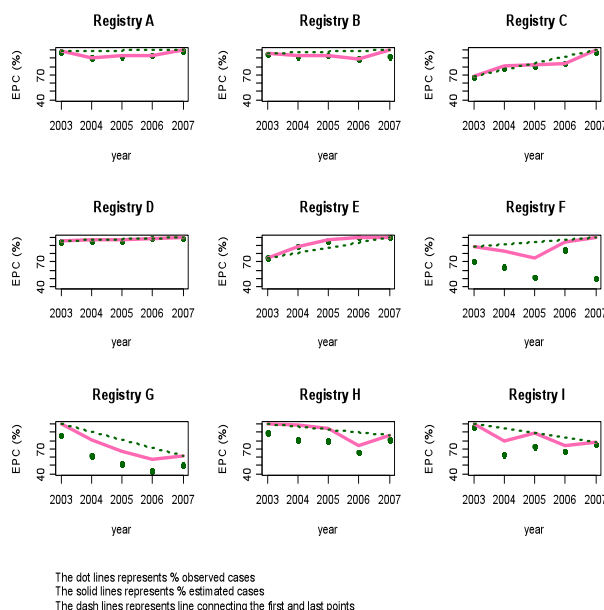
registries, the percentage of completeness cannot exceed 100%. The percentage difference can be calculated by 100 - EPC. The estimates of total number of cancer cases in 5 years and annual cases were calculated. The graphs for consecutive 5-year observed and estimated cases were plot to see the trend pattern. A linear model was used to determine the predictors of EPC in case ascertainment. Three main potential factors included in the model were the percentage of morphological verified cases (%MV), percentage of death certificate only cases (%DCO), a linear time effect in year, and the individual registries. The first two predictors were chosen because they represented two main sources used in the capture-recapture model. Even though the two factors were calculated from the same set of data where the capture-recapture procedure was applied, here in the linear regression model, they were summary parameters of cancer registries. The coefficients of these two parameters would express the degree of dependency upon which EPC lies. The last variable reflected implicit difference among registries.

**Results**

In this study the Poisson2 with bias correction model usually produced the smallest AIC. The number of observed and estimated cases and 95% confidence intervals estimated using Poisson2 modeling technique is shown in Table 3. The EPC in 5 registries was greater than 97%. The worst was observed in registry F, where the EPC was only 70% (95% CI 66.1-72.8).

The percentage difference between the actual cases and the estimated cases was less than 5% in registries A to E. Of the other five registries, it was greater than 10%, and the worst was observed in registry F, where it was about 30%.

The EPC values for each registry for the period 2003-2007 are shown in Figure 1. Comparison of relative



**Figure 1. Relative EPC Plots, 2003-2007**

**Table 3. Estimated Percentage of Completeness with 95% Confidence Interval of 9 Cancer Registries**

Reg	Observed	Estimated	95% CI*	EPC(%)	95 % CI**
A	12,067	12,105	12,092-12,119	99.7	99.5-99.8
B	13,958	14,054	14,033-14,077	99.3	99.2-99.5
C	12,175	12,512	12,464-12,564	97.2	96.9-97.7
D	8,389	8,450	8,433-8,470	99.3	99.0-99.5
E	8,137	8,150	8,143-8,159	99.8	99.7-99.9
F	5,275	7,588	7,244-7,986	70.0	66.1-72.8
G	2,764	3,511	3,370-3,673	78.7	75.3-82.0
H	4,676	5,295	5,197-5,407	88.3	86.5-90.0
I	42,608	48,140	47,838-48,456	88.5	87.9-89.1

\*multinomial profile likelihood confidence interval; \*\* EPC

**Table 4. Factors Affecting Estimated Completeness by Capture-recapture Method**

	Model 1			Model 2			Model 3		
	Coeff.	95% CI		Coeff.	95% CI		Coeff.	95% CI	
		LL	UL		LL	UL		LL	UL
Intercept	94.88	85.85	103.91	94.03	85.31	102.74	96.73	94.27	99.19
%MV	0.09	-0.03	0.22	0.09	-0.03	0.21	-	-	-
%DCO	-0.07	-0.33	0.18	-0.1	-0.34	0.14	-0.67	-0.95	-0.39
MV:DCO	-	-	-	-	-	-	0.02	0.01	0.02
Year	-0.38	-1.35	0.59	-	-	-	-	-	-
Registry(Ref: A-E)									
F	-24.42	-36.56	-12.28	-23.12	-34.73	-11.51	-15.57	-25.66	-5.48
G	-14.67	-23.2	-6.13	-13.83	-22.04	-5.61	-14.15	-20.89	-7.4
H	-8.18	-13.71	-2.65	-7.86	-13.29	-2.42	-8.94	-13.18	-4.71
I	-6.91	-17.09	3.27	-5.85	-15.01	3.9	-12.47	-21.05	-3.88
AIC	272.24			270.95			253.39		
Adj.R <sup>2</sup>	0.85			0.85			0.9		

CI= Confidence interval, AIC= An Information Criterion, LL=Lower limit, UL=Upper limit, Coeff= Coefficient, Ref= Reference

completeness across registries and time is possible since the unit are in percentages. The dashed line linking the first and last points in the estimated lines represents the overall trends over the entire 5-year period.

There are three aspects to consider in the EPC plots: 1) the difference between observed and estimated percentages, 2) the direction of the overall trends in the entire period, and 3) the pattern of the estimated line. The difference between the observed and estimated percentages are small in registries A to E and large in registries F to I. The overall trends from the 9 registries can be categorized into three directions. Increasing trends are observed in registries C, E and F, declining trends in registries G, H and I, and steady trends are observed in registries A, B and D.

The patterns of the estimated lines in relation to the line of overall trends can be classified into 4 patterns. The first pattern is a concave curve observed in registries A, B, F and G. The second one is mixed concave and convex curve present in registries C, H and I. The third pattern shows steadiness and closeness to the overall trend as seen in registry D. The last one is a convex arch in registry E.

Factors affecting estimated completeness were modeled using linear regression. Table 4 summarizes the coefficients and 95% CIs of parameters in three different models. The registries in which completeness was greater than 90% were grouped together as the reference. The first model includes %MV, %DCO, linear time effect, and registries. The linear time effect was omitted in the second model and a better fit in terms of AIC was obtained. %MV was replaced by the interaction of %DCO and %MV in the third model, which of the 3 models, was the best and could explain 90% of the variability in the data.

The third model has an intercept of 96.7 which is the average EPC of the reference registries A to E, where the average difference between observed and estimated values is minimal. A higher percentage of DCO induces a reduction in EPC. %MV better explains the EPC through its interaction with %DCO. However, the effect is minimal as the coefficient is only 0.02. Registries F to I have intrinsic effects of incompleteness in terms of large negative coefficients in the models compared to the reference registries.

## Discussion

We have demonstrated the use of a virtual capture-recapture method to estimate completeness of 9 cancer registries in Thailand during the period 2003 to 2007. Three virtual sources were generated from three actual variables in registry databases; basis of diagnosis, ICD-O morphology code and last known patient status. The estimated completeness of the registries ranged from 70% to 99.6%.

The virtual three-source model introduced in this article has some advantages to the actual data sources. It requires three core variables collected by all cancer registries as recommended by the IARC and IACR (<http://www.iarc.fr>; <http://www.iacr.com.fr>). The definition and codes of the three variables are standardized in all registries. Thus, the proposed virtual three-source model is eventually standardized and the estimated completeness provided by a capture-recapture analysis of the three-source model is comparable among registries. The model can be used even in a registry where death notifications are captured through any reliable process but official death certificates are not available.

Comparability is an important aspect of data quality in cancer registration (Parkin, 1994; Parkin and Plummer, 2002). Comparability of cancer statistics concern not only the differences in place or population but also in time. Hence, the consistency in source definition enables comparison of completeness of cancer registry data over time and across population.

The capture-recapture method in ascertainment of completeness of cancer registration has been adopted for many years. Limitations of this method using the real source of cases have been mentioned by many authors. Ballivet et al., (2000) used a combination of sources to handle the dependency of sources. Brenner et al., (1995) used actual clinical, pathological and mortality sources. Both assumed independency of sources and found that estimated cases could be captured by different pairs of sources. By this method the completeness of cancer registry cases could not be estimated.

In this study, we used Rcapture package in R statistical software in analysis of capture-recapture method. The

**Table 5. Comparison of EPC by Capture-recapture and %DCO and %MV in 9 Registries**

Registry	EPC, Capture-recapture	% DCO	% MV
A	99.6	5.0	52.0
B	99.3	9.0	72.0
C	97.2	11.0	41.0
D	99.3	7.0	68.0
E	99.8	3.0	81.0
F	70.0	55.0	25.0
G	78.7	38.0	36.0
H	88.3	20.0	39.0
I	88.5	46.0	45.0

authors of this package claimed to address the problems of dependency, temporal effect, and heterogeneity of sources. The approach tries to fit various log-linear models, selects the best one, and finally estimates the abundance of cases. Such an approach relaxes the constraint of source characteristics against traditional capture-recapture method in cancer registration.

Currently, the Thai Network of Cancer Registries is seeking a set of completeness indices to monitor cancer registries throughout the country. %DCO, %MV and, if possible, percentage of mortality incidence ratio (%MI) are good candidates. In Table 5, registries A, B, D and E have high level of completeness, rather high %MV and low %DCO, while registries F, G, H and I show the reverse. Registry C is intermediate in terms of %MV but rather good in terms of completeness estimated by the capture-recapture method and %DCO. Such a discrepancy suggests that additional information of the capture-recapture method is needed to improve the completeness aspect of quality of cancer registries.

The association between EPC and %DCO and the interaction between %DCO and %MV shown in table 4 (model 3), demonstrates that registries F to I are different from registries A to E, of which the percentage of under-registered cases was estimated at 9% or higher.

The model gives the average EPC of high quality registries in Thailand at around 96.7%. Although it is hard to draw a cut off point of acceptable EPC, a value of 90% is a reasonable level for completeness. Registries with EPC calculated by the virtual three-source model using the capture-recapture method that are lower than 95% should be reviewed for their performance. The magnitude of EPC is not dependent on the size of the population covered by the registry but on its precision, noted by the 95%CI. EPC is poor when the number of observed cases is much less than that expected by the model.

Figure 1 shows poor reliability of the method when the average EPC is lower than 90%. The trend and pattern analysis of the annual EPC plots is not relevant for registries F to I. The convex pattern observed in registry E may be caused by incomplete or late update of cases in the very last years, 2007, close to the time of analysis. This reason has been confirmed by the registrar of registry E by personal communication. The concave pattern observed in registries A and B may be explained by the registration method of entering suspicious cancer cases into the database as soon as they arrive and removing them later when a non-cancer diagnosis is confirmed. Another

reason for this pattern might be the change in registration method during the 5-year period. These registries adopted a computerized database of all hospitals within the registry network at some time during the 5-year study period. Both reasons were personally reported from the staff of the two registries.

There are some limitations of using this method in some cancer registries in other countries. A registry in which a significant number of cases are identified by morphology alone without any clinical data may have a biased estimate of completeness since the ratio of (MV alone) / (MV + clinical) would be too high and the model will produce an over estimate of the number of cases. Another problem may occur when the temporal assumption is violated, for example, in a registry where a large proportion of cases is identified by morphological diagnosis or death before clinical information is obtained. However, this situation is not common in Thailand.

In conclusion, in summary, the virtual three-source capture-recapture method applied to cancer registry data can be used to estimate the level of completeness of case ascertainment. The problem of source dependency, time and heterogeneity effects can be handled with sophisticated capture-recapture modeling strategies. The association of EPC with DCO and MV was observed, however, some implicit characteristics of source usage within cancer registries were also demonstrated.

The virtual three-source model with capture-recapture method can be used as a simple tool to evaluate the level of completeness for case underascertainment in cancer registries. The capture-recapture method should be established as another standard indicator to monitor the quality of case ascertainment of cancer registries.

## Acknowledgments

This study was part of the first author's thesis to fulfill their requirements for a PhD in Epidemiology at Prince of Songkla University. The researchers would like to express their thanks to the Health Information System Development Office (HISO) and Prince of Songkla University for grant support, Thai Cancer Information Network for encouragement and support, and the 9 cancer registries mentioned above for their data provided in this study. We would like to thank Edward McNeil for his assistance with English proofing.

## References

- Baillargeon S, Rivest L-P (2007). Rcapture: Loglinear models for capture- recapture. *J Statistical software*, **19**, 1-31.
- Baillargeon S, Rivest L-P. Rcapture: Loglinear models for capture - recapture. Package. [update 2009 Apr 28; cited 2009 Jun 30] Available from: <http://www.R-project.org>.
- Ballivet S, Salmi LR, Dubourdiou D (2000). Capture-recapture methods to determine the best design of a surveillance system. Application to a thyroid cancer registry. *Eur J Epidemiol*, **16**, 147-53.
- Brenner H, Stegmaier, Ziegler H (1995). Estimating completeness of cancer registration: an empirical evaluation of the two source capture-recapture approach in Germany. *J Epidemiol*

- Chao A (2001). An overview of closed capture-recapture models. *J Agric Biol Environ Stat*, **6**, 158-75.
- Cormack RM (1968). The statistics of capture-recapture methods. *Oceanography and Marine Biology: Ann Rev*, **6**, 455-506.
- Deerasamee S, Martin N, Sontipong S, et al (1999). Cancer in Thailand Volume II, 1992-1994. IARC Technical Report No.34 [2]. Lyon, IARC.
- Fritz A, Percy C, Jack A, et al (2000). International Classification of Diseases for Oncology. 3rd ed. World Health Organization, Geneva.
- International Agency Research on Cancer. [update 2010 ; cited 2010 April 15]; Available from: <http://www.iarc.fr> 16.
- International Association of Cancer Registries. [update 2010 ; cited 2010 April 15]; Available from: <http://www.iacr.com.fr>
- International working group for disease monitoring and forecasting (1995). Capture-recapture and multiple-record systems estimation II: applications in human diseases. *Am J Epidemiol*, **142**, 1059-68.
- Khuhaprema T, Srivatanakul P, Sriplung H, et al (2007). Cancer in Thailand, Volume IV, 1998-2000. 4th ed. Bangkok, National Cancer Institute.
- MacLennan R (1991). Cancer registration: principles and methods. Items of patient information which may be collected by registries. *IARC*, 43-63.
- Parkin DM, Bray F (2009). Evaluation of data quality in the cancer registry: principles and methods part II completeness. *Eur J Cancer*, **45**, 756-64.
- Parkin DM, Plummer M (2002). Comparability and quality of data. Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB, editors. Cancer Incidence in Five Continents. 8th ed. Lyon, France, IARC.
- Parkin DM, Chen VW, Ferlay J, et al (1994). Comparability and Quality Control in Cancer Registration. IARC Technical Report No. 19. Lyon, France, IARC.
- R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. [update 2010 Jan 31; cited 2010 May 31]; Available from: <http://www.R-project.org>.
- Rivest L-P (2008). Why a time effect often has a limited impact on capture-recapture estimates in closed populations. *Canadian J Statistics*, **35**, 75-84.
- Sriplung H, Sontipong S, Martin N, et al (2003). Cancer in Thailand Volume III, 1995-1997. 3rd ed. Bangkok, National Cancer Institute.
- Vatanasapt V, Martin N, Sriplung H, et al (1993). Cancer in Thailand Volume I, 1988-1991. IARC Technical Report No.16 [1]. Lyon, IARC.