RESEARCH COMMUNICATION

A Method to Adjust for Ascertainment Bias in the Evaluation of Cancer Registry Data

Mahdi Fallah*, Elham Kharazmi

Abstract

Using the trend of age-standardized incidence rate of cancers (ASR) is inaccurate for registration with incomplete reporting, especially in developing registries. The relative age-standardized ratio (RASR) is a new measure that takes ascertainment bias of registration into account. RASR is calculated from the ASR for each cancer divided by the ASR for leukemia. Leukemia was chosen as the reference because its ASR is rather constant over time in valid registries. The adjusted relative age-standardized rate (ARASR with same unit as ASR) is calculated by multiplying the RASR for a specific cancer in a particular year by the sum of ASRs of that cancer over the years for which a trend is being determined and then dividing result by the sum of RASRs of the cancer for those years. Two likely assumptions are behind use of ARASR, first, constant ASR of leukemia over time, second, if under/over-registration occurs, it happens for all cancers to the same extent (random under/over-reporting). Using the ARASR with empirical data of valid Finnish and SEER cancer registries proved that trend of ASRs for each cancer is almost equal to its ARASR. Using trends of ARASRs instead of ASRs in a registry with incomplete data collection in first years of registration demonstrated more realistic results. In conclusion, the ARASR is more accurate than the ASR for studying cancer incidence trends in registries with incomplete reporting. ARASRs in different countries or different times are comparable since they are age-standardized. Moreover, comparison between trends of ASRs and ARASRs can be used as a test for validity of registration.

Key Words: Neoplasms - incidence - trends - bias - registries

Introduction

Ideally, age-standardized incidence rate (ASR) is used in studying cancer incidence trends. ASR is the optimal measure of incidence for time trend analysis only when the registration is complete and there is no under-registration or over-registration (ascertainment bias) in the population-based cancer registries. However, not all registrations are perfect and the completeness of registration may vary over time especially at the beginning of new registries which may have either under-registration due to under-report and difficulty in collecting data, or over-registration because of including prevalent cases into registry instead of newly-diagnosed cancer cases. This article is about a new measure (adjusted relative age-standardized ratio) which can be used as an alternative to ASR to study time trend of cancer incidence in registries subject to ascertainment bias.

As an example, there was a large scale retrospective five year population-based registration in northern Iran (Fallah, 2007) in which numbers of new cases and ASRs were increasing substantially each year. This increase unrealistically happened to all cancers (including cancers with rather constant ASR over time such as leukemia) during a short period of time (5 years). The steep increase was more likely due to improvement in completeness of registration rather than a real rapid change in the incidence

of cancers in such a short period of time. Therefore, the time trend of ASR for such kind of data was not valid. In other words, as an example, if there is 10% under-registration in the first year (1996), 5% under-report in the second year (1997) and 0% under-estimation in the third year (1998), given no real change in the ASR of cancers over this short period of time, the trend analysis will show a steep increase in ASR of cancers. This is the situation that one can use the alternative measure, adjusted relative age-standardized ratio, instead of ASR to correct for variation in completeness of registration.

Method of Calculation

This is the first time that the term "relative age-standardized ratio (RASR)" is used in cancer epidemiology. This ratio is the ASR of each cancer divided by ASR of a specific reference cancer (Equation 1). Adjusted relative age-standardized rate (ARASR) is calculated by multiplying the RASR for a specific cancer in a particular year by the sum of ASRs of that cancer over the years for which a trend is being determined and then dividing the result by the sum of RASRs of the cancer for those years

$$RASR_{C_i\_y} = ASR_{C_i\_y} / ASR_{Leukemia\_y}$$

Equation 1. Relative Age-standardized Ratio

*Tampere School of Public Health, FIN-33014, University of Tampere, Tampere, Finland. * For Correspondence: Email, Mahd.Fallah@gmail.com, Mahdi.Fallah@uta.fi*

(Equation 2). The general formula for calculating RASR and ARASR are as following where "Ci" is the cancer site (C1 to Cn) and "y" is the calendar year (1 to k):

For this novel method, leukemia has been chosen as reference cancer because it has been proven in long term registries that its ASR is generally stable over time (Stewart and Kleihues, 2003). ASR of this cancer had no substantial

$$ARASR_{C_i\_y} = RASR_{C_i\_y} \times \sum_{y=1}^{k} ASR_{C_i\_y} \Bigg/ \sum_{y=1}^{k} RASR_{C_i\_y}$$

Equation 2. Adjusted Relative Age-standardized Rate

change during 50 years of registration in the highly valid Finnish Cancer Registry as shown in Figure 1 (Hakulinen et al, 2006; Teppo et al, 1994) similar to other long term valid registries such as in Norway 1955-2004 (Cancer Registry of Norway, 2006) (Figure 2) or in SEER Program in USA (0.0% change from 1975-2002; Figure 3) (Anonymous, 2005). Another study in Italy (10,946 analyzed cases aged 15+) showed rather stable the incidence rates trend for leukemia during the period 1986 to 1997(De Lisi, 2004). Therefore, in registries with under/
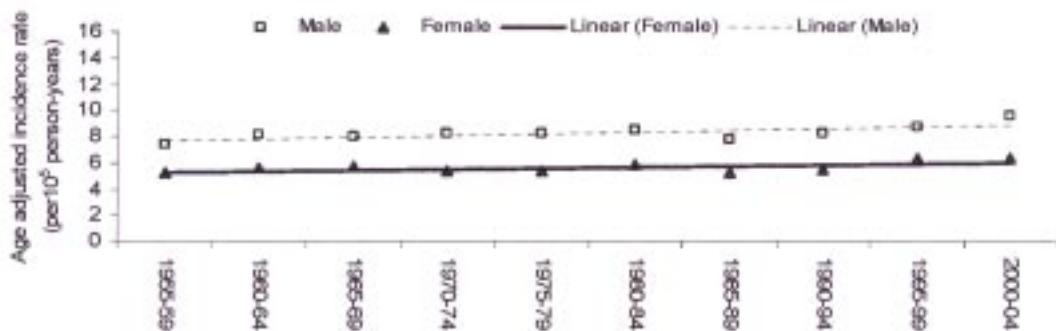
over-registration assuming that the proportion of under/ over-registration is equal for every cancer (random under/ over-reporting), the change due to variation in registration will be cancelled out if relative age-standardized ratio is used instead of ASR. In the previous example with 10, 5 and 0% under-registration in three consecutive years, if ASR is used, it shows rapid increase for all cancers, but if RASR or ARASR is used for trend analysis it does not show any false change in the incidence of cancers.
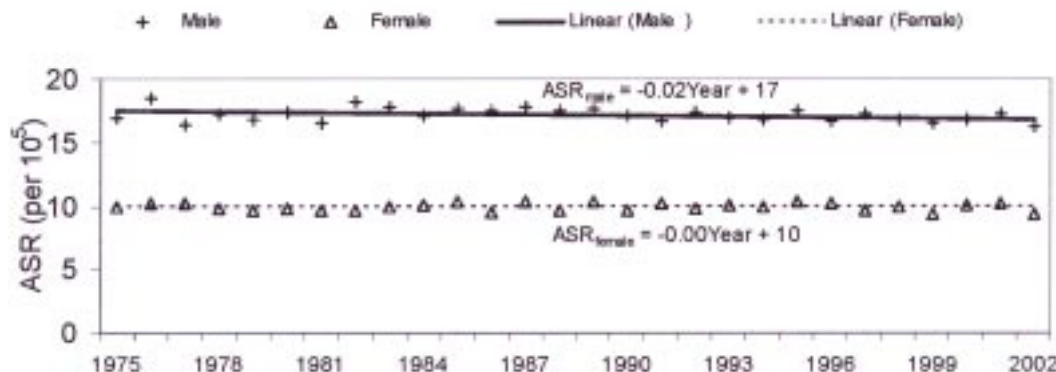
Testing and Simulations

Trend of RASR and ARASR was tested on empirical data of Finnish Cancer Registry with well-known validity (Teppo et al, 1994). The shape of time trend for RASR was similar to that of ASR (compare Part A and B in Figure 4). Results were very supportive to this theory that RASR can be used as an alternative estimate for ASR, but there was a difference in the scale of these two parameters that could be corrected by adjustment (Equation 2). Comparing coefficients of "Year" for ASR and ARASR in each part of Figure 5 showed that coefficients for the time trend of ASR of each cancer and the corresponding adjusted relative
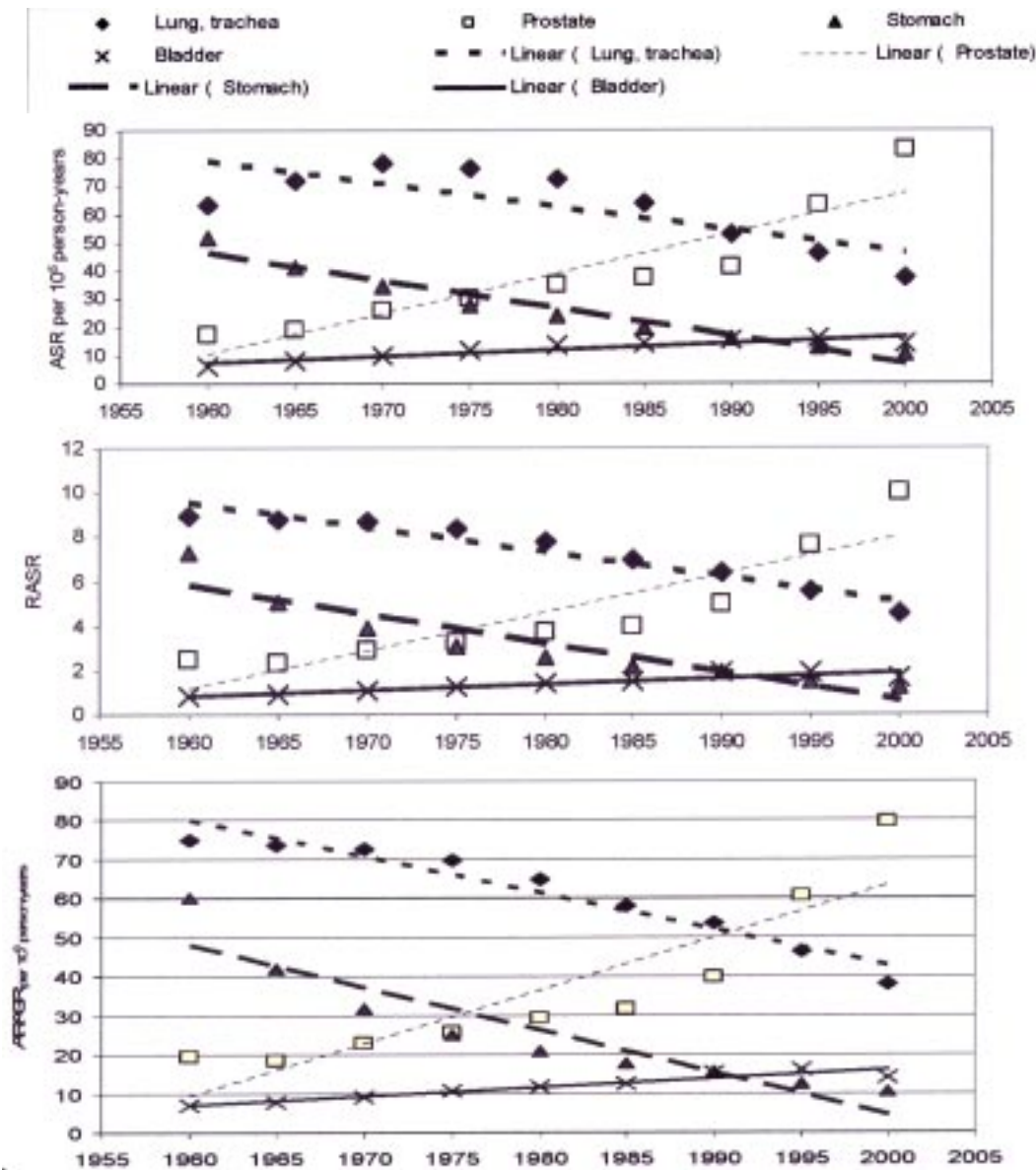

Figure 1. Time Trend of ASR of Leukemia in Finland, 1960-2000

Figure 2. Time Trend of Age-adjusted Incidence Rate of Leukemia in Norway, 1955-2004

Figure 3. Time Trend of Age-adjusted Incidence Rate of Leukemias in USA, 1975-2002

Figure 4. ASR, Relative Age-standardized Ratios and Adjusted Relative Age-standardized Rates for Four Leading Cancers in Finnish Men

age-standardized rate of that cancer were almost equal for every cancer both in men and women. For example, ASR of cervical cancer in female Finns was decreasing 0.31 (per 100,000 persons) each year which was exactly the same as annual change in ARASR of cervical cancer in Finnish women. The test for using ARASR instead of ASR on another large long term empirical data of SEER program in USA showed similar results (Figure 6).

This theory was true for registries with perfect cancer registration; however, it should also be tested in the registries subject to variation in completeness of registration. Figure 7 illustrates results of comparison between trend of ASR and ARASR for some cancers in a newly-established cancer registry in northern part of Iran (41 million person-years follow-up) with relatively incomplete data collection in the first years of registration (Fallah M, unpublished monograph). For instance, Part A1 in Figure 7 demonstrates that ASRs of all three cancers (leukemia, esophageal and rectal cancer in male) were increasing during five years of registration (1996-2000).

This amount of increase in such a short period of time for all cancers especially for leukemia seems to be exaggerated. This might have happened due to under-registration in the first years of registration (improvement of completeness by time). Based on other sources of information about the population from which this data were collected, in fact esophageal cancer was decreasing (Sadjadi et al, 2005; Yazdizadeh et al, 2005). This fact can be seen in Part A2 where ARASR was used instead of ASR. Part A2 elucidated that incidence of esophageal cancer was not increasing; in fact it was annually decreasing 0.38 per 100,000 persons. In the same parts, comparing slope of change in ASR and ARASR of rectal cancer in men showed that both ASR and ARASR of rectum were increasing, but slope of increase in ARASR was around half of the slope for ASR which sounds more realistic for such a short period of time.

Trend in ASR of rectal cancer in Iranian women (Part B1, Figure 7) showed that the incidence was increasing 0.17 per 10^5 each year and the shape and slope of trend was very similar to those of leukemia. However, ARASR
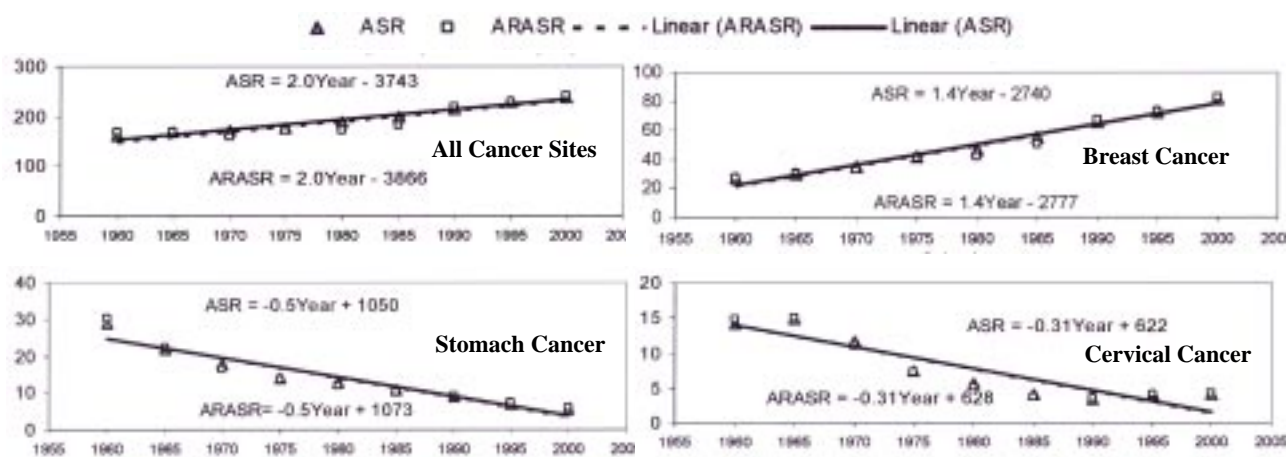
Figure 5. Time Trends of ASR and Adjusted Relative Age-standardized Rates (ARASR) of Some Cancers in Finnish Women and Fitted Linear Regression Line with Slope Coefficient
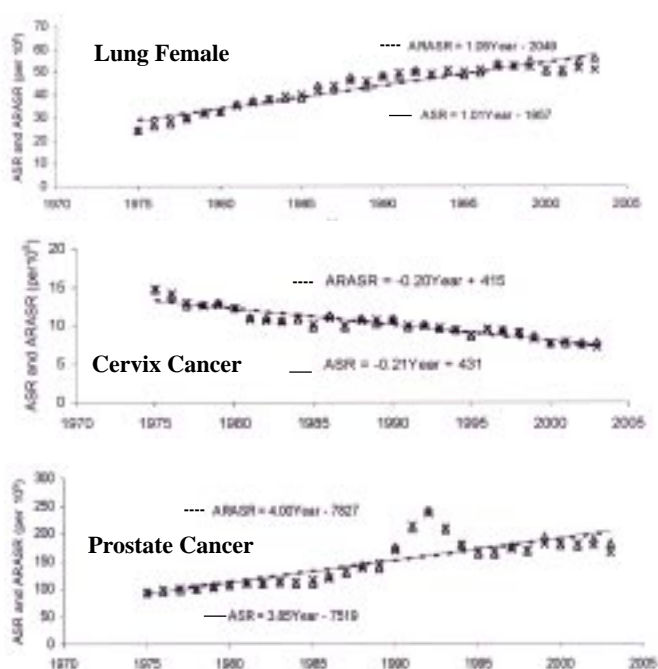

Figure 6. Time Trend by ASR (xx) and ARASR (ΔΔ of Some Randomly Selected Cancers in USA, 1975-2003

of rectal cancer in women in Part B2 showed almost no change over time. Same happened to "All sites excluding skin" in men (compare Part C1 and C2 in Figure 7).

To demonstrate the application of ARASR, two different scenarios of ascertainment bias have been simulated using data of cervical cancer from SEER program, USA, 1996-2000. First scenario is under-reporting in the first years and improvement in the completeness of registration by time (Figure 8). In this simulated scenario, both leukemia and cervical cancer were 25, 20, 10, 5 and 0% under-estimated in the years 1996 to 2000 respectively. Trend in ASRs of cervical cancer showed a slow increase (0.1 per 10^5 annually) whereas in SEER program (real population) the trend was in opposite direction (decreasing 0.5 per 10^5 annually). Unlike the ASR trend, ARASR trend (decreasing 0.4 per 10^5 annually) is very similar to the real situation in the USA population. Second scenario is 25, 20, 10, 5 and 0% over-estimation of ASRs for cervical cancer and leukemia in USA, 1996

to 2000 respectively. In this scenario also trend in ASR is misleading (coefficient -1.1) and trend of ARASR is very similar to the reality (both coefficients -0.5).

Discussion

Testing ARASR and ASR trends in complete registries showed that their amounts and trends were almost equal. However, in situations where a registration was subject to ascertainment bias like in an incomplete developing registry in northern Iran and in the simulated scenarios of under/over-registration of USA data, ARASR was a more accurate measure for time trend analysis conditional to two assumptions, first, rather constant incidence rate for leukemia over time (which was acceptable by empirical data of perfect registries), second, proportional variation in completeness of registration which means if any under/over-registration occurs, it happens to every cancer including leukemia in the same proportion. First condition was true for the simulated scenario (rather invariant incidence of leukemia in USA 1996-2000; coefficient -0.065). The second condition also seems to be true in this data because the shape of every increase and decrease in ASR of cancers is similar for these cancers (more under-registration are seen in first and third years in all male cancers and in first, second and fourth years in female cancers although generally they all are increasing; Figure 7, Part A1, B1, and C1).

There is another application for ARASR as well so that it can also be used to evaluate completeness of registration. When ARASRs and ASRs and their trends are similar, it indicates unbiased registration (no under/over-registration) and if they are very different, it may show a problem in validity of registration in terms of either completeness (no under-registration) or accuracy (no over-registration).

ARASR has the same unit as ASR as it is a ratio with no unit (RASR) multiplied by a rate with unit of number per person-time (θΘASR) divided by another ratio with no unit (θΘRASR). Using ARASR has another positive point so that it has already taken the change in population age structures over time into account because ASRs have been already directly standardized by a standard population such as world standard population. Therefore, ARASRs
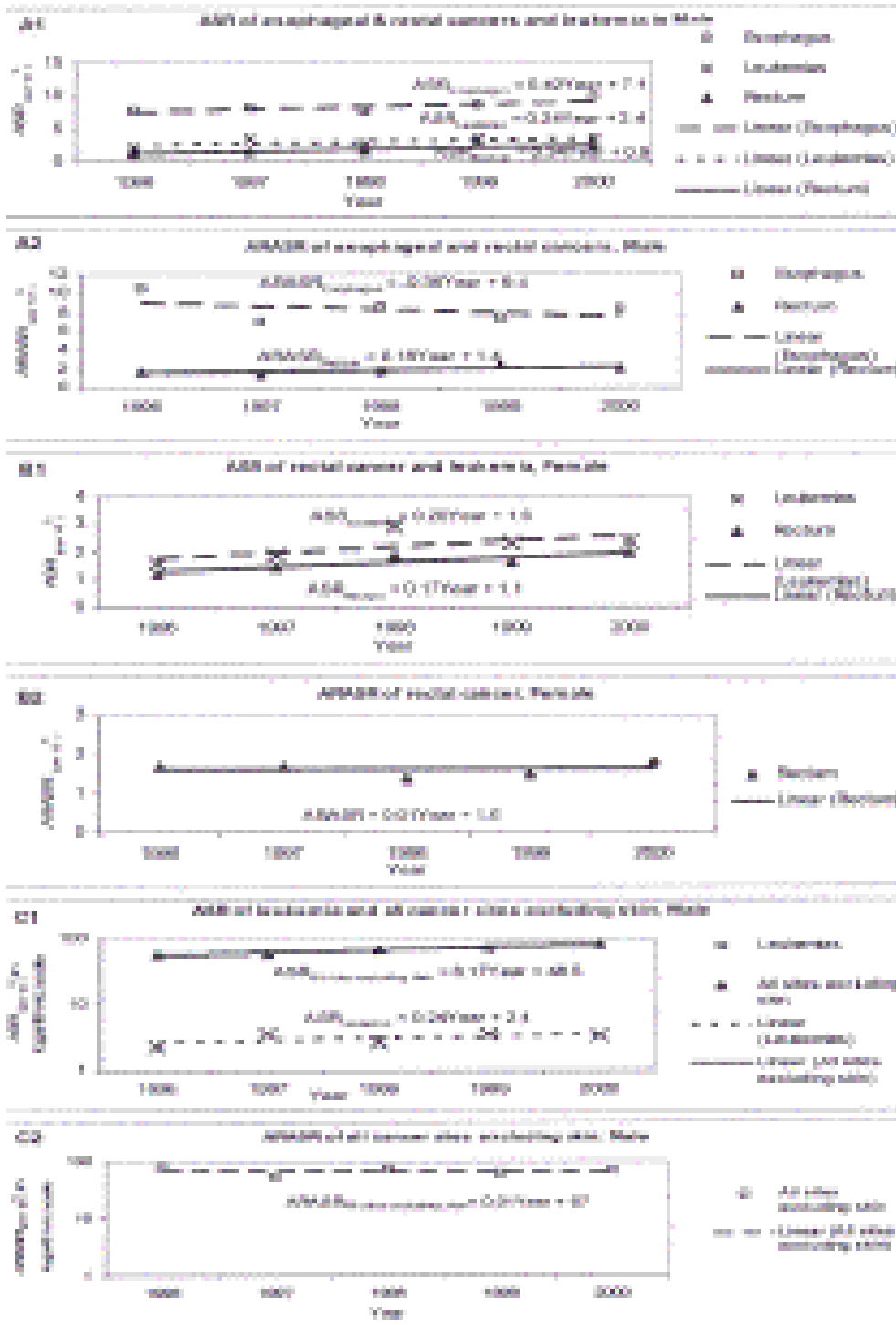
Figure 7. Comparison of Trends in ASR and ARASR of Some Cancers in a Registry with Incomplete Data Collection (unpublished data)
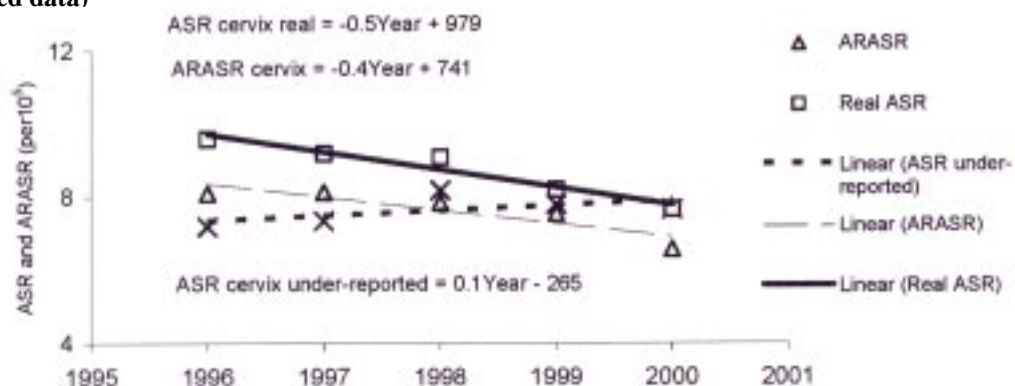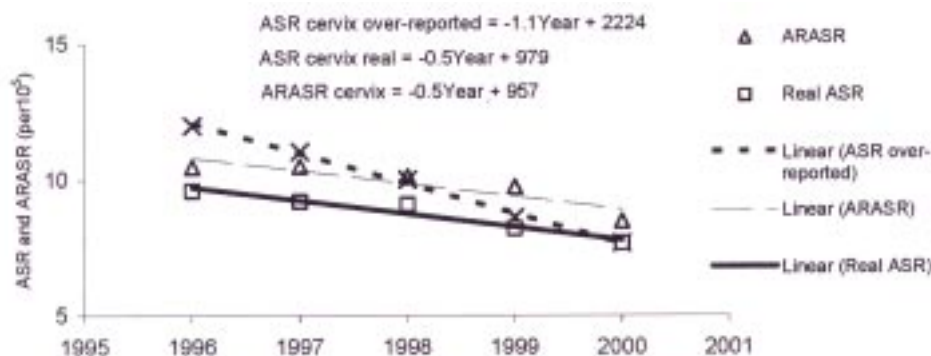

Figure 8. Simulated Trends of Under-reported ASR, Real ASR and Calculated ARASR for Cervical Cancer in the USA, 1996-2000

Figure 9. Simulated Trends of Over-reported ASR, Real ASR and Calculated ARASR for Cervical Cancer in the USA, 1996-2000

from countries are comparable to each other because they have the same unit and they have been already age-standardized in the first stage (using ASRs for calculation).

Something should be considered when using ARASR; one is two assumptions behind it that should be true before any judgment about results of time trend in ARASR. In spite of evidences for stable time trend of leukemia in long term valid registries, there are some controversial reports in the short term developing registries which might be due to improvement in the ascertainment of leukemic cases, change diagnostic tool or a real trend. However, in general, the annual change in the incidence of leukemia was not substantial (slope = 0.02-0.07 per 10^5) (Kamsa-ard et al, 2006) or there was only slow alteration in subtypes of leukemia sometimes in the price of reverse change in the non-specified type or other subtypes of leukemia(Gonzalez et al, 2001; Kroll et al, 2006; Linet and Cartwright, 1996; Magnani et al, 2003; McNally et al, 1999; McNally et al, 1999). Nevertheless, the trends in overall incidence of leukemia (all subtypes including non-specified type as well as combination of childhood and adulthood leukemias) have generally been stable or just slightly changing over time (Stewart and Kleihues, 2003). Meanwhile, simulated scenarios demonstrated that ARASR is more accurate measure even with small amount of variance in incidence of leukemia (up to 0.07 per 10^5). The other point is that this theory is new and should be tested in some other cancer registries although it seems to work wherever the assumptions behind it are not strongly violated.

Similar concept was used in another study where morbidity odds ratios (MOR) were compared between past and present situation of a cancer in the absence of accurate incidence estimates (Yazdizadeh et al, 2005). Yazdizadeh et al. used childhood cancers as the reference. They referred to another similar concept, morbidity or mortality odds ratio, as an alternative to proportional mortality ratio recommended by Miettinen to assess cancer risk in the absence of denominators (Miettinen and Wang, 1981). The concept of MOR does not seem appropriate when age structure of a population is changing over time.

In conclusion, ARASR is a more accurate than ASR in studying cancer incidence trends in registries for which there is incomplete reporting. ARASR in different countries or different times are comparable since it is already age-standardized. Moreover, comparison between time trend of ASR and ARASR can be used as a test for validity of registration.

References

Cancer Registry of Norway (2004) . Cancer in Norway 2004. http://www.kreftregisteret.no/frame.htm?english.htm.

De Lisi V (2004). Leukemias and multiple myeloma. *Epidemiol Prev*, **2 Suppl**, 97-101.

Fallah M (2007). Cancer incidence in five provinces of Iran: Ardebil, Gilan, Mazandaran, Golestan and Kerman, 1996-2000. Acta Universitatis Tamperensis 1214, Tampere University Press, Tampere. Available at :http://acta.uta.fi/978-951-44-6876-6.pdf

Gonzalez JR, Moreno V, Maria Borras J, et al (2001). Incidence and survival of leukemias according to the different histologic subsets, in Tarragona, Spain, between 1980-1994. *Med Clin,* **5**, 174-8.

Hakulinen T, Pukkala E, Sankila R (2006). Basic statistics. http://www.cancerregistry.fi/eng/statistics/9-36-122.html 4.7.2006.

Kamsa-ard S, Wiangnon S, Kamsa-ard S, et al (2006). Trends in incidence of childhood leukemia, Khon Kaen, Thailand, 1985-2002. *Asian Pac J Cancer Prev*, **6**, 75-8.

Kroll ME, Draper GJ, Stiller CA, et al (2006). Childhood leukemia incidence in britain, 1974-2000: Time trends and possible relation to influenza epidemics. *J Natl Cancer Inst*, **6**, 417-20.

Linet MS,Cartwright RA (1996). The leukemias. In Cancer epidemiology and prevention Eds Schottenfeld D and Fraumeni JF. Oxford University Press, New York, pp 846-7.

Magnani C, Dalmasso P, Pastore G, et al (2003). Increasing incidence of childhood leukemia in northwest italy, 1975-98. *Int J Cancer*, **4**, 552-7.

McNally RJ, Roman E, Cartwright RA (1999). Leukemias and lymphomas: Time trends in the UK, 1984-93. *Cancer Causes & Control*, **1**, 35-42.

Miettinen OS, Wang JD (1981). An alternative to the proportionate mortality ratio. *Am J Epidemiol*, **1**, 144-8.

Sadjadi A, Nouraie M, Mohagheghi MA, et al (2005). Cancer occurrence in Iran in 2002, an international perspective. *Asian Pac J Cancer Prev*, **3**, 359-63.

SEER cancer statistics review, 1975-2003. http://seer.cancer.gov.

Stewart BW, Kleihues P (2003). Leukaemia. In WORLD CANCER REPORT Eds Stewart BW and Kleihues P. IARC Press, Lyon, pp 242.

Teppo L, Pukkala E, Lehtonen M (1994). Data quality and quality control of a population-based cancer registry. Experience in Finland, *Acta Oncol*, **4**, 365-9.

Yazdizadeh B, Jarrahi AM, Mortazavi H, et al (2005). Time trends in the occurrence of major GI cancers in Iran. *Asian Pac J Cancer Prev*, **3**, 130-4.