Calculating Adjusted Survival Functions for complex sample survey data and application to vaccination coverage studies with National Immunization Survey (NIS).

Zhen Zhao[a, *], Ph.D.; Philip J. Smith[a], Ph.D.; David Yankey[a], MS
Kennon R. Copeland[b], Ph.D.

[a] National Center for Immunization and Respiratory Diseases
Centers for Disease Control and Prevention
1600 Clifton Road NE, Mail Stop A19, Atlanta, GA 30333, USA

[b] NORC at the University of Chicago
55 E. Monroe Street, Suite 3000
Chicago, IL 60603

[*] Corresponding author at: National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Mail Stop A19, Atlanta, GA 30333, USA.
Tel.: +1 404-639-8238; fax: +1 404-639-3266.
E-mail addresses: zaz0@cdc.gov (Zhen Zhao)

[‡] Disclaimer: The findings and conclusions in this article are solely the responsibility of the authors and do not necessarily represent the official view of Centers for Disease Control and Prevention.

Abstract

Background: In vaccination studies with complex sample survey data, non-parametric survival functions may be useful. Recent publications have proposed several methods for evaluating the adjusted survival functions in non-population-based studies. However, alternative methods for calculating adjusted survival functions for complex sample survey data have not been described.

Objectives: 1) Propose and describe two methods for calculating adjusted survival functions in the complex sample survey setting; 2) implement these two methods with SUDAAN software package; and 3) apply these two methods to 2011 National Immunization Survey (NIS) data.

Methods: (1) The inverse probabilities of being in a certain group are defined as the new weights and applied to obtain the inverse probability weighting (IPW) adjusted Kaplan-Meier survival function. (2) Survival functions are evaluated for each of the unique combination of all levels of covariates in a complex sample survey obtained from a single Cox proportional hazards (PH) model, and the weighted average of these individual functions is calculated, with weights equal to the weighted sample size of the individual function, to obtain the Cox corrected group (CCG) adjusted survival function.

Illustrative example: For illustration of the basic techniques rather than a thorough epidemiologic investigation of a specific research question, the two proposed methods were applied to 2011 National Immunization Survey (NIS) data. We estimated the adjusted survival function by age in days of children receiving the first dose of varicella vaccination by children’s family mobility status and by IPW, CCG, and crude Kaplan-Meier (KM) methods controlling for parents’ attitude toward vaccination, mother’s age group, and children first born status.

Conclusions: If the Cox PH assumption is not met, then the IPW adjusted KM method is the only good choice among the two proposed methods, if adjusted survival estimates are desired. If the Cox PH assumption is valid, either the IPW or CCG methods can be used.

Key words: complex sample survey, adjusted survival functions, inverse probability weighting, Cox corrected group, cumulative vaccination coverage, Kaplan-Meier methods.

1. Introduction

In vaccination studies with complex sample survey data, non-parametric Kaplan-Meier (KM) survival functions may be applied to account for time to vaccination in order to estimate cumulative vaccination coverage, assess the timeliness of vaccination, and compare cumulative vaccination coverage between any two levels of selected covariate [1-7]. To accurately estimate vaccination status for these purposes, it is important to develop methods of generating covariate adjusted survival curves which may reduce bias and increase precision when evaluating the effect of a particular "exposure" variable on trends over time. In medical literature, several methods for calculating adjusted survival functions have been proposed. The average covariate adjusted method is frequently used in biomedical papers, which applies the parameter estimates obtained from the Cox proportional hazards regression to the average value of the covariates of interest in the groups being compared [8]. The limitations of this method have been discussed; the major problem is that for categorical covariates, the meaning of the *adjusted* survival for individuals with the average covariate value is quite difficult to explain [9]. The corrected group prognosis method [10-11] was proposed to overcome the limitation of the average covariate adjusted method [12]. This method calculates the survival curve for each unique combination at all levels of the covariates with a Cox model and obtains the adjusted survival curve as a weighted average of those individual curves, in which weights are based on the sample sizes in each combination. Recently an adjusted Kaplan–Meier (KM) estimator using inverse probability of treatment weighting was proposed [13] and it was shown to be a consistent estimate of the survival function. A non-parametric covariate-adjusted survival curve approach was also proposed [14], but this method involves a loss of efficiency especially when the proportional hazards assumption was valid, and demonstrated lower power than the method for generating covariate adjusted survival curves from Cox proportional hazards model. The additive model uses a linear regression model to the adjusted survival function [15], but the additive assumption is not valid in some situations, and the hazard function is not naturally restricted to non-negative values. A direct adjustment method based on the KM survival estimates calculates a weighted average of the strata-specific KM estimates, weighting according to the baseline sample size of the study population in each stratum [16]. However this method produces very similar survival curves to those generated by the original KM method.

Many national public health surveys employ complex sampling schemes, such as the National Immunization Survey (NIS), Behavioral Risk Factor Surveillance System (BRFSS), National Health and Nutrition Examination Survey (NHANES), and the National Health Interview Survey (NHIS). Brogan [17, 18] has discussed the impact of sample survey design on data analysis and has illustrated the possible consequences of ignoring the survey design in analysis of national health survey data. Complex sample surveys are designed to yield population-based estimates and inferences. In the context of complex sample survey, any adjustment procedures need to incorporate the characteristics of complex sample survey designs which typically involve some combination of stratification, multistage sampling, clustering, weighting, and finite population adjustments; otherwise, the estimate and inference could be biased. The covariate adjusted methods described above are intended to be used for non-population-based studies. Alternative methods of calculating covariate adjusted survival functions for complex sample survey data have not been described.

Because the KM product limits estimate and Cox proportional hazards model are two popular procedures in survival analysis, we propose and describe two approaches for calculating covariate adjusted survival functions in the context of complex sample survey: the Inverse Probability Weighting (IPW) adjusted KM survival functions and the Cox Corrected Group (CCG) adjusted survival functions. The two methods are implemented with SUDAAN software package (Research Triangle Institute, Research Triangle Park, North Carolina) using the 2011 National Immunization Survey (NIS) data, a population-based complex sample survey.

2. Methods

2.1. Inverse probability weighting (IPW) adjusted KM survival functions for complex sample survey data.

We assumed that all of the variables, except the event time, considered in a complex sample survey survival data analysis were categorical. Let p_{ik} be the predicted probability for the *ith* individual being in the *kth* group of the complex sample survey data, i.e. the probability of the *ith* individual being in group *k*, which was calculated by use of the Logistic Procedure in SUDAAN [19-21]. These probabilities may depend on the covariate vector Z_i, i.e. $p_{ik} =P(X_i =k|Z_i)$, where X_i is the group index. To reduce the sample bias of different groups, we assigned a new weight $W_{ik} =1/p_{ik}$ for the *ith* individual in group *k*, then applied the new weights W_{ik} to SUDAAN KM procedure to obtain the inverse probability weighting (IPW) adjusted KM survival function for the *kth* group.

2.2. Cox corrected group (CCG) adjusted survival functions for complex sample survey data.

Again, all of the variables, except the event time, were considered categorical. First, the backward-selection method [22-23] was applied to the Cox proportional hazards model for complex sample survey survival data, to obtain the final model which contains the significant variables including the group variable for which the adjusted survival functions was evaluated for each of the levels, and the covariates to be controlled. All of the predictors in the right hand side of the model are assumed to be categorical, and Cox proportional hazard assumption was assumed to be valid for all of the variables. Second, the individual cumulative hazards functions $H(t)$ were obtained for each of the unique combination at all levels of the predictors including the group variable and the covariates in the final Cox model by applying SUDAAN Survival procedure and output the estimated cumulative hazard functions [21]. Third, the estimated individual survival functions $S(t)$ were calculated by $S(t)=Exp\ [-H(t)]$. Fourth, the weighted sample sizes for each of the individual survival functions were calculated. Fifth, when the group variable had *m* levels, all of the individual survival functions were separated into *m* subgroups. Finally, the CCG adjusted survival functions for each of the group level were estimated as a weighted average of those individual survival functions within each of the *m* subgroups with weighs equal to the weighted sample sizes obtained in the fourth step.

3. Illustrative example

In vaccination studies with complex sample survey data, the Inverse probability weighting (IPW) adjusted and Cox corrected group (CCG) adjusted methods may be applied to estimate adjusted cumulative vaccination coverage, assess the timeliness of vaccination, and compare the adjusted cumulative vaccination coverage between any two levels of selected covariate. In this example, we analyzed data from the 2011 National Immunization Survey (NIS) Child data to calculate adjusted cumulative vaccination coverage controlling for the selected socio-demographic factors. The analysis contained in this example was not intended to be a thorough epidemiologic investigation of a specific research hypothesis; rather, it is intended as an illustration of the methodology described in section 2.

The NIS is conducted annually by the Centers for Diseases Control and Prevention (CDC) to provide national, state, and selected urban-area estimates of vaccination coverage among U.S. children aged 19-35 months [24]. The NIS is a stratified clustered random-digit-dialed telephone survey of households with age-eligible children. The NIS landline sample frame was used for this example. Data for 19,534 children who had adequate provider vaccination information were analyzed. In 2011 the NIS landline household response rate based on Council of American Survey and Research Organizations (CASRO) guidelines was 61.5%; and among sampled children with a completed NIS telephone interview, 71.6% had adequate provider-reported vaccination history information.

Adjusted cumulative vaccination coverage curve for the first dose of Varicella.

IPW and CCG methods were applied to generate the adjusted cumulative vaccination coverage curve of children's age in days upon receiving the first dose of varicella vaccination by children's family mobility status (the state of residence at birth is different from current residence state: moved vs. not moved), and controlling three other significant covariates: parents attitude of refusal/delay vaccination (yes vs. no); mother's age group (≤29 years vs. >30 years); and child's first born status (yes vs. no). All four of the independent variables did not meet the Cox PH assumption based on the 2011 NIS child data. For comparison, the unadjusted cumulative vaccination coverage curve was estimated using the crude KM method (using original sampling weights).

Figure 1 presents the cumulative vaccination curves for receipt of the first dose of varicella vaccine by age in days among children whose family moved by the three methods. The unadjusted KM curve fell between the IPW and CCG adjusted curves as presented. Comparison of the cumulative varicella vaccination coverage curves for children whose family moved vs. not-moved by IPW and crude KM methods are shown in Figure 2; mobility not-moved curves of both IPW and unadjusted KM were higher than the corresponding mobility moved curves, as expected. The IPW adjusted curves were positioned between the unadjusted curves, and maintain the shape that was seen in the unadjusted curves. In addition, the adjustment made the curve for moved households closer to the curve for not-moved household, and this movement of curves might be explained as follows: the socio-demographic factors act as confounders, therefore the association of mobility with status of vaccination is attenuated when controlling for these factors via adjusted survival curve. Thus the IPW method in this example generated reasonable adjusted cumulative varicella vaccination coverage curves. However, as presented in Figure 3, the CCG adjusted curves were approximately located outside of the unadjusted curves. The CCG method requires the satisfaction of Cox PH assumption, which was not met in this example.

4. Discussion

The IPW method adjusts for confounding by using the inverse probability weights. It is a non-parametric method and easy to calculate. In addition, the IPW method provides marginal survival function estimates, does not require the validity of the Cox PH assumption which often may not be satisfied, and does not assume any semi-parametric or parametric survival model [13]. Thus, if the Cox PH assumption is not met, the IPW adjusted KM method is the only good choice among the two proposed methods. If the Cox PH assumption is valid, either IPW and CCG adjusted methods can be used, or the two methods could be used in combination (e.g., IPW as the primary method and CCG for subsequent adjustment). The Cox PH model is a "robust" model, reasonable estimates of adjusted survival curves can be obtained for a wide variety of data situations, and the results from using the Cox model will closely approximate the results from the correct parametric model [8]. The CCG is also a flexible tool for adjusting important covariates [14].

In practice, we recommend presenting the unadjusted survival curves first. The objectives of the study will determine if adjusted survival curves are needed. For example, in a study of disparities by race/ethnicity, the unadjusted curves are most important and need to be shown first. If researchers want to explain the disparity in terms of causal factors, the adjusted survival curves may be useful.

In the illustrative example presented in this report, the Cox proportion hazards assumption was not met for all the variables for the varicella vaccination data from 2011 NIS, so the IPW method is the only appropriate approach among the two proposed methods for adjusted survival functions. This example illustrates the two methods and provides only one of the many situations that may be encountered in complex sample survey survival data.

One limitation to these two proposed methods for calculating adjusted survival functions with complex sample survey data is the assumption that all variables considered in the analysis are categorical. However, if an important continuous covariate is necessary for inclusion in the adjusted survival analysis, one may categorize that covariate. This study is a statistical practice report that proposes and describes two methods for calculating adjusted survival functions in the context of complex sample survey data using the 2011 NIS and procedures in SUDAAN v11. Future theoretical study, comprehensive simulation studies and other illustrative examples are needed.

5. References

[1]. Blank PR, et al. Population access to new vaccines in European countries. Vaccine 31 (2013) 2862–2867.

[2]. Lu PJ, et al. Meningococcal conjugate vaccination among adolescents aged 13-17 years, United States, 2007. Vaccine 28 (2010) 2350-2355.

[3]. Strenga A, et al. Varicella vaccination coverage in Bavaria (Germany) after general vaccine recommendation in 2004. Vaccine 28 (2010) 5738–5745.

[4]. Clark A., Sanderson C. Timing of children's vaccinations in 45 low-income and middle-income countries: an analysis of survey data. Lancet. Vol. 373 May 2, 2009.

[6]. Akmatov MK, et al. Timeliness of vaccination and its effects on fraction of vaccinated population. Vaccine 26 (2008) 3805–3811.

[6]. Dayan GH, et al. Assessment of Delay in Age-appropriate Vaccination Using Survival Analysis. Am J Epidemiol (2006);163:561–570.

[7]. Peter A, et al. Vaccinia scars associated with better survival for adults. An observational study from Guinea-Bissa. Vaccine 24 (2006) 5718–5725.

[8]. Kleinbaum DG. Survival analysis –A Self-learning text, 3rd ed. 2012. Springer New York, NY.

[9]. Grouven U, Bender R, Schultz A, et al. Application of adjusted survival curves to renal transplant data. Methods Inf Med 1992;31:210-14.

[10]. Makuch RW. Adjusted survival curve estimation using covariates. J Chronic Dis 1982;35:437-43.

[11]. Chang IM, Gelman R, Pagano M. Corrected group prognostic curves and summary statistics. J Chronic Dis 1982;35:668-74.

[12]. Ghali WA et al, Comparison of 2 methods for calculating adjusted survival curves from proportional hazards models. JAMA, 26, 2001: 286, 1494-97.

[13]. Xie J. et al, Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. Statist. Med. 2005; 24:3089–3110.

[14]. Jiang H, et al. Covariate-adjusted non-parametric survival curve estimation. Stat. Med. 2011, 30: 1243–1253.

[15]. Aalen O. A linear regression model for the analysis of life times. Stat. Med. 1989, 8: 907-925. (1989).

[16]. Cupples LA, Gagnon DR, Ramaswamy R, et al. Age-adjusted survival curves with application in the Framingham Study. Stat Med 1995; 14:1731-44.

[17]. Brogan D. Software for sample survey data: misuse of standard packages. In: Armitage P, Colton T, eds. Encyclopedia of Biostatistics. 2nd ed. Chichester, United Kingdom: JohnWiley & Sons Ltd; 2005:5057–5064.

[18]. Brogan D. Sampling error estimation for survey data. (Chapter XXI and annex). In: Yansaneh IS, Kalton G, eds. Household Sample Surveys in Developing and Transition Countries. (Studies in methods, series F, no. 96). New York, NY: United Nations; 2005: 447–490. (http://unstats.un.org/unsd/HHsurveys/ pdf/Chapter_21.pdf, http://unstats.un.org/unsd/HHsurveys/pdf/ Annex_CD-Rom.pdf). (Accessed March 1, 2009).

[19]. Graubard BI, Korn EL. Predictive margins with survey data. Biometrics. 1999;55(2):652–659.

[20]. Korn E, Graubard B. Analysis of Health Surveys. New York, NY: John Wiley and Sons, Inc; 1999.

[21] Research Triangle Institute (2012). SUDAAN Language Manual, Release 11.0 Research Triangle Park, NC: Research Triangle Institute.

[22] Hosmer D and Lemeshow S, "Applied Survival Analysis: Regression Modeling of Time to Event Data. Vol. 1," John Wiley & Sons, New York, 1999.

[23] Thabut G, Christic JD, Kremers WK, Fourbier M and Halpern SD, "Survival Differences Following Lung Transplantation among US Transplant Centers," Journal of the American Medical Association, Vol. 304, No. 1, 2010, pp. 53-60.

[24]. Centers for Disease Control and Prevention. National, State, and Local Area Vaccination Coverage among Children Aged 19-35 Months – United States, 2011. September 7, 2012. MMWR, 61(35);689-696.

Figure 1. The first dose of varicella vaccination coverage by age in days among children whose family moved by the three methods, 2011 NIS-Child.

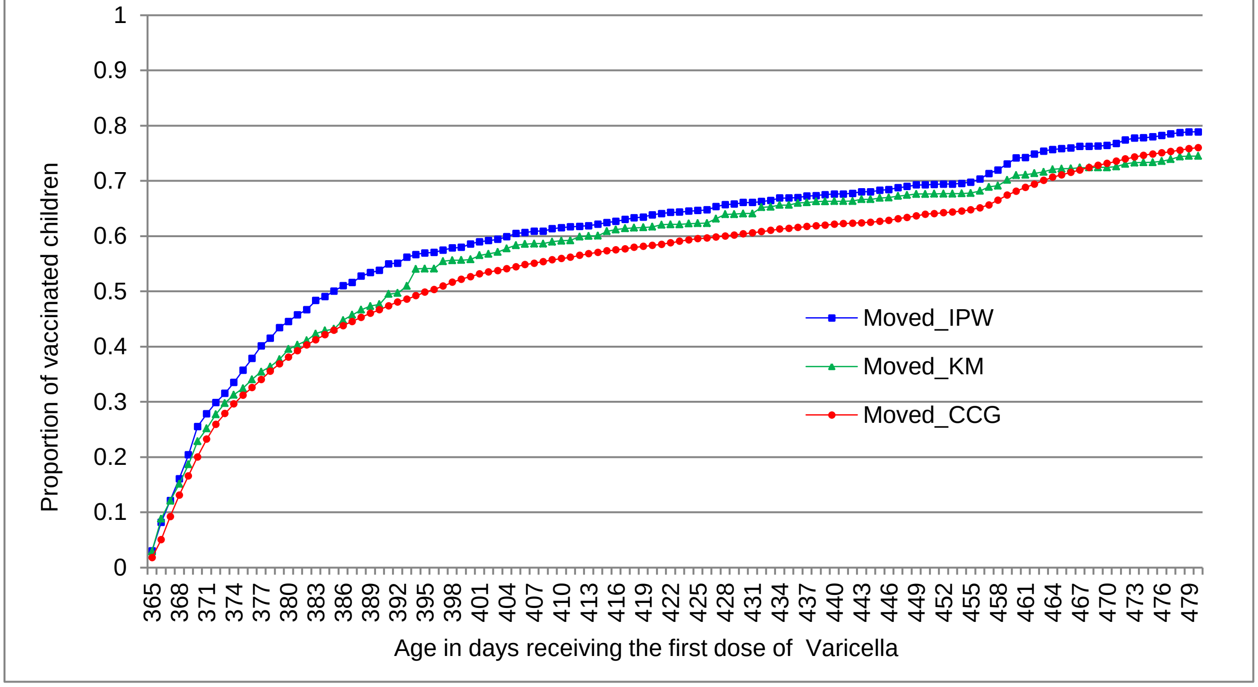


Figure 2. Comparison of the first dose of varicella vaccination coverage for children whose family moved vs. not-moved by IPW and crude KM methods, 2011 NIS-Child.

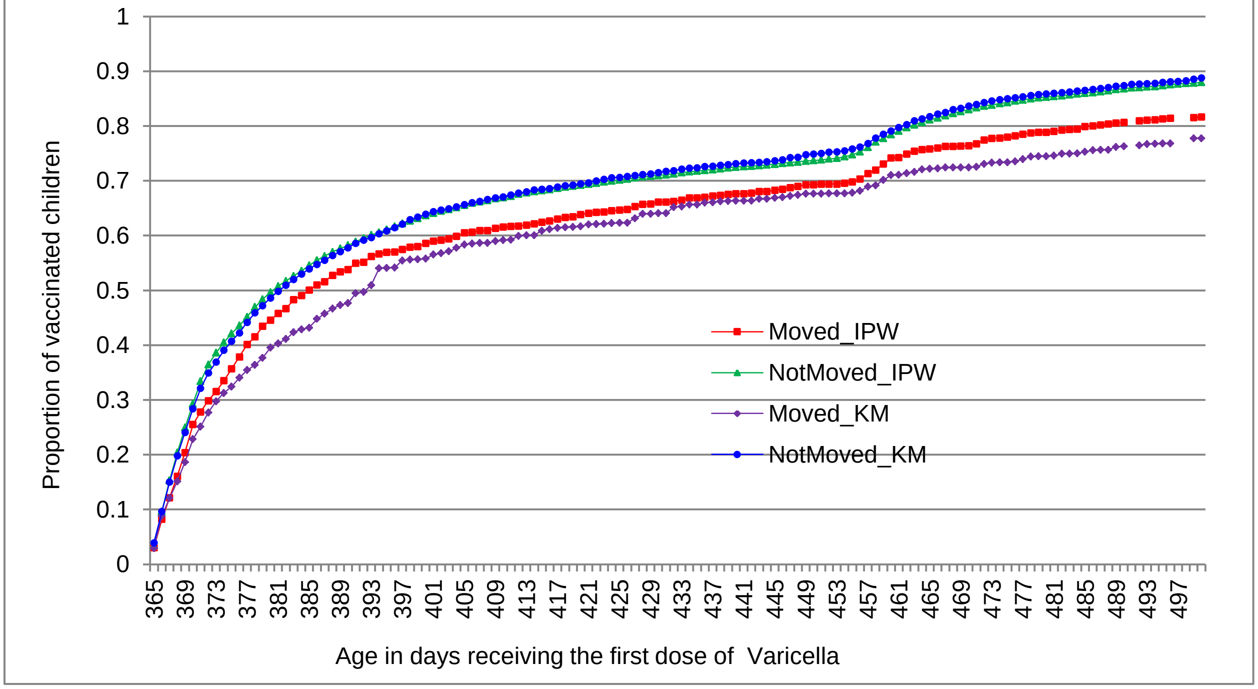
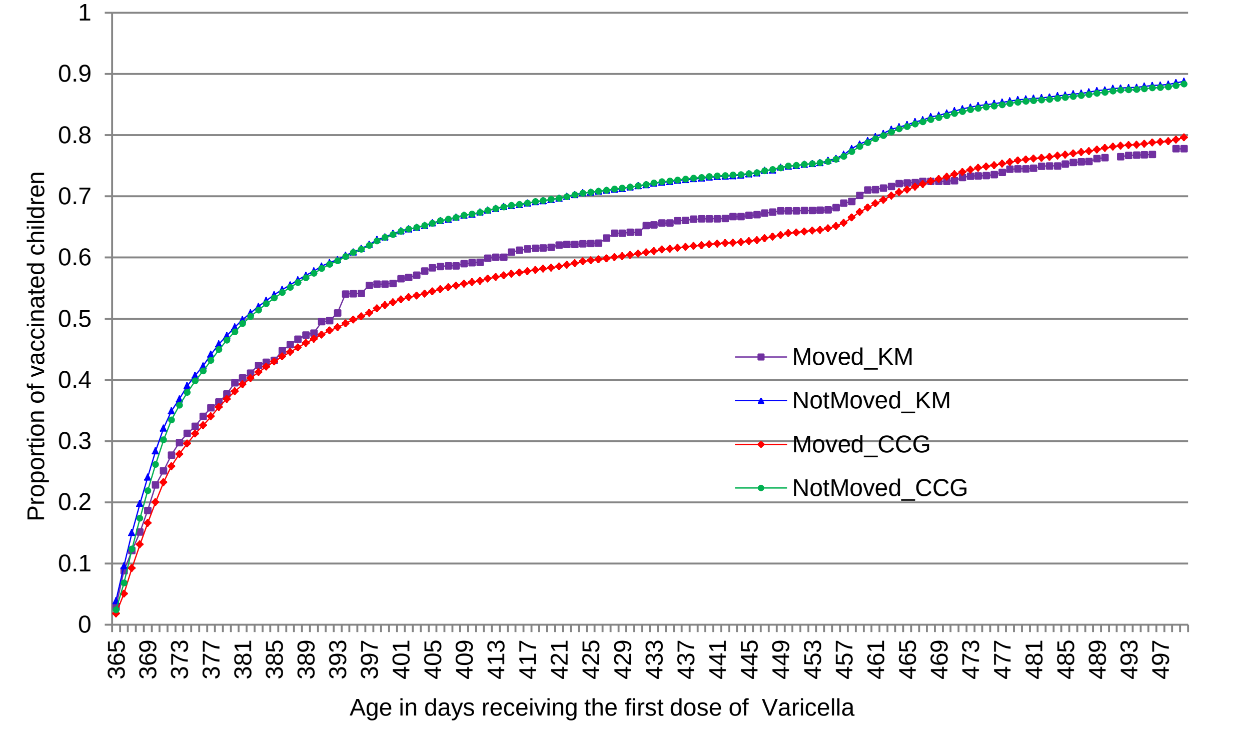


Figure 3. Comparison of the first dose of varicella vaccination coverage for children whose family moved vs. not-moved by CCG and crude KM methods, 2011 NIS-Child.


Proportion of vaccinated children
Age in days receiving the first dose of Varicella
Moved_KM
NotMoved_KM
Moved_CCG
NotMoved_CCG