universiteit hasselt

Homework 3

# Inverse probability in statistics and estimation of rare events

**Taught by**
**Pr. Dr. Geert Molenberghs**

By
Stijn Van Herck
Thomas Bullen
Tom Louwies
Soutrik Banerjee

*Master of Science in Biostatistics:1st year*
*2010 – 2011*

# Using inverse probability weighting in the analysis of survey data

## Introduction

The first of two topics we will discuss in this third assignment is the use of inverse probability weighting when analysing survey data. If one performs a survey one can encounter missing data as a result of for instance the non-response of interviewees on certain studied variables. Furthermore certain missing observations can also be induced by for instance the survey design. The occurrence of missing data in a survey is not a rare event. Almost all substantive data sets collected for social research contain missing data. One way to deal with this missing data is of course to restrict the analysis to individuals with no missing values. This is the easiest way to deal with the problem, although leaving out the units with partial observations and restrict the analysis to the remainder complete cases can induce bias. A commonly used method to handle this bias is the inverse probability weighting (IPW) method. By implementing this method missing values can be predicted, using complete data of other observations, to account for the non-response in the overall survey results. Another purpose of the IPW is that it adjusts for unequal sample fractions in sample surveys. The use of IPW to deal with missing values however has not always been a success. For many years, since the introduction of IPW in 1952, the method gained little acceptance in the missing data literature because of its imprecision in comparison with more efficient missing data methods, such as the multiple imputation method. The last decade however, there were major advancements in determining the inverse probability weighting estimators which made this method more useful and efficient (Vansteelandt *et al.*, 2009).

In the following paragraph the IPW method will be explained in more detail. Afterwards an example will be given to clarify this method in practice. Finally a conclusion will give an overview of the most important findings of the IPW method and its main advantages and disadvantages.

## Method

When outcomes are missing for some subjects in a random sample, it's possible to use inverse probability weighting, whereby missing values can be predicted with the values from observations with complete data.

Most standard analyses are restricted to subjects with complete data, but these subjects may form an unrepresentative group from whom biased conclusions may be obtained. This is because these subjects with complete data form a selective subgroup and thus the sample averages of their outcomes may systematically over- or underestimate the population mean.

The selection bias can however be corrected by weighting each responder's data by the reciprocal of the probability $\pi_i$ having observed outcome data $Y_i$ on the basis of his/her background characteristics $X_i$. In particular, having estimated the probabilities $\pi_i$ given the background characteristics $X_i$ there can be calculated:

$$\hat{\mu}_{IPW} = \frac{\sum_{i=1}^{n} R_i Y_i / \pi_i}{\sum_{i=1}^{n} R_i / \pi_i},$$

Where $R_iY_i$ is '$R_i$ x $Y_i$' and so is defined as zero whenever $R_i$ is zero. This estimator is called an inverse probability weighted (IPW) estimator. It is a variant of the Horvitz-Thompson estimator, which was introduced in the context of survey sampling in finite populations. In practice the IPW estimator is easily obtained with standard software. Important in the IPW is the need for suitable estimates of the probabilities. In addition the sample size must be big enough to give a good result. IPW estimators can behave very badly in examples where a few individuals receive a very large weight. This is likely to happen when measured background characteristics are strongly predictive of missing data in the outcome. In practice for these situations there are some other methods that can be used: imputation estimators, doubly robust estimators, etc. (Vansteelandt *et al.*, 2009).

## Example

In this paragraph an example of the IPW is given. The example is about the highly active antiretroviral therapy (HAART) that improves the acquired immunodeficiency syndrome (AIDS)-free survival of human immunodeficiency virus (HIV)-infected patients compared with no use of HAART.

Appropriate analysis of observational data is the best chance to obtain answers to many questions that involve dynamic treatment regimes. IPW can be similarly applied to compare either non-dynamic or dynamic treatments.

To decide for example whether it is better to start HAART therapy at CD4 cell count below 500 cells/µl or at 200 cells/µl artificial censoring/IPW can be applied to an observational cohort. Treatment A was 1 if the patient started HAART within one month of baseline (i.e. their first CD4 cell count below 500 cells/µl), and 0 otherwise.

Among patients in the group A=0 with CD4 cell count still greater than 200 cells/ml at time t after entry, those who initiate treatment and thus become censored at t will likely have (on average) lower CD4 cell counts than those who do not initiate and remain uncensored at time t. If so, individuals censored at t will have a worse prognosis than those who remain uncensored. Similarly, among those with a CD4 cell count less than 200 cells/ml for the first time, those who did not start HAART and thus become censored likely have, on average, a higher CD4 cell count than those who did start HAART and remained uncensored. This time-dependent selection bias in the artificially censored data may be fully adjusted for by IPW only if all time-varying risk factors for mortality that predict HAART initiation in the A=0 group were recorded (e.g. time since diagnosis of HIV-infection, sex, age, risk group, etc.). In this case the HAART regime is taken as an example, of course the IPW method can be applied to compare any treatment regimens using observational data (Hernan *et al.*, 2006).

## Conclusion

The above paragraphs give an overview of the use of the inverse probability weighting method in the analysis of survey data. IPW is a method for correcting for missing value

mechanisms that are not strictly MCAR (missing completely at random), that uses information about the missingness probabilities themselves. Although the use of IPW method for dealing with missing data wasn't always a success, the advancements made regarding this method brought along some big advantages. The major advantage is of course that the use of IPW removes bias because partial observations are taken into account instead of removing them out of the analysis. Another advantage of IPW is that it is robust for estimating approximate models, which is an important feature for causal analysis. Another major advantage of the IPW method is its simplicity because current software is available which makes the implementation of IPW easy. As we have seen in our example inverse probability-weighted estimators also allow researchers to deal with another issue common in observational data, namely artificial censoring. Next to the advantages there are some disadvantages of the IPW method. For example the use of IPW for dealing with missing data is inefficient compared to a model based approach. Another major drawback is that the use of IPW is also sensitive to the choice of weighting model. IPW estimators can for example behave poorly in examples where a few individuals receive major weights (Carpenter et al., 2006).
Overall we can conclude that the IPW method is a useful tool for dealing with missing values and observations. Although there are nowadays more efficient techniques of dealing with missing data such as the imputation-based 'augmented' inverse probability weighting (AIPW) method (Samii, 2011), the IPW method is rather preferred due to its simplicity when dealing with missing data.

### References

Carpenter J.R. & Kenward M.G. (2006). *A comparison of multiple imputation and inverse probability weighting for analyses with missing data.* London: London School of Hygiene & Tropical Medicine.
Hernan, M.A., Lanoy, E., Castagliola, D. & Robins, J.M. (2006). *Comparison of Dynamic Treatment Regimes via Inverse probability Weighting*. Basic & Clinical Pharmacology & Toxicology 2006, 98, 237–242.
Samii, C. (2011). *Weighting and Augmented Weighting for Causal Inference with Missing Data: New Directions.* Columbia: Columbia University.
Vansteelandt, S., Carpenter J. & Kenward M.G. (2009). *Analysis of incomplete data using inverse probability weighting and doubly robust estimators.* Ghent: Ghent University.

## Estimation of rare events: rare diseases prevalence and rare adverse events

Measuring rare diseases prevalence or rare adverse events (AE) detection has always posed challenges in choosing the most suitable method. Questions are raised if the study is well powered or the duration of the follow-up is long enough to capture rare diseases or rare safety signals. Although considerable work is done in the past few decades, no benchmark method exists until now and new methods are continually discovered. In the paragraphs below both absolute (prevalence rates) and relative (prevalence odds) approaches are described.

## *Prevalence of rare diseases*

**Standardisation**: Direct and indirect standardisation enables to standardise prevalence (or incidence) rates typically by age categories and gender. Direct standardisation employs the principle that the stratum-specific rates from the study population are multiplied by the stratum-specific weights obtained from a standard population and the total is summed to derive the *overall* rate. However this method does not work well for (often rare) diseases where the rates obtained from the study population are not stable due to very small population (denominator) size among certain strata. The advantage of direct standardisation is that information about the rates in the standard population is not required for its calculation. To counteract the problem of unstable rates, indirect standardisation is used. Here the principle is reversed; that is a standard population's stratum-specific rates are multiplied by the stratum-specific weights of the study population and the total is summed to derive the *expected* rate. In practice with indirect standardisation, the expected rates (or counts) by stratum and overall are obtained from a standard population, and then the observed rates (or counts) by stratum and overall are divided by the expected rates (or counts) by stratum and overall, respectively, in order to get the standardised mortality ratio (SMR) or standardised incidence ratio (SIR) with a null hypothesis value of 100 or 1.

**Logistic (or Poisson for rates) regression** (see King, 2001): This provides certain advantages over (direct) standardisation in that the precision of estimates are better [Roalfe *et al.*, 2008]. Further, logistic (or Poisson for rates) regression allows to adjust for more variables (other than age and gender), which can be continuous, allows clustering effect and interactions to be taken into account, and can perform a weighted analysis in case of missing data. Non-significant variables in the model indicate that stratification would not be necessary for these variables. ***Conditional*** (on the matched-pairs) ***logistic regression*** is another alternative for matched (1:*k*) case-control study [Agresti, 2002] although there is a loss of precision due to sampling from the available population.

**Time-to event analysis**: Case-control studies can be viewed as a nested follow-up study. This implies that the estimated odds ratios although provide a valid measure of risk ratios, but have less precision than those obtained from a follow-up study, because the study population becomes a sample from the entire population at risk [Rothman, 2002, see pp. 77-78]. A precision gain can be obtained from Cox's proportional hazards regression, where the denominator is the entire cohort population. As the statistical power in this case increases only with the numerator (*i.e.*, number of events), by including the entire population as the denominator indirectly increases the chance of finding more rare events, consequently more gain in statistical power, than that obtained by selecting only a sample as in a case-control design. The hazard ratios thus obtained have greater precision and provide valid estimates of risk ratios.

**Disease mapping**: Although standardisation methods are currently seldom applied in most of the disciplines in statistics, in the field of disease mapping this is still widely used due to the fact that the observational units are not individual subjects, but geographical areas with aggregated population characteristics like age and gender. Although indirect standardisation is a nice way to overcome the difficulties as mentioned earlier, often rates from the standard population are not available, which are thus replaced by the so-called 'internal' indirect standardisation method to compute the crude standardised rates. Use of

***Bayesian hierarchical models*** by assuming exchangeability circumvents the problem of sparse data in certain geographical areas by smoothing or 'filling-the-gaps'; here the areas with sparse data borrows strength from the other areas to provide more stable and uniform confidence intervals for the geographical areas [Best *et al.*, 2005] in model-based approaches at the cost of shrinkage of the posterior estimates towards the prior means.

**Meta-analysis**: This is a nice method to pool the results of several studies (and gain in statistical power, especially in case of sparse data), which may be homogeneous or heterogeneous in characteristics, by applying weights inversely proportional to the within-study variance (± between-study variance). Fixed-effects meta-analysis by Mantel and Haenszel's method [Mantel & Haenszel, 1959] or Peto's method [Danesh *et al.*, 1998] and Random-effects meta-analysis [DerSimonian & Laird, 1986] approaches provide estimates by assuming without or with between-study variability ($\tau^2$) in computing the odds ratio$_{meta}$, respectively. This also applies for the subsection below on rare AEs.

**Confidence intervals of proportions**: When the rates are close to zero, there have been debates and discussions on which method would be the most appropriate. Historically, Clopper-Pearson method is used in many instances [Agresti, 2002; Clopper & Pearson, 1934]; a comparison of 20 methods are discussed by Pires and Amado [Pires & Amado, 2008].

**Databases**: One strategy to overcome the sparse data in clinical trials is the use of large databases (*e.g.*, PHARMetrics, Disease Analyser, GPRD, etc.), which follows patients for a longer duration than on average during a clinical trial usually having a larger number of subjects (sample size) in a 'real-world' scenario, although issues regarding the validity of the data, data quality, missing data, and duplicates are common.

## *Detection of rare adverse events*

In the pharmacovigilance or drug safety domain, AEs or safety signals detection (also known as adverse drug reactions (ADR)), new research methods are underway since the past two decades. Typically used methods for mining the WHO adverse event reporting system (AERS) databases in detecting drug-event associations (DEA) in the form of a 2 x 2 contingency table design (see below) are: the reporting odds ratios (ROR), proportional reporting ratio (PRR), and Bayesian confidence propagation neural network (BCPNN) derived information component (IC), which appears to be the most popularly used methods [for details, see Hauben & Zhou, 2003; Hauben *et al.*, 2005; Purcell & Barty, 2002].

| | Specific drug | All other drugs |
|---|---|---|
| Specific adverse event | A | B |
| All other adverse events | C | D |

## *Conclusion*

Amongst the wide variety of available methods, the scope for new methods is still quite open in this domain [Chan & Hauben, 2005; Evans 2000]. New research methods or algorithms, for instance those borrowed from the field of data mining, are welcome or needs to be tested.

# References


1. Agresti A. *Categorical Data Analysis*, 2nd ed., Wiley & Sons, 2002
2. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res* 2005;**14**:35-59 (doi: 10.1191/0962280205sm388oa)
3. Chan KA, Hauben M. Signal detection in Pharmacovigilance: empirical evaluation of data mining tools. *Pharmacoepidemiology Drug Safety* 2005;**14**:597-9
4. Clopper C, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934;**26**:404-13 (doi:10.1093/biomet/26.4.404)
5. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177-87
6. Evans SJW. Pharmacovigilance: a science or fielding emergencies? *Statist Med* 2000;**19**:3199-209
7. King G. Logistic regression in rare events data. *Pol Analysis* 2001;**9**:137-63
8. Mantel N, Haenszel W. Statistical aspects of the analysis of data from the retrospective analysis of disease. *J Nat Cancer Institut* 1959:**22**:719-48
9. Hauben M, Zhou X. Quantitative methods in Pharmacovigilance. *Drug Safety* 2003;**26**:159-86
10. Hauben M, Madigan D, Gerrits CM, Walsh L, van Puijenbroek EP. The role of data mining in pharmacovigilance. *Expert Opin Drug Saf* 2005;**4**:929-48
11. Danesh J, Collins R, Appleby P, Peto R. Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies. *JAMA* 1998;**279**:1477-82
12. Pires AM, Amado C. Interval estimators for a binomial proportion: comparison of twenty methods. *Revstat* 2008;**6**:165-97
13. Purcell P, Barty S. Statistical techniques for signal generation: the Australian experience. *Drug Safety* 2002;**25**:415-21
14. Roalfe AK, Holder RL, Wilson S. Standardisation of rates using logistic regression: a comparison with the direct method. *BMC Health Services Research* 2008, **8**:275 (doi:10.1186/1472-6963-8-275)
15. Rothman K. *Epidemiology: an introduction*, Oxford University Press, USA, 2002