Marginal and conditional approaches to confounder control

Niels Keiding
January 2007

Overview

Direct and indirect standardisation

Clayton IBC Freiburg July 2002

- Conditional approach ~ indirect standardisation ~ stratification and regression
- Marginal approach ~ direct standardisation ~ randomisation

Potential outcomes
No unmeasured confounders
Inverse probability weighting

Direct and indirect standardization

k age groups

	Age distribution	Death rates
Standard population	$s_1, \cdots, s_k$	$\lambda_1, \cdots, \lambda_k$
Study population	$a_1, \cdots, a_k$	$\alpha_1, \cdots, \alpha_k$

No. of deaths in standard population $\sum s_i \lambda_i$

No. of deaths in study population $\sum a_i \alpha_i$

Expected number of deaths in study population if standard death rates applied: $\sum a_i \lambda_i$

$$SMR = \frac{\sum a_i \alpha_i}{\sum a_i \lambda_i}$$ Indirect standardization

SMR: unnecessary to know age distribution in standard population, death rates enough

Expected number of deaths in standard population if study population death rates applied: $\sum s_i \alpha_i$

$$SRR = CMF = \frac{\sum a_i \alpha_i / \sum a_i}{\sum s_i \alpha_i / \sum s_i}$$ Direct standardization

SRR: unnecessary to know age distribution in study population.

Clayton IBC Freiburg July 2002

- There are two paradigms for control for extraneous influences in experimentation:

  1. Hold all other relevant factors constant, or
  2. Randomly allocate subjects to treatments

- Statistical approaches to control for confounding mirror these two paradigms and likewise fall into two main approaches:

  – Classically these are represented by “indirect” and “direct” methods of standardization
  – In the modern model-based view of statistics these correspond with “conditional” or “marginal” modelling approaches

Conditional approach

- Model exposure effects by contrasting responses conditional upon values of confounding variables
- In general this approach requires constancy of effect (no effect modification)
- But if this and *no unmeasured confounders* hold, then the results are generalizable to other populations with different distributions of confounders
- Nowadays this is done in regression models (logistic, Poisson, Cox, $\cdots$) with regression coefficient as effect measure

SMR and contemporary statistical models

k age groups, death intensity in age group i:

in standard population λ_i

in study population α_i

Assume $\alpha_i = \theta\lambda_i$ (proportional hazards assumption!)

Assume individual j observed from age u_j to age t_j; let T_{ij} be the time lived by j in age group i, let $D_{ij} = 1$ if j died while in age group i. If λ_i is known the likelihood

$$L(\theta) = \prod_i \prod_j (\theta\lambda_i)^{D_{ij}} e^{-\theta\lambda_i T_{ij}} \cdot \text{factor}(\lambda_i) \quad \sim \quad \theta^D e^{-\theta A}$$

$D = \sum_i \sum_j D_{ij}$ = number of deaths in study population

$A = \sum_i \sum_j \lambda_i T_{ij}$ = "expected" number of deaths in study populatio

$$\hat{\theta} = D/A \qquad \widehat{\text{Var}}(\hat{\theta}) = D/A^2$$

Two study populations

same standard population with known death rates $\lambda_1, \cdots, \lambda_k$

$\alpha_{1i} = \theta_1 \lambda_i$ $\qquad$ $\alpha_{2i} = \theta_2 \lambda_i$

$\hat{\theta}_1 = SMR_1$ $\qquad$ $\hat{\theta}_2 = SMR_2$

$\left(\frac{\hat{\theta}_1}{\theta_2}\right) = \frac{\hat{\theta}_1}{\hat{\theta}_2} = \frac{SMR_1}{SMR_2}$ estimates relative mortality

$$\alpha_{1i} = \frac{\theta_1}{\theta_2} \alpha_{2i}$$

but this is *not* the SMR of study pop_1 wrt study pop_2 (Yule).

$\theta_1 = \theta_2$ test: ordinary χ^2 or likelihood ratio.

k study populations; standard mortality unknown: similar; ordinary statistical analysis.

Regression models

If the study population is partitioned into sub-groups, the k-sample approach may be used.
Alternatively, assume that for each person j in the study group there are covariates $z_{j1}, \cdots, z_{jp}$ (sex, income, blood pressure, urbanization, $\cdots$) and that the death intensity

$$\begin{aligned} \alpha_i(z_{j1}, \cdots, z_{jp}) &= \theta_1^{z_{j1}} \cdots \theta_p^{z_{jp}} \lambda_i \\ &= e^{\beta_1 z_{j1} + \cdots + \beta_p z_{jp}} \lambda_i \end{aligned}$$

The *regression coefficients* $\beta_1, \cdots, \beta_p$ express age-standardized versions of the effects of the covariates on mortality.

If λ_i known: standardized by known (e.g. population) mortality

If λ_i unknown: internally standardized

"Poisson" regression model, analogies to Cox regression model.

Marginal effects under simulated randomization

- Randomization ensures equal distribution of potential confounders between treatment (exposure) groups. Effects of treatment are then measured by contrasting the *marginal* distributions of response

- "Direct" standardization by age simulates such an experiment by fixing the distribution of the confounder (age) to be equal across exposure groups. We then compare marginal measures of response between exposure groups

- This measure of effect is less dependent on modelling assumptions but is less generalizable to other populations with different distributions of confounders

Potential outcomes

$Y_i^{(x)}$ outcome if treatment x followed

$Y_i^{(0)}$ outcome if treatment $x = 0$ followed

$Y_i^{(1)}$ outcome if treatment $x = 1$ followed

Outcomes other than that observed are called *counterfactual*

Causal effect of **1** vs. **0**

$$Y_i^{(1)} - Y_i^{(0)}$$

cannot be observed directly since we only observe one potential outcome per subject.

Estimation of causal effects in randomised studies

average response of treated subjects $E\left(Y_i \mid x=1\right)$

is an unbiased estimate of mean of $Y_i^{(\mathbf{1})}$ in entire population

and similarly for untreated subjects

$$E\left(Y_i \mid x=0\right)$$

estimates mean of $Y_i^{(\mathbf{0})}$ in entire population

Causal effect $\delta = \mu^{(1)} - \mu^{(0)}$ $\mu^{(x)} = E\left(Y^{(x)}\right)$

Covariates z

Causal effect $\delta(z) = \mu^{(1)}(z) - \mu^{(0)}(z)$ $\mu^{(x)} = E\left(Y^{(x)} \mid Z=z\right)$

Estimation of causal effects in observational studies

We assume that we have recorded so much confounder information that the study can be considered essentially randomised given these confounders

No unmeasured confounders (= ignorable treatment (or exposure) assignment)

Assignment of treatment X is conditionally independent of potential outcomes (r_T, r_C) given covariates z.

Propensity score and inverse probability weighting

Propensity score $e(z) = P(x = 1|z)$

the probability of being assigned to treatment 1 given covariates z.

Causal effect from observational study:

$$\frac{XY^{(1)}}{e(z)} - \frac{(1-X)Y^{(0)}}{1-e(z)}$$

is unbiased estimate of treatment effect $\delta = \mu^1 - \mu^0$.

Note. $XY^{(1)}$ and $(1-X)Y^{(0)}$ are always observable (often = 0).

Note. Response weighted with inverse probability of treatment assignment.

Causal effect from inverse probability weighting: proof

$$E\left(\frac{XY^{(1)}}{e(z)} - \frac{(1-X)Y^{(0)}}{1-e(z)}\middle| z\right) = \mu^{(1)}(z) - \mu^{(0)}(z).$$

Proof $\quad E\left(XY^{(1)}\middle| z, X\right) = X\, E\left(Y^{(1)}\middle| z, X\right)$

$$= X\, E\left(Y^{(1)}\middle| z\right) \quad \text{because of no unmeasured confounders}$$

$$\Rightarrow \quad E\left(XY^{(1)}\middle| z\right) = E\left(X\middle| z\right)E\left(Y^{(1)}\middle| z\right) = e(z)\,\mu^{(1)}(z).$$

Example: one-dimensional categorical confounder

Confounder stratum	Exposed subjects		Unexposed subjects	
	Number	Events	Number	Events
⋮				
i	n_{i1}	d_{i1}	n_{i0}	d_{i0}
⋮				

- Estimate of $\Pr\left(\text{response observed} \,\middle|\, \text{confounder}\right)$ is n_{i1}/n_{i+} or n_{i0}/n_{i+}
- Correcting for unobserved counterfactual responses by inverse probability weighting, the probability of event given exposure is estimated by

$$\frac{\sum_i \left(n_i/n_{i1}\right)d_{i1}}{\sum_i \left(n_i/n_{i1}\right)n_{i1}} = \frac{\sum_i n_{i+}\left(d_{i1}/n_{i1}\right)}{\sum_i n_+} = \text{“Direct” standardized measure of risk}$$