Published in final edited form as:

Biometrics. 2011 March ; 67(1): 319–323. doi:10.1111/j.1541-0420.2010.01451_1.x.

Estimation and Inference for the Causal Effect of Receiving Treatment on a Multinomial Outcome: An Alternative Approach

Stuart G. Baker

Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, EPN 3131, 6130 Executive Blvd MSC 7354, Bethesda, MD 20892-7354, USA.

SUMMARY

Recently Cheng (Biometrics, 2009) proposed a model for the causal effect of receiving treatment when there is all-or-none compliance in one randomization group, with maximum likelihood estimation based on convex programming. We discuss an alternative approach that involves a model for all-or-none compliance in two randomization groups and estimation via a perfect fit or an EM algorithm for count data. We believe this approach is easier to implement, which would facilitate the reproduction of calculations.

Keywords

All-or-none compliance; Causal effect; Multinomial outcomes; Noncompliance; Perfect fit; Principal stratification; Randomized trials

1. Introduction

Cheng (2009) proposed a causal analysis of a randomized trial to investigate the effect of seeing a specialist on the development of depression symptoms among elderly patients. The role of the specialist was to increase adherence to the depression treatments via education and assessment. Twenty primary care practices were randomized to either seeing a specialist or usual care which involved no access to the specialist. For simplicity, group randomization was ignored, and randomization was considered at the level of the individual. An important outcome was depression symptoms (none, minor, and major) at four months. Some subjects randomized to seeing a specialist did not see the specialist and so essentially received usual care. To analyze these data, Cheng (2009) specified a causal model with latent compliance classes for multinomial outcomes with all-or-none compliance in one randomization group. Cheng (2009) computed maximum likelihood (ML) estimates using a grid search and a convex programming algorithm. For a more general causal model with all-or-none compliance in two groups, we propose ML estimation via a perfect fit estimate or a simple EM algorithm for count data. Our estimation approach is simple, which is helpful if others wish to reproduce the results. Reproducibility of research is an important desideratum in statistical analyses (Peng, 2009).

Supplementary materials

Computer code in Mathematica 7.0 (Wolfram Research, Inc., 2008) to reproduce these results is available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org>.

2. Model

All-or-none-compliance, sometimes called all-or-none switching of interventions, has a precise meaning with important implications for the formulation of a causal model (Baker, 1997, Baker and Kramer 2005). Suppose subjects are randomly assigned to either T0 or T1, which denote different programs of interventions starting at randomization. All-or-none compliance in one randomization group means that all subjects randomized to T0 receive T0, and some subjects randomized to T1 receive T0 and some receive T1. All-or-none compliance in two randomization groups means that in each randomization group some subjects receive T0 and some receive T1. Under all-or-none compliance, one can use latent compliance classes, randomization, and reasonable assumptions to estimate the causal effect of receiving intervention (e.g. Baker and Kramer, 2005). In a more general setting, latent compliance classes are called principal strata (Frangakis and Rubin, 2002).

Models for latent compliance classes with all-or-none compliance in one group were formulated by Bloom (1984) for continuous outcomes and Sommer and Zeger (1991) and Connor et al (1991) for binary outcomes. Similar models for latent compliance classes with all-or-none compliance in two groups were formulated by Baker and Lindeman (1994) for binary outcomes (with before-and-after time periods treated like randomization groups) and Angrist Imbens and Rubin (1996) for continuous outcomes.

We generalize the model of Cheng et al (2009) to multinomial outcomes with all-or-none compliance in two groups instead of one. Let r index randomization group with $r = 0$ for assignment to intervention T0 and $r = 1$ for assignment to intervention T1. Let a index intervention received, where $a = 0$ for receipt of T0 and $a = 1$ for receipt of T1. Let $j = 1, 2, \dots, J$ index multinomial outcomes. The counts are denoted

n_{raj} = number of persons in randomization group r who receive intervention a and have outcome j .

In the terminology of Angrist et al (1996), there are four latent compliance classes:

- A , always-takers, who would receive intervention T1 if randomized to either group;
- C , compliers, who would receive the intervention to which they were randomized;
- N , never-takers, who would receive intervention T0 if randomized to either group;
- D , defiers, who would receive the opposite intervention to which they were randomized.

Following Cheng (2009), we make the same assumptions listed in Angrist et al (2006) and described below without mathematical notation.

ASSUMPTION 1: *Stable unit treatment value assumption. The outcome for a subject is unaffected by the particular assignments of treatments to the other subjects (Rubin, 1980).*

ASSUMPTION 2: *Random assignment to randomization groups.*

ASSUMPTION 3: *Exclusion restriction. For always-takers and never-takers, the distribution of outcomes does not depend on randomization group.*

ASSUMPTION 4: *The fraction of subjects who receive each intervention varies by randomization group.*

ASSUMPTION 5: *Monotonicity. There are no defiers.*

Extending the notation in Cheng (2009), we define the following parameters under the latent compliance model:

π_C , the probability of being a complier,

π_A , the probability of being an always-taker,

π_N , the probability of being a never-taker,

v_j , the probability a complier in randomization group 0 has outcome j ,

t_j , the probability a complier in randomization group 1 has outcome j ,

s_j , the probability a never-taker has outcome j ,

b_j , the probability an always-taker has outcome j ,

where $\pi_C + \pi_A + \pi_N = 1$ and $\sum_j t_j = \sum_j s_j = \sum_j v_j = \sum_j b_j = 1$. See Table 1 for the distributions of n_{raj} in terms of these parameters.

The quantity of interest is the change in a weighted sum of outcomes due to receiving T1 instead of T0 among compliers,

$$\theta_{CACE} = \sum_j w_j (t_j - v_j), \quad (1)$$

where w_j denotes a weight for outcome j based on the relative severity of the outcome. This quantity is called a complier average causal effect (CACE) (Angrist, Imbens, and Rubin, 1996).

2. Estimation

Estimation is via maximum likelihood with the count data. The kernel of the likelihood is

$$L = \prod_{j=1}^J [(\pi_N s_j + \pi_C v_j)^{n_{00j}} (\pi_A b_j)^{n_{01j}} (\pi_N s_j)^{n_{10j}} (\pi_C t_j + \pi_A b_j)^{n_{11j}}], \quad (2)$$

which generalizes the likelihood in Baker and Lindeman (1994) from binomial to multinomial outcomes and the likelihood in Cheng (2009) from all-or-none compliance in one group to all-or-none compliance in two groups. Let $\hat{\theta}_{CACE}$ denote the ML estimate of θ_{CACE} . There are two cases for $\hat{\theta}_{CACE}$: $\hat{\theta}_{CACE(\text{PerfectFit})}$ if ML estimates of all parameters lie in the interior of the parameter space, and $\hat{\theta}_{CACE(\text{Boundary})}$ if ML estimates of some parameters lie on the boundary of the parameter space.

A perfect fit estimate is possible with this model because the number of independent parameters equals the number of independent cell counts. There are $4J - 2$ independent parameters, corresponding to $J - 1$ values for each of s_j , b_j , t_j , and v_j plus the 2 parameters π_N and π_A . There are also $4J - 2$ independent cells counts, with $4J$ counts for $\{n_{raj}\}$ and a fixed total in each randomization group. Thus the model is saturated, and one can compute $\hat{\theta}_{CACE(\text{PerfectFit})}$ by setting observed counts equal their expected values. As derived in Appendix A,

$$\begin{aligned} \hat{\theta}_{CACE(\text{PerfectFit})} &= \sum_{j=1}^J w_j (\hat{t}_j - \hat{v}_j), \text{ where} \\ \hat{t}_j &= \frac{n_{11j} - n_{01j}}{\hat{\pi}_C}, \hat{v}_j = \frac{n_{00j} - n_{10j}}{\hat{\pi}_C}, \hat{\pi}_C = 1 - \frac{n_{01+}}{n_{0++}} - \frac{n_{10+}}{n_{1++}}. \end{aligned} \quad (3)$$

When $\hat{\pi}_C$, $\hat{v}_j$, and $\hat{t}_j$ are each between zero and one, $\hat{\theta}_{CACE(\text{PerfectFit})}$ is the ML estimate. This is because the observed counts are the ML estimates for a different saturated model in which each expected count is a parameter, there is a one-to-one function relating the parameters in the model for latent compliance classes and the aforementioned parameters corresponding to expected counts, and a theorem states that estimates derived from a one-to-one function of ML estimates are also ML estimates (Rohatgi, 1976).

The perfect fit estimate of causal effect equals the intent-to-treat estimate, $\hat{\theta}_{ITT}$, divided by the difference in the fraction receiving T1 in each group,

$$\hat{\theta}_{CACE(\text{PerfectFit})} = \frac{\hat{\theta}_{ITT}}{\hat{\pi}_C}, \text{ where } \hat{\theta}_{ITT} = \sum_{j=1}^J w_j \left(\frac{n_{0+j}}{n_{0++}} - \frac{n_{1+j}}{n_{1++}} \right), \quad (4)$$

which parallels the perfect fit estimate for a causal effect with binary outcomes (Baker and Lindeman, 1994) and the instrumental variables estimate for a causal effect with continuous outcomes (Angrist Imbens and Rubin, 1996). The estimated asymptotic variance of $\hat{\theta}_{CACE(\text{PerfectFit})}$ can be computed using the Multinomial-Poisson transformation (Baker, 1994),

$$\widehat{\text{var}} \{ \hat{\theta}_{CACE(\text{PerfectFit})} \} = \sum_{r=0}^1 \sum_{a=0}^1 \sum_{j=1}^J \left\{ \frac{\partial \hat{\theta}_{CACE(\text{PerfectFit})}}{\partial n_{raj}} \right\}^2 n_{raj}, \quad (5)$$

where the derivatives are readily computed using software for symbolic algebra.

If either $\hat{\pi}_C$, $\hat{v}_j$, or $\hat{t}_j$ in (3) is less than zero or greater than one, the perfect fit estimate is not admissible, and we require the ML estimate of θ_{CACE} when some parameters lies on the boundary of the parameter space. In this case ML estimates can be computed using the following EM algorithm. The E-step computes the complete-data counts $\{e_{ruj}\}$ corresponding to the expected the number of persons in randomization group r who are in latent compliance class $u = N, C, A$ and have outcome j . The M-step estimates the parameters from the complete data on the E-step. Let superscript (i) index each iteration of the combined E-step and M-step. Based on the parameter estimates from the previous M-step, iteration (i) of the E-step is

$$\begin{aligned} e_{0Nj}^{(i)} &= n_{00j} \frac{\hat{\pi}_N^{(i)} \hat{s}_j^{(i)}}{\hat{\pi}_N^{(i)} \hat{s}_j^{(i)} + \hat{\pi}_C^{(i)} \hat{v}_j^{(i)}}, e_{0Cj}^{(i)} = n_{00j} \frac{\hat{\pi}_C^{(i)} \hat{v}_j^{(i)}}{\hat{\pi}_N^{(i)} \hat{s}_j^{(i)} + \hat{\pi}_C^{(i)} \hat{v}_j^{(i)}}, \\ e_{0Aj}^{(i)} &= n_{01j}, e_{1Nj}^{(i)} = n_{10j}, \\ e_{1Cj}^{(i)} &= n_{11j} \frac{\hat{\pi}_C^{(i)} \hat{t}_j^{(i)}}{\hat{\pi}_C^{(i)} \hat{t}_j^{(i)} + \hat{\pi}_A^{(i)} \hat{b}_j^{(i)}}, e_{1Aj}^{(i)} = n_{11j} \frac{\hat{\pi}_A^{(i)} \hat{b}_j^{(i)}}{\hat{\pi}_C^{(i)} \hat{t}_j^{(i)} + \hat{\pi}_A^{(i)} \hat{b}_j^{(i)}}, \end{aligned} \quad (6)$$

and iteration (i) of the M-step is

$$\begin{aligned} \hat{\pi}_N^{(i+1)} &= \frac{e_{+N+}^{(i)}}{e_{+++}^{(i)}}, \hat{\pi}_C^{(i+1)} = \frac{e_{+C+}^{(i)}}{e_{+++}^{(i)}}, \hat{\pi}_A^{(i+1)} = \frac{e_{+A+}^{(i)}}{e_{+++}^{(i)}}, \\ \hat{s}_j^{(i+1)} &= \frac{e_{+Nj}^{(i)}}{e_{+N+}^{(i)}}, \hat{b}_j^{(i+1)} = \frac{e_{+Aj}^{(i)}}{e_{+A+}^{(i)}}, \hat{v}_j^{(i+1)} = \frac{e_{0Cj}^{(i)}}{e_{0C+}^{(i)}}, \hat{t}_j^{(i+1)} = \frac{e_{1Cj}^{(i)}}{e_{1C+}^{(i)}}. \end{aligned} \quad (7)$$

The parameter estimates in (7) are used for iteration $(i + 1)$ of the E-Step. The ML estimate of θ_{CACE} is obtained by substituting the final M-step estimates at convergence into (1). Starting values for parameter estimates equal the perfect fit parameter estimates in (3) with any perfect fit estimate of π_C , v_j , or t_j that is less than zero set equal to zero, and any perfect fit estimate of π_C , v_j , or t_j that is greater than one set equal to one. Imbens and Rubin (1997) also computed ML estimates for a multinomial model via an EM algorithm but did not provide details.

Consider the important special case of a binomial outcome with all-or-none compliance in one randomization group. The counts in group $r = 0$ are $\{n_{00j}\}$ and the counts in group $r = 1$ are $\{n_{1aj}\}$, for $a = 0, 1$ and $j = 0, 1$. In this case, both ML perfect fit and ML boundary estimates can be written in closed-form,

$$\begin{aligned} \hat{\theta}_{CACE(\text{PerfectFit})} &= \hat{t}_1 - \hat{v}_1, \quad \text{when } 0 < \hat{v}_1 < 1, \\ \hat{\theta}_{CACE(\text{Boundary})} &= \begin{cases} \hat{t}_1, & \text{when } \hat{v}_1 \leq 0, \\ \hat{t}_1 - 1, & \text{when } \hat{v}_1 \geq 1, \end{cases} \\ \text{where } \hat{t}_1 &= \frac{n_{111}}{n_{11+}}, \hat{v}_1 = \frac{\frac{n_{001}}{n_{00+}} - \frac{n_{101}}{n_{1+}}}{\pi_C}, \hat{\pi}_C = \frac{n_{11+}}{n_{1++}}. \end{aligned} \quad (8)$$

To our knowledge, the above ML boundary estimate, which is derived in Appendix B, is a new result. A previous ML estimate for this particular boundary scenario involved an iterative algorithm (Cheng et al, 2009).

4. Re-analysis of depression data in Cheng (2009)

For the analysis of the depression data from Cheng (2009) in Table 2, we modified the formulas for the special case of all-or-none compliance in one randomization group, so there are no always-takers. *Assumption 1* holds because a patient's outcome did not depend on the treatment of another patient. *Assumption 2* holds because of random assignment. *Assumption 3* holds because a person who would receive usual care in either group (a never-taker) receives the same usual care if randomized to either group. *Assumptions 4* and *5* hold because seeing a specialist was not possible in the control group.

We computed $\hat{\theta}_{CACE}$ for the weights in Cheng (2009) of $(w_1, w_2, w_3) = (1, 2, 3)$ and $(1, 3, 4)$. We also investigated weights $(0, -0.5, -1.0)$ under the premise that minor depression is half the severity of major depression, with a negative sign indicating a decrease in these detrimental outcomes. With these latter weights, $\hat{\theta}_{CACE}$ can be interpreted as the expected reduction in major depression "equivalents" per person due to seeing a specialist instead of receiving usual care. Confidence intervals computed asymptotically and by bootstrapping were similar (Table 3), with all but one of ten thousand bootstrap iterations yielding a perfect fit estimate in the interior of the parameter space. For the proposed weighting, $\hat{\theta}_{CACE(\text{PerfectFit})} = 0.009$ with a 95% confidence interval of $(-0.14, 0.16)$. Thus, there is little evidence that seeing a specialist reduces the probability of developing depression symptoms, but the wide confidence intervals may preclude a definitive conclusion.

5. Conclusion

With the growing emphasis on reproducible research (Peng, 2009), this simpler approach to ML estimation in causal models for multinomial data should be attractive to many researchers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The author thanks Jing Cheng for helpful comments and kindly providing the data.

APPENDIX A

Perfect fit estimates

Using the following procedure we derived the perfect fit estimates in (3) for all-or-none compliance in two randomization groups with multinomial outcomes. Setting observed counts equal to their expected values yields

$$n_{00j} = n_{0++} \{ \pi_N s_j + (1 - \pi_N - \pi_A) v_j \}, \quad (\text{A.1})$$

$$n_{01j} = n_{0++} \pi_A b_j, \quad (\text{A.2})$$

$$n_{10j} = n_{1++} \pi_N s_j, \quad (\text{A.3})$$

$$n_{11j} = n_{1++} \{ (1 - \pi_N - \pi_A) t_j + \pi_A b_j \}. \quad (\text{A.4})$$

Summing (A.2) and (A.3) over j yields $n_{01+} = n_{0++} \pi_A$ and $n_{10+} = n_{1++} \pi_N$. Therefore $\hat{\pi}_A = n_{01+}/n_{0++}$, $\hat{\pi}_N = n_{10+}/n_{1++}$, and $\hat{\pi}_C = 1 - \hat{\pi}_A - \hat{\pi}_N$. Substituting $\hat{\pi}_A$ and $\hat{\pi}_N$ into (A.2) and (A.3) gives $\hat{b}_j = n_{01j}/n_{01+}$ and $\hat{s}_j = n_{10j}/n_{10+}$. Substituting these estimated parameters into (A.1) and (A.4) gives the perfect fit estimates for v_j and t_j in (3). For estimates under all-or-none compliance in one randomization group, one substitutes $n_{01j} = 0$ into the aforementioned estimates of π_C , s_j , t_j , and v_j .

APPENDIX B

ML boundary estimate with binary outcomes

Using the following procedure we derived the closed-form ML boundary estimates in (8) for all-or-none compliance in only one randomization group and a binomial outcome. The counts in group $r = 0$ are $\{n_{0a}\}$, and the counts in group $r = 1$ are $\{n_{1aj}\}$ for $j = 0, 1$. The kernel of the log-likelihood is

$$\begin{aligned} \mathcal{L} = & n_{000} \log\{\pi_N (1 - s_1) + \pi_C (1 - v_1)\} \\ & + n_{001} \log(\pi_N s_1 + \pi_C v_1) \\ & + n_{100} \log\{\pi_N (1 - s_1)\} \\ & + n_{101} \log(\pi_N s_1) \\ & + n_{110} \log\{\pi_C (1 - t_1)\} \\ & + n_{111} \log(\pi_C t_1). \end{aligned}$$

Let $n = n_{0+} + n_{1++}$. Using software for symbolic algebra (Wolfram Research, 2008) to solve $\partial \mathcal{L} / \partial \pi_C = \partial \mathcal{L} / \partial s_1 = \partial \mathcal{L} / \partial t_1 = 0$ yields $\hat{t}_1 = n_{111} / n_{11+}$ and

$$\begin{aligned} \widehat{\pi}_C &= \frac{n_{11+} (n_{000} + f)}{n f}, \widehat{s}_1 = \frac{f (n_{001} + n_{101})}{n_{000} n_{100} + f (n_{001} + n_{10+})}, \text{ for } v_1 = 0, \\ \widehat{\pi}_C &= \frac{n_{11+} (n_{001} + k)}{n k}, \widehat{s}_1 = \frac{n_{101} (n_{001} + k)}{n_{001} n_{101} + k (n_{000} + n_{10+})}, \text{ for } v_1 = 1. \end{aligned}$$

where $f = n_{100} + n_{11+}$ and $k = n_{101} + n_{11+}$.

References

- Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*. 1996; 91:44–455.
- Baker SG. The Multinomial-Poisson transformation. *The Statistician*. 1994; 43:495–504.
- Baker, SG. Compliance, all-or-none. In: Kotz, S.; Read, CR.; Banks DL, DL., editors. *The Encyclopedia of Statistical Science, Update Volume*. New York: John Wiley and Sons, Inc.; 1997. p. 134-138.
- Baker SG, Kramer BS. Simple maximum likelihood estimates of efficacy in randomized trials and before-and-after studies, with implications for meta-analysis. *Statistical Methods in Medical Research*. 2005; 14:1–19. correction 14, 349.
- Baker SG, Lindeman KS. The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine*. 1994; 13:2269–2278. correction 14, 1841. [PubMed: 7846425]
- Bloom HS. Accounting for no-shows in experimental evaluation designs. *Evaluation Review*. 1984; 8:225–246.
- Cheng J. Estimation and inference for the causal effect of receiving treatment on a multinomial outcome. *Biometrics*. 2009; 65:96–103. [PubMed: 18373714]
- Cheng J, Small DS, Tan Z, Ten Have TR. Efficient nonparametric estimation of causal effects in randomized trials with noncompliance. *Biometrika*. 2009; 96:19–36.
- Connor R, Prorok PC, Weed DL. The case-control design and the assessment of the efficacy of cancer screening. *Journal of Clinical Epidemiology*. 1991;44.1215–44.1221.
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58:21–29. [PubMed: 11890317]
- Peng RD. Reproducible research and *Biostatistics*. *Biostatistics*. 2009; 10:405–408. [PubMed: 19535325]
- Imbens GW, Rubin DB. Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*. 1997; 64:555–574.
- Rohatgi, VK. *An Introduction to Probability Theory and Mathematical Statistics*. New York: John Wiley and Sons, Inc.; 1976. p. 383
- Rubin DB. Comment on 'Randomization analysis of experimental data: The Fisher randomization test' by D. Basau. *Journal of the American Statistical Association*. 1990; 75:591–593.
- Sommer A, Zeger SL. On estimating efficacy from clinical trials. *Statistics in Medicine*. 1991; 10:45–52. [PubMed: 2006355]
- Wolfram Research, Inc. *Mathematica*, Version 7.0. Champaign, IL: Wolfram Research, Inc.; 2008.

Table 1

Model for latent compliance classes

Randomization Group (r)	Treatment Received (a)	Latent compliance classes	Joint probability of latent compliance classes and outcome j	Count
0	0	N and C	$\pi_N s_j + \pi_C v_j$	n_{00j}
	1	A	$\pi_A b_j$	n_{01j}
1	0	N	$\pi_N s_j$	n_{10j}
	1	C and A	$\pi_C t_j + \pi_A b_j$	n_{11j}

 A = always-taker; C =complier, N =never-taker

Table 2

Data from Cheng (2009)

Randomization Group (<i>r</i>)	Treatment Received (<i>a</i>)	Depression symptoms (<i>j</i>)		
		None	Minor	Major
Usual Care	Usual Care	38	19	28
Specialist	Usual Care	5	2	3
	Specialist	32	17	23

Counts are n_{raj}

Table 3

Comparison of results

Method	Quantity	$\hat{\theta}_{\text{CACE}}$ with specified weights		
		(1,2,3)	(1,3,4)	(0,-.5,-1)
Cheng(2009)	estimate	-0.020	-0.023	
Perfect Fit	estimate	-0.019	-0.024	0.009
	standard error	0.15	0.23	0.077
	lower bound	-0.32	-0.49	-0.14
	upper bound	0.28	0.44	0.16
Bootstrap	mean	-0.18	-0.022	0.009
	standard error	0.15	0.23	0.076
	lower bound	-0.32	-0.47	-0.14
	upper bound	0.28	0.44	0.16

On each bootstrap iteration, a perfect