On-line Supplement to “Dissonant Conclusions When Testing the Validity of an Instrumental Variable”

Fan Yang, José R. Zubizarreta, Dylan S. Small, Scott Lorch, Paul R. Rosenbaum¹

University of Chicago, Columbia University and University of Pennsylvania

Abstract: This on-line supplement to “Dissonant Conclusions When Testing the Validity of an Instrumental Variable” documents some rather small technical innovations that are used in the manuscript. These innovations are not central to the point being made in the manuscript itself, but they are relevant to reproducing the details of the analysis in the manuscript or to producing a similar analysis in a different context.

Appendix I: A new bipartite matching algorithm for strengthening an instrumental variable

Following Baiocchi et al. (2010) and Zubizarreta et al. (2013), we used matching to strengthen the instrumental variable while balancing observed covariates. However, we changed, simplified, and in some contexts improved, a key element. These two papers both took a single population, discarded part of the population, split the remainder into pairs, where the pairs balance covariates while being far apart on the instrument. Discarding a middle portion, an ambiguous portion, of the population makes the instrument stronger, improving its design sensitivity, making the study less sensitive to bias from nonrandom assignment of encouragement; see Small and Rosenbaum (2008). Traditionally, splitting a single population into pairs is called by the awkward name “nonbipartite matching” which means “not two parts.” The history of the awkward name involves the fact that optimal two-part matching (e.g., treatment versus control matching), so called optimal bipartite matching, was studied and solved first; see Korte and Vygen (2008) for a textbook discussion of both problems with comprehensive references. Baiocchi et al. (2010) used an algorithm and Fortran code for optimal nonbipartite matching created by Derigs (1988), as implemented in Lu et al.’s (2011) R package `nbpmatching`; it minimizes the total distance within pairs formed from a single population, discarding a portion of the population using a technical trick called “sinks”. Zubizarreta et al. (2013) used integer programming, specifically Zubizarreta’s (2012) `mipmatch` package, to impose additional linear constraints

¹ *Address for correspondence:* Department of Statistics, The Wharton School, University of Pennsylvania, Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340 USA. E-mail: ds-small@wharton.upenn.edu. 8 Dec 2013.

on the nonbipartite match, such as requiring nominal covariates to be perfectly balanced or requiring means of continuous covariates to be close. Also, Zubizarreta et al. (2013) changed the optimized objective function along the lines suggested in Rosenbaum (2012), so as to optimize the number of individuals discarded. A feature of the nonbipartite approach is that individual pairs are far apart on the instrument, but the high baby in one pair may be lower on the instrument than the low baby in some other pair. Depending upon the nature of the instrument and the covariates, that feature may or may not be reasonable. It might be reasonable if the meaning of the instrument changed with the levels of the covariate. In the current study, with an instrument defined in terms of a hospital's rate of use of c-sections in older babies, this feature did not seem reasonable.

We wanted each and every baby in the high group to have a higher value of the instrument than each and every baby in the low group. This change was implemented in a simple way using bipartite matching. We cut the population into three groups based on the value of the instrument, V , where the middle group, $0.29 \leq V \leq 0.31$ contained 10% of the population and was discarded. Write $\{\alpha_1, \dots, \alpha_h, \dots, \alpha_H\}$ for the H remaining babies in the high group and $\{\beta_1, \dots, \beta_\ell, \dots, \beta_L\}$ for the L remaining babies in the low group, noting that $V_{\alpha_h} > V_{\beta_\ell}$ for every h, ℓ . We then matched babies in the high group to babies in the low group to be close in terms of a covariate distance, $\delta_{h\ell}$, measuring how similar baby α_h and baby β_ℓ were in terms of covariates, and far apart on the instrument, with $\delta_{h\ell} = \infty$ if $V_{\alpha_h} - V_{\beta_\ell} < \omega$ for an $\omega > 0$. The covariate distance combined a robust Mahalanobis distance for covariates with $\delta_{h\ell} = \infty$ for mismatches on the variables in Table 1. Write $a_{h\ell} = 1$ if baby α_h in the high group is paired with baby β_ℓ in the low group, $a_{h\ell} = 0$ otherwise, so that we require $a_{h\ell} \in \{0, 1\}$, $\sum_{i=1}^H a_{i\ell} \leq 1$, $\sum_{j=1}^L a_{hj} \leq 1$, for each h, ℓ . In principle, one could simply minimize the total distance within matched pairs, $\sum_{h=1}^H \sum_{\ell=1}^L a_{h\ell} \delta_{h\ell}$, subject to $\sum_{h=1}^H \sum_{\ell=1}^L a_{h\ell} = \min(H, L)$, and this could be done using the optimal assignment algorithm — e.g., Bertsekas' (1981) auction algorithm as made available in the `pairmatch` function of Hansen's (2007) `optmatch` package in R. Alternatively, one could make ω larger, as we did, to further strengthen the instrument, discarding some babies to achieve this more stringent objective. This can be done using the same software for the assignment algorithm without constraining $\sum_{h=1}^H \sum_{\ell=1}^L a_{h\ell}$ and instead minimizing $\sum_{h=1}^H \sum_{\ell=1}^L a_{h\ell} \delta_{h\ell} - \lambda \sum_{h=1}^H \sum_{\ell=1}^L a_{h\ell}$ for specified $\lambda > 0$, and this determines an optimal number of babies to discard; see Rosenbaum (2012) for extensive specifics.

As in Zubizarreta (2012) and Zubizarreta et al. (2013), we used integer programming, not the optimal assignment algorithm, to minimize $\sum_{h=1}^H \sum_{\ell=1}^L a_{h\ell} \delta_{h\ell} - \lambda \sum_{h=1}^H \sum_{\ell=1}^L a_{h\ell}$

but with additional linear constraints. As in these references, these added constraints forced the fine balance in Table 2 and the close mean match seen in Table 3. Moreover, we added a new constraint to further strengthen the instrument. Setting $\delta_{h\ell} = \infty$ if $V_{\alpha_h} - V_{\beta_\ell} < \omega$ forces each matched pair to differ by $\geq \omega$ in terms of the instrument. The new additional constraint forced the mean difference in the instrument V to differ by a larger number, $\Omega > \omega$, so every pair meets the minimum requirement of ω , but on average a larger difference of Ω is achieved. The new constraint was $\sum_{h=1}^H \sum_{\ell=1}^L a_{h\ell} (V_{\alpha_h} - V_{\beta_\ell} - \Omega) > 0$.

Appendix II: Confidence intervals and sensitivity analyses for A/D

Section 2.4 of the paper reported confidence intervals for ratios of survival effects to differences in the frequency of use of c-sections. These intervals are new but are a direct extension of an existing method. This appendix describes the new method and briefly indicates its justification. There are I matched pairs, $i = 1, \dots, I$, of two subjects, $j = 1, 2$, one encouraged, $Z_{ij} = 1$, the other not, $Z_{ij} = 0$, so $Z_{i1} + Z_{i2} = 1$ for each i . In §2.3, there are $I = 1489$ pairs of two babies, one at a high c-section hospital, $Z_{ij} = 1$, the other at a low c-section hospital, $Z_{ij} = 0$. Pairs were matched for observed covariates $\mathbf{x}_{ij}$, so $\mathbf{x}_{i1} = \mathbf{x}_{i2}$ for each i , but the matching may have failed to control an unobserved covariate u_{ij} , so possibly $u_{i1} \neq u_{i2}$ for many or all i . Baby ij has two potential binary responses (r_{Tij}, r_{Cij}) , one r_{Tij} if encouraged with $Z_{ij} = 1$, the other r_{Cij} if unencouraged with $Z_{ij} = 0$. In §2.4, $r_{Tij} = 1$ if baby ij would survive at the high c-section hospital in the i^{th} pair, $r_{Tij} = 0$ otherwise, and $r_{Cij} = 1$ if baby ij would survive at the low c-section hospital in the i^{th} pair, $r_{Cij} = 0$, otherwise. Fisher's (1935) sharp null hypothesis of no treatment effect asserts $H_0 : r_{Tij} = r_{Cij}$ for all babies ij — in words, switching from a low c-section hospital to a high c-section hospital does not change any baby's survival. In a randomized paired experiment with binary response, McNemar's test is the randomization test of Fisher's H_0 . Each baby is observed under one treatment, so the effect of the treatment, $r_{Tij} - r_{Cij}$, is not observed for any baby; see Neyman (1923), Welch (1937) and Rubin (1974). An important unobservable quantity in §2.4 is the attributable effect $A = \sum_{i=1}^I \sum_{j=1}^2 Z_{ij} (r_{Tij} - r_{Cij})$; it is the unobservable number of babies caused to survive by virtue of delivering at the high c-section hospital. A two-sided interval for A may be constructed from two one-sided intervals. Under Fisher's null hypothesis of no effect, every $r_{Tij} - r_{Cij} = 0$, so $A = 0$ no matter how treatments Z_{ij} are assigned.

Similarly, (d_{Tij}, d_{Cij}) is the binary indicator of delivery by cesarean section or vaginal

delivery (1 for c-section, 0 for vaginal delivery) at the high and low c-section hospital. Baby ij is said to be a complier if encouragement shifts the baby's delivery in the encouraged direction, that is, if $1 = d_{Tij} > d_{Cij} = 0$, so this baby would be delivered by c-section at the high c-section hospital in pair i and would be delivered vaginally at the low c-section hospital in pair i . Baby ij is said to be an always taker if $d_{Tij} = d_{Cij} = 1$, a never taker if $d_{Tij} = d_{Cij} = 0$, and a defier if $0 = d_{Tij} < d_{Cij} = 1$, and we follow the usual practice of assuming there are no defiers, $d_{Tij} \geq d_{Cij}$, so a baby who would be delivered by c-section at a low c-section hospital would also be delivered by c-section at a high c-section hospital; see Angrist, et al. (1996) for discussion of this terminology. In constructing one-sided confidence intervals for A , we additionally assume $r_{Tij} \geq r_{Cij}$; this would follow from the assumption that a 23-24 week baby who would survive with the stress of a vaginal delivery would also survive with the reduced stress of a cesarean delivery, combined with the monotonicity assumption of Angrist et al. (1996), namely $d_{Tij} \geq d_{Cij}$, and the exclusion restriction, namely $r_{Tij} = r_{Cij}$ whenever $d_{Tij} = d_{Cij}$. Write $\mathcal{F} = \{(r_{Tij}, r_{Cij}, d_{Tij}, d_{Cij}, \mathbf{x}_{ij}, u_{ij}), i = 1, \dots, I, j = 1, 2\}$ and $\mathcal{Z}$ for the event that $Z_{i1} + Z_{i2} = 1$ for each i . In a randomized paired encouragement design, encouragement Z_{ij} is assigned by $\Pr(Z_{i1} = 1 \mid \mathcal{F}, \mathcal{Z}) = 1/2$, $Z_{i2} = 1 - Z_{i1}$, and assignments in distinct pairs are independent. A simple model for sensitivity analysis in observational studies has $1/(1 + \Gamma) \leq \Pr(Z_{i1} = 1 \mid \mathcal{F}, \mathcal{Z}) \leq \Gamma/(1 + \Gamma)$ for specified $\Gamma \geq 1$, $Z_{i2} = 1 - Z_{i1}$, with independent assignments in distinct pairs, so randomization inference corresponds with $\Gamma = 1$; see Rosenbaum (1987; 2002, §4) for discussion of this method of sensitivity analysis, and for other methods, see Cornfield et al. (1959), Rosenbaum and Rubin (1983), Gastwirth (1992), Marcus (1997), Small (2007), Yu and Gastwirth (2005), Hosman et al. (2010), and Liu et al. (2013). Write R_{ij} for the baby ij 's observed survival response, $R_{ij} = Z_{ij} r_{Tij} + (1 - Z_{ij}) r_{Cij}$, and D_{ij} for the observed delivery, $D_{ij} = Z_{ij} d_{Tij} + (1 - Z_{ij}) d_{Cij}$. Appendix Table 1 rennumbers the two babies in a pair so $Z_{i1} = 1$, $Z_{i2} = 0$, and then records the joint distribution of $(R_{i1}, D_{i1}, R_{i2}, D_{i2}) = (r_{Ti1}, d_{Ti1}, r_{Ci2}, d_{Ci2})$.

As $I \rightarrow \infty$ in a randomized encouragement design, for fixed α , $0 < \alpha < 1$, conventionally $\alpha = 0.05$, it is possible to find an observed random variable $\tilde{A}$ such that $\Pr(A \geq \tilde{A} \mid \mathcal{F}, \mathcal{Z})$ tends to a probability $\geq 1 - \alpha$, so that $A \geq \tilde{A}$ holds with 95% confidence, that is, the unobserved attributable effect A is at least equal to $\tilde{A}$ except in at most $100\alpha\%$ of experiments; see Rosenbaum (2002) for specifics and Weiss (1955) for general discussion of confidence sets for unobserved random variables in terms of observed random variables. Moreover,

in a sensitivity analysis in an observational study, if the bias in treatment assignment is at most $\Gamma \geq 1$, then there is an observed random variable $\tilde{A}_\Gamma$ such that $\Pr\left(A \geq \tilde{A}_\Gamma \mid \mathcal{F}, \mathcal{Z}\right)$ tends to a probability $\geq 1 - \alpha$ as $I \rightarrow \infty$; again, see Rosenbaum (2002).

The exclusion restriction says that encouragement that does not change the delivery (d_{Tij}, d_{Cij}) does not change the response (r_{Tij}, r_{Cij}) , that is, $r_{Tij} = r_{Cij}$ whenever $d_{Tij} = d_{Cij}$. Stated informally, the exclusion restriction says that if high c-section hospitals sometimes save the lives of babies, then they do it by performing c-sections not by doing something else. The exclusion restriction could easily be false: high c-section hospital could be more aggressive in many ways in trying to save the lives of babies of 23-24 weeks gestational age, and c-sections may produce only a part or even none of the survival effect of generally more aggressive treatment. The exclusion restriction places a series of constraints on the relationship between the observed appendix Table 1 and the unobservable table recording $(r_{Ci1}, d_{Ci1}, r_{Ci2}, d_{Ci2})$. The unobserved table is called the pivot table. Consider, for example, the 44 pairs in the first row and first column of the observed appendix Table 1. Because the exclusion restriction says $r_{Tij} = r_{Cij}$ whenever $d_{Tij} = d_{Cij}$, those 44 pairs could be in the same place in the pivot table or some could move to the third and fourth row of the first column, but none could move to the second row. Also, pairs in the second row could move to the fourth row. In fact, the only differences that can exist between the observed and pivot tables are the movements just described. Under the exclusion restriction, A is the total number of pairs that are in the first row of the observed table and in the fourth row of the pivot table.

Let $b_{ij} = 1$ if $r_{Tij} > r_{Cij}$ and $d_{Tij} > d_{Cij}$, and $b_{ij} = 0$ otherwise. If $b_{ij} = 1$, then baby ij would survive receiving a c-section at the high c-section hospital in pair i and would die without a c-section at the low c-section hospital in pair i . Using the exclusion restriction, $r_{Tij} - b_{ij}(d_{Tij} - d_{Cij}) = r_{Cij}$, and the attributable effect is $A = \sum_{i=1}^I \sum_{j=1}^2 Z_{ij}(r_{Tij} - r_{Cij}) = \sum_{i=1}^I \sum_{j=1}^2 Z_{ij}b_{ij}(d_{Tij} - d_{Cij})$. The mean difference in survival is:

$$\begin{aligned}
T_r &= \frac{1}{I} \sum_{i=1}^I \sum_{j=1}^2 Z_{ij} R_{ij} - (1 - Z_{ij}) R_{ij} = \frac{1}{I} \sum_{i=1}^I \sum_{j=1}^2 Z_{ij} \{r_{Cij} + b_{ij} (d_{Tij} - d_{Cij})\} - (1 - Z_{ij}) r_{Cij} \\
&= \frac{1}{I} \sum_{i=1}^I \sum_{j=1}^2 (2Z_{ij} - 1) r_{Cij} + \frac{1}{I} \sum_{i=1}^I \sum_{j=1}^2 Z_{ij} b_{ij} (d_{Tij} - d_{Cij}) \\
&= \frac{1}{I} \sum_{i=1}^I \sum_{j=1}^2 (2Z_{ij} - 1) r_{Cij} + \frac{A}{I}.
\end{aligned}$$

In a randomized paired encouragement experiment, $E(Z_{ij} = 1 \mid \mathcal{F}, \mathcal{Z}) = 1/2$ so that

$$E \left\{ \frac{1}{I} \sum_{i=1}^I \sum_{j=1}^2 (2Z_{ij} - 1) r_{Cij} \mid \mathcal{F}, \mathcal{Z} \right\} = 0, \text{ and } E \left(\frac{A}{I} \mid \mathcal{F}, \mathcal{Z} \right) = \frac{1}{2I} \sum_{i=1}^I \sum_{j=1}^2 (r_{Tij} - r_{Cij}) = \tau, \text{ say,}$$

so that T_r and A/I are both unbiased for the average effect of encouragement, τ ; however, departures from random assignment (i.e., failures of (i) in §1.1) can introduce bias. The observable random variable $T_d = \frac{1}{I} \sum_{i=1}^I \sum_{j=1}^2 Z_{ij} D_{ij} - (1 - Z_{ij}) D_{ij}$ is the difference in the proportion of babies delivered by c-section at high c-section hospitals and the proportion of babies delivered by c-section at low c-section hospitals; in Table 4 it is $T_d = 396/1489 - 194/1489 = 202/1489$. It is a descriptive, not a causal quantity: it describes what happened, not what would happen. The Wald estimator is T_r/T_d . For the Wald estimate to work, encouragement must increase the frequency of what is encouraged so that T_d converges in probability to a strictly positive quantity $\delta > 0$ as $I \rightarrow \infty$, and that is assumed here; therefore, with high probability, high c-section hospitals have done more c-sections among the I pairs than low c-section hospitals for sufficiently large I , and $\Pr(T_d \leq 0 \mid \mathcal{F}, \mathcal{Z})$ is negligible for large I . The quantity

$$W = \frac{\sum_{i=1}^I \sum_{j=1}^2 Z_{ij} b_{ij} (d_{Tij} - d_{Cij})}{\sum_{i=1}^I \sum_{j=1}^2 Z_{ij} D_{ij} - (1 - Z_{ij}) D_{ij}} = \frac{A}{\sum_{i=1}^I \sum_{j=1}^2 Z_{ij} D_{ij} - (1 - Z_{ij}) D_{ij}} = \frac{A/I}{T_d}$$

is the number of babies caused to survive by a c-section in a high c-section hospital as a fraction of the number of additional c-sections performed by high c-section hospitals. Now, W is the ratio of an unobservable random variable A/I , a causal quantity, and an observed random variable T_d , a descriptive quantity, so W is unobservable. The quantity W is

directly interpretable on its own; however, it might reasonably be regarded as the intended finite sample estimand of the Wald estimator, in the sense that T_r/T_d and W both converge in probability as $I \rightarrow \infty$ to the average effect of c-sections on compliers if encouragement is randomized within pairs; see Angrist et al. (1996) for discussion of this estimand. Given the large sample confidence interval, $A \geq \tilde{A}_\Gamma$ with $\Pr\left\{A \geq \tilde{A}_\Gamma \mid \mathcal{F}, \mathcal{Z}\right\} \geq 1 - \alpha$ for sufficiently large I , and continuing to regard $\Pr(T_d \leq 0 \mid \mathcal{F}, \mathcal{Z})$ as negligible for large I , we have $\Pr\left\{A/T_d \geq \tilde{A}_\Gamma/T_d \mid \mathcal{F}, \mathcal{Z}\right\} = \Pr\left\{W \geq \tilde{A}_\Gamma/T_d \mid \mathcal{F}, \mathcal{Z}\right\} \geq 1 - \alpha$ for sufficiently large I . The confidence interval $W \geq \tilde{A}_\Gamma/T_d$ was reported in §2.4.

Without assuming the exclusion restriction, the 95% confidence interval for A is $A \geq 133$ obtained by assuming that 133 of the 158 concordant pairs in Table 5 of the paper contained a high baby caused to survive by virtue of delivering at a high c-section hospital, $r_{Tij} - r_{Cij} = 1$, where McNemar’s one-sided, uncorrected test compares discordant counts of 360 and $318 = 185 + 133$ with P -value 0.053; see Rosenbaum (2002). In contrast, assuming the exclusion restriction is true, yields only $44 + 37$ concordant pairs that might have been affected based on the first row of Appendix Table 1. If 55 discordant pairs were also affected, McNemar’s test compares discordant counts of $305 = 360 - 55$ and $266 = 185 + 44 + 37$ with P -value 0.051, so the interval for A becomes $A \geq 136 = 55 + 44 + 37$. This is divided by the observed, noncausal variable $T_d = 202$ to yield $A/T_d \geq 136/202 = 0.67$. This was incorrectly reported near the end of §2.4 as $A/T_d \geq 133/202 = 0.66$.

Additional References for the On-line Supplement

- Bertsekas, D. P. (1981), “A new algorithm for the assignment problem,” *Mathematical Programming*, 21, 152-171.
- Cornfield, J., Haenszel, W., Hammond, E., Lilienfeld, A., Shimkin, M., Wynder, E. (1959), “Smoking and lung cancer,” *Journal of the National Cancer Institute*, 22, 173-203.
- Derigs, U. (1988), “Solving nonbipartite matching problems by shortest path techniques,” *Annals Operations Research*, 13, 225-261.
- Fisher, R.A. (1935), *The Design of Experiments*, Edinburgh: Oliver & Boyd.
- Gastwirth, J. L. (1992), “Methods for assessing the sensitivity of statistical comparisons used in Title VII cases to omitted variables,” *Jurimetrics* 33, 19-34.
- Hansen, B. B. (2007), “Optmatch,” *R News*, 7, 18-24. R package `optmatch`.
- Hosman, C. A., Hansen, B. B., and Holland, P. W. H. (2010), “The sensitivity of linear regression coefficients’ confidence limits to the omission of a confounder,” *Annals of*

- Applied Statistics*, 4, 849-870.
- Korte, B. and Vygen, J. (2008), *Combinatorial Optimization: Theory and Algorithms*, New York: Springer.
- Liu, W., Kuramoto, S. J., and Stuart, E. A. (2013), “An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research,” *Prevention Science*, 14, 570-580.
- Lu, B., Greevy, R., Xu, X., and Beck, C. (2011), “Optimal nonbipartite matching and its statistical applications,” *American Statistician*, 65, 21-30. R package `nbpmatching`.
- Marcus, S. M. (1997), “Using omitted variable bias to assess uncertainty in the estimation of an AIDS education treatment effect,” *Journal of Educational Statistics*, 22, 193-201.
- Neyman, J. (1923, 1990), “On the application of probability theory to agricultural experiments,” *Statistical Science*, 5, 463-480.
- Rosenbaum, P. R. (1987), “Sensitivity analysis for certain permutation inferences in matched observational studies,” *Biometrika*, 74, 13-26.
- Rosenbaum, P. R. (2002), *Observational Studies*, New York: Springer.
- Rosenbaum, P. R. (2012), “Optimal matching of an optimally chosen subset in observational studies,” *Journal of Computational and Graphical Statistics*, 21, 57-71.
- Rosenbaum, P. and Rubin, D. (1983), “Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome,” *Journal of the Royal Statistical Society B*, 45, 212-218.
- Rubin, D. B. (1974), “Estimating causal effects of treatments in randomized and nonrandomized studies,” *Journal of Educational Psychology*, 66, 688-701.
- Small, D. (2007), “Sensitivity analysis for instrumental variables regression with overidentifying restrictions,” *Journal of the American Statistical Association*, 102, 1049-1058.
- Small, D. S. and Rosenbaum, P. R. (2008), “War and wages: the strength of instrumental variables and their sensitivity to unobserved biases,” *Journal of the American Statistical Association*, 103, 924-933.
- Weiss, L. (1955), “A note on confidence sets for random variables,” *Annals of Mathematical Statistics*, 26, 142-144.
- Welch, B. L. (1937), “On the z-test in randomized blocks,” *Biometrika*, 29, 21-52.
- Yu, B. B. and Gastwirth, J. L. (2005), “Sensitivity analysis for trend tests: application to the risk of radiation exposure,” *Biostatistics*, 6, 201-209.
- Zubizarreta, J. R. (2012), “Using mixed integer programming for matching in an observational study of kidney failure after surgery,” *Journal of the American Statistical Association*

ation, 107, 1360-1371. (R software `mipmatch` at www-stat.wharton.upenn.edu/~josezubi/)

Appendix Table 1: Mortality R_{ij} and mode of delivery D_{ij} (C = C-section, V = vaginal) in 1489 matched pairs of babies of 23-24 weeks gestational age. For the high baby with $Z_{ij} = 1$, mortality is $R_{ij} = r_{Tij}$ and delivery is $D_{ij} = d_{Tij}$, whereas for the low baby with $Z_{ij} = 0$, mortality is $R_{ij} = r_{Cij}$ and delivery is $D_{ij} = d_{Cij}$. To avoid notational ambiguity, in this table j is changed so the first baby, $j = 1$, is the high baby. The table counts pairs, not babies.

	Low Baby, $Z_{i2} = 0$			
	C-Alive	C-Dead	V-Alive	V-Dead
High Baby, $Z_{i1} = 1$	$r_{Ci2} = 1$	$r_{Ci2} = 0$	$r_{Ci2} = 1$	$r_{Ci2} = 0$
	$d_{Ci2} = 1$	$d_{Ci2} = 1$	$d_{Ci2} = 0$	$d_{Cij} = 0$
C-Alive, $r_{Ti1} = 1, d_{Ti1} = 1$	44	54	37	144
C-Dead, $r_{Ti1} = 0, d_{Ti1} = 1$	37	38	36	179
V-Alive, $r_{Ti1} = 1, d_{Ti1} = 0$	31	35	46	127
V-Dead $r_{Ti1} = 0, d_{Ti1} = 0$	47	81	65	488