Attributable Effects in Experiments and Observational Studies

Paul R. Rosenbaum

Department of Statistics

Wharton School

University of Pennsylvania

Philadelphia, PA 19104-6340

1 Outline of Talk

1. Sir Ronald Fisher devoted 15 pages of his *Design of Experiments* to discussing his randomization test for a 2×2 table. Why 15 pages?

2. Review of randomization inference and sensitivity analysis for additive effects.

3. Attributable effects in 2×2 tables.

4. Attributable effects in for quantile displacements.

2 Progress in Statistics Since 1935

- Fisher published *Design of Experiments* in 1935; hence, that date as an anchor.

- How would statistics today look to statisticians in the 1930's, 1940's, 1950's?

- If we look back at papers written at that time, we find a great deal of attention to computing formulas – articles were constrained by what could be computed. The first thing you notice, the dominant impression.

- Looking back we see enormous progress facilitated by computing.

- But I am not asking what the 1930's look like to us. I am asking what we would look like to statisticians back then.

3 Two Quotes About Assumptions

Welch (1937, p. 21): "Of especial interest are the cases of experimentation into which randomization enters as part of the structure. R. A. Fisher has pointed out that, in any such case, it is possible to carry through arithmetical calculations, from which the hypothesis under test may be judged, without making any assumptions whatever."

Cornfield and Tukey (1956, p908): "The question of what assumptions to make seems, at first glance, to be a purely empirical question ... But closer study shows that the choice of assumptions depends on more than empirical questions about the behavior of the experimental material. It depends on the nature of the sampling and randomization involved in obtaining the data ..."

4 The 'Really Modern' Idea

- From Cochran and Cox (1950, 1957, p. 7) *Experimental Designs*:

> "This important result has been illustrated in detail by Fisher (1.2), who has shown how tests of significance and confidence limits can be constructed, using only the fact that randomization has been properly applied in the experiment. Randomization is one of the few characteristics of modern experimental design that appears to be really modern."

- The premodern idea was that inferences were *formally* developed from assumptions (linear model, Gaussian errors, etc).

- The "really modern" (ie post-Fisher) idea is that inferences are *formally* developed from research design.

5 Return to the Original Question

- How would statistics today look to statisticians in the 1930's, 1940's, 1950's?

- Doubtless impressed by the computer and analytical tools it has made possible.

- Doubtless pleased to see that we conduct randomized experiments and probability samples.

- Perhaps disappointed that in our fondness for analytical tools, we have downplayed a 'really modern' idea of which that era was extremely proud, namely that statistical inferences are *formally* developed from, and *warranted* by, research design.

6 The 'Really Modern' Idea Today

- Unless inferences are *formally* developed from research design, it is hard to *formalize* the important point that the uncertainty is greater with weaker research designs.

- If it isn't formalized — if it isn't in the significance level or the confidence interval — it is a footnote to an addendum to an appendix.

- Fisher's argument encourages randomized experimentation when it is ethical and feasible.

- It also forces nonrandomized studies to *formally* acknowledge greater uncertainty — e.g., longer confidence intervals — than would be present in a randomized study.

7 A Limitation

- Much of randomization inference tests the null hypothesis of no treatment effect. But interval estimates are needed in practice.

- Can invert tests obtaining confidence intervals for an additive treatment effect; Lehmann (1963), Hodges & Lehmann (1963), Moses (1965).

- Additive model often inapplicable: binary responses, distributions clearly not shifted, effects vary as a function of other outcomes.

- Common solution abandons the 'really modern idea' of formally developing inferences from research design.

- GOAL: Exact, distribution free, design-based, non-null randomization inferences and sensitivity analyses.

8 Review of Randomization Inference & Sensitivity Analysis for Additive Effects

8.1 Notation

- S strata defined by pretreatment covariates, $s = 1, \ldots, S$, with n_s subjects in stratum s, and $N = \sum n_s$ subjects in total

- $Z_{si} = 1$ if the i^{th} subject in stratum s is assigned to treatment, $Z_{si} = 0$ if assigned to control. $\mathbf{Z} = \left(Z_{11}, Z_{12}, \ldots, Z_{S,n_S}\right)^T$

- $m_s = \sum_{i=1}^{n_s} Z_{si}$ treated subjects in stratum s.

- Eg, treated/control matched pairs is the special case with $n_s = 2$, $m_s = 1$ for $s = 1, \ldots, S$.

8.2 Review: Treatment Effects

- Each subject has two potential responses, a response r_{Tsi} that would be seen under treatment and a response r_{Csi} that would be seen under control. (Neyman 1923, Rubin 1974). $\mathbf{r}_T = \left(r_{T11}, \ldots r_{TS,n_S}\right)^T$, $\mathbf{r}_C = \left(r_{C11}, \ldots r_{CS,n_S}\right)^T$. Fixed features of the finite population of N subjects. Just $2N$ numbers.

- Causal effect compares r_{Tsi} and r_{Csi}. Challenge: we see one or the other, never both.

- *Null hypothesis of no effect*: $H_0 : r_{Tsi} = r_{Csi}$ for all s, i, or $\mathbf{r}_T = \mathbf{r}_C$.

- Model of an *additive effect*, $r_{Tsi} = r_{Csi} + \tau$ for all s, i or $\mathbf{r}_T = \mathbf{r}_C + \tau\,\mathbf{1}$. Shifted distributions.

- The treatment has a *nonnegative effect* if $r_{Tsi} \geq r_{Csi}$ for each s, i

8.3 Review: Randomization

- There are $K = \prod_{s=1}^{S} \binom{n_s}{m_s}$ possible values $\mathbf{z}$ of the $N-$dimensional treatment assignment $\mathbf{Z}$ with $m_s = \sum_{i=1}^{n_s} z_{si}$ for $s = 1, \ldots, S$.

- Collect these K vectors $\mathbf{z}$ in the set Ω.

- In a randomized experiment, $\mathbf{Z}$ is picked at random from Ω, that is, $\Pr(\mathbf{Z} = \mathbf{z}) = \frac{1}{K}$ for each $\mathbf{z} \in \Omega$.

- The difficulty in an observational or nonrandomized study is that this may not be true, and $\Pr(\mathbf{Z} = \mathbf{z})$ is typically unknown.

- Randomization inference uses $\Pr(\mathbf{Z} = \mathbf{z}) = \frac{1}{K}$ as the basis for inference. Sensitivity analysis asks how far $\Pr(\mathbf{Z} = \mathbf{z})$ must move from $\frac{1}{K}$ to change the conclusions.

8.4 Review: Observed Responses

- Observe responses to treatment, r_{Tsi}, from treated subjects, $Z_{si} = 1$, and responses to control, r_{Csi}, from control subjects, $Z_{si} = 0$.

- Observe $R_{si} = Z_{si}\, r_{Tsi} + (1 - Z_{si})\, r_{Csi}$. A random variable: it depends on Z_{si}. $\mathbf{R} = \left(R_{11}, \ldots, R_{S,n_S}\right)^T$.

- However, under the null hypothesis of no treatment effect, $H_0 : \mathbf{r}_T = \mathbf{r}_C$, the observed responses are $\mathbf{R} = \mathbf{r}_C$ and are fixed, unchanging with treatment assignment, *and* observed.

- Under the model of an additive effect, $\mathbf{r}_T = \mathbf{r}_C + \tau\, \mathbf{1}$, the observed responses are $\mathbf{R} = \mathbf{r}_C + \tau\, \mathbf{Z}$ which are random variables depending on treatment assignment $\mathbf{Z}$ whenever $\tau \neq 0$.

8.5 Review: Randomization Test of No Effect

- Under the null hypothesis of no treatment effect, $H_0 : \mathbf{r}_T = \mathbf{r}_C$, the observed responses are $\mathbf{R} = \mathbf{r}_C$ fixed and observed.

- Any test statistic, $T = t(\mathbf{Z}, \mathbf{R})$, is a function of treatment assignment $\mathbf{Z}$ and observed responses $\mathbf{R}$.

- Under the null hypothesis, $H_0 : \mathbf{r}_T = \mathbf{r}_C$, in a randomized experiment, things are simple, because $T = t(\mathbf{Z}, \mathbf{r}_C)$ is a function of fixed, observed constants, $\mathbf{R} = \mathbf{r}_C$, and a random variable $\mathbf{Z}$ with a known distribution, $\Pr(\mathbf{Z} = \mathbf{z}) = \frac{1}{K}$ for each $\mathbf{z} \in \Omega$.

- Hence the distribution of $T = t(\mathbf{Z}, \mathbf{r}_C)$ is known in this case because it was created by the randomization, which forms the "reasoned basis for inference".

8.6 Review: Fisher's Exact Test

- Binary responses, $r_{Ti} = 1$ or 0, $r_{Ci} = 1$ or 0. One stratum, $S = 1$, drop the s.

R_i	Treated	Control	Total
1	$\sum Z_i R_i$		$\sum R_i$
0			$\sum (1 - R_i)$
Total	$m = \sum Z_i$	$\sum (1 - Z_i)$	n

- Under the null hypothesis of no effect saying H_0 : $\mathbf{r}_T = \mathbf{r}_C = \mathbf{R}$, table equals:

r_{Ci}	Treated	Control	Total
1	$\sum Z_i r_{Ci}$		$\sum r_{Ci}$
0			$\sum (1 - r_{Ci})$
Total	$m = \sum Z_i$	$\sum (1 - Z_i)$	n

and $\sum Z_i r_{Ci}$ has the hypergeometric distribution.

8.7 Review: Other Randomization Tests

- The same logic works for any test statistic, $T = t(\mathbf{Z}, \mathbf{R})$.

- Under $H_0 : \mathbf{r}_T = \mathbf{r}_C = \mathbf{R}$, in a randomized experiment, $T = t(\mathbf{Z}, \mathbf{r}_C)$ is a function of fixed, observed constants, $\mathbf{R} = \mathbf{r}_C$, and a random variable $\mathbf{Z}$ with a known distribution, $\Pr(\mathbf{Z} = \mathbf{z}) = \frac{1}{K}$ for each $\mathbf{z} \in \Omega$.

- One obtains in this way the exact distributions of Fisher's exact test for a 2×2 table, the Mantel-Haenszel-Birch test for a $2 \times 2 \times S$ table, the Wilcoxon rank sum test, the Wilcoxon signed rank test, the Hodges-Lehmann aligned rank test, and many others.

- Welch (1937): null randomization distribution of an anova test, $T = t(\mathbf{Z}, \mathbf{R})$, say the F-test, converge in distribution to the usual F-distribution.

8.8 Review: Point and Interval Estimates by Inverting the Test

- If the treatment had an additive effect, $\mathbf{r}_T = \mathbf{r}_C + \tau\,\mathbf{1}$, the observed responses are $\mathbf{R} = \mathbf{r}_C + \tau\,\mathbf{Z}$.

- To test $H_0 : \tau = \tau_0$, compute the (observed) adjusted responses, $\mathbf{R} - \tau_0\,\mathbf{Z}$, which equal $\mathbf{r}_C$ if H_0 is true, so randomization creates the null distribution of $T = t\,(\mathbf{Z}, \mathbf{R} - \tau_0\,\mathbf{Z}) = t\,(\mathbf{Z}, \mathbf{r}_C)$.

- Eg, With one stratum, $S = 1$, to test $H_0 : \tau = \tau_0$, subtract τ_0 from responses in the treated group, obtaining $\mathbf{R} - \tau_0\,\mathbf{Z}$, and apply the rank sum test to them, $T = t\,(\mathbf{Z}, \mathbf{R} - \tau_0\,\mathbf{Z})$.

- The set of values τ_0 not rejected by a 0.05 level test is the 95% confidence interval, and the value $\widehat{\tau}$ that equates $t\,(\mathbf{Z}, \mathbf{R} - \widehat{\tau}\,\mathbf{Z})$ to its null expectation is the Hodges-Lehmann point estimate.

8.9 Review: Sensitivity Analysis - Cornfield Inequality

- In an observational study, treatments are not randomly assigned, so $\Pr(\mathbf{Z} = \mathbf{z})$ may not equal $\frac{1}{K}$ for each $\mathbf{z} \in \Omega$, and the previous argument breaks down.

- First sensitivity analysis by Cornfield, et al. (1959) in the *Journal of the National Cancer Institute*, concerned smoking as a cause of lung cancer. The treatment, smoking, is not randomly assigned to people. However, they found the departure from random assignment would need to be enormous to alter the conclusion that heavy smoking causes lung cancer.

- "association does not imply causation" $\Rrightarrow$ "to explain away the observed association, hidden biases would have to be this magnitude."

8.10 Review: Sensitivity Analysis Model

- Model for sensitivity analysis: Z_{si} initially independent, and two subjects i, j in the same stratum s differ in odds of treatment by at most $\Gamma \geq 1$

$$\frac{1}{\Gamma} \leq \frac{\Pr(Z_{si}=1) \,/\, \Pr(Z_{si}=0)}{\Pr\left(Z_{sj}=1\right) \,/\, \Pr\left(Z_{sj}=0\right)} \leq \Gamma \text{ for all } s, i$$

- Condition on $m_s = \sum_{i=1}^{n_s} Z_{si}$ returns the distribution to Ω. With $\Gamma = e^{\gamma}$, equivalent to

$$\Pr(\mathbf{Z}=\mathbf{z}) = \frac{\exp\left(\gamma \mathbf{z}^T \mathbf{u}\right)}{\sum_{\mathbf{b}\in\Omega} \exp\left(\gamma \mathbf{b}^T \mathbf{u}\right)}$$

for a $\mathbf{u}$ with $0 \leq u_{si} \leq 1$, $\forall\ s, i$; (Rosenbaum 1995, 2002, §4). $\Gamma = 1 \Longrightarrow \Pr(\mathbf{Z}=\mathbf{z}) = 1/K$.

8.11 Review: Sensitivity Bounds

- Each $e^{\gamma} = \Gamma \geq 1$ produces a set of possible treatment assignment distributions,

$$\Pr(\mathbf{Z} = \mathbf{z}) = \exp\left(\gamma \mathbf{z}^T \mathbf{u}\right) / \sum_{\mathbf{b} \in \Omega} \exp\left(\gamma \mathbf{b}^T \mathbf{u}\right)$$

  and each of these distributions yields a inference quantity, say a significance level, a point estimate, or the endpoint of a confidence interval.

- For several values of Γ, the sensitivity analysis computes the max and min of the inference quantity, say the significance level.

- How large must Γ be before the range of, say, significance levels includes values that leave qualitatively different impressions?

- For the 2×2 table, bound from the extended hypergeometric distribution with parameter Γ. Other bounds for Wilcoxon rank sum, etc. (Rosenbaum 2002, §4).

9 Attributable Effects

- When the model of an additive treatment effect is inappropriate, as for binary responses, we would like to use a similar argument to obtain randomization inferences and sensitivity analyses.

- Suppose first there is one stratum, $S = 1$, dropping the s, and the responses are binary, $r_{Ti} = 1$ or 0, $r_{Ci} = 1$ or 0.

- Assume nonnegative effect, $r_{Ti} \geq r_{Ci}$, so (r_{Ti}, r_{Ci}) is $(1,1)$, $(1,0)$, or $(0,0)$. Write $\delta_i = r_{Ti} - r_{Ci}$ which is 0 or 1. Observe $R_i = r_{Ci} + Z_i \, \delta_i$.

- Can't use an additive model. Common to drop the logic of Fisher's exact test and introduce an infinite population model. Must we do this?

9.1 Attributable Effects: Small Example

- In the London Underground, some stations have a "pit" about a meter deep. If a passenger waiting for a train falls, or is pushed or jumps onto the tracks — what the British Railway Regulations Act of 1893 calls an "incident" — then the pit is a place to avoid a train. Act requires "incidents" to be recorded. Coats & Walter (*Brit Med J* 1999) looked at 53 "incidents":

	No Pit	Pit
Dead	16	14
Alive	5	18

- Test no effect using Fisher's exact test, the one-sided P-value is 0.0193, and the upper bound on the P-value is 0.0885 for $\Gamma = 1.5$.

- Can we obtain an estimate of effect without introducing random sampling from an infinite population of incidents?

9.2 Attributable Effects: The Non-null 2×2 Table.

- Observed 2×2 table is $R_i = r_{Ci} + Z_i\,\delta_i$ by Z_i

Response	Treated	Control
1	$\sum Z_i\, r_{Ti}$	$\sum(1 - Z_i)\, r_{Ci}$
0	$\sum Z_i\,(1 - r_{Ti})$	$\sum(1 - Z_i)\,(1 - r_{Ci})$
Total	m	$n - m$

which does not generally have a hypergeometric or extended hypergeometric distribution, even in a randomized experiment.

- Unobserved table of potential responses to control

Response	Treated	Control
1	$\sum Z_i\, r_{Ci}$	$\sum(1 - Z_i)\, r_{Ci}$
0	$\sum Z_i\,(1 - r_{Ci})$	$\sum(1 - Z_i)\,(1 - r_{Ci})$
Total	m	$n - m$

is hypergeometric in a randomized experiment.

9.3 Attributable Effects: Compatible Hypotheses.

- Suppose we had a hypothesis, $H_0 : \delta_1 = 1, \delta_2 = 0, \ldots, \delta_n = 0$, say, or $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$, where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^T$. Some hypotheses are clearly wrong, that is, logically impossible given what we assumed.

- If we observe a control subject, $Z_i = 0$, with $R_i = r_{Ci} + Z_i\,\delta_i = r_{Ci} = 1$, then a hypothesis which asserts $\delta_i = 1$ is clearly false.

- We assumed $r_{Ti} \geq r_{Ci}$. If we observe a treated subject, $Z_i = 1$, with $R_i = r_{Ci} + Z_i\,\delta_i = r_{Ti} = 0$, then a hypothesis which asserts $\delta_i = 1$ is clearly false.

- Call a hypothesis *compatible* if it is logically possible, and *incompatible* if logically impossible. A logically impossible hypothesis can be rejected with type 1 error rate of zero.

9.4 Attributable Effects: A Pivot

$$A = \sum_{i=1}^{n} Z_i \, \delta_i = \sum_{i=1}^{n} Z_i \, (r_{Ti} - r_{Ci})$$

is the number of events among treated subjects that were caused by the treatment. (Eg, the number of deaths in stations without a pit that were caused by the absence of a pit.)

- A is a random variable, not a parameter, but one we can never observe.

- We observe the number of events among treated subjects, $T = \sum_{i=1}^{n} Z_i \, R_i = \sum_{i=1}^{n} Z_i \, r_{Ti}$.

- Notice that $T - A = \sum_{i=1}^{n} Z_i \, r_{Ti} - \sum_{i=1}^{n} Z_i \, (r_{Ti} - r_{Ci}) = \sum_{i=1}^{n} Z_i \, r_{Ci}$, the corner cell in the table with no treatment effect. A pivot!

9.5 Attributable Effects: The Adjusted 2×2 Table

Because $\sum_{i=1}^{n} Z_i \, r_{Ti} - A = \sum_{i=1}^{n} Z_i \, r_{Ci}$, the adjusted table

Response	Treated	Control
1	$\sum Z_i \, r_{Ti} - A$	$\sum (1 - Z_i) \, r_{Ci}$
0	$\sum Z_i \, (1 - r_{Ti}) + A$	$\sum (1 - Z_i) \, (1 - r_{Ci})$
Total	m	$n - m$

equals the table for potential responses to control,

Response	Treated	Control
1	$\sum Z_i \, r_{Ci}$	$\sum (1 - Z_i) \, r_{Ci}$
0	$\sum Z_i \, (1 - r_{Ci})$	$\sum (1 - Z_i) \, (1 - r_{Ci})$
Total	m	$n - m$

which has a hypergeometric distribution in a randomized experiment.

9.6 Attributable Effects: Testing a Hypothesis About δ

- If $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$, is incompatible, reject it with certainty, that is, with type 1 error rate of zero.

- Otherwise, compute $A_0 = \sum Z_i \, \delta_{0i}$, which equals A if H_0 is true. Then compute the adjusted 2×2 table,

Response	Treated	Control
1	$\sum Z_i \, r_{Ti} - A_0$	$\sum (1 - Z_i) \, r_{Ci}$
0	$\sum Z_i \, (1 - r_{Ti}) + A_0$	$\sum (1 - Z_i) \, (1 - r_{Ci})$
Total	m	$n - m$

  which has a hypergeometric distribution when the null hypothesis is true. Perform a one-or-two tailed test using the corner cell.

- Invert the test to get a confidence set $\mathcal{C}$ for $\boldsymbol{\delta}$. Easy to describe the confidence set, because whether a $\boldsymbol{\delta}_0$ is in $\mathcal{C}$ depends solely on the value of $A_0 = \sum Z_i \, \delta_{0i}$.

9.7 Attributable Effects: Example

	No Pit	Pit
Dead	16	14
Alive	5	18

$A_0 = 0$
$P - value = .0193$

	No Pit	Pit
Dead	15	14
Alive	6	18

$A_0 = 1$
$P - value = .0439$

	No Pit	Pit
Dead	14	14
Alive	7	18

$A_0 = 2$
$P - value = .0875$

Confidence set $\mathcal{C}$ for $\boldsymbol{\delta}$ includes all compatible $\boldsymbol{\delta}_0$ with $A_0 \geq 2$ deaths attributable to the absence of a pit, that is, $\geq 2/30$ of the deaths.

- Sensitive to hidden bias: For $\Gamma = 1.5$, $A_0 = 0$ has upper bound on the $P - value$ of 0.0885, and $A_0 = 1$ has upper bound 0.16.

9.8 Attributable Effects: Other Situations

- Other cases with *exact, non-null, design based* randomization inferences and sensitivity analyses.

- Can look at McNemar's test and the Mantel-Haenszel test in matching with one or more controls. Requires asymptotic separability *a la* Gastwirth, Krieger & Rosenbaum (2000).

- A quantity related to Mann-Whitney statistic, proportion of treated responses above control responses attributable to treatment.

- Number of treated subjects caused to have responses above a certain quantile of the potential control responses, somewhat analogous to control median procedure of Gart (1963) and Gastwirth (1968).

- Can do something similar for matched pairs, and will illustrate this.

9.9 Attributable Effects: Continuous Response

Thun et al (1989) compared male workers at a cadmium recovery plant to unexposed male workers at a local hospital in term of kidney function, matching for age. β-2-microglobulin in micrograms per gram of creatine. 23 Pairs Matched for Age.

Kidney Function of Cadmium Workers and Unexposed Controls.

From Thun, et al. (1989).

Pair	Cadmium Worker	Hospital Control
1	107,143	311
	⋮	
8	211	242
	⋮	
23	941	247

9.10 Attributable Effects: Order Statistics

- S strata/matched sets, n_s in stratum s, $N = \sum n_s$. Here, $S = 23$, $n_s = 2$, $N = 46$.

- Let (y_{Tsi}, y_{Csi}) be the potential responses under treatment, y_{Tsi}, and control, y_{Csi}. See y_{Tsi} if $Z_{si} = 1$, y_{Csi} if $Z_{si} = 0$. The (y_{Tsi}, y_{Csi}) are fixed features of the finite population of N subjects. Will assume $y_{Tsi} \geq y_{Csi}$.

- Let $y_{T(1)} < \ldots < y_{T(N)}$ be the 46 order statistics of potential responses to treatment, and $y_{C(1)} < \ldots < y_{C(N)}$ be the 46 order statistics of potential responses to control. Also fixed. (Easy to allow for ties.)

- We do not observe these order statistics. We observe $Y_{si} = Z_{si}\, y_{Tsi} + (1 - Z_{si})\, y_{Csi}$ with order statistics $Y_{(1)} \leq \ldots \leq Y_{(N)}$. Random variables.

9.11 Attributable Effects: Displacements

- Fix a k so $y_{C(k)}$ is the (*unobserved*) k/N quantile of potential responses y_{Csi} to control.

- Let θ be any (*unknown*) value between $y_{C(k)}$ and $y_{C(k+1)}$, so $y_{C(k)} < \theta < y_{C(k+1)}$.

- Subject (s, i) has a displacement if $y_{Tsi} > \theta > y_{Csi}$.

- Write $r_{Tsi} = 1$ if $y_{Tsi} > \theta$, $r_{Tsi} = 0$ otherwise; $r_{Csi} = 1$ if $y_{Csi} > \theta$, $r_{Csi} = 0$ otherwise; so there is a displacement if $\delta_{si} = r_{Tsi} - r_{Csi} = 1$. Let $R_{si} = Z_{si}\, r_{Tsi} + (1 - Z_{si})\, r_{Csi}$, which indicates whether $Y_{si} > \theta$. *Not observed*.

- $A = \sum_{s,i} Z_{si}\, \delta_{si}$ is displacements attributable to treatment.

9.12 Attributable Displacements: Key Fact

- Because we don't see the $y_{C(j)}$'s, we don't know θ and can't compute things.

Lemma: If $a = \sum_{s,i} Z_{si}\,\delta_{si}$, then

$$Y_{(k+1-a)} > \theta > Y_{(k-a)}.$$

Proof: $N-k$ subjects have $y_{Csi} > \theta$, and since $y_{Tsi} \geq y_{Csi}$, these subjects have $Y_{si} > \theta$. Because $a = \sum_{s,i} Z_{si}\,\delta_{si}$, there are a other subjects with $Y_{si} = y_{Tsi} > \theta > y_{Csi}$. The remaining $k-a$ subjects have $\theta > Y_{si}$. So $N-k+a$ of the $Y_{si} > \theta$ and $k-a$ of the $Y_{si} < \theta$, proving the lemma.

9.13 Displacements: Procedure

- Consider displacements about the 80^{th} percentile of control responses, that is, since $0.8 \times 46 = 36.8$, let $r_{C(36)} < \theta < r_{C(37)}$ and subject (s, i) would be displaced, $\delta_{si} = 1$, about the 80^{th} percentile of responses to control if $y_{Tsi} > \theta > y_{Csi}$.

- Consider testing a compatible hypothesis $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$. Compute $A_0 = \sum_{s,i} Z_{si}\, \delta_{0si}$. If the hypothesis is true, then $Y_{(k+1-A_0)} > \theta > Y_{(k-A_0)}$, so $R_{si} = 1$ if $Y_{si} \geq Y_{(k+1-A_0)}$ and $R_{si} = 0$ if $Y_{si} \leq Y_{(k-A_0)}$.

- Therefore, the earlier methods for binary responses can be applied, in this case, for McNemar's test.

9.14 Attributable Displacements: Example

- Test 1 $= A_0 = \sum_{s,i} Z_{si}\,\delta_{0si}$ about the 80^{th} percentile $y_{C(36)}$. Find $Y_{(k+1-A_0)} = Y_{(36+1-1)} =$ 892 and $Y_{(k-A_0)} = Y_{(36-1)} =$ 700, determines R_{si}.

		Cadmium Worker	
		$Y_{si} \geq 892$	$Y_{si} \leq 700$
Matched	$Y_{si} \geq 892$	0	0
Control	$Y_{si} \leq 700$	11	12

Remove the 1 displacement:

		Cadmium Worker	
		$Y_{si} \geq 892$	$Y_{si} \leq 700$
Matched	$Y_{si} \geq 892$	0	0
Control	$Y_{si} \leq 700$	10	13

McNemar's test: $\left(10 - \frac{10}{2}\right) / \sqrt{10/4} = 3.16$, $P - value = 0.00078$.

9.15 Displacement Example, Continued

- In the same way, McNemar's test rejects hypotheses $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$ with $A_0 = \sum_{s,i} Z_{si}\, \delta_{0si} \leq 9$ and accepts $A_0 = \sum_{s,i} Z_{si}\, \delta_{0si} \geq 10$.

- So we are 95% confident that at least 10/23 cadmium workers had displacements above the 80^{th} percentile $y_{C(36)}$ that would have been observed had all subjects been spared cadmium exposure.

- Fairly insensitive to hidden bias:

Γ	95% confident that $A \geq k$
1	10
2	7
3	7
4	0

  so the null hypothesis of no effect is barely plausible with $\Gamma = 4$, the upper bound on the $P - value$ for $A_0 = 0$ being 0.057.

10 Attributable Effects for the Mann-Whitney Test

- The Mann-Whitney $U-$statistic is the number of times a treated subject had a higher response than a control.

- Sometimes this happens by accident, that is, by how subjects were randomly assigned to treatment or control.

- Other times it happens because of effects of the treatment.

- The attributable effect in this case is the number of times a treated subject had a higher response than a control because of effects of the treatment.

11 Attributable Effects for Wilcoxon's Signed Rank Test

- Wilcoxon's signed rank statistic for matched treated - control pairs equals the number of positive Walsh averages.

- Some Walsh averages are positive by accident, by random assignment of treatment or control within each pair.

- Other Walsh averages are positive because of effects of the treatment.

- The attributable effect in this case is the number of Walsh averages that are positive because of effects of the treatment.

12 Summary

- The "really modern idea" is that inferences depend explicitly on the research design.

- View encourages randomized experiments when ethical and feasible, and encourages explicit discussion of greater uncertainty from nonrandomized studies via sensitivity analysis.

- Possible to invert more randomization inferences to obtain confidence intervals. Illustrated for 2×2 tables and quantile displacements, but applicable in many other situations as well.

- Rosenbaum, P. R. (2001) Effects attributable to treatment: Inference in experiments and observational studies with a discrete pivot. *Biometrika*, **88**, 219-231.

- Rosenbaum, P. R. (2002) Attributing effects to treatment in matched observational studies. *JASA*, **97**, 183-192.

- Rosenbaum, P. R. (2002) *Observational Studies* (Second Edition). New York: Springer-Verlag, §5.

- Rosenbaum, P. R. (2003) Exact confidence intervals for nonconstant effects by inverting the signed rank test. *American Statistician*, **57**, 132-138.