

Portland State University

PDXScholar

Community Health Faculty Publications and
Presentations

School of Community Health

2014

Comment on “The Effect of Same-Sex Marriage Laws on Different-Sex Marriage: Evidence From the Netherlands”

Alexis Dinno

Portland State University, alexis.dinno@pdx.edu

Follow this and additional works at: https://pdxscholar.library.pdx.edu/commhealth_fac



Part of the [Community Health and Preventive Medicine Commons](#)

Let us know how access to this document benefits you.

Citation Details

Dinno, Alexis, "Comment on “The Effect of Same-Sex Marriage Laws on Different-Sex Marriage: Evidence From the Netherlands”" (2014). *Community Health Faculty Publications and Presentations*. 36.

https://pdxscholar.library.pdx.edu/commhealth_fac/36

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Community Health Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

**Comment on “The Effect of Same-Sex Marriage Laws on Different-Sex Marriage:
Evidence From the Netherlands”**

Alexis Dinno

RUNNING HEAD: Comment: = Effect of Same-Sex Marriage

A. Dinno

School of Community Health, Portland State University, PO Box 751DSCH, Portland, Oregon
97207-0751, USA

e-mail alexis.dinno@pdx.edu

Abstract In the recent *Demography* article titled “The Effect of Same-Sex Marriage Laws on Different-Sex Marriage: Evidence From the Netherlands,” Trandafir attempted to answer the question, Are rates of opposite sex marriage affected by legal recognition of same sex marriages? The results of his approach to statistical inference—looking for evidence of a *difference* in rates of opposite-sex marriage—provide an absence of evidence of such effects. However, the validity of his conclusion of no causal relationship between same-sex marriage laws and rates of opposite-sex marriage is threatened by the fact that Trandafir did not also look for *equivalence* in rates of opposite-sex marriage in order to provide evidence of an absence of such an effect. Equivalence tests in combination with difference tests are introduced and presented in this article as a more valid inferential approach to the substantive question Trandafir attempted to answer.

Keywords Same sex marriage; Equivalence test

Post-peer review manuscript.

The final publication is available at Springer via <http://dx.doi.org/10.1007/s13524-014-0338-1>.

In a recent *Demography* article, Trandafir (2014) made an important contribution to a growing nascent literature (Badgett 2004; Dillender 2014; Dinno and Whitney 2013; Langbein and Yost 2009) on the effects of extending legal recognition of same-sex marriages on opposite-sex marriages. Such research is important because it clarifies that legally recognized same-sex marriage—which is a material social good, legal right, and determinant of health—is not denied to the population of same-sex couples and their children based on a potentially spurious argument that such a denial will protect marriages of opposite-sex couples. The fundamental question here is, Are rates of opposite-sex marriage affected by legal recognition of same-sex marriages? Trandafir attempted to answer this question by examining both (1) population marriage rates during an 18-year period in the Netherlands versus a synthetic control while adjusting for a variety of determinants of marriage rates, and (2) the behavior of individuals in a discrete-time event history model of marriage during a 10-year period while adjusting for a variety of determinants of the individual discrete hazard probability for marriage.

Unfortunately, Trandafir only half answered the *statistical* question of whether rates of opposite-sex marriage are affected by legal recognition of same-sex marriages, because he posed hypothesis tests for only *difference* in marriage rates and *difference* in marriage hazards but did not pose hypothesis tests for *equivalence* in marriage rates and marriage hazards. For example, in his analysis of aggregate rates, Trandafir actually tested whether differences between Netherlands marriage rates and synthetic control marriage rates were different from zero. However, he could have also tested whether the size of these differences was greater than some *a priori* level of relevance. As Altman and Bland (1995) noted, absence of evidence of an effect—such as the effect of same-sex marriage laws on rates or hazards of opposite-sex marriage—is not the same thing as evidence of the absence (or equivalence) of an effect. Trandafir looked for evidence of difference in marriage rates and did not find it. However, one cannot conclude there is equivalence without looking for evidence of such, and in interpreting lack of significance of a difference in marriage rates or in marriage hazards as evidence of the absence of “causal estimates” veers toward the fallacy of accepting the null hypothesis.

Why is looking for “evidence of absence” important for policy makers? If policy makers are concerned about whether legalizing same-sex marriage affects rates of opposite-sex marriage, then to be confident that there is no effect, they need sufficiently strong evidence that opposite-sex marriage rates with same-sex marriage laws and without same-sex marriage laws are equivalent. Tests for difference do not provide evidence of equivalence. Equivalence tests do.

The null hypotheses of equivalence tests (H_0^-), sometimes termed “negativist” null hypotheses (Reagle and Vinod 2003), take the general form $H_0^- : \theta \geq \Delta_1$ or $\theta \leq \Delta_2$, where Δ_1 and Δ_2 are upper and lower researcher-defined *a priori* levels of tolerance, with $\Delta_1 > 0$ and $\Delta_2 < 0$. (It is possible that $\Delta_1 = |\Delta_2|$.) By contrast, the more common null hypotheses of difference tests (H_0^+), termed “positivist” null hypotheses, take the general form $H_0^+ : \theta = 0$. Equivalence tests, originally motivated by demonstrating equivalence in therapeutic pharmacological effects, were developed in clinical epidemiology using a ‘two one-sided tests’ framework (Anderson and Hauck 1983; Hauck and Anderson 1984; Schuirmann 1987). However, evaluating evidence of equivalence is generally useful to the sciences because it allows the burden of evidence to be shared evenly between demonstrating the existence of a relationship and demonstrating the absence of a relationship.

One would reject H_{01}^- if $P(T \geq t_1) < \alpha$, would reject H_{02}^- if $P(T \leq t_2) < \alpha$ —see equations (4) and (6)—and would conclude equivalence within the interval $[\Delta_2, \Delta_1]$ for a given α only by rejecting both these null hypotheses.¹ As mentioned earlier, the equivalence interval $[\Delta_2, \Delta_1]$ itself is defined by the researcher, and the boundaries of this interval are the thresholds dividing relevantly large differences from equivalently small differences; a difference in this interval is too small to care about. The threshold values, Δ_2 and Δ_1 , are measured in the same units as θ (e.g., the same units as a mean difference).

$$H_{01}^- : \theta \geq \Delta_1 \tag{1}$$

$$H_{02}^- : \theta \leq \Delta_2. \tag{2}$$

¹ The overall Type I error is controlled with α , rather than with $\alpha/2$, because the intervals defined by H_{01}^- and H_{02}^- do not overlap.

Combining tests of H_0^+ (no difference) with test of H_0^- (no equivalence) gives four possible conclusions, as shown in Fig. 1. These possible conclusions are as follows:

1. Rejecting neither H_0^+ nor H_0^- indicates *indeterminacy* (i.e., an underpowered test). For example, we are unable to draw conclusions about the effects of same-sex marriage laws on rates of opposite-sex marriage.

2. Rejecting both H_0^+ and H_0^- indicates *trivial difference* (i.e., the difference is ignorable at the given tolerance). For example, differences in rates of opposite-sex marriage under same-sex marriage laws are found to be ignorable because they are smaller than we have said we care about.

3. Rejecting H_0^+ , but not rejecting H_0^- indicates a *relevant difference* (i.e., a difference that is large enough to fall outside the interval $[\Delta_2, \Delta_1]$). For example, differences in rates of opposite-sex marriage under same-sex marriage laws are found to be relevant because they are large enough to matter.

4. Not rejecting H_0^+ but rejecting H_0^- indicates *equivalence* (i.e., no difference within the interval $[\Delta_2, \Delta_1]$), as shown in Fig. 2. For example, rates of opposite-sex marriage under same-sex marriage laws are found to be equivalent; no evidence of a difference in rates was found.

The hypothesis (1) can be rearranged as in (3), and a Wald-type t test statistic is easily constructed, as in Eq. (4). Likewise, the hypothesis (2) can be rearranged as in (5), and a t test statistic is easily constructed, as in Eq. (6). This logic works for paired and unpaired data, and works for constructing Wald-type z test statistics. It is important to note that data where both $\Delta_1 \leq s_\theta t_\alpha^*$ and $|\Delta_2| \leq s_\theta t_\alpha^*$ are underpowered and will not reject any H_0^- (where t_α^* is the critical value of the test statistic for a given level of α). Nuances in the precise way that interval hypotheses like H_0^- alter the distribution of test statistics such as Eqs. (4) and (6) can motivate more sophisticated calculations using noncentrality parameters, and such tests are well established (Wellek 2010).

$$\begin{aligned}
 H_0^- : \quad & \theta \geq \Delta_1 \\
 & -\Delta_1 + \theta \geq 0 \\
 & \Delta_1 - \theta \leq 0.
 \end{aligned} \tag{3}$$

$$t_1 = \frac{\Delta_1 - \theta}{s_\theta} \quad (4)$$

$$H_{02}^- : \begin{aligned} \theta &\leq \Delta_2 \\ \theta - \Delta_2 &\leq 0 \end{aligned} \quad (5)$$

$$t_2 = \frac{\theta - \Delta_2}{s_\theta}. \quad (6)$$

Trandafir's research, and the policy implications drawn from it, would benefit from taking an equivalence testing approach to providing evidence of whether there is an absence of an effect of same-sex marriage laws. Because Trandafir tested only for differences in analysis of individual rates, we do not know if the differences he found (legalized same-sex marriage increasing opposite-sex marriage in the Bible Belt but decreasing opposite-sex marriage in the four largest cities) are relevant or trivial. But by providing conclusions based on combined inference from tests for difference with tests for equivalence, the burden of evidence is divided between evidence for both the existence of and the absence of an effect of same-sex marriage laws on opposite-sex marriage rates. Such an approach has previously been published in the literature on the effects of same-sex marriage legalization of rates of opposite-sex marriage (Dinno and Whitney 2013), albeit using a different form of equivalence test, as described in (Wellek 2010). The population sciences in general, and the policies informed by them, would benefit from this kind of division of the burden of evidence between evidence of effect and evidence of absence of effect.

References

- Altman, D. G., & Bland, M. J. (1995). Statistics notes: Absence of evidence is not evidence of absence. *British Medical Journal*, *311*, 485.
- Anderson, S., & Hauck, W. W. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Communications in Statistics—Theory and Methods*, *12*, 2663–2692.
- Badgett, M. V. L. (2004). Will providing marriage rights to same-sex couples undermine heterosexual marriage? *Sexuality Research and Social Policy*, *1*, 1–10.
- Dinno, A., & Whitney, C. (2013). Same sex marriage and the perceived assault on opposite sex marriage. *PLoS ONE*, *8*(6), e65730. doi:10.1371/journal.pone.0065730

- Dillender, M. (2014). The death of marriage? The effects of new forms of legal recognition on marriage rates in the United States. *Demography*, *51*, 563–585.
- Hauck, W. W., & Anderson, S. (1984). A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *Journal of Pharmacokinetics and Pharmacodynamics*, *12*, 83–91.
- Langbein, L., & Yost, M. A. (2009). Same-sex marriage and negative externalities. *Social Science Quarterly*, *90*, 292–308.
- Reagle, D. P., & Vinod, H. D. (2003). Inference for negativist theory using numerically computed rejection regions. *Computational Statistics & Data Analysis*, *42*, 491–512.
- Schuirman, D. A. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Pharmacometrics*, *15*, 657–680.
- Trandafir, M. (2014). The effect of same-sex marriage laws on different-sex marriage: Evidence from the Netherlands. *Demography*, *51*, 317–340.
- Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority* (2nd ed.). Chapman and Hall/CRC Press.

Fig. 1 Combining tests for difference (H_0^+) and tests for equivalence (H_0^-)

	Reject H_0^+	Not Reject H_0^+
Reject H_0^-	Conclude <i>trivial difference</i> (overpowered test)	Conclude <i>equivalence</i>
Not Reject H_0^-	Conclude <i>relevant difference</i>	Conclude <i>indeterminate</i> (underpowered test)

Fig. 2 . An illustration of the rejection regions for two one-sided tests. The region under the curve to the right of Δ_1 corresponds to H_{01}^- . The region under the curve to the left of Δ_2 corresponds to H_{02}^- . In this example, equivalence is defined symmetrically, so $\Delta_2 = -\Delta_1$. If one rejects both H_{01}^- and H_{02}^- , then must lie in the shaded region between Δ_2 and Δ_1 which is the joint rejection region defining the range of equivalence.

