Absence of evidence and evidence of absence in the effects of same sex marriage laws on rates of opposite sex marriage

Alexis Dinno

Assistant Professor, School of Community Health
Portland State University

Puzzling assertions about same sex marriage

Opponents to legalization of same sex marriage have positioned it as an "assault" seeking to "weaken," "destroy" and "undermine" *opposite sex* marriage.

In a recent ruling of the United States Court of Appeals for the Ninth Circuit in Perry vs. Brown (the repeal of CA Prop 8) the proponents argue "if the definition of marriage between a man and a woman is changed, it would fundamentally redefine the term from its original and historical procreative purpose. This shift in purpose would weaken society's perception of the importance of entering into marriage to have children, which would *increase the likelihood that* [opposite sex] *couples would choose to cohabitate rather than get married*."

This language has been deployed by representatives, judges and pundits.

(A) Ridiculous! (B) That's a testable hypothesis!

Are changes in state rates of opposite sex marriage different in states legalizing same sex marriage, than in states with no legal same sex marriage?

And while we are at it what about in states with strong same sex civil unions?

And what about weak same sex civil unions?

Dinno, A. and Whitney, C. (2013). Same sex marriage and the perceived assault on opposite sex marriage. *PLoS ONE*, 8(6):e8.

Data were obtained…

- Reported number of marriages by states and year from 1989-2009 from the National Center for Health Statistics

- Reported number of same-sex marriages for those states permitting them (communication with state health and vital records departments); these data permitted us to measure *opposite sex marriages* in each year.

- Estimated population age 18 and older by year and state from the US Bureau of the Census

- Proportion of the year during which same sex marriage and strong and weak civil unions were in effect from state legislative records.

Modeling marriage rates

Marriage rates are *non-stationary* processes (because $r_{ti} \approx r_{t-1i} + \zeta_i + \varepsilon_{ti}$), which makes application of traditional regression models invalid (i.e. observations are not i.i.d.).

No good! → $r_{ti} \sim \beta_0 + \beta_t t + \beta_m m_{ti} + \beta_s s_{ti} + \beta_w w_{ti} + \varepsilon_{ti}$

Modeling marriage rates

Marriage rates are *non-stationary* processes (because $r_{ti} \approx r_{t-1i} + \zeta_i + \varepsilon_{ti}$), which makes application of traditional regression models invalid (i.e. observations are not i.i.d.).

No good! $\rightarrow$ $r_{ti} \sim \beta_0 + \beta_t t + \beta_m m_{ti} + \beta_s s_{ti} + \beta_w w_{ti} + \varepsilon_{ti}$

(Excerpted from Dinosaur Comic 1897 by Ryan North http://www.qwantz.com/index.php?comic=1897)

Modeling (non-stationary) marriage rates

Fortunately, there's a not too involved analytic technique developed by the econometricians to deal with this type of situation termed a *single-equation generalized error correction model* which models *change* in marriage rates, rather than simply marriage rates:

$$\Delta r_{ti} = \beta_{0i} + \beta_c \left[r_{t-1i} - (m_{t-1i} + s_{t-1i} + w_{t-1i} + ms_{t-1i}) \right]$$
$$+ \quad \beta_{\Delta s} \Delta s_{ti} + \beta_s s_{t-1i} + \beta_{\Delta m} \Delta m_{ti} + \beta_m m_{t-1i}$$
$$+ \quad \beta_{\Delta w} \Delta w_{ti} + \beta_w w_{t-1i} + \beta_{\Delta ms} \Delta ms_{ti} + \beta_{ms} ms_{t-1i}$$
$$+ \quad \varepsilon_{ti} + \mu_{0i}$$

Where:
$x_{t-1} = x$ last year,

$\Delta x_t = x_t - x_{t-1}$ (the difference between x this year and last year)

Generalized error correction models

This type of model describes change in marriage rates in terms of three kinds of effects of each policy on opposite sex marriage rates:

1. There may be an *instantaneous short run effect*: the number of opposite sex marriages jumps as policy is goes into/out of effect

2. There may be a *lagged short run effect*: the linear trend in the number of opposite sex marriages changes while policy is in effect as opposed to when it is not

3. There may be a *long run effect*: the dynamic equilibrium implied by $r_{ti} \approx r_{t-1i} + \zeta_i + \varepsilon_{ti}$ may be shifted while the policy is in effect

Results presented in our first submission

We found no evidence of any relationship between rates of opposite sex marriage and same sex marriage or strong or weak same sex union laws.

Table 1. Effects of same sex marriage and union laws on opposite sex marriage rates (N=1071)

	estimate[a]	s.e.[b]	95% CI[c]	q-value[d]
Instantaneous short run effects of				
same sex marriage w/o strong unions	0.0001	0.0013	-0.0025, 0.0027	> 0.9999
same sex marriage & strong unions	-0.0007	0.0014	-0.0035, 0.0021	> 0.9999
strong same sex unions w/o marriage	-0.0003	0.0007	-0.0016, 0.0010	> 0.9999
weak same sex unions	-0.0004	0.0006	-0.0016, 0.0008	> 0.9999
Lagged short run effects of				
same sex marriage w/o strong unions	-0.0003	0.0015	-0.0031, 0.0026	> 0.9999
same sex marriage & strong unions	-0.0004	0.0031	-0.0064, 0.0056	> 0.9999
strong same sex unions w/o marriage	0.0000	0.0007	-0.0014, 0.0014	> 0.9999
weak same sex unions	0.0002	0.0007	-0.0011, 0.0015	> 0.9999
Long run run effects of				
same sex marriage w/o strong unions	-0.0037	0.0153	-0.0336, 0.0262	> 0.9999
same sex marriage & strong unions	-0.0279	0.0756	-0.1760, 0.1203	> 0.9999
strong same sex unions w/o marriage	-0.0067	0.0076	-0.0215, 0.0081	> 0.9999
weak same sex unions	-0.0036	0.0083	-0.0200, 0.0127	> 0.9999

[a] The arithmetic mean of the estimates from all ten imputed data sets.
[b] Combined standard errors account for both within- and between-imputation estimate variance.
[c] 95% confidence intervals are given by the estimate $\pm 1.96 * $ s.e..
[d] q-values are p-values adjusted upward to account for twelve multiple comparisons; compare to $\alpha/2$.

Reported effects and β estimates

The presented tables report *instantaneous short run* effects, *short run lagged* effects, and *long run* effects, but the GECM a few slides back estimated β terms. The reported *short run lagged* and *long run* effects are combinations of various β estimates.

Check it out, yo:

$$\text{instantaneous short run effect of } x = \beta_{\Delta x}$$

$$\text{lagged short run effect of } x = \beta_x - (\beta_c + \beta_{\Delta x})$$

$$\text{long run effect of } x = (\beta_c - \beta_x) / \beta_c$$
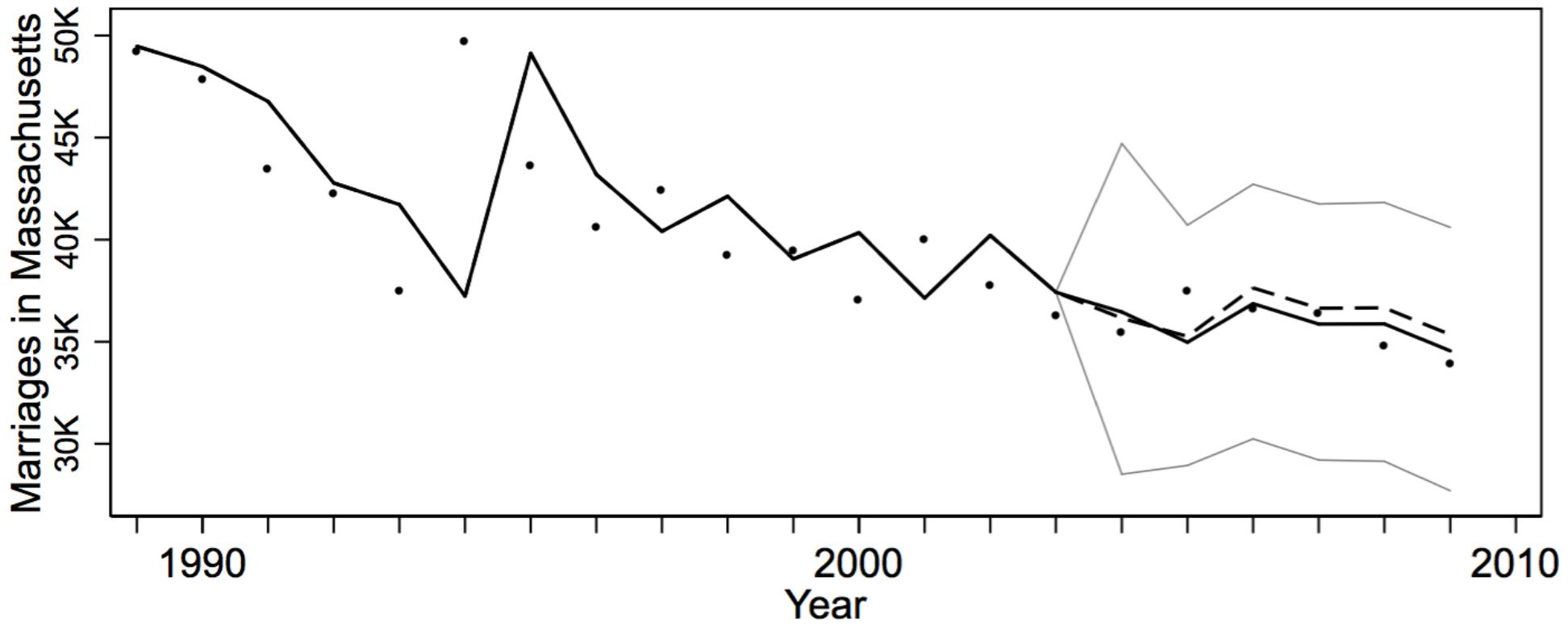
One reviewer took issue with our conclusion

Us: There is no relationship between implementation of same sex marriage or strong or weak same sex union laws and rates of opposite sex marriage.

Reviewer: Putting the premise of the paper aside, the empirics do not settle the argument about whether rates of OSM decline in response to the implementation of SSM. The tests have very low statistical power (as indicated by the wide confidence bands around the predictions) and thus the finding of a null result tells us little.

It's also worth noting that the findings are of the "wrong" sign: coefficients on most of the SSM predictors in Table 1 are negative and the dotted lines fall above the solid lines in Figure 1. So if we don't get hung up on statistical significance, this paper actually confirms the argument of those opposed to SSM that it leads to a relative drop in OSM rates.

Our responses to this critique (1st of 3)

1. We reject the idea that all effects are true effects and simply require a large enough sample size: *all* differences are statistically significant with a large enough sample size.

Our interpretation is than any apparent state-level effects are due to chance alone. (Also: N = 1071, with dozens of years with same sex marriage or union laws in effect.)

2. The issue of the "wrong sign" can be formally assessed: assuming that there is no relationship between same sex marriage and union laws (i.e. observed effects are due to chance alone), then we would expect 6 of the twelve effects reported in Table 1 to be positive, and 6 to be negative, in contrast to the 3 positive and 9 negative we did report.

We could formalize such an expectation as a null hypothesis with a binomial distribution, and p_0 would correspond to the probability that any reported finding is positive equals 0.5 and the number of tests $n=12$. Under these circumstances, the $P(X \leq 3) = 0.073 \geq 0.025$ which fails even liberal willingness to make a Type I error (α would need to equal 0.146 to conclude that enough of the reported effects were of the "wrong" sign).

Our responses to this critique (3rd of 3)

Absence of evidence is not evidence of absence.

—Altman, 1995

We would like either to provide *evidence of absence of an effect* of same sex marriage and same sex union laws on rates of opposite sex marriage, or to revise our conclusion.

But, how might we provide *evidence of absence of an effect*? Or, to put it another way, how do we provide evidence that two quantities are *equivalent*?

Altman, D. G. and Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *British Medical Journal*, 311(7003):485.

Equivalence testing: evidence of *sameness*?

Thankfully, I did not need to invent a new statistical machinery to provide evidence of equivalence.

Equivalence testing has its origins in pharmacology and clinical epidemiology, where a drug manufacturer does not want to be held to a higher standard than another manufacturer with respect to bringing a new drug to market.

Instead, they want to provide evidence that the new drug performs *equivalently* to the existing, FDA-approved, drug.

Equivalence is defined by ranges of tolerance

The methodological basis originates with *one-sided hypothesis tests*, and the logic that, if one selects a *tolerance* (Δ)—some quantity which the researchers or regulators value as "making no difference" between two measures—one can reject or fail to reject that a quantity is not greater (or less) than Δ.

For example, a decision to apply for a grant might change if the award was $1000 versus $1,000,000 (the *difference* between those quantities is *relevant*). But a researcher might find a grant awarding $999,000 to be *equivalent* to one awarding $1,000,000 if their budget tolerance is, say, ± $5,000.

The actual value of Δ varies depending on the nature of the data and/or hypothesis test, on the nature of the question, and even on regulatory guidelines (e.g. the FDA requires Δ to be a factor of 1.25 by an existing standard in bioequivalence trials: standard/1.25 and standard*1.25).

Tolerances can also be framed in terms of the test statistic under H_0.

Framing null hypotheses in tests for equivalence

This idea of tolerance in mind, one might want to know if the difference between two quantities (e.g. rates of opposite sex marriage with and without same sex marriage) is between $-\Delta$ and Δ. This translates into what is termed the *two one-sided test* approach to equivalence testing.

The *negativist null hypothesis* H_0^-: $|\mu_1 - \mu_2| \geq \Delta$ (as opposed to the *positivist null hypothesis* H_0^+: $|\mu_1 - \mu_2| = 0$ in tests for difference), translates into two one-sided null hypotheses:

$$H_{01}^-: \mu_1 - \mu_2 \geq \Delta \qquad \text{or} \quad H_{02}^-: \mu_1 - \mu_2 \leq -\Delta; \text{ giving:}$$

$$H_{01}^-: \Delta - (\mu_1 - \mu_2) \geq 0 \quad \text{or} \quad H_{02}^-: (\mu_1 - \mu_2) + - \Delta \leq 0$$

If we reject *both* H_{01}^- and H_{02}^-, we conclude *equivalence* within Δ.

Combining tests for difference & for equivalence

When tests for difference and for equivalence are combined, four possibilities result: (1) conclude equivalence or (2) conclude difference based on congruent decisions, (3) find trivial difference—that is, is difference *is* present, but is so small as to be ignorable, and (4) indeterminate findings, where there is not enough power to reject either positivist or negativist null hypotheses.

		H_0^+ (positivist)	
		Reject H_0^+	**Not reject H_0^+**
H_0^-	**Reject H_0^-**	*Trivial difference* (overpowered)	Conclude *equivalence*
(negativist)	**Not reject H_0^-**	Conclude *relevant difference*	*Indeterminate* (underpowered)

Meanwhile… I'm teaching biostatistics

Why am I only teaching my students how to test for *difference* when I am at the same time teaching my epidemiology students about publication and researcher biases *against* negative findings?

Why am I not also stressing the importance of testing for *equivalence*?

Intro Biostats-level equivalence testing: TOST

As implied, the two one-sided tests approach to determining equivalence does not require radically new math for any of the basic hypothesis tests. For example, a paired test of *mean difference* uses the (hopefully) familiar *t* test:

H_0^+: $\mu_1 - \mu_2 = 0$, and the samples' means are assumed to be *t*-distributed, with some degrees of freedom, ν.

$$t_\nu = \frac{\bar{d}}{s_{\bar{d}}}; \text{ where } d = x_1 - x_2$$

The test statistics for $\Delta - (\mu_1 - \mu_2) \geq 0$ and H_{02}^-: $(\mu_1 - \mu_2) + - \Delta \leq 0$ are:

$$t_{1\nu} = \frac{\Delta - \bar{d}}{s_{\bar{d}}} \qquad \text{and} \qquad t_{2\nu} = \frac{\bar{d} + \Delta}{s_{\bar{d}}}$$

Where *both* tests are right-tail tests: $P(T \geq t_1)$ and $P(T \geq t_2)$.

Software for TOST

If you are using **Stata**, you can find the package `tost` in Stata by typing :

`net describe tost, from(http://www.doyenne.com/stata)`

If you are using **R**, you can find the package `equivalence` on CRAN:

`http://cran.r-project.org/web/packages/equivalence`

If you are using **SAS**, you can perform equivalence t tests using the `TOST` option of `PROC TTEST`.

Aside: stupid math tricks

So the tolerance for one's idea of relevance, Δ, has the same units as the variable one is testing. However, this can be problematic for a few reasons:

- Your audience may be unfamiliar with your measure, and the relevance of a specific number of it's units.

- Some forms of the TOST test for equivalence have *very* unintuitive units of Δ (e.g. sign-rank or rank-sum statistics, where the unit is weighted by N).

Fortunately, it is easy enough to translate Δ into units of the test statistic's distribution under the null hypothesis. For example:

$$t_{1\nu} = \frac{\Delta - \bar{d}}{s_{\bar{d}}} = \frac{\Delta}{s_{\bar{d}}} - \frac{\bar{d}}{s_{\bar{d}}}, \text{ so } \varepsilon = \frac{\Delta}{s_{\bar{d}}}, \text{H}_{\bar{0}}: |t| \geq \varepsilon, t_{1\nu} = \varepsilon - t_\nu \text{ \& } t_{2\nu} = t + \varepsilon.$$

ε is in units of t. We might consider $\varepsilon = t_\alpha + 0.5$ a liberal relevance threshold.

Equivalence test results in our resubmission

Opposite sex marriage were found equivalent with and without same sex marriage.

Using a *uniformly most powerful t* test of mean equivalence (not TOST), we rejected the negativist null hypotheses of difference given a liberal tolerance ($\varepsilon = 0.5$ standardized units), a strict tolerance ($\varepsilon = 0.25$ standardized units), and even a very strict tolerance ($\varepsilon = 0.125$ standardized units) across the board for same sex marriage and strong and weak same sex civil unions.

Equivalence test results

Table 2. Equivalence tests for dynamic effects on opposite sex marriage rates (N=1071)

	t^{a}	$\mathrm{P}(|t| < \tilde{C}_{0.5})^{\mathrm{b,c}}$	$\mathrm{P}(|t| < \tilde{C}_{0.25})^{\mathrm{b,c}}$	$\mathrm{P}(|t| < \tilde{C}_{0.125})^{\mathrm{b}} \ (q)^{\mathrm{d}}$
Instantaneous short run effects of				
same sex marriage w/o strong unions	0.0741	0.0000	0.0000	0.0078 (0.047)
same sex marriage & strong unions	-0.5095	0.0000	0.0000	0.0191 (0.023)
strong same sex unions w/o marriage	-0.4456	0.0000	0.0000	0.0176 (0.023)
weak same sex unions	-0.5782	0.0000	0.0000	0.0208 (0.023)
Lagged short run effects of				
same sex marriage w/o strong unions	-0.1730	0.0000	0.0000	0.0108 (0.032)
same sex marriage & strong unions	-0.1435	0.0000	0.0000	0.0099 (0.040)
strong same sex unions w/o marriage	0.0181	0.0000	0.0000	0.0051 (0.061)
weak same sex unions	0.3044	0.0000	0.0000	0.0141 (0.028)
Long run run effects of				
same sex marriage w/o strong unions	-0.2426	0.0000	0.0000	0.0126 (0.030)
same sex marriage & strong unions	-0.3700	0.0000	0.0000	0.0270 (0.027)
strong same sex unions w/o marriage	-0.8857	0.0000	0.0000	0.0286 (0.029)
weak same sex unions	-0.4364	0.0000	0.0000	0.0260 (0.026)

[a] The quotient of the Table 1 estimates and their standard errors.
[b] The critical value $\tilde{C}_\varepsilon = F_{\alpha=0.05,1,\mathrm{df}=n-k,\varepsilon}$ where F is a quantile function of the noncentral F-distribution, the degrees of freedom are $n - k = 1060$ from equation 2, and ε is the noncentrality parameter of F, and the $\mathrm{P}(|t| < \tilde{\theta}_\varepsilon)$ is the cumulative density of $F_{1,\mathrm{df}=n-k,\varepsilon}$ at t [56]. Because under the null hypothesis of *difference*, one of the two single-tails of the tests *must* be rejected, these p-values should be compared to α rather than to $\alpha/2$ for the common interpretation of false rejection under null hypotheses of difference [56, 60].
[c] The q-values for $\varepsilon = 0.5$ and $\varepsilon = 0.25$ are not explicitly reported because the figures remain just as the p-values within the precision of this table.
[d] $q = 12p/i$, where i is the position of ordered p-values from smallest to largest. When stepping down from largest to smallest i, all hypotheses are rejected including and subsequent to the first with $q \leq 0.05$ to control the FDR for twelve multiple comparisons.

And then… what?

We have helped delegitimize the narrative that extending legal recognition to same sex marriages will harm state-level rates of opposite sex marriage.

This paper was the first published to attempt a longitudinal analysis of such rates of opposite sex marriage and legalizing same sex marriage. There are now at least two others:

Trandafir, M. (2014) The Effect of Same-Sex Marriage Laws on Different-Sex Marriage: Evidence From the Netherlands *Demography*. 51:317–340.

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.

Our paper has been cited in U.S. District Court lawsuits for at least seven states, including here in Oregon (Perriguey, et al., v. Kitzhaber), where last month legal recognition of same sex marriages was established in the courts.

Fin

Questions?

Some citations of interest

A seminal paper on *two one-sided tests of equivalence*: Schuirmann, 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(6):657–680.

On analogous confidence intervals supporting inference about equivalence and the interpretation of both *tests of equivalence* and *tests of difference*: Tryon, W. W. and Lewis, C. (2008). An inferential confidence interval method of establishing statistical equivalence that corrects Tryon's (2001) reduction factor. *Psychological Methods*, 13(3):272–277.

The current state of the art in *uniformly most powerful tests of equivalence*: Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*. Chapman and Hall/CRC Press, second edition.